This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.
A Concept Is A Container

Robert O’Shaughnessy

PhD in Philosophy
The University of Edinburgh
January 2016
Chapter 8

Barsalou

8.1 Introduction

8.2 Container theory vs Barsalou and Fodor

8.2.1 The perceptual area, the conceptual area, and the working memory area

8.2.2 Marking up perceptions

8.2.3 How the three areas feature in the DA-finst account and Fodor's account

8.3 Barsalou's perceptual symbol systems

8.3.1 Summary of Barsalou's account

8.3.2 Perceptual symbols do not get their meaning from what they resemble

8.3.3 Barsalou opposes Fodor's account

8.4 Simulations, simulators, frames and perceptual symbols

8.4.1 Simulators

8.4.2 Perceptual symbols

8.4.3 Frames

8.4.4 Simulations

8.5 Complex propositions in Barsalou's account

8.5.1 Perceived scenes

8.5.2 Imagined/simulated scenes

8.6 Conclusion

Chapter 9

Containers and intentionality: private concepts

9.1 Introduction

9.2 On interpretation

9.3 The owl and private concepts

9.4 Recognitional and functional information and how error occurs
Declaration of own work

I declare that this thesis is my own work, and has not been submitted for any other professional degree or qualification:

Robert O'Shaughnessy
Abstract

This thesis puts forward a theory which I call container theory for how a single notion of concept can satisfy the two desiderata that Machery (2009) sets out for concepts: (a) the Judgement Desideratum - a concept must permit categorisation, typicality and inferential judgements; and (b) the Propositional Desideratum: a concept must be capable of being used as a unit or constituent in compositions of propositional thoughts.

Achieving this depends on two core claims: (a) that a concept is a container of stored knowledge that pertains to a single category and (b) that double-armed finsts (DA-finsts), which keep track of both objects and the categories they have been recognised as, are composed into propositional thoughts. In addition to satisfying the desiderata, the framework allows a full explanation of what a concept is, how one is acquired, how it comes to have meaning, and how they are involved in matters of sense and reference and wide and narrow content.

The first core idea is that a concept is a container which gathers together information pertaining to a single category. Recognitional information (e.g. exemplars and prototypes) in the container allows categorisation and typicality judgements and functional information facilitates inferential judgements. Together these satisfy Machery’s judgement desideratum. I stress that a concept is the container itself and not the information stored in the container and not the label or symbol of the container.

The second core idea is that DA-finsts, which are an extension of Pylyshyn’s (1988) notion of a finst, are what are composed into propositional thoughts, thus satisfying Machery’s propositional desideratum. These finsts have no intrinsic meaning; they inherit their meaning from what concept/container they are currently keeping track of. DA-finsts are disposable symbols in that they dispose of their meaning once they stop pointing to a container or point to a new container.

In addition to satisfying Machery’s desiderata this framework has three major benefits. First, because DA-finsts can be substituted for Mentalese symbols in Fodor’s (2008) LOT account and for perceptual symbols in Barsalou’s (1999) account an attractive fusion of these two accounts can be achieved.
Second, a full account can be given of how new concepts/containers are acquired and come to have the meanings they do. In container theory concepts are individuated by their container structure rather than by their semantic structure. Information that is not recognised as belonging to any existing container causes a new container to be opened. What a container/concept pertains to is determined by what information is in the container, but the concept itself is determined by the container structure. Error (misrepresentation) occurs when the expected outcomes of inferences and interactions based on the stored information in a container do not match the actual outcomes experienced.

Finally, the framework is used to make sense of puzzles that give rise to distinctions between wide and narrow content and between sense and reference.
Acknowledgements

Many thanks to my supervisors Mark Sprevak and Tillman Vierkant, and especially to Mark who, as my primary supervisor, bore the brunt of the work. His endless patience, encouragement and suggestions had a huge influence on the project. I would also like to thank my Masters supervisors Christopher Mole and Julian Kiverstein, without whose help and encouragement I would not have pursued the doctorate.

Thanks also to everyone who made my stay at Edinburgh so enjoyable, but especially to Andrea, Francesca, Mario and Richard, fellow members of the Mind and Cognition Discussion Group, who were all incredibly supportive to me.

My family are very important to me and were a huge help at all stages. I am very grateful to all of them. A special mention must also go to Betty O’Connor for all the help and support she provided.

Finally, I would like to dedicate this thesis to my mother and father, and to the newest arrival in the family, my nephew Ben.
Chapter 1
Introduction

The primary thesis that I will defend is that a concept is a container of stored information that pertains to a single category. The concept/container refers to a category in the world, but it also contains recognitional and functional information for identifying and interacting with members of the category. The secondary thesis that I will defend is about the nature of the symbols that I claim are the direct constituents of propositional thoughts. Building on Pylyshyn’s (1989) notion of a finst, I will claim that the constituents of thought are DA-finsts that operate as disposable symbols. This secondary thesis is not essential to the primary thesis but it is central to my vision of what I call ‘container theory’ and greatly enhances the container theory’s explanatory power.

The notion of a concept plays a vital role in both cognitive science and philosophy of mind. Concepts are what allow us to recognise, infer, interact, and think; they are what allow creatures to semantically experience the world. But despite agreement on the importance of concepts, there is no consensus on what concepts are, or on what structure they have, or on how they come to have content/meaning.

Here are the principal questions answered by the container theory:

A. What is the difference between perception and (propositional) cognition and how do concepts fit in?

B. Is there a single notion of concept or is the term applied to different things?

C. How do concepts allow us to recognise, infer, predict, and interact?

D. How do concepts allow propositions to be expressed?

E. How are concepts and their meanings acquired?
F. Can concepts and their meanings be naturalised?

G. Do concepts have senses, narrow content, inferential roles?

Adopting what I will call ‘container theory’ allows one to give clear and, I hope, convincing answers to all of these questions and more.

1.1 What is container theory?

The central element of the container theory is that a concept is a container of stored knowledge pertaining to a single category. According to container theory, it is vital to realise that a concept is a container that is a distinct entity from what it contains.

This is different from the relation between, for example, a sentence and its words. A spoken sentence contains a number of words, but the sentence is also entirely constituted by those words. Similarly, a lamp contains a number of parts (a shade, a stand, a bulb holder, a switch, and so on), but it is also constituted by those parts. That is not the sense of containment in container theory. A concept is to be regarded as being like a box or a DVD. It is entirely distinct from its contents and exists even if it has no contents.

According to container theory, what a concept contains are pieces of information that are ways of recognising members of a category and ways of making inferences and predictions about members of that category. But those pieces of information do not constitute the concept, any more than the movie on a DVD constitutes that DVD. We will see, in Chapters 3 and 4, that those pieces of information should be regarded as constituting the subject’s conception of a category not her concept.

The other important element of container theory is the claim that the constituents of expressions of propositional thoughts are not concepts/containers, rather they are symbols for those concept/containers. This is in contrast to theories, like Fodor’s (1975) language of thought hypothesis, for which such symbols are concepts themselves. That is, the symbols stand directly for categories in the world rather than stand for concepts (which stand for categories). I will propose in Chapters 5 and 6 that those symbols are what I call double-armed fins (DA-finsts).

I will now give an outline of how container theory addresses the above questions.
A. What is the difference between perception and (propositional) cognition and how do concepts fit in?

I argue in Chapter 2 that perceptions record many pieces of information simultaneously in a dimensional space, and, as a consequence, many propositions are represented in a perception simultaneously. In contrast, propositional cognition has the ability to pick out a single proposition, which can be evaluated for truth. I argue that concepts/containers can feature in both perception and propositional cognition; however, in principle, perception can be devoid of concepts, but propositional cognition cannot.

B. Is there a single notion of concept or is the term applied to different things?

I argue, in Chapter 3, that there is a single notion of concept: a concept is a container of stored knowledge pertaining to a single category, and this notion can be seen to apply to all the things theorists take concepts to apply to.

C. How do concepts allow us to recognise, infer, predict, and interact?

According to container theory, a concept/container for category X serves to gather together recognitional and inferential information about X. Once something is recognised as being an X, the inferential information in the container is made available for inferring, predicting and interacting.

D. How do concepts allow propositions to be expressed?

I do not propose that concepts/containers are the direct constituents of propositional expressions; rather I claim that working memory symbols that keep track of point to concepts/containers are the direct constituents. I call these symbols DA-finsts. Expressing propositions is then a matter of juxtaposing or composing these symbols.

E. How are concepts and their meanings acquired?
Container theory has a straightforward answer to the question of how new concepts are acquired. There is nothing more to the acquisition of a concept than for the cognitive system to open a new container. There is no need for the system to know what the new concept is a concept of; the mere fact that there is a distinct container allows it to be gathering information about something different from the rest of its concepts. The answer to how the concept/container acquires a meaning is not so straightforward (this being one of the most vexed questions in the philosophy of mind). According to container theory, the fact that containers are distinct entities makes all the difference. Containers have the purpose of gathering recognitional and functional information that goes together. The system does not first work out to which category the container pertains, and then endeavours to ensure that only information pertaining to that category is put in the container. Instead, it endeavours to ensure merely that the functional information applies to what is recognised by the recognitional information. Outside interpreters or theorists can interpret what the category is, but the system does not need to. It follows from this that interpreters may have different interpretations of the category to which a concept refers, but I show that that is irrelevant to the concept possessor, who will be able to recognise, infer, predict and interact with members of a category without having to be able to also specify what that category is. This is the case for what I call ‘private’ concepts, which are concepts that do not have to be shared with other members of a community. These concepts can also be said to have original intentionality rather than derived intentionality. In Chapter 10, I contrast private concepts with ‘social’ concepts and word concepts.

F. Can concepts and their meanings be naturalised?

Can concepts and their meanings be naturalised? I argue that it is straightforward to naturalise concepts themselves. Whatever vehicles implement the container can be reduced to non-semantic constituents. It is also, in container theory, straightforward to naturalise the information in containers. In principle, it should be possible to point to the vehicles (for example, the neurons) which realise the stored information.

G. Do concepts have senses, narrow content, inferential roles?
In the final chapter, I deal with the related topics of senses, narrow content, and inferential role. I note how the fact that information is contained in discrete entities (containers) and the fact that DA-finsts have two arms, one pointing to entities and the other pointing to concepts, means that container theory can account for the puzzles surrounding these semantic properties.

1.2 Increasing complexity

In Chapters 2 to 6, each chapter builds upon the previous and adds new functionality to the capacities described in the earlier chapters. There is a gradual increase to the ability to express complex propositional thoughts. Sterelny (2003) notes how very simple creatures can get by with having a very small number of detection capacities and responses. For example, a bacterium can just detect and follow a chemical trail. As creatures become more complex, they develop more flexible ways of detecting and responding. Chapter 2 sets out how creatures that have the ability to iconically represent what they are sensing can record whole scenes at once. Creatures may then develop new ways of detecting what is in these recordings (rather than directly detecting what is currently in the world), and develop appropriate automatic responses to recognised entities. For example, if a snake shape is detected, this invariably causes the response of jumping into the nearest tree. Some recognition/detection capacity is required for this response to occur, but it does not require the full-blown concept/container system, merely a response associated with a stimulus type.

Flexibility can be added by adopting the concept/container system described in Chapters 3 and 4. The container system adds the capacity to gather together in the one place (a) many different ways of recognising members of the same category, and (b) many different ways of responding (inference and interaction) to the recognised instances. This range of responses can then be selected from flexibly. Perhaps now, when a snake is encountered, in addition to the fleeing into a tree response, there is also information about how snakes with certain markings can be made to flee, or even how this type of snake can be killed and eaten. Concepts/containers add flexibility to how a creature responds to what it encounters.

Chapter 5 sets out how working memory devices, which I call DA-finsts (double-armed finsts), add the functionality of tracking which concept/container has been associated with which entity in perception. This facilitates, for example, keeping track of which entity is best for carrying out a task, when comparing a number of different ones, so that it may be
chosen once the comparison is complete; or keeping track of an item and its category during a delay period when the entity is no longer present (e.g., a predator or prey goes behind a clump of trees); or facilitates returning to an item, after having attended to something more urgent.

Chapter 6 then sets out how DA-finsts can come to be constituents of expressions of complex propositions, i.e. propositions involving more than one concept. A creature may first develop just the capacity to represent, for example, conjunctions and disjunctions of DA-finsts. Then it may develop the capacity to combine DA-finsts with representations of simple relationships (e.g. above/below, bigger/smaller). And so on until it has developed a complex productive compositional system.

1.3 Container theory and the literature on concepts

In the literature on concepts there is much discussion on what concepts are and what structure they have. The container theory not only takes a particular stand on these questions but it can also explain why others have differing views. The primary source of disagreement is a common confusion between the contents of concepts and concepts themselves. According to container theory concepts are representations that stand for categories and they have the structure of a container.

There are three main views on what concepts are. One view is that concepts are mental representations, entities that have meanings (Fodor 1998, Margolis & Laurence 2007). Others take concepts to be senses, abstract entities that are meanings (Frege 1892, Peacocke 1992). A third view is that concepts are not objects or entities at all but abilities to recognise and infer (Wittgenstein 1953, Dummett 1993).

According to container theory, a concept is a representation that has a meaning (the category the container pertains to). Concepts are not senses but rather they contain information that plays an inferential role, which is what Fregean senses are often now taken to be. Concepts are not abilities but contain information that gives the possessor of the concept the ability to recognise and infer. So concepts/containers are mental representations that contain senses and give rise to abilities. This will be a recurring theme: that theorists often confuse the contents of concepts for concepts themselves.
This confusion arises also in disputes over the structure of concepts. Margolis and Laurence (2012) say concepts are variously taken to have ‘definitional structure, prototype structure, exemplar structure, theory structure, no structure at all, or some more complex combination of these options’ (2012, p. 291). According to container theory, concepts/containers have no semantic structure, they just have the structure of being able to contain information. However, the information they contain can have definitional, descriptional, prototype, exemplar, or theory structure, or indeed any other structure (e.g. emulator/simulator structure or affordance structure) that facilitates recognition, inference and interaction. That is, concepts contain exemplars, prototypes, theories, definitions, etc., but concepts themselves are not constituted by any of these entities.

Another dispute in the literature concerns what concepts are for. Fodor (2004) claimed that concepts are for thinking: concepts are the constituents in propositional thoughts. In reply, Prinz and Clark (2004) claimed that concepts are for doing: concepts give the ability to interact successfully with the world. Fodor insists that concepts are identified by their relation to propositional thoughts; in contrast, concept pragmatism (which Prinz and Clark argue for) takes concepts to be constituted by their role in behaviour. Container theory shows how to fuse these two viewpoints: concepts are for both thinking and doing. However, according to container theory, concepts are not constituted by their role in behaviour but concepts contain such roles; and concepts are not the symbols that make up propositional thoughts, but concepts are what such symbols refer to. In some respects, I side more with Prinz and Clark than with Fodor on this issue in that I take it that concepts can be possessed by creatures that cannot engage in complex propositional thinking (i.e. thoughts that involve more than one concept), and that concepts (at least private concepts as opposed to social concepts) are identified by the contents they contain (this will be discussed fully in Chapters 9 and 10). However, as discussed in Chapter 3, according to container theory, Fodor and Prinz & Clark are talking past each other because they have different aspects of concepts in mind. Fodor, I maintain, is concerned with symbols of concepts and Prinz & Clark are concerned with the contents of concepts. Container theory shows how both these aspects fit together.

Container theory allows for a single notion of a concept to satisfy the two desiderata Machery (2009) lays out for concepts:
a) The Judgement Desideratum: a concept must permit us to make categorisation, typicality and inferential judgements;
b) The Propositional Desideratum: a concept must be capable of being used as a unit or constituent in compositions of propositional thoughts.

Machery claims that Fodor and Prinz & Clark are talking past each other because the two parties mean different things by the term 'concept'. Machery claims that psychologists (into whose camp we could put Prinz & Clark) are concerned with a notion of concept that satisfies the Judgement Desideratum, while philosophers (into whose camp we could put Fodor) are concerned with a notion of concept that satisfies the Propositional Desideratum. Machery argues that generalisations that are made about the entities that satisfy the Judgement Desideratum have nothing in common with generalisations made about the entities that satisfy the Propositional Desideratum, and vice versa. Container theory allows us to say, however, that a single notion of concept satisfies the two desiderata. The contents of concepts satisfy the Judgement Desideratum, while the symbols of concepts satisfy the Propositional Desideratum.

Container theory also has suggestive connections to work on neural localisation. My concern in this thesis is not a claim about neural localisation, but a claim about cognitive functional architecture. My claim is that making sense of concepts requires us to distinguish between three functionally distinct cognitive capacities. But I also tentatively suggest that there is evidence that these capacities may also, predominantly but by no means exclusively, be associated with activity in, respectively, the sensory cortex, associative cortex, and pre-frontal cortex:

1. An area in the sensory cortex (call it the ‘perceptual area’) where occurrent perceptions and simulations of perceptions (e.g. imaginings, dreams, episodic memories) are recorded; for example visual perceptions are a result of activity in the visual sensory cortex. This is discussed in Chapter 2.

2. An area in associative or infero-temporal cortex (call it the ‘conceptual area’) where containers of stored knowledge pertaining to single categories are located. This is discussed in Chapters 3 and 4.

3. An area in the pre-frontal cortex (call it the ‘working memory area’) where DA-finsts keep track of and selectively attend to parts of the perceptual area and containers in
the conceptual area. Here, discrete symbols can also be composed/juxtaposed into expressions of complex propositions. This is discussed in Chapters 5-8.

Container theory claims that activity in the working memory area (DA-finsts) points to and keeps track of activity in the perceptual area (e.g. of what is being occurrently perceived) and activity in the conceptual area (containers of long-term stored information). Because DA-finsts have two arms, one DA-finst can simultaneously point to a perceived object, and a container full of inferential information, which the cognitive system has associated with that object. The claim is then that these DA-finsts are, in effect, disposable symbols, and, as such, can be the constituents of compositions that express propositional thoughts, while also pointing to the elements in the perceptual area and the conceptual area from which they inherit meaning.

1.4 Chapter by chapter summary

In Chapter 2, I argue that the difference between perception and propositional cognition is that the former represents many propositions simultaneously, while the latter picks out a single proposition in a form evaluable for truth. I adopt Haugeland’s principle for what constitutes an iconic representation: it is plotting variable information against independent dimensions (a reference frame). When information is plotted against a reference frame, one can always express many propositions about the information, and how it relates to the reference frame. On this basis I claim that perception is iconic representation, and, consequently, always represents many propositions. I then note that concepts can be operative in both perception and propositional cognition, but perception (unlike cognition) can, in principle, take place without any concepts being deployed.

In Chapter 3, I detail what I mean by a concept being a container of stored knowledge pertaining to a single category, and how this single notion of concept can satisfy both Machery’s Judgement and Propositional Desiderata.

Chapter 4 is concerned with how the Judgement Desideratum is satisfied: how it is information in containers that allows the required judgements to be made. I discuss the recognition process, which is how container/concepts are activated and associated with parts of perceptions. The result of the recognition process is that information in a container becomes available for making inferences about, and interacting with, what has been recognised. I set out how it comes about that information is placed in containers, and how it
can, in principle, be entirely perceptual and sensorimotor in nature. For this, I draw on the work of Barsalou and Prinz.

In Chapter 5, I introduce the notion of DA-finsts (double-armed finsts) which are based on Pylyshyn’s notion of a finst. I describe how they have both subject and predicate arms: the subject arm keeps track of points to a part of a perception, and the predicate arm keeps track of points to the associated concept/container. These items, which keep track of what has been recognised as what in the recognition process (described in chapter 4), can be regarded as disposable symbols, which stand for concepts. I also set out experimental evidence that supports the existence of DA-finsts.

In Chapter 6, I describe how DA-finsts, in their guise as disposable symbols that stand for concepts, can be the constituents of propositional thoughts. I argue that they can be substituted for Mentalese symbols in Fodor’s LOT account, and for perceptual symbols in Barsalou’s perceptual symbol account without losing anything essential from either picture. Consequently, I argue that container theory, with its incorporation of DA-finsts, can fuse the best parts of both Fodor’s and Barsalou’s account.

In Chapter 7, I look more closely at Fodor’s LOT account, and argue that it is wrong to think that a symbol manipulation system requires a stock of symbols with meanings that are the same each time they are used or encountered, as Fodor and Pylyshyn assume. I note how DA-finsts (disposable symbols), whose meanings depend on what concept/container they are currently pointing to/tracking, are equally as effective as word-like symbols in compositions of propositional thoughts.

In Chapter 8, I look at Barsalou’s perceptual symbol account and argue that when it comes to the Propositional Desideratum, perceptual symbols are redundant. They can be replaced by DA-finsts. Container theory shares with Barsalou’s account the idea that information is stored in containers, which he call frames. But container theory differs from Barsalou’s theory because it claims that these containers should be equated with concepts and not ‘simulators’, as in Barsalou’s account.

In Chapter 9, I distinguish between three types of concept: ‘private’ concepts, ‘social’ concepts, and word concepts. In this chapter, I focus on private concepts (concepts that are not shared with other members of a community). I set out how they are acquired, how their meanings are identified, and how error/misrepresentation occurs. I argue that it is possible to naturalise concepts/containers and the information they contain (for example, one could
reduce talk of them to talk of neural elements), but it is not always possible to naturalise an unambiguous and precise *interpretation* of the intentional content of a concept/container. However, I argue that a precise interpretation of which category a container pertains to is unnecessary for subjects to be able to successfully deploy such concepts.

In Chapter 10, I deal with the intentionality of social concepts and word concepts. Social concepts are concepts that are shared by members of a community; the most paradigmatic of these, I say, for humans are concepts/containers that are 'headed' by a public language word. I set out how they are acquired, how their meanings are identified, how error occurs, and I note how these processes differ for private concepts and word concepts.

Finally, in Chapter 11, I discuss sense and reference, wide and narrow content, and inferential role semantics, and note how helpful the notions of containers and DA-finsts are for explaining these semantic properties and resolving existing disputes.
Chapter 2
Perception

2.1 Introduction

In this chapter, I claim that the difference between perception and (propositional) cognition is not that one is nonconceptual and the other is conceptual, but that one plots information so that it represents many propositions simultaneously, and the other plots information in a way that picks out just a single proposition, which can be evaluated for truth. I will claim that perception involves iconic representation and (propositional) cognition involves logical representation.

I will adopt Haugeland’s (1991) criterion that what makes representation iconic is that information is plotted against independent dimensions. I will then claim that in neuroscience the representations that are taken to underlie perception also plot information against independent dimensions in an iconic manner. This gives rise to two further consequences.

First, perception does not have to be conceptualised. By this, I mean that the system does not have to have associated (or indeed even to possess) concepts of what is being perceived in order to plot information against dimensions. However, we will see that parts of perceptions can be conceptualised, that is, associated with concepts of what is being perceived.

Second, because iconic representations plot information against dimensions, iconic perceptions are not capable of having as their content just a single proposition. Rather, iconic representations always embody many propositions simultaneously, and there is no canonical way of interpreting them so that they pick out just one. A corollary of this is that a perception by itself cannot distinguish between, for example, the two propositions that is a dog and that is a pet made about the same object in the perception. There is no canonical way of interpreting the perception alone that makes it about the former rather than the latter or vice versa. We will see below, and in Chapter 8, that it is possible to combine an iconic perception
with a method of interpreting it, so that a single proposition is picked out, but the point I am making here is that the perception by itself always embodies many propositions. Iconic representations are to be contrasted with logical (Haugeland, 1991) or discursive (Fodor, 2008) representations – a paradigmatic example of which is a public language sentence – which are individuated by their ability to pick out a single proposition. There is a canonical way of interpreting logical representations that determines which single proposition is being asserted.

This applies both to simple propositions – ones involving a single concept – and to complex propositions – ones involving more than one concept. The same perception can embody, for example, the propositions That is a dog or That is a pet (simple propositions) or the propositions The balloon is above the jet and The cloud is to the left of the jet (complex propositions). All of these propositions can be true of the perception simultaneously, and there is no canonical way of interpreting the perception alone that singles out just one of these propositions. Something further is required to pick out just one proposition and that something further, I claim, is a logical representation (for example a sentence), a representation capable of picking out a single proposition from the very many that could be signified.

The rest of the thesis is devoted to explaining the nature of the concepts that are associated with parts of perceptions and that are incorporated in these logical representations that are capable of picking out and asserting single propositions – propositions capable of picking out that That is a dog rather than that That is a pet; or that The balloon is above the jet rather than that The cloud is to the right of the jet even though in each case both propositions could be true of a perception.

2.2 Iconic representation

I am adopting Haugeland’s (1991) principle for what characterises iconic representations. In an iconic representation information is plotted against independent dimensions. Each piece of information is not standalone; each piece of information only makes sense when it is known which dimensions it is plotted against. If one imagines these dimensions as axes of a graph they fix all the possible points where variable information can be plotted.
Here is a simple example: imagine a graph with two axes and for which each locational position can either be on or off. You have the information that location (20, 4) is on. What does this mean? It does not have any meaning by itself; rather you must first know what the axes (the independent dimensions) represent. Let us say that the dimensions are temperature in Celsius and day. Then one can interpret the information in the iconic representation as the temperature was 20 degrees Celsius on day 4. Change the independent dimensions to e.g. units sold and months and the interpretation changes.

A key point about iconic representations is that multiple pieces of information can be plotted at the same time. The information in iconic representation is relative information (as opposed to standalone information) because (a) it gets plotted relative to the independent dimensions and (b) because it gets plotted relative to other pieces of information. The temperature on day 5 will be plotted at a particular location in the graph which is relative both to the independent dimensions and to the temperature on day 4. An interpreter of the graph can make statements about a particular piece of information on the graph and also about how two or more pieces of information relate to each other. Because a number of pieces of information are plotted on the one graph and against independent dimensions, the graph represents many propositions simultaneously (The temperature on day 4 was 20 degrees Celsius; The temperature on day 5 was higher than on day 4; The average temperature of the two days was 21 degrees Celsius, and so on). The graph is a representation of all of these propositions. But to take the graph to be representing the proposition that The temperature on day 4 was 20 degrees Celsius requires an observer of the graph to be able to single out that particular proposition from all the others that the graph represents. To do this the observer must have the necessary concepts. We will return to this issue in Section 2.5 below.

A slightly more complex example of an iconic representation is a map. Let us suppose that the independent dimensions of a map are longitude and latitude. These dimensions fix all the possible locations where information can be plotted. Now the variable information is not simply on or off but can have a number of different values (e.g. green for land, blue for water, black line for road, a symbol for church, and so on).

Haugeland notes that ‘Iconic representations are distinctive in representing their contents by virtue of being somehow isomorphic to them’ (Haugeland, 1991, p. 174). He notes that sometimes the isomorphism is so obvious that the observer will take it to be
resemblance. However, it is generally agreed that isomorphism or resemblance is not sufficient for determining the meaning of a representation (Goodman 1976, Crane 2003). As Haugeland puts it, everything can be interpreted as resembling something else at least in some respects. Haugeland concludes that isomorphism or resemblance is not the defining characteristic of iconic representations. The defining characteristic of iconic representations is that variable information is plotted against fixed independent dimensions — a frame of reference (e.g. temperature or longitude) — and it is a consequence of this fact that it will be isomorphic to, or resemble, what it is about. For example, in visual perception the location of the perceptual elements is isomorphic to their locations in actual space, and because each part of a perceived object is plotted isomorphically to its location in space the perceived object will resemble the actual object.

The independent dimensions could be any independent dimensions, but some dimensions are more useful than others. As Haugeland notes there is no point plotting people’s telephone numbers against their social security numbers because there is no useful relationship between these two dimensions. In contrast, the dimensions of height, width, and depth which make up three dimensional space contain useful relationships of which to keep track. Most obviously, a subject can use information about where objects are located, relative to these dimensions, to approach or retreat from those objects.

Chapter 6 of Fodor (2008) distinguishes between discursive representations — logical representations — and iconic representations. Fodor says that the two types of representation are mutually exclusive (p.171). He also claims that ‘it’s in the nature of iconic representations not to be conceptual’ (p. 170). As we will see this is not correct. If we adopt Haugeland’s criterion for iconic representations, it is possible to plot conceptual information against independent dimensions. The information plotted in an iconic representation can be either conceptualised or non-conceptualised.

Fodor distinguishes between the two types of representation as follows. Logical representations have a canonical decomposition, a way of breaking them up into constituent parts that determines the meaning. If the constituents of a logical representation are ‘John’, ‘pushed’, and ‘Jim’, then there is a canonical way of interpreting the representation so that it means John pushed Jim rather than Jim pushed John, or any other proposition. By contrast, Fodor says, iconic representations have no canonical decomposition; one way of breaking up iconic representations into parts is as good as any other.
Laurence and Margolis (2012) cite Figure 1 as a counterexample to Fodor’s criteria. Consider, they say, a representation system that uses dots to represent people. This means that each constituent is conceptual. Each dot is a token of the concept PERSON. One can, for example, count the number of people represented. Yet the representation does not have a canonical decomposition. Laurence and Margolis say, ‘the whole can be decomposed by grouping the dots any way we like or by treating them individually. However we do it, each part will represent part of what the whole represents.’ (Laurence & Margolis, 2012, p. 303)

Fodor claims (a) that logical/discursive representations are conceptual and iconic representations are non-conceptual, and (b) that logical representations have a canonical decomposition while iconic representations do not. Laurence and Margolis produce a counterexample in which the constituents of the representation are conceptual but the representation does not have a canonical decomposition. I agree with them that this is a counterexample. My claim, however, is not that iconic representations must be non-conceptual but that they can be. And I claim that the distinction between logical and iconic representations is that the former pick out a single proposition while the latter iconic representations always embody many propositions. Laurence and Margolis, in their paper, are concerned only with the scope of the conceptual, and not with how to distinguish between logical and iconic representations. They take no stance on whether Figure 1 is a logical or iconic representation. They just take themselves to have shown that it is possible for a representation with conceptual constituents to have no canonical decomposition. But it is instructive to use Figure 1 to illustrate how I see the difference between logical and iconic representations.

First, note that it is not clear that the dots in the diagram are plotted against independent dimensions as Haugeland’s criterion requires. In other words, where the dots are
located in the diagram appears to be random; one place is as good as any other. What would make Figure 1 an iconic representation would be if the dots were plotted against spatial dimensions so that, for example, the distance between two dots was proportional to the distance between the corresponding people in the environment. Or, it could be that the X axis represents weight of a person and the Y axis represents height of a person.

Under which conditions would Figure 1 count as a logical representation? I claim that what is distinctive about logical representations is that they are capable of picking out single propositions. If the only rule according to which Figure 1 can be interpreted is 'there are X number of people' where X is the number of dots in the diagram, then Figure 1 has a canonical interpretation that specifies a single proposition. Consequently, it fits my requirement for being a logical representation. Logical representations will generally be accompanied by a canonical interpretation. This is not the same as Fodor's claim that what distinguishes logical/discursive representations is that they have a canonical decomposition, but it is worth remarking that the property of having a canonical decomposition is one way to allow a single proposition to be picked out by a representation.

Laurence and Margolis do not specify either a canonical way of interpreting Figure 1 that would make it a logical representation, nor the independent dimensions that would make it an iconic representation. But it is possible for an interpreter to provide his own way of interpreting, or set of independent dimensions. For example, he could treat it as an iconic representation by relating the dots to the dimensions of the page, and, consequently, form many propositions about the Figure (for example, There are the same number of dots in the top half of the diagram as there are in the bottom half or All of the dots are the same size). This is like encountering a page of text in a language one does not understand. One does not know the canonical way of interpreting the logical representations on the page, but one can still treat it as an iconic representation, and express many propositions about it, such as There are more words in the top half of the page than in the bottom half and Some of those symbols are punctuation marks.

2.3 Perception is iconic

I now want to say why I will be taking perceptual representations to be iconic representations. It is widely accepted in neuroanatomy that each sense makes use of topographic maps, and that what we perceive will depend on the information in those maps.
We do not perceive the information in the map; we perceive shapes and colours out there in the world. What is assumed is that the shapes and colours we perceive depend on what information is plotted in the maps. Topographic maps are ordered projections of the contents of spaces onto a representational schema. The visual system maintains maps that are retinotopic; the auditory system maintains maps that are tonotopic; and the somatosensory system maintains maps that are somatotopic.

By means of the cornea and lens of the eye, images of the world are focussed onto the retina. Activation of retinal cells causes signals to be sent via nerves which eventually arrive in region V1 of the cortex. The cells in V1 are said to be retinotopic in that they preserve the relative locational information in the retina (Tootell, Silverman, Switkes, & De Valois, 1982). That is, the information in adjacent retinal cells will be transmitted to adjacent cortical cells. The cells in V1 provide a map of the light reflectance information which fell at a certain angle on the retina. We can say that the independent dimensions of these maps are horizontal and vertical relative to the observer’s body.

The human cochlea and auditory cortex is organised tonotopically (Neff 1961, Talavage, et al. 2004). High frequency sounds are transduced at the base of the cochlea and low frequency sounds are transduced at the apex. Approximately 3,500 hair cells in the cochlea will vibrate at different rates (the variable information) and each cell corresponds to a particular frequency (the independent dimension) ordered from low at the apex to high at the base. This information is eventually transmitted to the auditory cortex where the mapping is preserved (adjacent hair cells in the cochlea correspond to adjacent sound frequencies and are transmitted to adjacent neurons in the auditory cortex).

In the somatosensory cortex, adjacent neurons represent adjacent nerve cells (Penfield & Boldrey, 1937). Each bodily part is represented in a distinct part of the cortex and the relative position of nerve cells in the body part is preserved in the relative position of neuronal cells because they are plotted against the independent dimensions of a space.

To simplify greatly, each sensory cell can have a variable value which is located in a particular ordered position relative to the independent dimensions of the larger iconic representation. The information picked up by these sensory cells is projected onto cortical maps. Much processing of the information can take place before it reaches the cortical map but relative locational position is preserved. In fact, it has been discovered that there are four
such somatotopically arranged maps (Kaas, et al., 1979), which record different types of tactile stimulation.

What I will be calling perceptual representations in this thesis will be representations which plot variable information against independent dimensions. When we visually perceive, there will always be a left and right part, and an up and down part, of the perception. If we perceive something out there in the world as being to the left of us, then we will find that it is plotted correspondingly against the left/right dimension in the internal retinotopic map. When we auditorily perceive, there will always be a pitch of the sound that is relative to the frequency dimension of the tonotopic map. When we feel a touch sensation it will be represented as being at a particular body location in the somatotopic map. Compare these perceptual representations to propositional or logical representations such as ‘I see a frog’, ‘I hear a piano’, ‘I feel a breeze’. There is no left/right/up/down information; there is no pitch information; there is no body location information. And this is because the information is not plotted against independent dimensions.

Unlike logical representations, perceptual representations have no standalone meaning; perceptual representations derive their meaning (what they are representations of) from the independent dimensions of the iconic representation. To see this, consider sensory substitution (Bach-y-Rita & Kercel, 2003). One of the best known forms of sensory substitution arises when a camera records a scene before a blind person and transmits the recorded signal into output on a pad on the blind person’s tongue. The pad contains sensors arranged in a 12x12 grid which correspond to vertical and horizontal independent dimensions (Kaczmarck, et al., 1985). With a little practice, the subject comes to treat the variable pressure on the tongue as imagistic because of the dimensions of the grid. She will treat, for example, stimulation at one corner of the grid as being at the top left corner, and stimulation at another corner as being at the bottom right corner. After a while, it is found that the subject can use these stimuli on the tongue to ‘see’ what the camera is recording. For example, she can use the information to pick up a moving ball. So what is going on here? I maintain that the cognitive system learns that it should treat the relative locations of the stimuli as being locations in a horizontal/vertical space out there in the world, rather than as locations on the tongue. In other words, over time, the system comes to change the reference frame against which the stimulation information on the tongue is plotted. At first it is plotted, in the usual manner, against the reference frame of the internal body and the subject experiences sensation on the tongue. After practice the system learns to plot the information against a
reference frame that corresponds to the spatial dimensions in front of the subject. The variable information (differing pressure on the tongue) stays the same but the independent dimensions it is plotted against changes.

Indeed, there is evidence from PET scans (Ptito, et al., 2005) that the stimulation of the tongue information that arrives in somatosensory cortex, after some practice, is sent to the visual cortex, where it can be plotted against the independent dimensions of extra-bodily space that may be hard-wired for vision. The point here is that the same variable perceptual information means different things depending on the independent dimensions against which the information is taken to be plotted. If the touch information is plotted in the normal fashion against a somatosensory map of the tongue, the information will be perceived as touch sensations on the tongue. If, however, touch information is plotted against a reference frame corresponding to extra-bodily physical space, the information will be perceived as objects out there in the world (for example, a looming ball). The sensory information does not have a standalone meaning: it derives its meaning from what independent dimensions it is plotted against. And plotting information against independent dimensions is what, according to Haugeland’s criterion, makes a representation be an iconic representation.

2.4 Perceptions could be devoid of concepts

Above I noted that I consider it possible for parts of iconic representations to be conceptualised, and below we will see that it is possible for all of the constituents of an iconic representation to be symbols standing for concepts. But, in this section, I will set out how it is also possible for an iconic representation to be entirely non-conceptual.

What do I mean by conceptualised? In this thesis I make a specific claim about what a concept is: it is a container of stored knowledge pertaining to a single category. So to be consistent with this claim, for a part of a perception to be conceptualised is either for the perception to include a concept (a container full of stored knowledge), or for it to be associated with a concept/container. Associating a concept/container with a part of a perception could be achieved via a symbol for the container, or just by the system recognising that the part of the perception in question is a token of the category a particular concept/container pertains to. Given that this is what I mean by ‘conceptualised’, it should be plain that many parts of perceptions are often conceptualised. As soon as I recognise that
what I am perceiving in front of me belongs to, for example, the dog category, I have conceptualised that part of the perception.

But, in principle, the kind of iconic representation that I take to be involved in, for example, visual perception could be entirely non-conceptualised. This is because iconic representations involve plotting information against fixed dimensions. It is therefore possible to mechanically fill each possible data point in the representation with a value without the system having a concept of what that value pertains to.

Take a recording device like a digital camera. There is no need for it to have, or to associate, any concept of what it is recording in order for it to produce an iconic representation. In principle, this can be the case for perception also. A subject would still perceive something (a shape, a noise) even if he has not associated a concept with that part of the perception.

A digital camera will record a picture of a scene (an iconic representation) by recording a variable colour value for each point in the space that is determined by the independent dimensions (the number of pixels in the horizontal and vertical axes). The camera will record a representation of the scene without representing it as anything. The camera does not require any concepts at all (not even the concept of colour values) to create the iconic representation. It has no idea that a conglomeration of those pixels constitutes, for example, a giraffe, and it has no need to in order for it to record the colour values for each point.

I am proposing that, as I am using the term, perception could be just like this. A visual perception, at a particular moment, could be a set of retinotopic variable values in (let us hypothesise) region VI of the visual cortex. Of course this would only get one to the blooming, buzzing, confusion stage as James (1890) puts it. It would (to simplify greatly) only give the system a map of what the colour at each point is. It is widely acknowledged that a great deal of processing takes place before the signals arrive in V1 allowing the system, for example, to segment the scene into objects and add a depth dimension. But even at this more advanced perceptual stage, there is no requirement for any (personal level) concepts to be deployed (e.g., there is no need for the system to recognise that a particular object in the perception is a giraffe). As we will see in Chapter 5, this is consistent with Pylyshyn’s (2001) theory that finsts capture preconceptual proto-objects in visual perceptions. Rensink (2000), puts forward a similar theory: early vision segments a perceptual scene into volatile ‘proto-
objects' which are grabbed by focussed attention. At the same time, a further system performs gist and spatial layout analysis on the whole scene.

I do not claim that what we perceive is always devoid of conceptual analysis. In visual perception, for example, there are many intermediate stages between the signals being transmitted from the retina and arriving in area V1 during which time the system endeavours to segment the scene into objects and proto-objects. It may well be the case that the kind of concepts I will be discussing in the following chapters play a role in determining what is perceived. What I do wish to claim is that signals could arrive in, for example, area V1 without the system having determined what high-level categories they belong to. In fact, I will be arguing in Chapter 5 that, in typical cases, the evidence is that we have not conceptualised large amounts of perceptual scenes, and that it often requires a DA-finst to capture an object in a perception before it can be recognised as being a member of a particular category.

2.5 Representing of, representing as

In effect what I am arguing for endorses Fodor’s (2008) distinction between representing of and representing as. A perception can be a representing of X without representing it as X. Fodor notes that representation of is transparent, i.e. you can substitute anything co-extensive at the X position. Discussing a photograph rather than a perception as an example of an iconic representation, he says

To be sure, a photograph may show three giraffes in the veld; but it likewise shows a family (one family) of giraffes; and an odd number of Granny’s favorite creatures; and a number of Granny’s favorite odd creatures; and a piece of veld that’s inhabited by any or all of these (Fodor, 2008, pp. 176-7).

All of these propositions are about the same X, the same group of giraffes. The picture can be said to show all of these propositions simultaneously regardless of what concepts an observer has. The picture is a representation of all of these propositions. By contrast, to take the picture as representing one specific proposition (e.g., There are three giraffes in the veld) requires the observer to have the concepts THREE, GIRAFFE, IN, and VELD, and to be able to form them into an expression that singles out that proposition from any others that could be made about the scene (e.g. There is a giraffe on the left and two on
The observer must commit to an expression of a single proposition from all the ones that could be made. It may be the case that such expressions must be sentential in form (perhaps only a representation with a sentential form can be evaluated for truth) or it may be that there are other ways of expressing single propositions (in Chapter 8, we will see that Barsalou (1999) proposes a way in which a proposition can be expressed in iconic perceptual form). But what I am claiming is that perceptions, just by themselves, as iconic representations, cannot express single propositions. Something over and above (what I am calling) perception is required to pick out an explicit single propositions — for example, a logical representation, or a canonical way of interpreting the iconic perceptual representation.

2.6 Perceiving, perceiving that

The distinction between representing of and representing as has much in common with Dretske's (1988) distinction between mere perceiving (of what something appears like) and perceiving that. The idea is that there can be something present in one's visual field that one will see (perceive what it appears like visually) without recognising what category it belongs to. This can be either because one does not have the concept (so cannot deploy it) or because, even though one possesses the concept, the system has not yet in fact deployed it.

One can be staring straight at a blicket (Gopnik & Sobel, 2000) but because one does not possess the concept of a blicket one cannot see that it is a blicket. In Dretske's terms, one sees the blicket (what it appears like), but one cannot see that it is a blicket. If one sees that it is a blicket one can bring to bear stored knowledge about that category. Dennett (1991) uses the example of the children's game Hunt the Thimble. A child can be staring straight at the thimble but fail to notice that she is. The child possesses the concept of a thimble but she fails to deploy it. She sees the thimble but she does not see that it is a thimble. In Fodor's terms, there is a representation present of a thimble but it is not represented as a thimble. The idea is that when she finally notices which part of her perception is the thimble, nothing necessarily changes about the perceptual information. What changes is that she deploys her THIMBLE concept; she comes to associate her THIMBLE concept with the part of the perception that corresponds to the thimble.

It often occurs that at a meal someone will ask for the salt to be passed, only to be told that it is already in front of him. What container theory claims is going on here (I will be discussing these matters much more fully in Chapters 4 and 5) is that the subject has in mind
a certain category that he is searching for (the *salt cellar* category). His SALT CELLAR container contains the means to recognise objects in the world that are salt cellars. But the system has not yet compared the part of the perception that is actually a salt cellar to the information in the container. We will see in Chapter 5 that I propose that if the subject's attention does not fall on the right part of the perceptual scene, then the comparison process which underwrites recognition does not take place. In Dretske's terms, the salt cellar is visible to the subject in his perceptual field (there isn't a gap in the perception at the salt cellar's location) but it has not yet been associated with/recognised by information in the SALT CELLAR container.

The subject perceives the salt cellar – there is visual information that corresponds to the salt cellar in his perceptual field. But in order to perceive *that* it is a salt cellar, the container for that category must be associated with the percept. It follows that I can see a salt cellar without seeing *that* it is a salt cellar. We will see later that I propose that DA-finsts (double-armed finsts) keep track of such associations. One arm tracks the part of the perception that corresponds to the salt cellar, and the other arm keeps track of the concept/container with which it has been associated.

Supporting evidence for the claim that we perceive even if the further stage of deploying concepts in the recognition process does not occur comes from the syndrome of visual agnosia (Farah, 2004). There are two kinds of visual agnosia: apperceptive and associative. In apperceptive visual agnosia, the subject reports being able to see colours and textures, but she is impaired in segmenting the scene into discrete objects. Consequently, she cannot recognise the categories to which the presented objects belong. In associative visual agnosia, the subject can correctly segment the scene into objects, and can describe the features of those objects, but she is unable to identify or recognise which category the objects belong to.

In these cases the subject possesses the necessary concepts (e.g. the subject can correctly identify the object if presented with the sound it makes) to identify the objects in the scene. But the subject cannot deploy them in visual perception (often the subject has suffered damage from a lesion). Nevertheless, the subject reports seeing things. In associative visual agnosia, the subject can describe what the object looks like. In apperceptive agnosia, the subject can report the colours she sees.
So, the view I am adopting can distinguish three kinds of perception. First, one can perceive without even being able to segment the scene into discrete objects; the result would be a blooming buzzing confusion (James, 1890). Second, one can perceive discrete objects in a scene but be unable, or have failed, to associate concepts/containers with the objects. Third, one can pick out one or more objects in one’s perception and associate concepts with them – recognise what category they belong to. This third kind corresponds to a transition from a representing of to a representing as (as set out in Section 2.5), or, equivalently, to a transition from just perceiving to perceiving that. I propose that it is the recognition process, which will be discussed in detail in Chapters 4 and 5, that implements this transition.

The main benefit of recognising to which category an object belongs (associating a concept/container with an object) is that it makes available the stored knowledge in the concept/container for inference or guided interaction with the perceived object. For example, before the DOG container is associated with a perceived object, the subject sees a perceptual token that is a dog. After the DOG container is associated with the perceived object, the subject sees that the perceptual token is a dog, and can bring to bear the stored knowledge about dogs in the container for inference and interaction.

This is not to say that a creature that has no concepts is effectively incapacitated or unable to act intelligently. Container theory is about the nature of the concepts we can associate with parts of perceptions and that we can utilise to express propositions. These deliberations can be grounds for carrying out actions. For example, a subject can compare a number of different propositions and select the most valued one. But many actions do not rely on such deliberations. When I leave a building, I do not manage my actions with explicitly formed propositional decisions. I don’t decide to walk to the door, I don’t decide to avoid hitting the sides of the door, as I walk through, and I don’t decide to put one foot in front of the other, and so on. Most of my behaviour occurs automatically and without any apparent deliberation. It is clear from the study of reinforcement learning in behavioural psychology (Rescorla & Wagner, 1972) and artificial intelligence (Sutton & Barto, 1998) that highly intelligent behaviour can be trained into a system using associations between objects and rewards/punishments or affordances (Gibson, 1979). For these associations to be effective, the system does not need to gather information in containers/concepts that pertain to single categories. Nor does the system need a capacity to express propositions. A system without concepts can associate fixed responses to appropriate stimuli. Nonetheless, a system that could, in addition, associate containers of stored knowledge (concepts) with perceived
objects, and make propositional claims about what is perceived, before deciding what action to take, would be a more flexible system than one that could not.

2.7 Iconic representations represent many propositions; logical representations represent one

In this section I return to the distinction between iconic representations and logical representations. I have argued that perceptions are iconic representations, and that they can either be conceptualised or non-conceptualised. I would point out that this claim is different to the claim that perception/iconic representation has a different type of content, namely nonconceptual content, to logical representations, which has conceptual content (McDowell 1994, Bermudez 1995, Peacocke 1998, Heck 2000, etc.).

My claim is that concepts/containers can be associated with both iconic and logical representation, and that what distinguishes iconic representation from logical representation is that the former always embodies many propositions simultaneously, and the latter picks out a single proposition. I claim that this is also what distinguishes perception from propositional cognition. Nonetheless there is a very important distinction between iconic and logical representations which also corresponds to a distinction between perception and (propositional) cognition.

It is because iconic representations plot information against a reference frame (independent dimensions) that they always simultaneously embody many propositions. An interpreter can express many propositions about how the plotted elements relate to the reference frame, and about how the plotted elements relate to each other given the reference frame. Iconic representations cannot, by themselves, express only a single proposition. Say X is plotted to the left and above Y in a graph. One cannot say that this graph only expresses the proposition that X is to the left and above Y, for it is also the case that it is true of the graph that Y is below and to the right of X, and that X is higher than Y. All of these propositions, and many more, are true of the graph. There is no canonical way of interpreting an iconic representation that makes it represent just a single proposition. Interpreters are free to pick out any proposition that they happen to notice and assert it.

Even if an interpreter recognises both X and Y in the graph (associates a concept/container pertaining to X and a concept/container pertaining to Y with the
corresponding parts of the graph), that does not mean that the interpreter has recognised that $X$ is above and to the left of $Y$. To recognise that, I claim, requires a logical representation that expresses that single proposition. So, even if every single constituent that has been plotted in an iconic representation has been conceptualised (had a container associated with it), that does not mean that every proposition that could be made about the iconic representation has been conceptualised. Far from it. There will still be a potentially infinite number of propositions that can only be said to be recognised by an interpreter once she has expressed them in a logical representation.

Logical representations are distinguished by the fact that they express a single proposition. The public language sentence ‘$X$ is to the left of $Y$ in this graph’ has a canonical interpretation. In a certain context, it picks out just one proposition that is either true or false. Often graphs will contain a caption (a logical representation) that expresses the proposition that the author wants to single out and bring to the attention of the reader. There is a canonical way for the reader to interpret the caption but there is no canonical way for the reader to interpret the graph. The reader may notice other propositions that are true of the plotted information and that are more interesting to him.

Perception, then, is a way of representing information so that many propositions are true simultaneously. Propositional cognition is a way of representing information that picks out a single proposition. This thesis explores how concepts/containers can be associated with iconic representations/perceptions (Chapters 3, 4, and 5) and with logical representations that pick out single representations (Chapters 6, 7 and 8).

### 2.8 Example: Chess positions

In this section I use chess positions to illuminate the difference between iconic and logical/propositional representations, and how both types of representation, nonetheless, go hand in hand with each other. I use chess positions to demonstrate that iconic representations can involve the plotting of entirely conceptual information.

#### 2.8.1 Representing a chess position iconically vs logically

One can represent a chess position iconically or logically. A chess position occurs in a space, namely a two-dimensional 8x8 grid. There are 64 possible locations (squares) for
pieces and only one piece can occupy each square at a time. To represent a position iconically is to plot, at each location where there is a piece, its type and colour. That is, to plot the variable information of the chess pieces against the independent dimensions that are the ranks and files. The result is a single iconic representation about which many propositional claims can be made.

The position can be represented logically by listing in propositional form where each piece is on the board (e.g. 'There is a white queen on square e2', 'There is a black bishop on square d7'). These are logical representations because each one picks out just a single proposition.

When a chess position is represented iconically in a diagram, all of the information about where each piece is located is available at once. An observer can immediately see all of the pieces and how they relate to each other and the other squares of the board. It is, however, a further step for an observer to notice or pick out useful propositional statements about the position.

It is less useful to represent a chess position as a list of pieces at locations because the set of logical representations do not reveal the relations between the pieces. Each logical representation picks out just one proposition. In contrast, an iconic representation represents a multitude of propositions simultaneously. This gives an observer the opportunity to notice the most useful ones.

One of the key elements in being a good chess player is to be able to form a small number of useful propositions that are true of the chess position out of the large number represented simultaneously. For example, 'I can capture the black queen' or 'There is a weakness in the queen side pawn structure'. It does not help a player if the iconic representation of the position is a representation of a pawn structure weakness. He must be able to represent it as a pawn structure weakness. That is, he must be able to deploy concepts that let him express this fact as a logical representation, and then know how to act to exploit the weakness.

Of course the most talented chess players do seem to be able to 'see' the right moves to make. They have the capacity to simulate the movement of the pieces on the board so that the end result is an iconic image of the position that results from, say, sacrificing a queen to achieve checkmate. They are not conscious of having to form propositional statements about which pieces to move; they just see them moving in their imagination. But even so, the player
still must be able to pick out single propositions such as *That position is checkmate*, or *The first move I make is to sacrifice the queen*, from the many propositions that hold true in each position.

When you are a beginner at chess, it can happen that you continue on playing even though checkmate has been achieved. You have perceived (iconically) a position in which one of the very many propositions that is true of the position is that it is checkmate, but as you have not singled out this proposition, and evaluated it as true, you do not notice it. The mere *iconic* representation of a position that happens to be checkmate does not result in the subject representing it as checkmate. Perceiving a position that is checkmate is not the same as perceiving that a position is checkmate. An *iconic* representation of a position that is checkmate needs to be accompanied by a *logical* representation that picks out the proposition that *it is checkmate*.

### 2.8.2 Concepts in iconic representations

Chess diagrams are a counterexample to Fodor’s claim that ‘it’s in the nature of iconic representations not to be conceptual’ (Fodor, 2008, p.170). Chess diagrams plot conceptual information in an iconic representation. I have already said that iconic images can be entirely devoid of concepts (e.g., the image created by a digital camera can come about without the camera deploying any concepts). But it is also possible for every piece of information in an iconic representation to be conceptual.

To represent a chess position iconically, in, say, a book, one option is to simply take a photograph of the board and use that in the book. There is no need for the camera, the photographer, or the book printer to have any understanding of chess in order for the photograph representing the chess position to appear in the book. What is plotted in the photograph are just light shadings. It is a matter for an observer to associate concepts with appropriate parts of the photograph.

However, usually a chess position in a book is represented by a diagram. In order to form such an iconic image, someone had to plot symbols of the pieces in their correct locations. The creator of the diagram will need to possess at least the concepts of the six different kinds of pieces, and how to plot symbols of them against the independent dimensions of the diagram that correspond to ranks and files. Each piece of information
plotted in the diagram (the iconic representation) is a symbol of a concept (PAWN, KING, QUEEN, etc.).

2.9 Conclusion

What I have attempted to do in this chapter is to show that perception is a useful way of presenting many pieces of information simultaneously and representing many propositions simultaneously. It is also useful, in certain contexts, to pick out single propositions in logical representations from the many that are present in iconic representations.

I claimed that perception can, in principle, be completely non-conceptualised. But perception becomes more useful when processed in the following two ways: (a) when concepts are associated with percepts of objects in the perception (which we will see, in Chapter 3, corresponds to Machery’s Judgement Desideratum); and (b) when concepts enable single propositions to be picked out about the perception (which we will see in Chapter 3 corresponds to Machery’s Propositional Desideratum). Going beyond non-conceptualised perception is what much of the rest of this thesis is about. I now turn to the nature of those concepts that can be associated with percepts and that will be constituents of logical representations.
Chapter 3
A Concept is a Container

3.1 Introduction

In this chapter, I will set out Machery’s Judgement Desideratum and Propositional Desideratum for concepts, and examine his claim that the notion of concept that satisfies the former is a different notion to the one that satisfies the latter. I will argue, to the contrary, that the same notion of concept (that a concept is a container) satisfies both desiderata.

In the last chapter, I set out how parts of perceptions could be conceptualised by associating a concept/container with them, and, in Chapter 4, we will see how this corresponds to the Judgement Desideratum being satisfied. I also set out how parts of perceptions can be conceptualised by the forming of a logical representation that pick out a single proposition by combining concepts/containers. In Chapter 6, we will see how this corresponds to the Propositional Desideratum being satisfied.

In this chapter, I will set out Machery’s desiderata, and his argument that they require different notions of what a concept is. In reply, I will put forward the central claim of this thesis: that a concept is a container. And I will argue that this notion of concept is operative in both desiderata.

Machery (2009) claims that philosophers and psychologists mean entirely different things when they use the term ‘concept’. If this claim is correct, it would be a problem for philosophers of cognitive science, for their theorising about concepts would concern an entirely different topic to the one cognitive scientists study. It would, equally, be a problem for cognitive scientists, when engaging with the arguments of philosophers.

Machery, in effect, identifies two desiderata (roles or purposes to be fulfilled) for concepts:
A. The Judgement Desideratum: a concept must permit us to make categorisation, typicality and inferential judgements;

B. The Propositional Desideratum: a concept must be capable of being used as a unit or constituent in compositions of propositional thoughts.

Machery’s claim is that the term ‘concept’ means a different thing when used in connection with the first desideratum than it does when used in connection with the second desideratum.

My claim will be that the term ‘concept’ has a single core meaning that applies to both desiderata, and that meaning is: *a container of stored knowledge pertaining to a single category*. If I am right, it will emerge that philosophers and psychologists are talking about the same thing, when they talk about concepts.

This single core meaning of ‘concept’ satisfies the first desideratum: the primary purpose of a concept/container is to store knowledge/information that permits categorisation, typicality and inferential judgements. And this core meaning satisfies the second desideratum because the container is a unit pertaining to a single category, and this is what is wanted in a constituent or unit of propositional thought.

Machery, in effect, claims that ‘concept’ for philosophers means *constituent of propositional thoughts*, and for psychologists means *bodies of knowledge used to categorise and infer*, and that these two extensions cannot be the same thing. He claims that the kind of things that form bodies of knowledge are not the kind of things that feature in propositional thoughts. My claim is that a single kind of thing, a *container of stored knowledge pertaining to a single category* is what (a) contains bodies of knowledge, and (b) is a unit that can be a constituent of a propositional thought.

I will argue that psychologists and philosophers do extend the term ‘concept’ to incorporate the meanings Machery sets out, but that this is a case of a very familiar form of polysemy that arises whenever containers are in play. It is like the case, for example, of the term ‘DVD’ (a container of stored information), where the term for the container is

---

1 I will be using the term ‘knowledge’ in this thesis the way Machery and psychologists do to mean “any contentful state that can be used in cognitive processes” (Machery, 2009 p. 8). As Machery points out such states do not have to be true or justified, nor do they have to be explicit or propositional, they can also be imagistic (perceptual) or procedural (sensorimotor).
commonly used as the term for the contents (e.g., when one says, 'that DVD was boring' meaning the movie on that DVD was boring), and where the term for the container is very commonly used as the term for the label of the container (as in, for example, handing someone a list of movie titles, and telling them to 'pick one of those DVDs', when in fact what is meant is pick one of those titles of (or labels for) DVDs). Even though 'DVD' is often used when one means movie or movie title, still, no one would think that a DVD is a movie or a movie title (or vice versa). What a DVD is, is a container of stored information.

My claim is that, similarly, a concept is a container of stored knowledge (more specifically, one that pertains to a single category), and that, even though psychologists extend the term and use it to mean the contents of the container, and philosophers extend it to mean the label of the container, still, no one should think that a concept is the contents or the label, any more than in the case of the DVD. My claim then is that what is playing the role of the concept in both desiderata is the container not the contents, and not the label of the container.²

3.2 Machery’s claim

Machery’s (2009) book, ‘Doing without Concepts’, caused something of a stir by proposing that psychologists should eliminate the term ‘concept’. A crucial step in his argument was the claim that philosophers and psychologists mean very different things by the term ‘concept’. Philosophers mean constituents of propositional thoughts, and psychologists mean bodies of knowledge used to categorise and infer. In effect, Machery argues that the

² What I am proposing is a version of what might be called the ‘different aspects’ response to Machery’s claim (I am grateful to Machery for pointing this out to me). That is, philosophers and psychologists are talking about different aspects of the same thing, and not about different things (see, for example, Margolis & Laurence, 2010 and Piccinini, 2011). However, insofar as these responses take there just to be two entities involved, Machery can reply by saying that the entities that are posited to satisfy the Propositional Desideratum play no role in generalisations about the entities that satisfy the Judgement Desideratum, and vice versa. That, Machery claims, strongly suggests that they are different things, not different aspects of the same thing. My account differs from the usual ‘different aspect’ responses, in that I take there to be a third entity involved (a container of stored knowledge), and I claim that psychologists are interested in one aspect of the container, and philosophers are interested in another aspect of the container. This third entity (the container) does feature in generalisations concerning both desiderata.
output of philosophising about concepts is irrelevant to the interests of psychologists, for it concerns an entirely different topic.

This would be especially disappointing, if true, for philosophers of cognitive science working on concepts, for they should be theorising about the same kind of concepts as cognitive scientists. It would also be puzzling: when I form the propositional thought ‘the dog chases the cat’, it seems clear that what is central to my understanding of the proposition is that I can make the kinds of judgements that psychologists are interested in about the individual constituents ‘dog’, ‘cat’ and ‘chase’. So how could there be no relation between the concepts philosophers talk about and the concepts psychologists talk about?

Chalmers (2011) distinguishes between verbal disputes and substantive disputes. Verbal disputes usually arise when the parties take themselves to both mean the same thing by a term, but, in fact, they have different meanings in mind. The parties then dispute facts surrounding the proposition, failing to realise that they are talking about different things. Once it is discovered that they were using a term in different ways, the dispute often vanishes. As Chalmers says, it is characteristic of verbal disputes that they are pointless. For a substantive dispute to arise over a term, the parties need to be using it to refer to the same proposition, and have differences concerning the underlying facts.

Machery (2009), in effect, is claiming that disputes between philosophers and psychologists over concepts are merely verbal disputes. They both use the same term ‘concept’, and assume that they are talking about the same thing. But, according to Machery, they are using the same term for different things. As a consequence, there is no point in, for example, Fodor (1998, p. viii) claiming, ‘I do think that most of what contemporary cognitive science believes about concepts is radically, and practically demonstrably, untrue; and that something pretty drastic needs to be done about it’. There is no point if Fodor has in mind a completely different meaning for ‘concept’ than cognitive scientists do. The dispute is merely verbal and verbal disputes are pointless. Or, as Machery (2010a, p. 199) puts it, ‘many controversies between philosophers and psychologists about the nature of concepts are thus vacuous’. The aim of this chapter, however, is to claim that because philosophers and psychologists do share a meaning for the term ‘concept’, there exists the possibility of substantive, and not merely verbal, disputes between them over the nature of concepts.
Having argued that philosophers and psychologists mean different things, Machery then proceeds to propose that psychologists should eliminate the term 'concept' because, as it is used by them, it does not refer to a natural kind. His argument is roughly as follows:

For something to be a natural kind it must be possible to make useful generalisations about the kind. Concepts for psychologists can be exemplars, prototypes and/or theories. One can make useful generalisations about each of these individually, but not generalisations that apply to all three simultaneously. It is normal practice in science to eliminate terms that do not refer to natural kinds, unless there are overriding reasons for retaining the term. There are no such advantages in retaining the term 'concept' for psychologists, therefore they should eliminate the term.

In this chapter, I am not going to be concerned with that argument, for I claim that exemplars, prototypes and theories are not concepts but the contents of concepts. Rather, I am going to focus my attention on the stepping stone Machery first needs: that philosophers and psychologists are talking about different things when they use the term ‘concept’. He needs this claim because philosophers do appear to make useful generalisations about concepts, and therefore do treat it as a term for a natural kind. For example, they can propose and dispute the generalisation that concepts must be word-like, or must be capable of referring to all and only Xs. So if philosophers mean by ‘concept’ the same thing that psychologists do, then there may be useful generalisations that can be made about concepts, and concepts would be a natural kind.

Machery can be taken to be claiming that ‘concept’ as used by philosophers and psychologists is a term of art. It therefore means whatever the practitioners of the discipline take it to mean. He claims that for psychologists

a concept of x is a body of information about x that is stored in long-term memory and that is used by default in the processes underlying most, if not all, higher cognitive competences when they result in judgments about x. I call this characterization “C” (Machery, 2009, p. 12).

This corresponds to the Judgement Desideratum. For philosophers ‘concept’ refers to:

that which allows people to have propositional attitudes (beliefs, desires, etc.) about the objects of their attitudes. The concept of a triangle is therefore that
which allows people to have propositional attitudes (beliefs, desires, etc.) about triangles (Machery, 2010a, p. 199).

This corresponds to the Propositional Desideratum.

I propose to simplify this formulation to the more specific ‘concepts are constituents of propositional thoughts’. There are, of course, many philosophers who do not think that the realizers of our attitudes are formed out of constituent building blocks (e.g., connectionists such as Clark 1993, and Cummins 2002, and autonomy theorists such as Davidson 1980). For these philosophers, the realisation of the mental states whose objects are the two propositions the dog chased the cat and the cow chased the cat would not share identifiable, discrete, constituent parts corresponding to chased and cat, and would not differ just in the part corresponding to dog or cow. The arguments of this chapter are primarily relevant to those who think that the realizers of propositional thoughts do have constituent structure (are composed out of discrete building blocks). The aim is to show how, under such circumstances, the meaning of ‘concept’ in the Judgement Desideratum can be the same as in the Propositional Desideratum.

3.3 Polysemy

Polysemy arises when an existing term is extended to have a related but new meaning. The example Murphy (2002) uses is the term ‘table’, for which he lists fourteen different meanings. So, for example, what furniture makers mean by ‘table’ (a piece of furniture with a flat top), what geographers mean by ‘table’ (a flat or level area), and what jewellers mean by ‘table’ (a facet of a cut precious stone) are very different things, although they seem to be related by the notion of a flat surface. It seems likely that the original central meaning of table is the one that we would naturally think of first (according to Murphy: four-legged piece of furniture), and that geographers and jewellers extended the existing term when they came to need a word for flat or level area and facet of a cut stone.

Machery is, in effect, arguing that ‘concept’ is polysemous in the same way.3 It doesn’t matter whether the philosophical or the psychological term came first. What matters is that they mean entirely different things, the way flat piece of land and piece of furniture

---

3 I take it that it is not plausible that the two uses of ‘concept’ are ambiguous or homonyms, as is the case for (river) bank and (money) bank which have nothing to do with each other.
mean entirely different things. There is no point in furniture makers and geographers entering into disputes over claims such as ‘tables usually have four legs’, or ‘tables often have grass growing on them’. These disputes are entirely verbal. Once the parties realise that they are using the same term to refer to entirely different things, the dispute vanishes. So, there is no more point in Fodor telling cognitive scientists that everything they say about concepts is wrong, than in furniture makers telling geographers that everything they say about tables is wrong. It’s not wrong, because the parties are simply referring to completely different things. Equally, there is no point in psychologists telling philosophers that concepts must have features that give rise to categorisation and typicality judgements, if philosophers are simply talking about something completely different. To resolve the dispute, all the parties need to do is realise that they mean different things when they use the same term.

Two questions arise for Machery’s claim: first is his characterisation of what psychologists and philosophers take concepts to be correct; and, second, if it is correct, does this mean that they are talking about two entirely different things? There is certainly scope for raising doubts on both counts, but, at least broadly speaking, I take Machery to be correct. Although many exceptions could be cited, broadly speaking, philosophers do normally seem to be examining concepts in the context of propositions, and psychologists in the context of making categorisation and inferential judgements. The case for these being entirely different things is essentially Fodor’s argument that the contents of bodies of knowledge (exemplars, prototypes and theories) are not the right kind of things to be constituents of thought, because they do not compose in the right kind of way. Challenges can be posed for this argument (e.g. Chapter 11 of Prinz (2002)), but for the purposes of this chapter, I propose to proceed on the basis that Machery is right on both counts.

What is certainly correct is Machery’s claim that psychologists do not generally concern themselves with how bodies of knowledge could be constituents of propositional thoughts. For the kind of experiments they carry out, and the kind of models and theories they produce, they don’t have to take into account how the entities they propose (exemplars, prototypes and theories) could feature as constituents in propositional thoughts. Nor do they concern themselves with how those entities could refer to ‘all and only Xs’ (i.e. pertain to a single category). Those matters simply don’t arise when psychologists try to work out what mechanisms allow us to judge that robins are more typical birds than emus; what mechanisms allow us to identify a penguin as a bird when it is so atypical; or what mechanisms allow us to know that most birds have wings and feathers. Psychologists take as their starting point
that knowledge is already grouped according to categories (e.g. the bird category above), and concern themselves with what that knowledge must be like (exemplar-like, prototype-like or theory-like) to give rise to the typicality and categorisation judgements found in the experiments. But as Machery (2009, p. 34) says, 'what psychological theories of concepts do not do, and are not supposed to do, is to explain what makes it the case that we can have propositional attitudes about the objects of our attitudes'.

But even though I will grant that Machery is right that philosophers and psychologists do use the term 'concept' polysemously, I will now claim that Machery is wrong to think that the kind of polysemy involved is the same kind that gave rise to the different meanings geographers and furniture makers have for 'table'.

For there is another way that polysemy comes about, and if I am right that this is the appropriate model in the case of 'concept', then it will show how philosophers and psychologists do, in fact, mean the same thing by the term. Take two other meanings together with usage examples that Murphy (2002, p. 404) gives for 'table':

(a) the company of people eating at a table: 'the entire table shared the pate',
(b) a painting, sculpture or photograph of a table: 'the table is painted very soulfully'.

I propose that psychologists use the term 'concept' polysemously in the first way and philosophers in the second way.

It is common for us to use the term 'table' when we, in fact, mean person or group of persons at the table. For example, a waiter might say, 'table six wants a coffee refill'. Here 'table' means person or persons, but note that there really is a table (with the meaning piece of furniture) involved. You can never use 'table' just to mean person(s). You can only use it to mean person(s) seated at the table. I can't point to a person walking down the street and say 'Look at that table'. Whenever 'table' means persons, there must be a table being referred to. The understander must identify an actual table, in order to know which person or persons are being referred to. Compare this with the case of the geographer's table; there is no piece of furniture involved. Apart from when the new term was first being coined, the meaning of the original term is irrelevant.

But in the case of 'table' meaning person, the original meaning of the term is essential to our understanding. If we do not know what a table is, in the furniture sense, we will have
no idea which person is being referred to. Moreover, it is because the table serves the function of grouping together the people that are to be picked out that it can be used to polyseemously refer to those people.

So, I suggest that this is exactly how psychologists use the term ‘concept’. The kind of polysemy is not like the geographer’s ‘table’, where the extended term comes to simply have an entirely different meaning. It is like the waiter’s ‘table’, where the original meaning of table is being used to identify something linked to, or contained by, or grouped together by that item. When psychologists use the term ‘concept’, what they mean, I claim, is not bodies of knowledge but ‘bodies of knowledge contained in (or stored in) a concept, where ‘concept’ means container of stored knowledge pertaining to a single category.’

Now let us turn to ‘concept’ meaning constituent of thought. As Murphy notes, polysemy often arises when we use the term for the object being represented as the term for the representation of the object. So we say, ‘the table is painted very soulfully’, and it might well take a moment’s reflection to even realise that this is an example of polysemy, so natural is the usage. What is actually meant is that the image or representation of the table is painted soulfully. It is not the table itself that is painted.

Note that Machery cannot make the same response to my claim as in his (2010) reply to Margolis and Laurence (2010). Margolis and Laurence suggest that we should think that concepts are mental symbols akin to words, and that psychologists are interested in an aspect of this concepts (the exemplars, prototypes and theories linked to such concepts). Machery responds on the following lines. Mental symbols that are constituents of thought play no role in, and cast no light on, the theories of what psychologists are calling concepts. Furthermore, Margolis and Laurence’s characterization make most of what psychologists say about concepts literally false. One should take a scientific term to be a failure only if most of what scientists say when they use the term is mistaken (Machery notes ether and force as examples). This certainly doesn’t seem to be the case for psychologists’ theories of concepts. If, on the other hand, ‘concept’ has the meaning set out in Machery’s characterization C, then most of what they say when they use the term comes out true. Such a result is surely to be preferred. But Machery cannot make the same response to container theory. First, concepts, conceived of as containers, can play a role in psychological theories. Note that, in Machery’s characterization C, he cannot say that concepts are just bodies of knowledge. He must say that they are bodies of knowledge activated by default as pertaining to a single category. An excellent explanation of why certain pieces of information are activated by default and others are not, is that only the ones contained in the container are activated. Second, what container theory proposes does not mean that what psychologists say about concepts is false, if it is understood that they are using the term as shorthand for bodies of knowledge contained in (or stored in) a concept, where ‘concept’ means container of stored knowledge pertaining to a single category.
Imagine that a restaurant manager has a computer drawing program which lets her arrange tables on the screen to match the bookings for that evening. When she moves about outlines of tables, perhaps composing them into greater wholes, it is natural for her to refer to the representations on the screen as ‘tables’, even though they are merely proxies for tables.

In the same way, according to container theory, when philosophers use the term ‘concept’, what they actually mean is label or proxy standing for a concept (where ‘concept’ means the same thing as it did for psychologists: a container of stored knowledge pertaining to a single category).

What is distinctive about this kind of polysemy and the waiter’s table kind is that it involves a form of shorthand. ‘Table’ is shorthand for representation of table or person(s) seated at a table. Just as I cannot use ‘table’ to mean simply person but only person at a table, so I cannot use ‘table’ to refer to a representation, but only to a representation of a table. This is entirely different to the type of polysemy involved in the geographer’s term ‘table’, which is the kind Machery needs to be involved in the case of ‘concept’.

So, just as the waiter and restaurant manager can use ‘table’ to mean such different things as persons and representation, but, nonetheless, only by virtue of there being an actual table involved, so I claim that psychologists and philosophers can mean such different things as body of knowledge and word-like constituent of thought, but only by virtue of there being an actual concept (meaning container of stored knowledge pertaining to a single category) involved.

3.4 Containers

It seems to be a general principle of polysemy that, whenever we have a term for a container or a thing that groups things together, we can use that term to mean the contents thus grouped together or contained. The term ‘table’ is used to mean persons sitting at the table because it is the table that groups them into a single category.

Consider the statement: ‘the x is boring’ (cf. Murphy, 2002, p. 438). Note how we can substitute any number of containers for x when we in fact mean the contents of the container. So we can substitute for x any of the following: DVD, newspaper, CD, video, TV, file, website, room, bottle (e.g. of wine), Christmas stocking, book, table, etc., and in each case we
can mean not x itself but the contents of x. I submit this is how psychologists are using the term 'concept': they mean the contents of the concept (i.e. the stored bodies of knowledge).

This kind of polysemy is even more common in the case of using x to mean representation of x. Indeed, it is the source of very many verbal disputes in the philosophy of mind, where one party is thinking of the vehicle of the representation, and the other party is thinking of the content represented, while both use the same term. Even when one is acutely aware of the danger, it is still easy to fall into the trap.

Let us examine a specific example, mentioned by Murphy, of a container of stored information, a DVD, and see how the example compares with the idea of a concept being a container of stored information.

As already mentioned, we often use the term 'DVD' as shorthand for the movie that the DVD contains (what the DVD contains). By the same token, if concept means container of stored knowledge, we should be able to use it as the term for the bodies of knowledge that the concept contains.

If I hand to you a list of movie titles and say, 'pick one of those DVDs; we'll watch it tonight', it doesn't occur to you to say, 'what you should say is: pick one of those labels/titles standing proxy for a DVD containing the movie pertaining to that title'. Similarly, we should be able to use the term 'concept' as shorthand for label or proxy representing a concept, where 'concept', again, means container of stored knowledge.

So, just as we do not for a moment think that a movie or a title/label is a DVD, even though we often use the word 'DVD' to mean those things, so, I claim, we should not think that bodies of knowledge or labels/proxies are concepts, even though psychologists and philosophers respectively often use the word to mean those things. And just as we should think that a DVD is a container of stored knowledge, so we should think that a concept is a container of stored knowledge.

In other words, we should think that when language is being used without any shorthand, philosophers and psychologists both think that concepts are containers/units of stored knowledge pertaining to a single category, the way we think (when we use language without shorthand) that DVDs are container discs of stored (digital) information pertaining to a single movie.
Good evidence that psychologists do use the term ‘concept’ in this way comes from Machery himself. In his characterization C mentioned above he says that for psychologists, a concept of x is a body of information about x that is stored in long-term memory. . . (Machery, 2009, p. 12).

But just two paragraphs before this he says,

the knowledge that is *stored in a concept* x is preferentially available when we think reason and so on about x. So to speak it spontaneously comes to mind. By contrast, the knowledge about x that is *not stored in a concept of* x is less available – it does not spontaneously come to mind (Machery, 2009, p.11-12, italics added).

Note how Machery shifts from saying that knowledge is *stored* (i.e. contained) in a concept to saying a concept *is* a body of information (knowledge). He shifts from saying that a concept *is* stored knowledge to a concept *contains* stored knowledge. This is precisely, of course, the kind of polysemy we expect whenever containers are involved: we use the term for the container as the term for the contents of the container. Indeed, six lines after setting out ‘C’, Machery is once more referring to ‘. . . the nature of the knowledge stored in concepts . . . ’ (2009, p.12). That is, he is once more treating concepts as containers that store knowledge, rather than the body of knowledge itself.

### 3.5 Concepts are like rooms

Let me try to drive home the point that it is the container (the establisher of boundaries for what belongs and what does not), and not the contents or the label of the container, that is the concept by using the analogy of concepts as rooms. Philosophers as diverse as Margolis (1998), Prinz (2005), Papineau (2006), Fodor (2008), and Recanati (2012) have made use of the notion of mental files when discussing concepts, and, of course, files are excellent examples of *containers* of information. They are suited to gathering together information pertaining to a single topic. We will return to the notion of files later, but, in order to emphasise my point that it is the container, and not the contents of the container or the label for the container, that should be taken to be the concept (what pertains to a single category), I will now use the analogy of a room and its furnishings, which is of
course suggested by the title of Prinz's (2002), 'Furnishing the Mind' which he in turn derives from Locke (1690), 'How come [the mind] to be furnished?'

It is natural to think that concepts are what furnish the mind, but, according to container theory, concepts are not the furnishings, but are in fact equivalent to the rooms, the containers of the furnishings. What furnishes the mind are not concepts, but pieces of knowledge/information that concepts contain. In other words, just as the room contains furnishings, so concepts contain information or knowledge. The room does not contain types of room, it contains types of furniture. Concepts do not contain types of concepts, but contain types of information or knowledge (in exemplar, prototype, theory, perceptual, sensorimotor, or factual form). Rooms contain bodies of furniture, but are not identical to such bodies, and concepts contain bodies of knowledge, but are not identical to such bodies. We gather in rooms pieces of furniture that belong together, and we gather in concepts pieces of information that belong together. We would not consider a piece of information, such as ‘cows have udders’, to be a concept of COW. It would be information contained in my concept pertaining to the single category COW. The same goes for an exemplar (say a visual image) of a cow, or any other type of information in my concept.

Notice how important the furniture is in determining the kind of room. The furniture in, for example, the bedroom tells you what kind of room it is, but it does not tell you what all and only bedrooms are. Similarly the information in a concept tells you what kind of thing the concept refers to, but it does not tell you (unless it is a definition) what all and only that kind is.

Let us imagine that we label the doors of rooms so that we can tell what type they are, without having to look inside. But the label is just a useful proxy, it does not determine the reference of the room. If I label the dining room door ‘bedroom’, that does not turn the room into the bedroom. If I swap all the furniture from the dining room with the furniture from the bedroom, the fact that I do not change the labels on the doors does not affect the fact that what used to be the dining room is now the bedroom. We will see in Chapter 9 that, in container theory, the contents of private concepts generally determine the intentionality of the concept.

One drawback of the room analogy is that buildings do not normally contain upwards of ten thousand rooms, with each one dedicated to a different purpose. This is why the notion of mental files is more common, for it is easy to imagine having many thousands of files and,
moreover, files typically contain pieces of information or knowledge. But it is important to realise that it is because files are examples of containers that they are useful as an analogy for concepts: they gather together information pertaining to a single topic; information is either contained in the file or it is not. That is what, I claim, concepts do: they gather pieces of information that pertain to a single category in the world.

So I am claiming in effect that rather than there just being two entities in the vicinity of concepts, as Machery would have it, there are in fact three: the two set out by Machery, and a third one, the container, which unites the other two. Whatever notion of a container one uses, there will be a distinction between (1) the container, (2) what the container contains, and (3) container labels or proxies which can stand for the container and the contents. My claim is that philosophers focus on the labels or proxies, and largely ignore stored contents, while the reverse is the case for psychologists. But both ignore what can unite them: that a container can simultaneously pertain as a unit to a single category in the world, and gather together knowledge that pertains just to that category.

We saw these three items when we discussed DVDs, and we noted how we use the term for the container for all three. It is the same with rooms. We can say, ‘Your room is a mess, tidy it up,’ meaning tidy up the contents of the room. In a hotel, we might look at a list of rooms with the name of their occupants (e.g. Room 233: Smith). We might say, ‘look at this room’ pointing to one of the items on the list, but we are not pointing to a room we are pointing to a label standing proxy for a room. I claim that it is the same for concepts.

I now want to look more closely at the two desiderata, starting with the Judgement Desideratum, and note how well the single core notion of a container can deal with both desiderata.

3.6 The Judgement Desideratum: conceptions

I will set out in the next chapter the specifics of the recognition process and how the contents of containers facilitate the judgements required by the Judgement Desideratum. Here I will make more general comments. In container theory, a concept is a container. A container satisfies the Judgement Desideratum because the contents of the container allow a conceptual creature to identify what category an instance belongs to, so that it can apply its stored knowledge about that category to the present instance. For each category for which a subject has the ability to make judgements, there are two kinds of information required (a)
recognitional information that allows recognition/identification of instances that belong to the category, and (b) functional information that allows inference and interaction with instances of the category. Storing the knowledge in a container is useful is that the two kinds of information are gathered together. It is of little use to me to have many discrete pieces of information about cows, if I cannot activate and bring to bear that knowledge when I have identified a cow. And it is of little use to me to activate a generalisation about cows, if I do not know how to recognise a cow so as to apply that information. The Judgement Desideratum entails that the cognitive system must endeavour to group together recognitional information that goes with functional/inferential information. As we will see in Chapter 9, according to container theory, the system does not have to first establish the category that a container pertains to, and then ensure that the information gathered applies to that category; rather, it just has to ensure that in each container the functional/inferential information does apply to what is recognised by the recognitional information.

In container theory, for every category that we can identify we have a concept, a container in which information pertaining to that category can be stored. For every concept, we have a conception (Rey 1985, Millikan 2000, Carey 2009, Gauker 2011). A conception is simply the sum of one’s stored knowledge pertaining to a single category. My conception of, for example, cows is made up of potentially many pieces of knowledge that guide me in recognising, drawing inferences about, and interacting with members of that category. This is what psychologists study: how it is possible to form such bodies of knowledge in a purely causal mechanistic fashion.

In container theory concepts contain conceptions of the category that the concept targets. This is what gives the oft-noted flexibility that conceptual creatures have: the conception is a guide for the kind of interactions with the world, and inferences in thought, that can arise in connection with members of the category. Your conception of x does not represent the category x, it represents what you know about xs. It is the container of the conception, the concept, that represents the category x; it is the container that pertains to all and only xs.

Compare this to a system that only has at its disposal word-like labels or proxies that are nomically activated, but does not have a conception of what category that label pertains to. As Brandom (2000, p. 65) puts it, ‘You do not convey to me the content of the concept gleeb by supplying me with an infallible gleebness tester which lights up when and only
when exposed to gleeb things. I would in that case know what things were gleeb without knowing what I was saying about them when I called them that, what I had found out about them or committed myself to.'

Furthermore, it seems that much of the knowledge in a conception is likely to be perceptual or sensorimotor in nature and not propositional. Here is Cummins arguing that 'mental words' can at best be cues (in effect proxies) for conceptions, stored bodies of knowledge:

The point can be brought out by a simple example. You are asked to go milk the cow. You make a plan to carry out this request. Among your early subgoals is the subgoal to find the cow. You decide to look in the barn... To recognize cows, you need to know something about them. You need, at least, to know how they look... You need to deploy your knowledge of cows in order to recognize a cow. It is your knowledge of cows, including tacit knowledge about the sort of retinal projections they tend to produce, that makes it possible for you to token a |cow| when you encounter a cow... But that is not all. The |cow| tokening triggers the next step in the plan. Now that you have located the cow and are on the spot, you need to locate the udder. Here, something like a picture of a cow, an image, say, would be very helpful, whereas a mental word is totally useless... Faced with actually having to deal with a cow, the burden therefore shifts... to your stored knowledge... you get the milking done... because you have... stored knowledge about cows (Cummins, 2002, p. 181).

Cummins notes how inadequate a proxy for a category, such as a mental word, is for letting us deal with the real world. Instead, things like mental images of a cow would be helpful to know what cows look like, and to know where the udder is. We need stored knowledge that barns are good places to look for cows; we need stored sensorimotor knowledge about how to milk a cow, and so on. We need many pieces of stored knowledge, and we need them gathered together, and activated together by default, when we are thinking about the category COW. In other words we need a conception of a cow: a collection of abilities to recognize and make inferences about things that pertain to the single category COW.
A conception should ideally pick out all and only Xs, store only true information about Xs, and store information only about Xs. But clearly, for most concepts, that is something our cognitive system cannot achieve. For most categories we cannot pick out all of its members, and for most categories we cannot pick out only its members (we often falsely identify an instance as belonging). So the extension of what the information in a conception can pick out rarely coincides with the extension of a concept (all and only Xs). The extensions would coincide if the conception amounted to a definition but, as, for example, Wittgenstein (1953) and Fodor (1981) have argued, there are very few definitions available. A conception, in most cases, amounts simply to the cognitive system’s ‘best efforts’ to pick out all and only Xs (the members of the single category X). It will consist of a collection of ways of identifying (different ways for different circumstances), and a collection of useful information (which can be context dependent). The cognitive system will delete information that turns out to give negative results, and add new information that promises to be predictive. The result is not picking out all and only Xs, but just some, many or most Xs. And the result is not having only true information that applies only to Xs, but just some or mostly such true information. But it is far better to have fallible incomplete information rather than none at all as Millikan notes:5

The ideal capacity to identify a substance would allow correct reidentification under every physically possible condition, regardless of intervening media and the relation of the substance to the perceiver. The ideal capacity also would be infallible. Obviously, there are no such capacities. If the cost of never making an error in identifying Xavier or ice or cats is almost never managing to identify any of them at all, then it will pay to be less cautious. But if one is to recognize a substance a reasonable proportion of the time when one encounters it, one will need to become sensitive to a variety of relatively reliable indicators of the substance, indeed, to as many as possible, so as to recognize the substance under as many conditions as possible (Millikan, 2000, p. 7).

Millikan then notes that there is no such thing as ‘the’ conception of e.g. CAT. Each person might have a different collection of abilities to identify. That is, according to psychologists, a different set of exemplars, prototypes and theoretical knowledge. Indeed, the same person will often have different conceptions at different times. But what will remain the

---

5 She restricts her inquiry to just substance categories.
same is the person’s concept of CAT. So a person’s conception cannot be the same as their concept.

She notes that, by contrast, the traditional view is that,

if two people each had a concept of cats, they would necessarily have both ‘the same concept’ and also ‘the same conception’ in our defined senses......I am opposing this tradition. There is no such thing as either ‘the’ conception of a substance nor as ‘the’ conception that corresponds to a public language term for a substance. Different people competently speaking the same language may have quite different – indeed, non overlapping – conceptions corresponding to the same substance term, and a single person may have quite different conceptions corresponding to the same substance at different times... I reserve the term ‘concept’ then for what we do have only one of per person per substance, and only one of per word for a substance, namely, for abilities to recognize substances and to know something of their potential for inductive use (Millikan, 2000, p. 12).

If we take Millikan to mean that we have a single ability (which remains the same) to recognise and make inductions per person per substance, while the various ways of recognising and making inductions may vary, this is the same as saying that the single ability contains the various pieces of stored knowledge that amount to the conception. What she claims is a concept, and calls an ability, is in fact a container of stored knowledge pertaining to a single category. Or, at least, it is the fact that there is a single container that gives rise to the single overarching ability that is comprised of the various sub-abilities to categorise and infer. We have many ways to recognise and make inductions per person per substance. What we have only one of per person per substance is, in container theory, a container or way of grouping together all of these diverse ways as pertaining to a single category. So we should reserve the term ‘concept’ for the container, which establishes what category a piece of information pertains to. Of course our cognitive system can only aspire to having one container/concept per person per substance; in certain circumstances it can arise that we have different containers for the same category (Frege cases such as Hesperus and Phosphorus), and in unusual circumstances it can arise that we have a single container when there are in fact two categories (e.g. water and twater).
3.7 The Propositional Desideratum

I will be dealing in detail with the Propositional Desideratum in Chapters 6, 7, and 8. What I say here will abstract away from many of the specifics mentioned in those chapters. The second purpose of a concept is to feature as a unit or constituent of a propositional thought. This is what allows us make claims or statements about the world involving more than one category, claims which can be true or false. This capacity is an enormous step forward in that it allows us to contemplate (make judgements about) states of affairs and events, rather than just single categories, and, as a further consequence, allows our conceptions to contain propositional thoughts. The claim of this chapter is that the exact same thing that satisfies the Judgement Desideratum also satisfies the Propositional Desideratum, namely a container of stored knowledge pertaining to a single category.

As Machery says psychologists are most interested in how we can make categorisation, typicality, and inferential judgements, and this generally requires associating just a single concept/container with the object in question. But philosophers are most interested in how we can combine a number of concepts into an endless number of explicit statements or claims (single propositions) about states of affairs and events in the world. One way we know that works is to use amodal atoms with fixed meanings for this is how public language operates. Fodor's (1975) language of thought hypothesis is that we form propositional thoughts using internal mental labels. His view is that the labels are concepts: they directly stand proxy for categories in the world. In contrast, the view being put forward in this thesis is that such labels would not be concepts, but proxies for concepts (i.e. proxies for containers of stored knowledge which are what directly refer to categories in the world).

I will now demonstrate that such containers can perfectly well satisfy the Propositional Desideratum. Containers can be the atomic constituents of propositional thoughts, and, moreover, there is no necessity for labels or word-like atoms to be involved.

We can see this most easily by adapting the very common idea that we form propositional thoughts or statements by combining atomic building blocks. Let us take building blocks to be the type that young children play with. They normally have letters on them, and can be combined in the right order to form words. But let us imagine that the blocks have words (labels) on them, and so can be combined into statements or sentences. The crucial idea is that each block is atomic. Each block is the smallest unit that can partake in the whole. Within the whole, it retains its discrete identity, and it can be manipulated as a
single item. So for example one could remove the block with the label ‘cow’ on it, and substitute it with the block with ‘dog’ on it, without having to do anything to the rest of the blocks. The blocks are atomic because they cannot be broken up into smaller wholes the way, for example, the sentence can be broken up into individual blocks.

But we can easily envisage these building blocks as containers (for example, as crates with lids). For each concept there is a separate building block. Inside the COW building block is stored all of our information pertaining to that category, giving the ability to recognise instances and make inferences about them. In other words, each block is a container that contains stored knowledge pertaining to a single category. Note how each block satisfies both the judgement and propositional desiderata. I can use the stored knowledge in the blocks to make categorisation, typicality and inferential judgements, and I can use the blocks as constituents (or smallest units) of propositional thoughts (just line them up so that they compose a statement).

The fact that these blocks contain conceptions doesn’t affect their ability to be atoms composable into larger wholes. But it does mean that labels are no longer necessary. On the standard picture, each building block is labelled in a way that distinguishes it from the other blocks. It would be pointless forming a composition of building blocks that had no identifying labels on them, for how would it be known which block referred to which category. But if the blocks contain conceptions, this does away with the need for labels because now the conceptions can be used to tell the blocks apart. You just look inside to see recognitional and functional information about the kind of thing that this particular block refers to. Note that you are not composing the contents of the blocks (you do not combine together the conceptions contained in the blocks); you are composing the discrete blocks each of which pertains to a single category (stands for all and only xs).

So we can see that a single container block can be a ‘concept’ that satisfies both desiderata: a concept that contains a conception that I use to identify and understand instances, and a concept that can be a constituent of a propositional thought.

However, further reflection shows that this would not be an efficient way for the cognitive system to operate. To form building blocks into endlessly different complex wholes requires a space in which to order and juxtapose them, i.e. a working memory. In order to form a statement with the concept COW in it, (a copy of) the COW container block (with all of its stored contents, the myriad pieces of information pertaining to the category COW)
would have to be transported into working memory. This is extremely inefficient: it would be far better to use labels or proxies for the container, for that will do just as well for the purpose of forming propositional thoughts. But note that these labels directly stand proxy for concepts (containers of stored knowledge), and not for categories in the world as, for example, LOT theorists propose. This will be discussed in more detail in Chapter 7.

I have shown that the constituents or units used in composing propositional thoughts do not have to be labels or word-like. Instead, containers can play the role of being atoms. However containers do not have the desirable features of being compact and easily manipulable. For that reason, it is plausible that labels/proxies are in practice what are used to compose propositional thoughts. But, in container theory, those labels are proxies for the container of the conception. As was seen in the arguments of Cummins and Brandom above, labels or proxies by themselves give no understanding; all the understanding comes from the conception contained in the container; mental words or labels should be seen as cues for accessing the conception. But, nonetheless, using labels or proxies, rather than containers packed with information, makes composition much easier.

One thing that I want to stress is that it is almost certainly the case that concepts are functional containers not physical containers. So whatever plays the role of gathering together all the various pieces of information that pertain to a single category is a functional container. The pieces of information do not have to be literally contained in a physical container. Compare a movie that is stored on a DVD with a movie that can be downloaded from the internet via a torrent. In the former case the information that comprises the movie is contained in a physical container, a discrete physical storage disc. In the latter case, the torrent file acts as a functional container. Torrents work by dividing up the information for, say, a movie into very small chunks, with each chunk being (physically) stored on a different host or computer. This can be with the aim of circumventing copyright laws, but it is also a technology used by legitimate content providers. The torrent file contains data that indicates how to retrieve all the various chunks of information, so that they can be downloaded as a movie. The torrent file can be said to functionally contain the movie but not physically contain it. In the same way, a concept does not have to physically contain the pieces of information pertaining to a single category, it can functionally contain them by storing data about where the information can be found in the brain (or indeed in one's notebook or other external storage device (cf. Clark & Chalmers, 1998)).
Another point is that our working memory is of decidedly limited capacity (Cowan, 2010). It is easy for experimenters to devise tasks to overload it (Cowan, 2012). So there will be a benefit in keeping the information worked on in it as compact as possible. It is true that functional containers are far more compact than physical containers but they still contain potentially large amounts of information (e.g. the list of how to retrieve each piece of information contained in the concept) most of which would be irrelevant to the forming of propositional statements. We can see this again with the example of torrents. If you visit a torrent site and view what movies they have available what you will be presented with on your screen is a list of movie titles together with perhaps a short description of the movie. In order to display (compose) the page that you view all that needs to be retrieved from each torrent file (the functional container of the movie) and sent to the working memory of your computer is the label (movie title) and the short description. There is no need to transmit to your computer each complete torrent file containing the long list of locations that host the small chunks of information into which the movie has being divided. If you then decide that you want to download a particular movie the label (movie title) that you click on can be used to cue the correct torrent file which in turn contains the information used to retrieve all the small chunks of information that comprise the movie you download. Similarly when composing a propositional thought out of concepts it would be efficient to use labels or proxies for the concept (the functional container of stored knowledge) rather than clogging up working memory with a long list of the information functionally contained in the concept. We will see in Chapter 6 that I propose that these proxies are DA-finsts (or disposable symbols) rather than Mentalese word-like symbols as in Fodor’s LOT hypothesis.

3.8 Conclusion

Machery claims that disputes between philosophers and psychologists about the nature of concepts are mostly vacuous because they simply mean different things by the term. I think that he is quite correct that they have often been talking at cross-purposes because psychologists have been focussing on what satisfies the Judgement Desideratum (which I say are the contents of concepts/containers) and philosophers on what satisfies the Propositional Desideratum (which I say are symbols or proxies for concepts/containers). So he is quite right that if psychologists mean contents of concepts and philosophers mean symbols for concepts they are indeed talking about completely different things. But I claim that neither the contents of concepts nor the symbols for concepts are concepts. As a consequence I disagree with
Machery's claim that the same notion of 'concept' cannot satisfy both desiderata. I have shown that a single core notion, where concept means *container of stored knowledge pertaining to a single category*, can indeed satisfy both desiderata. Moreover, I have shown how it is a typical feature of polysemy for container-like words that their meanings get extended in just the way philosophers and psychologists have extended them for concept.

The upshot is that philosophers and psychologists do after all mean the same thing by the term 'concept'; they can enter into substantive disputes over the nature of concepts, over the nature of units or containers of stored knowledge that pertain to a single category. We thus achieve the desirable result that philosophers concerns should be of interest to psychologists and vice versa. And we also have an explanation for the fact that we are able to make the kind of judgements required by the Judgement Desideratum about the very constituents of propositional thoughts that feature in the Propositional Desideratum.

In the next chapter, I will explore in more detail how containers come to be filled with stored knowledge and how those contents satisfy the Judgement Desideratum and, in Chapter 6, I will turn to the Propositional Desideratum.
Chapter 4
Recognition and inference: The contents of containers

4.1 Introduction

In this chapter, I will expand on the claim that a concept is a container, and on how Machery’s Judgement Desideratum is satisfied. The Judgement Desideratum is that concepts should give the capacity to make the kind of categorisation, typicality, and inferential judgements found in psychologists’ experiments. I will describe how what I will call the recognition process involves associating a concept/container with (typically) a part of a perception. Information contained in a concept/container allows a subject to make categorisation judgements (recognise an entity as being a member of a category), typicality judgements (determine how typical a member a particular instance is of a category) and inferential judgements (bring to bear stored information about the category).

A container is associated with a percept when information in the container is sufficiently similar to the percept information. For example, a percept of what is, in fact, a dog will be recognised as a dog if it is associated with the DOG container. Recognition occurs when recognitional information (e.g. exemplars or prototypes) in the DOG container is sufficiently similar to the percept of the dog.

In Chapter 2, I claimed that percepts of objects can be either conceptualised or non-conceptualised. In Chapter 3, I claimed that containers/concepts pertaining to a category gather together recognitional information (ways of recognising members of the category) and functional information (ways of inferring and interacting with members of the category). In container theory, a percept is conceptualised if it is associated with a container/concept, and it is non-conceptualised if it is not associated with a container/concept.

I claimed that, for example, visual information about an object could be mechanically plotted against independent dimensions without the system having to associate any
concept/container with that information. However, should the visual perceptual information be matched to recognitional information in a particular container (e.g. an exemplar or a prototype in the DOG container), then the system can associate that container with the visual information (the percept). This association of the container with the percept gives the benefit that the functional information in the container is made available to guide inference and interaction with the entity recognised (the object of the percept).

Recognition (conceptualisation) is a process: a non-conceptualised percept is matched to recognitional information in a particular container, and, as a result, functional information in that container, which facilitates inferential judgements about the object of the percept, becomes available. According to container theory, a container becomes activated when information in the container matches perceptual information. The container becomes activated in the sense that all of the information in the container becomes available to the system for dealing with the object of the percept. It is like inputting the word ‘dog’ into an internet search engine, and, once a match is found, a website containing stored knowledge about dogs being activated.

This chapter will discuss the nature of the information in containers, how information gets into containers, how information is used to recognise and infer, and how container theory can be neutral about whether the recognition process is top-down or bottom-up.

4.2 Exemplars, prototypes and theories

According to container theory, containers contain pieces of information that enable the recognition of members of a category, and enable the drawing of inferences about, and interaction with, members of a category. Psychologists have many theories for how we categorise and infer, but most centre on the notions of exemplars (Medin & Schaffer, 1978), prototypes (Hampton, 1979) and theories (Murphy & Medin, 1985). In fact, many theorists now accept that all three types of information are involved (Machery 2009, Prinz 2002).

An exemplar is a stored example. It is like, for example, determining that the colour before you is crimson by comparing it to a sample of crimson on a colour chart, or recognising a person by comparing them to a photograph. According to exemplar theory, entities that are encountered are compared with a stored representation of a particular instance or example of a category. It is assumed that a subject will have many exemplars of,
for example, dogs, and, when a dog is perceived, it just has to be sufficiently similar to any one of those stored exemplars in order to be recognised as a dog.

Many psychologists take exemplars to be made up of a list of features, and if a percerpt of an object shares, say, 75% of those features, then it is recognised as being a member of the category of the exemplar. There are two problems with this. First, the list of features required to describe, for example, any particular dog is likely to be very long indeed. Second, it requires the cognitive system to be able to recognise (and so have the concept of) each of those features before the system can determine whether, for example, the dog currently being perceived possesses those features or not. And, in turn, each feature may have a long list of sub-features that have to be recognised before it can be determined if the feature is present; and the same goes for those sub-features, and so on.

Below, I will be looking at how Barsalou and Prinz claim that recognitional and functional information has a perceptual basis. The advantage of ways of recognising being perceptual in nature is that it avoids the problems just mentioned for features. Plotting the recognitional information against independent dimensions – the mechanical way perceptual information is plotted – allows the system to plot the features concerned without having to conceptualise or recognise them.

The traditional approach to recognition via exemplars involves ticking off, from a long list, the features that are present. But a perceptual visual exemplar of, for example, a dog would plot the information of what it appeared like against the same dimensions it is plotted against in visual perception. Recognising a newly encountered dog would simply be a matter of comparing the percepts of the dog (plotted in a dimensional space) with the exemplar (plotted against the same dimensions). The system does not have to recognise any of the features of the dog in order to successfully compare it. The features are present in the perceptual representation, in the sense that they could be labelled by outside observers (e.g. psychologists), but they do not have to be recognised as such by the system. The system just has to match iconically stored information with the iconic information currently being perceived. This does not rule out that features are also stored explicitly, as we will see below. But it does mean that concepts/containers can be filled with ways of recognising that do not require the system to have concepts of the features of members of the category.

A prototype is essentially a summary, average, or central tendency of a number of exemplars. A prototype of a bird would be formed by averaging the features of all the birds
previously encountered that resulted in the formation of exemplars. Prinz (2002, p. 53) gives an example adapted from Smith (1989) of a feature list for a prototype of a bird:

has a beak, has wings, has two legs, flies, walks, paddles, lives in trees, lives at lakes, eats worms, eats seeds, eats fish.

Each of these features can be given a different weight depending on how commonly it is found in exemplars. In Prinz’s example, ‘flies’ is given a weight of 0.8, while ‘paddles’ is given a weight of 0.1, reflecting the fact that most birds fly, while only some paddle. Once again, prototypes do not have to be comprised of a list of features. They can be averages of plotted information against independent dimensions derived from exemplars.

Prototypes are particularly useful for explaining how we form the typicality judgements mentioned in Machery’s Judgement Desideratum. Psychologists have found that subjects are more likely to judge an instance as being a typical member of a category the more closely it matches a prototype (the average or central tendency of all previously encountered exemplars). Indeed, some theorise that what we find beautiful are instances that most closely match prototypes. A beautiful face would be one which closely matches the average or central tendency of all faces (Rhodes et al., 2003).

Theory theorists propose that theories explain how we recognise and make inferences. While exemplars and prototypes offer good explanations for how we can recognise instances based on superficial perceptual similarities, theories are useful for dealing with information that goes beyond the simply perceptible. As Prinz puts it,

many psychologists now say that concepts encompass beliefs about causal mechanisms, teleological purposes, hidden features, and fundamental divisions in ontology (Prinz, 2002, p. 76).

Theories can account, for example, for ‘hidden essences’ that determine the category to which something belongs. Theories can allow one to recognise ducklings as the offspring of ducks, even though they do not look like their parents.

While theoretical information might be best set out in propositional form (e.g., ‘water is constituted by H₂O’), which would require the subject to have the concepts involved, and the means to compose them into an expression, Barsalou (1999) and, particularly, Prinz (2002) suggest ways in which the theoretical information could be in perceptual and sensorimotor form. Take, for example, the category APPLE. In addition to exemplars and
prototypes of apples, the creature might also store theoretical information (e.g. the inside of apples are white; sometimes they contain worms; apples grow on trees; etc.). According to Prinz, such information might be in the form of perceptual episodic memories. However, insofar as one regards episodic memories as sequences of iconic images, there arises the problem set out in Chapter 2 that iconic representations, by themselves, cannot single out just one proposition. They always represent many propositions simultaneously. As I will argue in Chapter 8, there has to be something over and above the perception or episodic memory that picks out the single proposition that constitutes the theoretical information.

Nonetheless, the point remains that the perceptual and procedural stored information is unlikely to be confined to static images; it is likely to also be sensorimotor (episodic) in nature. According to Barsalou and Prinz, there will not just be a set of static images of dogs in the DOG container. Instead, there will be a capacity to generate simulations and object-centred models based on the stored information. This would allow the subject to imagine, for example, a dog rotating before him (which requires the knowledge of how it appears from many angles), or imagine himself walking around the dog, thus knowing what it looks like from behind. This same capacity would enable the subject to recognise the dog from many viewing angles. Noe (2004) describes this kind of ability in terms of sensorimotor contingencies. Barsalou and Prinz put flesh on what the vehicles and processes that give rise to those contingencies might be like, and I will now discuss their views in more detail.

4.3 Barsalou

I will devote the whole of Chapter 8 to discussing Barsalou’s (1999) account and Machery’s Propositional Desideratum, but Barsalou’s claims are also relevant to the Judgement Desideratum. Barsalou discusses how pieces of perceptual information (what he calls ‘perceptual symbols’) are organised together in working memory in a manner that is consistent with container theory, when a subject views a particular car:

As one looks at the car from the side, selective attention focuses on various aspects of its body, such as wheels, doors, and windows. As selective attention focuses on these aspects, the resulting memories are integrated spatially, perhaps using an object-centered reference frame (Barsalou, 1999, p. 586).
Using Haugeland's principle, we can say that the reference frame is comprised of the independent dimensions that the information is plotted against. Barsalou continues:

Similarly, as the perceiver moves to the rear of the car, to the other side, and to the front, stored perceptual records likewise become integrated into this spatially organized system. As the perceiver looks under the hood, peers into the trunk, and climbs inside the passenger area, further records become integrated. As a result of organizing perceptual records spatially, perceivers can later simulate the car in its absence. They can anticipate how the car would look from its side if they were to move around the car in the same direction as before; or they can anticipate how the car would look from the front if they were to go around the car in the opposite direction (Barsalou, 1999, p. 586).

Note that the information that is gathered in the container can be used to recognise future instances of this type of car, but it can also be used as functional information thereby allowing the subject, for example, to infer or predict how the car will look from different viewpoints. The information placed in the CAR container does not have to be conceptualised. The subject may have no concept of what, for example, a trunk is, but she can still plot the information that is about a trunk in the CAR container by plotting the information against the dimensions of the reference frame. For example, she can plot information about the trunk's appearance at the part of the reference frame that corresponds to the rear of the car. She can plot what it appears like open, and what it appears like closed. She can plot information about the location of the button that opens the trunk, and interactional information about how to open it, and so on. Using that information the subject can predict that something like this will be found at this location of the car, upon which can be performed these actions, without having to have a concept of what this is.

In Chapter 9, I discuss the circumstances in which the cognitive system might open new container/concepts. One such set of circumstances would be when the system finds it has sufficient pieces of information about the same sub-part (e.g. the trunk) to warrant the opening of a new container. This new container will then contain information that allows the recognition of trunks and the inferences that can be made about them. Once this happens, the part of the reference frame, and the information plotted against it that corresponds to the trunk, can be associated with the TRUNK container, thus conceptualising that perceptual information.
Barsalou goes on to describe how perceived event sequences can be stored, so as to be available as functional information in the CAR container:

Imagine that someone presses the gas pedal and hears the engine roar, then lets up and hears the engine idle. Because the perceptual information stored for each subevent is not stored independently but is instead integrated temporally, the perceiver can later simulate this event sequence. Furthermore, the simulated event may contain multimodal aspects of experience, to the extent that they received selective attention. Besides visual information, the event sequence might include the proprioceptive experience of pressing the pedal, the auditory experience of hearing the engine roar, the haptic experience of feeling the car vibrating, and mild excitement about the power experienced (Barsalou, 1999, p. 586).

And he notes that after processing many cars, much multimodal information will be stored that specifies what it is like to experience cars sensorially, proprioceptively, and introspectively. In other words, the frame for *car* contains extensive multimodal information of what it is like to experience this type of thing (Barsalou, 1999, p. 586).

Recognising a car in one’s visual field is a matter of the percept of the car being sufficiently similar to stored recognitional information in the CAR container. That recognitional information came from previous perceivings of cars. When recognition occurs, the container is activated, and any information in the current perceiving that is sufficiently novel or salient gets added to the container. In addition, all the information currently in the container becomes available, allowing the subject to make predictions and inferences about, for example, how to get into the car, how to access the trunk, how the engine will sound, and so on. All of this inferential knowledge *could* in principle be perceptual and unconceptualised in nature (i.e. iconic plotting of information against independent dimensions). However the information could equally be conceptualised, which would happen if, for example, the part of the reference frame in the CAR container that pertains to the trunk is associated with the TRUNK container.
4.4 Prinz

Prinz (2002) argues in favour of concept empiricism: that concepts have a largely perceptual basis. His account agrees in many respects with Barsalou's. However, one major difference is that Prinz posits that concepts are identical to what he calls proxytypes, which are not container-like; Barsalou considers concepts to be what he calls simulators (see Chapter 8 for discussion), which can be interpreted, in some respects, as being container like.

Prinz notes that psychologists take visual perception to involve cells being activated in 'a two-dimensional array not unlike a snapshot of a perceived scene' (2002, p. 139). Information is passed forward to area V1, where cells are arranged retinotopically, as was discussed in Chapter 2. Prinz considers two competing theories for how objects are recognised: a viewpoint-invariant theory (Marr 1982, Biederman 1987) and a viewpoint-specific theory (Tarr & Pinker, 1989). He notes that there is evidence to support both theories in different contexts. We can relate this distinction to container theory by saying that if the recognitional information stored in a container is viewpoint-invariant then it is plotted against a three dimensional reference frame centred on the object. If the information is viewpoint-specific then the information is plotted against a two dimensional reference frame, as seen from a particular perspective. On the latter option, the whole object is not modelled; only what is visible from the particular perspective is modelled. Viewpoint-specific information may also contain depth information, which is why Marr (1982) refers to it as a two-and-a-half dimensional reference frame.

Prinz (2002, p. 141) suggests that both theories may be involved in perception. Initially, representations are plotted from a particular perspective, but, once there are a number of representations of parts of objects, these are plotted against a reference frame of the whole object (e.g., the wheels and trunk are plotted against a reference frame of the whole car rather than just from a particular viewpoint). Prinz also argues that viewpoint invariant models can 'abstract away from details of position, scale, metric proportion, and viewpoint' (ibid., p. 144), so one should not think of perceptual representations as being just like simple pictures in the head. Information plotted against an invariant three dimensional reference frame does not have to be plotted in any particular location in space, nor does it have to be plotted from any single point of view. In addition, the scale of the object depends on what value units of the dimensions are given, and that value could be easily altered.
From the point of view of container theory, both viewpoint-specific and viewpoint-invariant reference frames allow information to be plotted in containers in a non-conceptualised form; that is, without the subject already having to possess a concept of what category of thing is being plotted.

According to Prinz, over time, as more instances of a category are encountered, a large long-term memory network of linked information is built up. Different pieces of information can be given greater or lesser weight for recognising instances of a category. He uses the example of gathering information about gorillas. Information about the whole body can be plotted against a reference frame, and then links can be used to zoom in on more detailed information about, for example, only the head of the gorilla. Links can be used to record information about how a gorilla transforms from being in a standing state to engaging in beating its chest. Links can be used to bind information from different modalities. For example, the sound made by a gorilla beating its chest can be linked to the visual image of its arm movements while doing this. Images of a gorilla eating a banana can be stored that are generated by the inferences that can be made, and so on.

Prinz notes that all of these pieces of information pertaining to gorillas can be linked together into a long-term memory network. But one big difference between Prinz’s view and the container theory (and indeed between Barsalou and Prinz) is that Prinz doesn’t see the need for containers. In Chapter 3, I set out how a functional container could be implemented by links in long-term memory. Functionally, what would amount to a container would be a node from which all links to the contained information emanated. But for Prinz pieces of information are simply linked to each other and there is nothing that corresponds to the container. So if there is no container to play the role of a concept, what does Prinz take a concept to be? He considers the idea that the long-term memory network corresponds to a concept, but notes two objections. First, he finds that it is difficult to determine where knowledge of one category begins and that of another category ends. Does my knowledge that dogs make good pets belong to my dog concept or my pet concept? (Prinz, 2002, p. 148).

In contrast, in container theory, the container is what marks the boundary between categories. Second, there is the difficulty of having long-term memory networks fulfil Machery’s Propositional Desideratum for concepts, that is, the difficulty of having long-term memory networks be constituents of propositional thoughts.
Instead, Prinz opts to identify concepts with what he terms ‘proxytypes’, which are similar to Barsalou’s perceptual symbols, and roughly the same kind of thing as perceptual exemplars, prototypes and theories as set out in Section 4.2 above. Prinz’s strategy then, for having a single notion of concept satisfy both Machery’s Judgement and Propositional Desideratum, is to say that proxytypes (pieces of information in long-term memory networks) can be used to make recognitional and inferential judgements on the one hand, but can also, on the other hand, be used as proxies/symbols for categories in propositional compositions in working memory.

According to container theory, however, Prinz is extending the meaning of concept here in the way it was argued in Chapter 3 that philosophers and psychologists do. In container theory, proxytypes would be the contents of containers and not concepts themselves. Prinz claims that proxytypes can also be used, in working memory, as concepts in propositional expressions but, in container theory, they would be symbols for concepts (i.e. symbols for containers) and not concepts themselves.

Prinz notes that ‘it seems odd to say that a concept used in working memory ceases to be a concept when it is inactively stored in long-term memory’ (Prinz, 2008, p. 149), so he argues that,

... concepts are mental representations of categories that are or can be activated in working memory. I call these representations ‘proxytypes,’ because they stand in as proxies for the categories they represent (Prinz, 2002, p. 149).

And he says, ‘proxytypes are perceptually derived representations that can be recruited by working memory to represent a category’ (ibid. p. 149). He claims that proxytypes are anything in long-term memory that could be used to detect members of a category (we can extend this to include inferential as well as recognitional information).

There are two main problems with identifying proxytypes with concepts. The first is that it doesn’t seem right that a piece of information about, say, dogs (e.g. an exemplar of a dog) is a concept of a dog. A particular way of recognising a dog does not amount to a dog concept; a particular inference about a dog (that it wags its tail) does not amount to a dog concept. Container theory claims, instead, that it is correct to say that a concept contains all the many ways one has of recognising dogs, and all the many inferences one can make about
dogs, and that it is not correct to say that just one of these ways (or even all of them together—the whole long-term memory network) constitutes the concept of, for example, DOG.

The second problem is that proxytypes, and pieces of iconic information generally, cannot, by themselves, stand for just one category. How can the system determine whether a proxytype stands for the category dog or the category pet? This problem arises for Prinz whether the proxytype is used to recognise an instance, or whether it is used as a constituent in a propositional thought. Prinz says that where the proxytype originated from may provide an answer to the question (2002, p. 151). By this, I presume he means which long-term memory network the proxytype emanates from. But, first, he has already said that there may be no way of knowing whether a piece of information counts as being in the PET network or in the DOG network. And, second, even if it were possible to decide, this would just make networks constitutive of concepts rather than proxytypes, and then there still would remain the problem he noted of having networks as constituents of thoughts. In container theory, as I will set out in the next section (Section 4.5), origin is precisely what determines to which category a piece of information pertains. I claim, in contrast to Prinz, that neither information in memory networks, nor memory networks themselves, can be constitutive of concepts. What is needed are containers to establish boundaries between one concept and another. For example, a piece of information pertains to the dog category if it is in the DOG container, and it pertains to the pet category if it is in the PET container (and the same piece of information may be in both).

4.5 The container determines the category a piece of information pertains to

In container theory, the idea is that for each concept possessed there is a container. If I have the concepts of apples and bananas then all of my stored knowledge about apples goes in my APPLE container, and all my stored knowledge about bananas goes in my BANANA container. If I attend to an object, the category I recognise that object as depends on which container provides the closest matching recognitional information. It does not depend just on which exemplar (or prototype or theory) provides the closest match, but also on which container the exemplar was located in. What is so advantageous about this mechanism is that the system can use the exact same exemplar (say an image of a banana – call it 'bananaimage') to not only recognise bananas but also to recognise, e.g., fruits or food. If bananaimage is placed in the FRUITS and the FOOD container, then bananaimage counts as
a fruit and food exemplar, as well as a banana exemplar. So if I attend to an object that happens to be a very close match to bananaimage, the category I take the object to be in will depend not on any intrinsic properties of bananaimage itself, but on which container contains the version of the image that provides the match. If the exemplar is the version of bananaimage in the FRUIT container, then I will take the object to be a fruit. If the exemplar is the version in the BANANA container, then I will take the object to be a banana. And, of course, which container is activated determines which set of stored knowledge becomes available.

This overcomes a problem with taking exemplars or prototypes to be concepts. When I perceive an object that is a banana, what determines whether I see it as a banana rather than as, say, a Cavendish, a fruit, a food, or a container of seeds? There has to be something that makes it a banana exemplar. That something extra in container theory is that the exemplar that matched the object was in the BANANA container rather than in a container pertaining to any of the other categories. So the identical exemplar will be an exemplar of a banana, if it causes the activation of the BANANA container, and an exemplar of a fruit, if it causes the activation of the FRUIT container.

What do I mean by a container being ‘activated’? I mean that the default information in the container becomes available. If I find I now have all of my stored knowledge about bananas available, then I have activated my BANANA container. And if I find all of my stored knowledge about fruit available, then I have activated my FRUIT container. Think of it like a search: I input an image of a banana into a search engine that can take as input images rather than words. If the search engine returns a webpage for BANANA (a container of stored information about bananas), then I will take the object of the image to belong to the banana category, and I will have available useful information about bananas. If the engine returns the FRUIT webpage (a container of stored information about fruit), then I will find myself considering the object of the image as a fruit, and be able to make inferences based on the stored information about fruit. It is likely that the banana page would contain a link to the fruit page, and vice versa, so that I can easily access stored information about both categories. Or, it could be that the search engine activates both webpages at the same time (one half of the screen shows the banana page, and the other half shows the fruit page), making available information about both categories at once. The key thing to note is that, in the example, webpages (containers of stored knowledge) stand for single categories, and the category pieces of information pertain to is determined by the page they are on. Potentially, the same
information could be on a number of different pages (the same exemplar image could be, for example, an exemplar of the categories *granny smith*, *apple*, *fruit*, *foodstuff*, etc.). There has to be something that determines the category a piece of information such as an exemplar pertains to, and, in container theory, that is the container in which the information is contained.

4.6 **Acquiring new concepts**

In Chapters 9 and 10, I discuss how new concepts/containers are acquired and how existing concepts refer. How existing concepts refer is complex, and requires distinguishing between private, social, and word concepts, so I will not discuss it now. However, I will give a summary here of how new concepts are acquired. According to container theory, concepts are containers. The number of concepts an individual possesses is the number of distinct containers she has. If an individual possesses the concept X, then she possesses a container that pertains to X.

Where do these containers come from? The basic principle in container theory is that whenever an entity is encountered that cannot be recognised by any recognitional information in an existing container, then the cognitive system opens a new container to gather together information about the category of this new entity. The information that could not be recognised is stored in the container, and can then be used to recognise future objects that are sufficiently similar.

Consider Barsalou’s car example above. Imagine a creature encountering a car for the first time. The visual percept of the car does not match the recognitional information in any existing container, so the creature cannot recognise it. As a result, the system opens a new container. As Barsalou describes, the creature selectively attends to various aspects of its body, such as wheels, doors, and windows. As selective attention focuses on these aspects, the resulting memories are integrated spatially, perhaps using an object-centered reference frame (Barsalou, 1999, p. 586).

In other words, this new container now contains recognitional information based on what was attended to. This allows the creature to recognise other cars that are sufficiently similar to the stored information. The container also gathers together functional information.
about how the doors are opened, and so on. This information, as I have previously claimed, may be non-conceptualised: it may just be the mechanical plotting of information against independent dimensions (a reference frame). As time goes by, and more cars are encountered, the container will become filled with many useful ways of recognising and making inferences about cars.

Note that, in opening the new container, the cognitive system did not have to know the category the perceived object pertained to, nor specify the category the new container pertained to. It just had to, in effect, regard the object as belonging to some new category.

According to container theory, the same occurs when you hear a new word. Your cognitive system attempts to recognise the word using information in existing containers, but fails. As a consequence it opens a new container for gathering information about the category of that word. The system does not have to know the category the new word pertains to; it just has to take the word to pertain to some new category.

One question that may be in the mind of the reader is what is there to prevent, for example, information about hovercrafts getting into the CAR container? What is there to prevent the container coming to pertain to both cars and hovercrafts? I will not answer this now, but promise to deal with the question in Chapters 9 and 10.

There is no requirement that a creature must already possess any concepts in order to open a new container. The new container/concept is just an empty structure for gathering functional and recognitional information that pertains to the same kind of thing. The first pieces of information placed in the container may be entirely non-conceptualised perceptual and procedural iconic information that is plotted against a reference frame. In principle, a creature could be born with no concepts at all and build up new containers from a tabula rasa. However container theory is neutral about whether any concepts are innate. Babies may, for example, be born with a container that contains recognitional information for human faces. As far as container theory is concerned, creatures could be born with a large repertoire of containers filled with recognitional and functional information. But equally, creatures could be born with no, or very few, concepts. Each new concept/container that is opened is done so on the basis of an encounter with an entity that cannot be recognised by existing containers.
4.7 Top-down or bottom-up?

On the face of it, container theory seems to tell a bottom-up story about conceptualisation. First, there is sensory perception, which, I claimed in Chapter 2, results in information being plotted against independent dimensions (a reference frame), and that information may be non-conceptualised. Second, as set out in this chapter, containers/concepts pertaining to single categories are associated with parts of percepts, thereby conceptualising them. Third, as I will claim in Chapter 6, there is propositional thought, which combines a number of concepts/containers, as associated with parts of percepts, in working memory.

Typically, the order of events will be: first, perception; then, association of concepts with parts of percepts; then, combination of associated concepts in propositional thought. But in container theory, this order is not necessary. It may be the case that top-down mechanisms are important. Take visual perception. There must be some sensory content forming the basis of what is currently seen. But it may be that I am puzzled by the visual scene, and I remain so, until I form the propositional hypothesis that a dog is chasing a cat. Having formed this hypothesis, I may subsequently recognise the objects in the scene as a dog and a cat. This recognition may, in turn, cause the system to alter how I perceive the scene based on the stored information of what cats and dogs appear like. Unlike the bottom-up model, where the order is perception, recognition, and proposition, here, the model is top-down: first, there is a propositional hypothesis; then, there is recognition; and finally, there is segmenting of the perceptual scene into objects.

Furthermore, there is no reason why top-down influences cannot have a large impact on sensory signals before they result in a perception. Whereas some top-down influences may belong to what Fodor (1983) calls proprietary systems, and so be sub-doxastic (i.e., not available for expressing propositions), there is no reason, in container theory, why information gathered in the container system cannot be brought to bear at early sensory stages, with the result that parts of percepts are already conceptualised when they are formed.

As far as container theory is concerned, both bottom-up and top-down processing (or a mixture of the two) are plausible. But according to container theory, perception, the association of concepts with parts of percepts (recognition), and the combination of associated concepts (the expression of propositions) are three distinct processes.
What I emphasise in this thesis are the different generalisations that can be made about these distinct processes. A different approach, which emphasises similarity between these processes, can be seen in the claims of predictive process theorists (e.g., Clark 2013) that a single process (Bayesian inference) underlies all three mechanisms. Similarly, Barsalou and Prinz claim that a single type of representation (perceptual) underlies all cognition. Others (e.g. Clark, 1993) posit that all cognition involves connectionist processing. Container theory can remain neutral about these claims. Cognition could be Bayesian, perceptual, and connectionist at the same time. But, the kinds of generalisation that container theory is most interested in are those that distinguish between what is required to perceive without conceptualisation, what is required to conceptualise perceptions, and what is required to express complex propositional thought.

Assume that the cognitive system engages solely in Bayesian inference. While this discovery would be of great interest, container theory would remain primarily interested in the differences between how Bayesian inference brings about (a) perceptions in which many pieces of information are plotted simultaneously, (b) recognition of the category that perceived entities belong to, and (c) the expression of complex propositions. Even if Bayesian inference underlies each of these processes, the generalisations about how each are brought about will, according to container theory, be different. The same goes for Barsalou’s claim that cognition makes exclusive use of perceptual representation. In order to give a full account, Barsalou (1999) makes different claims about how, for example, stored perceptual information is used to recognise objects in perceptions, and how stored perceptual symbols can be combined to form complex propositions.

4.8 Conclusion

I have shown in this chapter how pieces of information gathered together in containers satisfy Machery’s Judgement Desideratum, i.e., how information contained in concepts allow categorisation judgements, typicality judgements, and inferential judgements. I set out that the recognition process involves using recognitional information in containers to associate concepts with parts of perceptions (which is to make categorisation judgements), and that once a container/concept is associated with a part of a perception, the functional information in the container enables typicality and inferential judgements.
I explained, following Barsalou and Prinz, how information gets to be in containers as a result of encounters with members of a category, and how new concepts/container come about. I detailed the importance of it being the container and not the contents of the container that is to be identified with a concept. The same piece of recognitional or functional information will pertain to different categories if it is in different containers. And different individuals can have very different conceptions (stored information) while having the same concepts.

In the next chapter, I turn to the notion of a DA-finst, and explain that these are representations that keep track of the associations of concepts with percepts that were covered in this chapter. And we will see in Chapter 6 how DA-finsts, functioning as disposable symbols, explain how Machery's propositional desideratum is fulfilled.
Chapter 5

DA-Finsts

5.1 Introduction

In this chapter, I will describe Pylyshyn’s (1989) notion of a finst, and I will expand on that notion with the idea of double-armed finsts (DA-finsts). I will describe how DA-finsts keep track of the concepts/containers that are associated with percepts, and how DA-finsts can operate as disposable symbols. I will also describe experiments that support the existence of DA-finsts.

In Chapter 2, I discussed how perception was the plotting of many pieces of information simultaneously against a reference frame. In Chapter 3, I introduced the claim that concepts were containers of stored recognitional and inferential knowledge pertaining to single categories. In Chapter 4, I claimed that the recognition process typically involved (a) a perceptual object being recognised by recognitional information in a container, and (b) the functional information in that container becoming available for inference and interaction. In this chapter, I posit the existence of DA-finsts: devices that keep track of the input and output of the recognition process, where the input is the entity to be recognised, and the output is the category it has been recognised as. I will describe how DA-finsts implement selective attention – perceptions plot many pieces of information simultaneously, but a DA-finst can ‘keep its finger’ selectively on a single object from the many in the scene. I will also discuss how a DA-finst represents or symbolises an entity and the concept/container associated with it. For example, when an object is no longer visible, a DA-finst can stand in for it, or represent it.

The account that I am laying out involves increasing complexity: each chapter builds upon the previous one, and adds new functionality. There can be creatures that only have the capacity to record information iconically, and to respond automatically to recognised entities. For each recognised stimulus, there is a predefined response – for example, whenever a snake shape is detected this is responded to by the creature climbing the nearest tree. Some
recognition/detection capacity is required but not a container system. A container system adds flexibility by giving the capacity to gather in the one place (a) many different ways of recognising members of the same category, and (b) many different ways of responding (ways of inferring and interacting with) to recognised instances. Now, when a snake is encountered, the creature can select from a number of options; in addition to responding by jumping into the nearest tree, the container may contain information about when and how to make the snake flee, or how the snake can be killed.

This chapter builds on these two stages. Double-armed finsts (DA-finsts) keep track of (a) the entity in a perception that is being recognised, and (b) what concept/container provided the recognitional information. DA-finsts facilitate, for example, keeping track of which entity is best for carrying out a task when comparing a number of different options, so that that entity may be chosen once the comparison is complete; or keeping track of an item, and its recognised category, during a delay period during which the entity is no longer present; or, it allows returning to an item, after having attended to something more urgent.

Although the additional functionality of being able to keep track of entities that DA-finsts provide may not be regarded as major in itself, the true significance of DA-finsts will emerge in the next chapter. I argue that DA-finsts, in their role as disposable symbols, are the constituents of propositional expressions. DA-finsts will be central to showing how the same concept/containers that fulfil Machery’s Judgement Desideratum can fulfil the Propositional Desideratum.

5.2 DA-finsts have a subject-predicate format

DA-finsts are working memory devices based on Pylyshyn’s notion of a finst. They have two arms that keep track of the input and output of the recognition process. Recognisings, paradigmatically, have a subject-predicate format: the subject matter is the entity to be recognised (the input to the process), and what it is recognised as is what is predicated of the entity (the output of the process). I can recognise the same subject matter in many different ways; that is, I can predicate many different categories of the same input. I can recognise the same visual percept as a dog, or as an animal, or as friendly, and so on. To recognise something is to predicate a category of it. So in any recognising there are always two strands to the recognition: a) the subject matter and b) the predicate.
DA-finsts are posited as devices that keep track of these two strands in recognisings. The finger of one arm keeps track of the subject matter, and the finger of the other arm keeps track of what category is predicated of the subject matter. If I recognise a particular part of my visual perception as a dog, a DA-finst is capable of keeping track of (a) where in my visual perception the percept of the dog is currently located, and (b) which container/concept provided the recognitional information (e.g. whether it was the DOG container or the PET container).

5.3 Subject-predicate form and propositional thought

A propositional thought is one that is evaluable as true or false. Its simplest form is a subject-predicate statement. Smith (2007) notes that, ‘In *On Interpretation*, Aristotle argues that a single assertion must always either affirm or deny a single predicate of a single subject’. Subjects and predicates are also central to Evans’ (1982) Generality Constraint, which claims that if one can think that a is F, and b is G, then one can think that a is G, and b is F.

Neither subject matters nor predicates are true or false in themselves, but a category predicated of a subject matter can be evaluated, or marked, as true or false. Concepts such as New York, or Mary, or dog are not markable as true or false on their own; nor or entities that happen to be New York, Mary, or a dog. But asserting that *That [thing or location] is New York*, that *That [thing or person] is Mary*, or that *There is a dog [in front of me]* are markable as true or false. What changed? There was added an instance to the concept or a concept to the instance. The term ‘New York’, by itself, simply refers to (is a symbol for) the category *New York*, and, by itself, is not true or false. But by adding an instance that is being claimed to be New York (the instance can be a photograph, the city itself (e.g., you are looking at it from an aeroplane window), or a marker on a map), one forms a claim or statement that is markable as true or false (it is either true or false that the photograph, city, or marker on the map is New York). DA-finsts are recordings of pairings of instances with categories, and so are markable as true or false. The instance captured by the subject arm either is, or is not, a member of the category captured by the predicate arm. So, in my terms, a DA-finst captures a thought (such as *that is a dog*).

And a system with the ability to think about, or take attitudes towards, its DA-finsts is, in effect, thinking about, and taking attitudes towards, expressions of propositions. If a
DA-finst points with its subject arm to a visual percept of a dog, and with its predicate arm points to the DOG container, then a system with the appropriate capacities can believe, or wonder, or doubt the proposition that That is a dog. For a creature with such capacities, a DA-finst operates as a mental object that it can think about; a mental object that expresses a simple proposition (such as that That is a dog). To sum up, I claim that DA-fints are concrete mental objects that express simple subject-predicate propositions, towards which attitudes can be taken.

5.4 DA-fints are disposable symbols

I claim that DA-fints are disposable symbols, and, in Chapter 6, I will claim that they can be combined to form complex propositional expressions. For now, I will just set out what I mean by ‘disposable symbols’. DA-fints are not just symbols for things/entities, or just symbols for categories. DA-fints are symbols for concepts predicated of entities. DA-finst symbols inherit their meaning from the representations that they are currently pointing to. But DA-fints only retain their meaning while their arms point to something. As soon as they stop pointing, they dispose of their meaning, or, if they point to another object and container, they inherit a new meaning.

5.5 What are finsts?

A number of experiments led Pylyshyn (1989) to posit his single-armed finst theory. Finst is short for ‘finger of instantiation’; Pylyshyn also uses the term ‘visual index’ (Pylyshyn, 2001). Finsts, as Pylyshyn conceives of them, are pointers or indexes that keep track of percepts of specific objects in the visual field, without regard to categorical or featural information. The idea is that a finst is like keeping a finger pointed at an object in the world. What singles out the object is not any information about it, such as what category it belongs to, or what features it has; rather, what singles out the object is just the fact that the finger is pointing at it.

You could keep track of something by noting to yourself that it is red and round. You could then attend to some other object and subsequently re-identify the first object by finding something before you that is red and round. Of course, this won’t be helpful if the object has, in the meantime, changed shape and colour. An alternative method is just to keep your finger
pointed at the object, without worrying about its category or features. That way, even if the object is moving about in the scene, and even if it is changing shape and colour, you keep track of it. According to Pylyshyn, this is how finsts keep track of an object in a perception: by, metaphorically, keeping a finger pointed at it. Pylyshyn proposed his theory to account for the results of multiple object tracking experiments carried out both in his laboratory (e.g., Pylyshyn & Storm 1988, Sears & Pylyshyn 2000), and in other laboratories (e.g. Intriligator & Cavanagh 1992, Yantis 1992 – for a review see Section 3.2 of Pylyshyn, 2001). These multiple object tracking experiments require subjects to keep track of objects moving around on a screen. For example, there might be ten identical circles on a screen. Four are highlighted as the ones the subject must keep track of. Then all ten circles start moving for a couple of seconds. When they come to rest, the task for the subject is to identify the four circles that were highlighted at the start. Clearly, the subject cannot keep track of the four circles with categorical or featural information, for all of the circles are identical (they are all circles of the same size and colour). What is found is that subjects can keep track of around four such objects at once, but, as the number of objects increases beyond four, the task becomes ever more difficult. Pylyshyn claims that the increased difficulty is explained by the fact that the cognitive system has four finsts at its disposal. When the number of objects to be tracked is four or less, subjects can easily perform the task by allocating one finst to each moving object. But, once the number of circles to be tracked exceeds the number of available finsts, it becomes very difficult to keep track of all of the moving objects.

Trick and Pylyshyn (1994) also claim that finsts are relevant to the phenomenon of subitizing. If up to four objects are briefly flashed on a screen, it is easy to report how many objects there are. Beyond four and the task becomes increasingly difficult. Pylyshyn claims that the system can easily report up to four objects by the simple means of counting how many finsts were assigned to the different objects. If four finsts captured objects, then there were four objects. But, once the number of objects goes beyond four, there are not enough finsts available.

5.6 Extending finsts: making them double armed

I support Pylyshyn’s finst theory, and his explanation of how subjects can keep track of, e.g., four out of ten identical circles. However, I propose that his theory can be extended, so that the finsts are double-armed. As in Pylyshyn’s theory, one arm (call it ‘the subject
arm') is devoted to capturing objects without regard to categorical or featural information. The second arm (call it 'the predicate arm') is dedicated to capturing the category that the object belongs to, or the category of the features it has. The idea here is that the information captured by the subject arm is just the kind of information that can be used as input to the container system in the recognition process described in the last chapter. While, in the tasks that arise in Pylyshyn's experiments, there is no need for the subject to determine the category that the object belongs to, in most everyday situations it is beneficial to know what category an object belongs to. That is what allows one to draw good inferences and interact appropriately with the object.

In container theory, the recognition process, in its paradigmatic form, is a two stage process. First, there is singled out an object to be recognised, and information about how it appears is input to the process. Second, there is singled out a category that the object is recognised as. This is the output stage of the process. This output is the concept/container that contains the closest matching piece of recognitional information (e.g. an exemplar or prototype) to the input. I claim that the subject arm of the DA-finst points to and keeps track of the input (what the object appears like, and where it is in the perceptual field), and that the predicate arm of the DA-finst points to and keeps track of the output (the container of stored knowledge pertaining to a particular category).

Sometimes the task before us does not require the predicate arm to be filled and only the subject arm captures an object, as in Pylyshyn's experiments. But, in the normal course of events, we do go on to recognise the category that an object belongs to, and the claim is that, when this occurs, the predicate arm of the DA-finst points to the container that contained the recognising information.

5.7 The subject arm and the predicate arm

The two arms of DA-finsts have very specific purposes. The first arm – the 'subject' arm – captures the subject matter of the representation (the particular object or entity or instance that is being thought of), and performs just the way Pylyshyn sets out. The subject arm keeps track of a particular entity without having to recognise the category that the entity belongs to or the features that it has. The second arm – the 'predicate' arm – captures a container, and so predicates the category of the container to the subject-matter (e.g. is a banana) or predicates a feature (e.g. is yellow).
If I selectively attend to a banana on the table in front of me what happens, according to container theory, is that the visual percept of the banana is captured by the ‘subject’ arm of a DA-finst together with information about the current location of the perceived object in the perceptual field. The subject arm essentially means ‘that’, where ‘that’ is the percept of the object being pointed to and by extension the object itself. The information in the subject arm is then compared to recognitional information in containers. When a sufficiently close match is found, the perceived object is recognised as belonging to the category of the container that contained the recognising information, which in this instance is the BANANA container. The predicate arm of the DA-finst then points to the BANANA container. If, instead, the recognising container had been the FRUIT container, then the predicate arm would have pointed to that container. Depending on what my current task or thought involves (e.g. ‘find a yellow object’ or ‘do you have any yellow fruit’), I can see the ‘that’ as just one of these categories, or as all three. The point of predicating a category is of course that it makes available the information in the container, so that useful predictions, inferences, and interactions can be made. If I activate the BANANA container then I make available my default stored knowledge about bananas which gives me understanding of what I am thinking of. What the predicate arm of a DA-finst adds is that it keeps the container active over time.

An important purpose of finsts is to keep track of things over time. With Pylyshyn’s single armed finsts the system keeps track of an object regardless of its categorical information. But double armed finsts have an extra arm dedicated to also keeping track of categorical information. The subject arm of a DA-finst captures a percept of an object. Once the percept has been recognised as, say, a banana (because the BANANA container contained the closest matching recognitional information), the predicate arm of the DA-finst is set to point to (keep its finger on) the BANANA container. Now the DA-finst tracks an object for the agent, and also tracks a container of stored knowledge. The information in the container allows the agent to make inferences about and interact with the object that has been recognised.

Imagine that you are hungry and you spot a banana across the room. A DA-finst captures the percept, and its location, with its subject arm, and captures the BANANA container with its predicate arm. You plan to eat the banana, but, first, you must spend a few seconds completing a task. As long as the DA-finst continues to point to the percept and the container of the object, then, as soon as you have completed the task, the system can use that DA-finst to bring your full attention back to the banana, so that you can pick it up, with the
full knowledge that it is a banana. The stored information in the BANANA container will inform you of how to pick up a banana, how to peel it, what it tastes like, and so on. The point is that, without having kept track of both the object and the container, the object would have to be recaptured and/or re-recognised all over again. We need something like DA-finsts to explain how we can keep track of goal objects during delays or distractions.

Note that during the delay period ‘I’ am not selectively attending to the banana (‘I am attending to finishing off another task), but my cognitive system is. Out of all the containers it could be keeping active, it is selectively keeping the BANANA container active. And out of all the objects and locations it could be selectively keeping a track of, it is keeping track of the object at location X.

Of course, often we find that the DA-finst has not been maintained. We may find ourselves coming into a room and saying ‘What did I come in here for – I forget now.’) This is, I will suggest below, because it is effortful to maintain a loop between the finst and what it is pointing to; the loop has to be actively maintained.

5.8 Single cell recording DMS experiments show recurrent loops that have the features of both arms of a DA-finst

I have talked a lot about how finsts ‘keep a finger on’ the entities they are tracking. I now want to evaluate experiments that suggest that ‘keeping a finger on’ is implemented by a form of recurrent or feedback loop processing (Elman 1990, Hopfield 1982, Zipser et al. 1993). These experiments support the idea of finsts having two arms, one devoted to ‘where’ information, and one devoted to ‘what’ information.

There is much evidence that working memory centres on the prefrontal cortex (pfc) (Baddeley 1996, Fuster 1989, Goldman-Rakic 1987, Miller & Cohen 2001, Smith & Jonides 1999). Working memory in computation is a dedicated location where the system can actively work on (manipulate – lay out, transform, rearrange, add, delete, combine, etc.) information in order to carry out tasks, solve problems, specify what the current state is, specify what the state ought to be, and so on. Working memory can be regarded as a temporary workspace in that, once the current task is finished, the information in the workspace is not retained, and will be overwritten by the information involved in the next task. Working memory differs from long term and short term stored memory in that its
contents are active, that is, currently in use. Stored memories are often loaded into working memory and operated on. The idea is that, when you are occurrently thinking, you are doing so in working memory; when you store what you have thought, you do so in short term or long term memory. Short term memories are available for a short time before fading away; long term memories are available for a long time, potentially a lifetime. When you remember what you have thought, you retrieve that memory from long or short term memory and make it active in working memory.

A number of theorists have found by performing single cell recording that prefrontal cortex (pfc) is vital to working memory and attention (Miller et al. 1996, Fuster 1995, Goldman-Rakic 1995, amongst others). The role of pfc is taken to be one of control and tracking. These theorists posit that, by and large, content is not passed forward to the pfc; rather the content information is kept in other dedicated areas of the cortex. For example, perceptual iconic information about what things appear like is maintained in the various sensory cortices (visual, auditory, etc.). Conceptual information is maintained in associative or inferotemporal cortex (ite). And spatial information about the location of perceived objects is maintained in post-parietal cortex (ppc). The claim is not that each of these kinds of information can only be found in these areas; the claim is just that often these kinds of information will be found in these areas.

As I have said previously, my concern in this thesis is not to make claims about neural localisation, but to make claims about cognitive functional architecture. In what follows there are localisation claims, which I tentatively support, but, as far as container theory is concerned, it is only the functional architecture that matters.

Consistent with what these single cell recording experiments find is that pfc organises, controls and tracks information in other areas by means of recurrent or re-entrant loops. If a network of neurons in pfc is attending to, or keeping a finger on, a particular perceptual object, it does so by activating the network that constitutes the percept of that object in, say, visual cortex. In turn, that visual cortex network reactivates the pfc network, which reactivates the visual cortex network, and so on in a continuous loop (the re-entrant or recurrent loop). So long as the loop remains in place the network in pfc 'keeps its finger on' the perceptual object (without needing any categorical information). In other words the sustained activity implements a finst.
Most of these experiments involve the experimenter recording the activity of single cells in various regions of a monkey’s cortex during delayed match to sample tasks. First, the subject will be presented with a sample (e.g. a red circle), which then disappears. After a delay, the subject is presented with a range of options, and, to get the reward, it must indicate which one matches the sample. In some experiments the task is to indicate the location of the sample (regardless, e.g., of its colour or shape), and in other experiments the task is to indicate the colour or shape of the sample (regardless of where the sample was shown). To solve these tasks, the subjects must be able to retain information about the sample during the delay period. To ascertain how this is achieved, the researchers examine the firing rate of single cells in pfc, ppc and itc during the delay period.

Ungerleider & Mishkin (1982) claim that information regarding where an object is located is maintained in a region of posterior parietal cortex (evidence for this is, for example, that patients suffering from visual neglect have damage in this region of ppc (Vallar, 1998)). And they claim that information regarding what an object is, is maintained in inferotemporal cortex (‘itc’)/associative cortex. A very similar idea comes from Milner & Goodale (1995). They distinguish between the dorsal stream (which coincides in part with ppc) and the ventral stream (which coincides with itc/associative cortex). Milner & Goodale propose that the dorsal stream is associated with the guidance of action, and the recognition of where objects are, while the ventral stream is associated with the recognition of the category and form of objects. Goldman-Rakic (1995) and others (e.g., Constantinidis & Steinmetz 1996, Funahashi et al. 1989), recording the activation of single cells in monkeys during delayed match to sample experiments, find that if a task requires maintaining information about where an object is located, then the activations are consistent with a loop being recurrently maintained between a pfc network and a posterior parietal cortex network. Fuster and Alexander (1971) and Miller and Desimone (1994) find that if a subject is engaged in a task that requires, e.g., re-identifying the colour or shape of an object, then the loop is between a pfc network and a network in infero-temporal cortex (itc; also known as associative cortex).

So these single cell recording experiments support the idea that during a delay period networks in working memory can engage in top-down control of information about either what or where an object is during a delay period. During the delay, when the object was no longer perceptible, what maintained the ‘what’ or ‘where’ information needed to succeed in
the task was sustained activity (recurrent loops) between pfc and the appropriate part of the cortex that contained the necessary information.

These recurrent loops have just the characteristics required by finsts that keep a finger pointed at something. The activity in the pfc network is, in effect, a finst that keeps track of (keeps a finger on, points to) the 'what' or 'where' content. Pylyshyn's one armed finsts could be fully explained by loops between pfc and ppc that maintain the where information.

Recall the multiple object tracking experiments, which required keeping track of four out of ten circles moving about on a screen. Pylyshyn proposed that finsts were what explained the ability to carry out this task. These finsts could be implemented by networks in pfc recurrently looping to locational information in the ppc map. Each finst network in pfc 'keeps a finger' on the locational information in the map by maintaining a recurrent loop. This looping activity between pfc and ppc corresponds to the 'subject' arm of a DA-finst.

But note that loops between pfc and itc are exactly what are required for the 'predicate' arm of a DA-finst. The principle is exactly the same for maintaining 'what' information during the delay period, as it is for maintaining 'where' information, except that the loop is now between pfc and itc, rather than pfc and ppc. What is required to be maintained is the predicative information that the object 'is red', for example, or 'is circular'. If we posit that itc is where much container information is to be found, and that it is knowing what container is activated is what allows recognition of 'what' category the object belongs to, then we can see how these experiments are compatible with the idea of DA-finsts having both subject arms and predicate arms. The idea is that the pfc finst network can simultaneously maintain two loops, one to ppc (the subject arm), and the other to itc (the predicate arm).

As we will see in the next section, for many tasks it is not sufficient just to know 'where' an object is, or 'what' an object is. It is often necessary to have both these pieces of information simultaneously bound together. DA-finsts achieve this binding of information: having two arms emanating from the one finst binds the 'what' and 'where' information together.

Note that there are 3 things that DA-finsts can keep track of: (a) what object (what it is like perceptually), (b) where the object is located in the perceptual field, and (c) what category the object belongs to. But we just have two arms for the finst. The predicate arm, of course, keeps track of (c). That leaves the subject arm to keep track of (a) and (b) which can
be implemented as follows. The subject arm of the finst is recurrently looped to the ‘where’ map in ppc, and the where map links, in turn, to the perceptual information in the appropriate sensory cortex. So, say a finst is keeping track of a running dog: the subject arm will loop to the ppc map, and it is the job of the ppc map to link to, let us say, the part of V1 of visual cortex that corresponds to that map location, and contains information about what the object at that location appears like. So, by the subject arm looping to the ‘where map’, the system is able to retrieve information about what the object at that location appears like.

5.9 Delayed-match-to-sample (DMS) experiments and DA-finsts

In this section, I will set out how delayed match to sample experiments can be explained by monkeys having double-armed finsts. First consider a task where the subject just has to retain information about what colour the sample had. Say the sample on the screen is a red circle, and it disappears after a few moments leaving the screen blank. After a short delay, the monkey will be presented with a number of different coloured circles on the screen, and it must select the previously shown sample colour. DA-finsts can explain how the monkey succeeds at the task. When the sample first appears, the subject arm of a DA-finst captures (forms a recurrent loop to) the percept of the red circle. This capturing allows it to input this perceptual information into the container system. The system finds matching recognitional information in the RED container, and this results in the predicate arm of the DA-finst capturing (recurrently looping to) the RED container. During the delay period, when the circle is no longer visible, the predicate arm of the DA-finst continues to track (keep a finger on) the RED container, thanks to the recurrent loop between the finst in pfc and the container in itc. Then, when the range of options appears on the screen (e.g. a red circle, a blue circle, and a yellow circle), the system just has to compare the perceptual information from each of the samples with the recognitional information in the RED container, until it finds a match.

If the task is to match the location of the sample (e.g. top left or bottom right), rather than the colour of the sample, now it is the subject arm of the DA-finst that will hold the information. Again, the monkey’s cognitive system captures the percept of the sample circle on the screen with the subject arm of a DA-finst, and the subject arm maintains the location al information by virtue of a recurrent loop between pfc and the part of the ppc map that corresponds to, say, the top left of the screen. To succeed at this task no categorical
information is necessary, so it doesn’t matter if the subject recognises the shape or colour of the object. The system just has to keep its metaphorical finger on the ‘where’ of the object. During the delay period the subject arm continues to track the locational information because of the recurrent loop. When the options are presented on the screen (e.g. the options are each of the four corners of the screen), the subject will succeed in the task by the cognitive system matching the where information maintained in the subject arm of the DA-finst with the location of one of the options on the screen.

Both of these experiments required just one of the two arms of the finst to maintain a loop during the delay period, but we can easily imagine an experiment where in order to be rewarded the subject must correctly specify both categorical and locational information. For example, say the options can be red, green or blue circles, and the circles can be either on the left or right of the screen. If the sample was a red circle on the left of the screen, the subject, in order to succeed, must indicate both that the circle was red and that it was on the left. With DA-finsts, this task is easily accomplished. When the sample is shown it is captured by the subject arm which also maintains its locational information. As described above, the predicate arm captures the activated RED container. After the delay, the subject just has to match the locational information in the subject arm, and the colour information in the predicate arm, to one of the options on the screen.

5.10 Chimpanzees outperform humans

Let us turn to an interesting experiment where chimpanzees can outperform humans (Inoue & Matsuzawa, 2007). What is significant about these experiments is that, in order to succeed, the subject needs both ‘where’ and ‘what’ information bound together. In the experiment a number of symbols are flashed on the screen for less than a second, and then replaced/‘covered’ by white squares (see Figure 2). The task is to point to the squares following the numerical order of the covered up symbols ‘under’ the squares. To succeed the chimpanzee requires both categorical and locational information and this kind of information is precisely what the two arms of DA-finsts keep track of.
Figure 2. (Taken from the BBC program, ‘Super Smart Animals’)

Note how, if the task had been to indicate which four of, say, the ten numerical symbols were flashing, this could be accomplished by the system capturing the four flashing symbols with the subject arms of four finsts. There would be no need to recognise what numbers the symbols represented, and, so, no need to fill the predicate arms of the finsts. All that would be required is the locational information, and that is given by the subject arm of the DA-finsts.

However, in the ordering task, recognition is essential. The task cannot be completed without recognising which category the different symbols represent. It is not necessary to have full-blown concepts of oneness, twoness, threeness, and so on. But, as a minimum, it is necessary to have the knowledge that ‘1’ comes before ‘2’, which comes before ‘3’, etc.; or, more basically, the interactional information that you point to the square that had contained the ‘1’ symbol before the one that contained the ‘2’ symbol, and so on. According to container theory, for the participant to succeed, it is necessary to recognise the symbols with recognitional information from containers, so as to make available the interactional functional information in the container. This, as was set out in the previous section, is a matter of capturing, with the subject arm of a DA-finst, the percept of a particular symbol (e.g. the symbol ‘1’), inputting that perceptual information into the container system, and outputting the container that contained the closest matching recognitional information. And the output container is captured by the predicate arm of the DA-finst.

For a human this container would, most likely, contain lots of information about the category one, which can be used for inferring and interacting; that is, it can be flexibly deployed in many different everyday tasks. But, for a chimpanzee it may be the case that its ONE container was created solely for the experimental task (following a long training regime), and the container may contain no information other than recognitional exemplars and prototypes, and the interactional information that this symbol is to be indicated first.
Recognition of the symbol allows the predicate arm of the DA-finst to now point to the ONE container (by virtue, I have suggested, of a recurrent loop between pfc and itc). The subject's other DA-finsts, similarly, capture percepts of symbols and the container/concepts the symbols have been recognised as.

To account for what is going on in this experiment, the picture is something like this. The symbols are flashed on the screen. Each of the subject's DA-finsts captures a different symbol with its 'subject' arm (the subject arm points to a location in the ppc map which in turn points to a location in sensory cortex, e.g., a location in area V1, where the percept of the object is recorded). Those percepts of symbols are input in parallel to the container system and allow the 'predicate' arm of each DA-finst to capture the container that contains the closest matching recognitional information. Once the symbols are 'covered' by white squares, the participant must indicate the squares in the correct order to succeed. With DA-finsts it is easy to see how this is done. The predicate arms track what the (now hidden) symbols have been recognised as. The system has to, first, find the DA-finst which points/loops to the ONE container, and then use the locational information in the 'subject' arm of that same finst to point to the right square. Then repeat that process for each of the other DA-finsts in the right order.

What is fascinating about this experiment is that humans can easily do the task for four numbers but rarely get past six numbers, whereas some chimpanzees can regularly get ten right. Human performance is what we would expect from having four finsts as Pylyshyn maintains. This might suggest that chimpanzees (or at least the chimpanzee involved in the experiment) have around ten DA-finsts.

Experiments involving delays show how valuable DA-finsts are. Whereas Pylyshyn's finsts can keep track of just 'where' information, DA-finsts keep track of both 'where' and 'what' information bound together. It would be no good in the above task to simply capture and recognise the ten symbols. After they are 'hidden', it is necessary to know where each recognised symbol was located.

5.11 DA-finsts and Sperling's experiments

Next, I turn to how well DA-finsts explain the results of Sperling's (1960) famous experiments. Three rows of four letters are very briefly (50ms) flashed on a screen. If subjects are asked to simply report as many letters as possible, they are generally able to
report about 4 or 5, which is exactly what we would expect, if a subject has about 4 DA-finsts that they can simultaneously deploy. In order to report what letters are present, it is necessary to recognise the symbols on the screen. The procedure is as before: the subject arm of each DA-finst captures a percept/image of one of the letters, and the predicate arm captures the container that contained the closest matching recognitional information. The subject can then report what letters she saw, even after they are no longer visible, by consulting the containers that the predicate arms of the DA-finsts have captured. Because, as Pylyshyn claims, we have only around four DA-finsts, subjects can, typically, only report four or five letters, despite insisting that they can see all of the letters at once. The explanation is that just seeing the letters does not tell one to which category they belong, so one cannot report that fact until one has carried out the recognition process. And that cannot start until a percept of a letter has been captured by the subject arm of a DA-finst. The fact that subjects can sometimes report more than four letters is explained by some of the letters being recognised speedily, thus allowing the reallocation of DA-finsts to other, so far, uncaptured percepts of letters, which are then also recognised speedily.

There is of course nothing surprising about this full report condition. This pattern is found in many experiments. What is surprising, however, is the partial report condition. Once again three rows of four letters are displayed on the screen for 50ms. However, now, after a short delay (and after the letters were no longer displayed), a cue is given telling the subject which of the three rows is to be reported on. It was found that, no matter which row was indicated, the subject could accurately report an average of about three of the letters in the cued row. This suggests that an iconic image of the letters on the screen persists in memory for a short time after the letters have disappeared.

Note here that I am taking this persistence to be just informational. The subjects do not claim that the letters are visible on the screen, after they have, in fact, disappeared. What persists is the information that a DA finst can capture with its subject arm, and, so, input to the container recognition system. It should be no great surprise that there are such persisting memory icons. In vision they would be very useful, for example, in determining the direction and speed that objects are moving and at what speed. Psychologists claim that visual ‘iconic’ memory lasts for about a second (Greene, 2007). It is likely that there are a number of iconic memory buffers, so that there will be a series of say five or more icons with 200ms gaps, with the oldest one being overwritten. In the case of audition, ‘echoic’ memory is found to last for about 4 seconds (Carlson, 2010). Subjects don’t just recall what happened 4 seconds ago;
they can recall all parts of the sound up to four seconds ago. This is useful for recalling, for example, what someone has just said, if you weren’t properly attending, or for experiencing just how a piece of music is flowing. Once again, it seems plausible that the entire auditory field is stored in memory every couple of hundred milliseconds or so, so that the system has more than just the current stimulus to refer to. Indeed this fits well with the work of Husserl (1964) and James (1890) on the perception of time.

The fact that the subjects can report an average of three letters from any row cued after the letters have disappeared from the screen means, either, that the subject already had recognised the categories of at least nine of the twelve letters before the cue, or else that an informational image persists in memory after the letters have disappeared, which the subject arms of DA-finst can use to capture objects. The former explanation seems unlikely. It is computationally complex to recognise the categories that entities belong to. It would be a very strange system that could recognise the categories of far more entities in a perceptual field than it could report. According to container theory, the explanation for why the subject can only report three or four characters is that it requires the subject arm of a DA-finst to capture a percept before it can be recognised by the container system, and there are a limited number of DA-finsts.

In the Sperling partial report paradigm, the letters are flashed on the screen. The available four DA-finsts capture, more or less randomly, four of the twelve percepts of letters, and these are input in parallel to the container/recognition system. Then the cue is heard informing the subject which row to attend to. Now, the cognitive system must reallocate any DA-finsts that have not captured letters in the correct row, thus terminating the recognition process for the letters those finsts had previously captured. These reallocated DA-finsts access the iconic buffer, which contains the perceptual iconic information about the letters that persists for a short time after the letters have disappeared. Because of the delay in capturing the letters from the correct row, often not all letters are recognised before the iconic memory fades. The longer the time before the cue for the row is given the less letters are recognised.

The reason we can only report around four letters, despite reporting that we see all of twelve of them, is a consequence of the number of DA-finsts we have. If we had more finsts (like the chimpanzee discussed in the previous section), we would be able to report more letters. We can recognise four things in parallel with four DA-finsts, but must wait for them
to become free again before more items can be recognised. Recognition is usually very quick, but it takes far longer than just recording information in sensory buffers. It is likely that we have four auditory finsts, as well as four visual finsts, as demonstrated by dual task interference experiments. These experiments find that an auditory memory task is less affected by a simultaneous visual task than by a simultaneous second auditory task, and vice versa (Wickens, 2002).

5.12 Change blindness and inattentional blindness

The claim that we have a limited number of DA-finsts also explains the results of change blindness and inattentional blindness experiments. For example, in one experiment, the task is for a subject to watch a video of basketball players, and count the number of passes being made (Simons & Chabris, 1999). Many of the subjects fail to notice that, halfway through the clip, a person in a gorilla suit wanders into the foreground of the scene. These subjects are surprised when they are shown a replay. According to container theory, the explanation is that all of the subject’s DA-finsts are allocated to the task of keeping track of the ball, the nearby players and the current pass count. That a gorilla has appeared on the basketball court is, of course, the kind of surprising occurrence that we expect to attend to as soon as it happens. But, we can only recognise that it is a gorilla after we have allocated a DA-finst to the percept of the gorilla and recognised it as such. Because of the ongoing complexity of the pass counting task, there are no free DA-finsts.

In change blindness, a subject is shown a picture followed by a brief mask, and then the picture is shown again, but now with a substantial change in it (Rensink, O'Regan, & Clark, 1997). The two images can be shown repeatedly for the subject to compare, provided there is a mask between each showing, and it can take subjects a long time to discover the change. The explanation involving DA-finsts is consistent with Dretske’s (2004) explanation of this phenomenon. As Dretske points out, if there is no mask between the two images, the eye is immediately drawn to where there are light transients. But, with the mask, these transients are no longer noticeable, and the subject must fall back on his recognitional system. He must recognise what the change was. According to container theory, he must selectively attend to different parts of the image and recognise the category of the entity present at that location. Then, when the second image appears, he checks to see if the entity at that same location belongs to the same category. In DA-finst terms, he captures the location with the
subject arm of the DA-finst, and he captures the container for the category with the predicate arm. When the second image appears, the system uses the information in the subject arm of the DA-finst to go to the same location in the picture, and the information in the predicate arm to test if the same category of thing is present there. Given that we can only allocate four DA-finsts at a time, it is matter of luck as to when the changed item is captured by a free DA-finst. Consequently, it often takes a long time before the change is spotted.

5.13 Conclusion

We often have to know (keep track of) ‘where’ something is as well as ‘what’ it is, in order to succeed at many tasks. Something must bind the ‘where’ and ‘what it is like’ information with the ‘what category’ information. DA-finsts with their subject arm and predicate arm naturally bind this information. They bind together the input and output of the recognition process: the subject matter that is to be recognised, and the predicate that is the category it is recognised as.

The evidence from single cell recording DMS experiments is consistent with the idea that DA-finsts are implemented by recurrent loops between top-down finst tracking devices in prefrontal cortex, and ‘content’ information in, for example, post parietal cortex, infero-temporal cortex, and sensory cortex.

In order for the system to recognise the category of something in the perceptual field, the system, typically, must first capture the percept of the object with the subject arm of a DA-finst. The percept is input to the container system, and the container that contains the closest matching recognitional information is output by the system. That container is captured by the predicate arm of the DA-finst. In order for the system to keep track of what the percept has been recognised as, the system has to maintain recurrent loops between the finst and what its two arms are pointing to. The small number of DA-finsts we have explains the results of many experiments in which we can easily notice around four pieces of information at once, but struggle when the task involves more than four pieces of information.

While Pylyshyn’s work revolves around capturing entities that require no categorical information, we have seen from the experiments above that the ‘around four’ rule applies equally to tasks that do require categorical information, which strongly supports the idea that finsts are double armed.
I have also set out how DA-finsts can be taken to be disposable symbols, and how they can be regarded, because of their subject-predicate form, as expressing very basic propositions. In the next chapter, I will examine these issues in more depth, for I shall be claiming that DA-finsts can be strung together, as disposable symbols, so as to express complex propositions.
Chapter 6
The Propositional Desideratum and DA-finsts

6.1 Introduction

This chapter sets out how I propose Machery’s Propositional Desideratum for concepts is satisfied. In Chapter 3, I considered Machery’s argument that psychologists and philosophers mean entirely different things by the term concept. Psychologists take the term to mean that which satisfies the Judgement Desideratum (a concept must permit us to make categorisation, typicality and inferential judgements). And philosophers take the term to mean that which satisfies the Propositional Desideratum (a concept must be capable of being used as units or constituents of propositional thoughts). Machery claims that these are two entirely different notions of concept, which use the same polysemous term. I argued that a single notion of concept (namely the notion that a concept is a container of stored knowledge pertaining to a single category) can satisfy both desiderata. I claimed that psychologists extend the term to mean the contents of the concept/container (as when we extend the meaning of DVD to mean the movie on the DVD), and in Chapter 4 I set out how the contents of containers satisfy the Judgement Desideratum. And I argued that philosophers extend the word concept to mean the symbol of the container (as when we refer to a title of a DVD as a DVD). In this chapter, I will set out how the DA-finsts described in Chapter 5 symbolise containers/concepts, and how they come to fulfil the Propositional Desideratum.

6.2 DA-Finsts and propositions

Propositional statements make claims. They assert something to be the case in a form that can be evaluated for truth, i.e. marked as either true or false. In Chapter 2, I noted that perceptions do not express single propositions, because a perception always represents many different propositions simultaneously. A perception, just by itself, cannot be marked as true
or false; it is necessary, first, to state a proposition about the perception. In the last chapter, I set out that the simplest form of proposition has a subject-predicate form, and that this form is mirrored by DA-finsts' subject and predicate arms. Consequently, I argued, DA-finsts can be regarded as expressing simple propositions about perceptions. For example, a DA-finst whose subject arm picks out a visual percept of a dog, and whose predicate arm picks out the DOG container, is, in effect, making the claim that *That is a dog*, which is either true or false. That object either does or does not belong to the category dog.

DA-finsts, in themselves, do not have any meaning, nor do they symbolise anything. However once their arms, and while their arms, are pointing to percepts and concepts/containers, DA-finsts inherit their meanings from the representational vehicles they are pointing to. I call DA-finsts 'disposable symbols' because they dispose of their meanings once their arms cease pointing. The same DA-finst can one moment be symbolising dogs, and the next moment be symbolising cats. Which concept the DA-finst is symbolising depends on which container its predicate arm is pointing to. In contrast, the word 'dog' persistently stands for/refers to the same category.

DA-finsts, by themselves, can be seen as expressions of simple propositions; simple because they only require one DA-finst and one container/concept to express the proposition. This chapter, however, concerns the expression of complex propositions, which require more than one concept/container and more than one DA-finst to express. Machery's Propositional Desideratum states that a concept must be capable of being used as a unit or constituent in propositional thoughts. In this chapter, I will claim that concepts, viewed as containers of stored knowledge, are not the direct units or constituents of propositional thoughts. Concepts/containers are not the entities directly manipulated in working memory. Rather, the direct constituents of propositional thoughts that are manipulated in working memory are DA-finsts: disposable symbols, whose predicate arms point to (symbolise, in effect) concepts/containers. Concepts are constituents of these propositional thoughts, but via DA-finsts, which operate as symbols of the containers.

DA-finsts can be combined as constituents of propositional thoughts. If one DA-finst is tracking an object that has been recognised as a dog (in effect asserting that it is true of the perception that *That object is a dog*), and another DA-finst is tracking an object that has been recognised as a cat (asserting it is true of the perception that *That is a cat*), then a system that can perform conjunctions can combine these two DA-finsts, so that the proposition that *That
is a dog and that is a cat is true of the perception. And the complex proposition (that That is a dog and that is a cat) is true only if the constituent DA-finsts are asserting truths. The perception contains many true propositions but DA-finsts and conjunctions of DA-finsts pick out single propositions that are true of the perception.

It seems plausible that many animals (mammals and birds, for example) possess concepts/containers that contain information that can be utilised to recognise the category of objects, and to interact intelligently with instances of the category. And it seems plausible that many animals possess DA-finsts that keep track of entities, and the category they have been recognised as, allowing them to carry out, for example, the delayed match to sample tasks described in the last chapter. It seems plausible, for example, that cats have containers of stored knowledge pertaining to categories such as dog, car, scratching pad, and so on that allow them to recognise instances of these categories, and interact appropriately with them. They may also utilise DA-finsts to, for example, keep track of prey that has temporarily disappeared from view, or to keep track of food that their owner has hidden away. DA-finsts, in effect, symbolise the claim, for example, that The food is at that location, even when the food cannot currently be perceived.

However, it is also plausible that only humans are able to form expressions of complex propositions, i.e., propositions that involve more than one concept. Container theory is neutral about this claim; perhaps other animals can as well. But according to container theory, expressing complex propositions requires a capacity to combine or manipulate discrete DA-finst symbols into larger wholes much the same way as symbols are combined in public languages.

This is, of course, not to claim anything new. It is precisely what, for example, Fodor's language of thought hypothesis claims (Fodor 1975, 2008). We will also see in Chapter 8 that Barsalou’s perceptual symbol theory (Barsalou, 1999) makes the same claim. What is different in container theory is the claim that the symbols that are combined or manipulated are disposable symbols (DA-finsts).
6.3 A comparison of Fodor’s account, Barsalou’s account, and the DA-finst account

In the next two chapters, I will be examining Fodor’s account and Barsalou’s account in detail, and describing how they fit into the DA-finst account. I argue that DA-finsts should be substituted both for Mentalese symbols in Fodor’s account, and for perceptual symbols in Barsalou’s account. The effect of substituting DA-finsts for Mentalese symbols and perceptual symbols is that Fodor’s and Barsalou’s accounts can be fused into a larger DA-finst account. For now, though, I will just compare how the three accounts deal with a particular example.

When we have a thought that expresses a proposition, it is usually the case that we have particular objects/entities in mind, and particular categories in mind. Imagine John and Mary are out for a walk. In the scene before them, there is a dog chasing a cat to their left, and another dog chasing a cat to their right. Mary is perceiving the scene iconically, as evidenced by the fact that she could express many propositions about what she perceives. However, she singles out just one, which she expresses in English as ‘the dog chases the cat’. Container theory assumes, as does Fodor’s account and Barsalou’s account, that creatures can express propositional thoughts without possessing a public language. Mary can express the propositional thought that The dog chase the cat without making use of any English symbols. She uses her public language to communicate the thought to John. What is the thought like that gave rise to the public language expression ‘the dog chases the cat’? And what is the thought like that John entertains after having heard Mary’s utterance?

Let us start with Mary. She is, in fact, attending to the dog chasing the cat on her left. According to Fodor’s account, what she perceives will cause to be formed a Mentalese sentence that is very similar to the English expression it gives rise to. So her perception will give rise to the Mentalese expression: M(dog) M(chases) M(cat), where M stands for Mentalese symbol. Perceiving the dog causes the activation of a token of M(dog); perceiving the cat causes the activation of a token of M(cat); and perceiving the chasing causes the activation of a token of M(chasing). In Fodor’s account these symbols are combined sententially, and a straightforward account can be given of how the Mentalese expression is converted into the English expression. There are three discrete Mentalese symbols and these are converted into the corresponding three English symbols (words).
According to Barsalou, Mary selectively attends to the dog on the left of her perception, which causes the activation of a perceptual symbol (perceptual because it gives information about what it appears like) in working memory. She, then, selectively attends to the cat and the event of chasing, which, in the same manner causes the activation of perceptual symbols. As we will see in Chapter 8, Barsalou posits that these symbols are arranged iconically rather than sententially, but for present purposes this makes no difference. Once again there are three discrete symbols (albeit perceptual symbols rather than Mentalese symbols), which can be converted into the English counterparts ‘dog’, ‘cat’ and ‘chases’.

To some extent, the DA-finst account fuses Fodor’s and Barsalou’s accounts. The subject arm of a DA-finst captures the percept of the dog in the left of the scene, which gives information about what the dog appears like, just as Barsalou’s perceptual symbols do. Also, the subject arm of a DA-finst causes the percept of the dog to be selectively attended to: the subject arm keeps its finger on the percept of the dog, as it moves about in the perception. After recognition, the predicate arm of the DA-finst captures the DOG container. The predicate arm gives information about the category that the dog has been recognised as, just as Fodor’s Mentalese symbols do. Just as M(dog) stands for the category dog, so does the DOG container.

DA-finsts will also capture percepts of the cat and the act of chasing with subject arms, and the CAT container and the CHASE container with predicate arms. In the DA-finst account, once again, there are three discrete symbols (disposable symbols rather than perceptual symbols or Mentalese symbols), which can be converted into English counterparts. For example, the DOG container will contain the information that ‘dog’ is the English word that pertains to the category of this container.

All three accounts posit that, in thought, complex expressions of propositions are underwritten by discrete symbols that are combined into larger wholes. The accounts differ in what they take those symbols to be like. For Fodor, they are word-like. For Barsalou, they are perceptual. And in container theory, they are disposable symbols – symbols that get their meaning from what they are currently pointing to. All three accounts take the expression of
complex propositions to be achieved much like they are in public languages: by juxtaposing discrete symbols. What differs is just the nature of the symbol in the three accounts.6

However, note that in Fodor's account, Mentalese symbols do not reveal the object that is being thought of; they reveal only the category being thought of. In the example, M(dog) contains no information about which of the two dogs the thinker has in mind. Recall the delayed match to sample experiments described in the last chapter. Imagine a monkey is shown a picture of the scene observed by Mary and John. One of the dogs is highlighted, and then the picture disappears for three seconds. When the picture reappears, the monkey must indicate the correct dog in the scene to obtain the reward. If the monkey is simply keeping active the Mentalese symbol M(dog), during the delay, it cannot tell which of the two dogs in the scene is the one that was previously highlighted. To succeed in the task, the monkey requires to also keep active the 'where' or 'what it is like' information that will allow it to select the correct dog. But this is what DA-finsts do automatically. The subject arm of the DA-finst contains the 'where' and 'what it is like' information, and the predicate arm contains the categorical 'what' information.

In Barsalou's account, perceptual symbols reveal what the object appear like, which would be enough for the monkey to succeed at the task, unless the two dogs in the scene were identical. But what perceptual symbols do not reveal is the category that is being thought of. The perceptual symbol, by itself, cannot reveal whether the thinker was thinking about the dog category or the pet category. Once again the DA-finst account is superior, because the two arms of the DA-finst contain information about what the object appears like, where it is in the scene, and the category that it belongs to.

I do not claim that Fodor's and Barsalou's accounts cannot deal with what is missing in their accounts. I just claim that their accounts would have to be augmented, in order to be as explanatory as the DA-finst account. In the DA-finst account there is no need for further disambiguation, because the subject arm picks out which object is being referred to, and the predicate arm picks out the category it is being thought of as.

---

6 We will see in chapter 8 that Barsalou's account also differs in that he takes the discrete symbols to be plotted iconically and not just sententially.
So on each of the above three accounts, before Mary formulated the public language expression ‘the dog chases the cat’, she formulated a thought composed out of three constituent symbols. In Fodor’s account, the symbols are Mentalese words; in Barsalou’s account, they are perceptual symbols; and in the DA-finst account, they are disposable symbols (DA-finsts). What is attractive about the DA-finst account is that the full interpretation of what each symbol is standing for is available from the DA-finsts (disposable symbols) themselves. Note that when Mary says ‘the dog is chasing the cat’, she has no doubt in her mind which proposition she is expressing. But because there are two instances of dogs chasing cats in the scene, the English expression, by itself, cannot reveal the proposition she has in mind. Further disambiguation is required. The same problem arises for Fodor’s Mentalese symbol account. The expression M(dog) M(chases) M(cat) does not specify which of the two instances is being referred to. In Barsalou’s account, the perceptual symbols may pick out the objects being thought of (the information in the perceptual symbols may resemble the dog and cat on the left more than the dog and cat on the right), but the perceptual symbols, without further disambiguation, do not reveal whether Mary has in mind the dog category or the pet category. In contrast, the DA-finst account explains how Mary is certain about just what proposition she is expressing: it is because the DA-finsts pick out both the objects she has in mind and the categories she is thinking of them as.

6.4 How a heard propositional expression is converted into a mental thought

We have seen how the three accounts posit discrete symbols as being the constituents of a propositional thought that can be converted into a public language expression. Now, let us see how John’s hearing of Mary’s expression works.

Before setting out what happens in the DA-finst account, let me give a short summary of some matters that will be covered in detail in Chapters 9 and 10. I claim that, for creatures with a public language, many containers/concepts come to be ‘headed’ by a public language word. By ‘headed’ I mean that the container contains information that can recognise the public language word for the category, as well as by perceptual tokens of the category. For example, for English speakers, the DOG container would contain the means to recognise the word ‘dog’, as well as recognise the sound of barking, or the visual image of a dog.

Also, as I will discuss in Chapter 9, while the account of reference I am offering can be regarded as an informational account (e.g. Dretske 1981, Fodor 1990), it differs from
traditional causal accounts, in that the interpretation of reference does not depend on what, for example, initially caused, or reliably causes, the activation of the concept/container; rather the interpretation of reference depends on what the recognitional and functional information in the container, considered together, is about. Containers gather any information that pertains to a single category, and not just information that is caused by members of the category. Very often the information in the container will have been caused by its members but that is not necessary. Part of the useful information that is gathered about the dog category, for English speakers, is that the word for the category is ‘dog’.

Having said that, let us return to how John understands Mary’s utterance of ‘the dog chases the cat’. According to the DA-finst account, John’s DOG container will contain the means not only to recognise percepts of dogs, but also percepts of the word ‘dog’. We can imagine that Mary’s utterance is captured iconically in something like Baddeley’s (1986, 1996) phonological loop. John’s cognitive system can retrieve sounds from this store for a short period of time. The perceptual sound of the word ‘dog’ can be captured by the subject arm of one of John’s DA-finsts, and recognised by the container which contains the closest matching exemplar/prototype (the DOG container). The predicate arm of the DA-finst is now set to looping to the DOG container. The same will apply for the words ‘chases’ and ‘cat’. John will end up with the predicate arms of three DA-finsts pointing to the DOG, CHASE, and CAT containers. Now John’s cognitive system’s task is to find something in his visual perception that would also activate those three containers. Roughly speaking the system must pick out something that is recognised by the DOG container, something that is recognised by the CAT container, and something that the relationship of CHASE applies to. His task is to end up selectively attending to the same entities that Mary had in mind when she expressed the proposition.

Note that if John is focussing on the right of the scene, he may capture in the subject arm of DA-finsts percepts of a different dog and cat to the ones Mary is thinking of. Such

---

7 I would remind the reader that containers in the brain are not literal containers, but functional containers. I do not maintain that exemplars and prototypes for the word ‘dog’ are located side by side with visual and auditory exemplars and prototypes of dogs. The information in the dog container can be distributed all around the brain. What makes it a functional container is the claim that if any one bit of information is activated, then all of the rest of the information in the container becomes available. A simple way of implementing this is that if the prototype for dog is activated, then there is activated a node which links to all of the rest of the information in the DOG container. From this node, any other piece of information that is stored can be retrieved.
misunderstandings do of course occur, and sometimes they have important consequences. But, in the present instance, all that will happen is that John will think he is thinking of the same proposition as Mary, but, in fact, he will have in mind a different one. This is fully explained by the DA-finst account. Although the predicate arms of the DA-finsts that are the constituents of Mary and John’s thought are the same, their respective subject arms are different.

In contrast, in Fodor’s account, the English words are converted into the corresponding Mentalese words. Both Mary and John will have an internal expression of a proposition in Mentalese that is the same (M(dog) M(chases) M(cat)), even though they are, in fact, thinking of different propositions. As we will see in Chapter 8, Barsalou does see the need for his perceptual symbols to link both to the entity being thought of, and the category it is being thought of as.

6.5 Substituting DA-finsts (disposable symbols) for Mentalese and perceptual symbols

In many respects, the DA-finst account is similar to both Fodor’s and Barsalou’s accounts. The main difference is in the nature of the discrete symbols that are constituents of thoughts. All three accounts say that the expression of complex propositions involves the combination of symbols in working memory. Container theory claims that the symbols are DA-finsts (disposable symbols). Fodor claims that they are Mentalese symbols that operate just like public language words. And Barsalou claims that they are perceptual symbols. But all three accounts agree that the expression of complex propositions involves the combination of discrete symbols each of which can be, for example, substituted with a different symbol whilst leaving the rest of the expression the same. My strategy, then, in the next two chapters is to claim that disposable symbols can be substituted for Mentalese symbols in Fodor’s account, and for perceptual symbols in Barsalou’s account, and that neither account will lose anything of great importance because of the substitution. As a consequence, the DA-finst account can be regarded, to a large extent, as a fusion of Fodor’s and Barsalou’s accounts.

The only thing we lose from Fodor’s account is that there must have evolved a brand new stock of word-like symbols (with persisting as opposed to disposable meanings) when the capacity to express complex propositions emerged. By substituting DA-finsts for Mentalese symbols, we remove what is, perhaps, the most controversial feature of Fodor’s
account. Evolution normally involves small incremental improvements. It can be argued that Fodor’s language of thought hypothesis requires a giant leap forward – a saltation (Gould, 1977) – in that there must be evolved simultaneously both the set of Mentalese symbols and the system that manipulates those symbols. In contrast DA-finsts (disposable symbols), as we saw in the last chapter, already serve a function in cognitive systems that cannot express complex propositions. DA-finsts keep track of single entities and the category they have been recognised as. A gradual story can be told of how these already functional DA-finsts come to be constituents of expressions of complex propositions. For example first there is just an ability to join together DA-finsts in conjunctions and disjunctions. Then, there emerges the ability to combine DA-finsts with simple relationships such as above/below, or bigger/smaller, and so on.

However, as far as container theory is concerned, most of Fodor’s account can be retained. In particular the claim that propositional thoughts are expressed, much as they are in public language, by combining constituents into larger wholes according to rules can be retained.

Again, once disposable symbols (DA-finsts) have been substituted for perceptual symbols, most of Barsalou’s account can be retained. DA-finsts are amodal in nature, and, so, the DA-finst account has no trouble dealing with abstract or non-perceivable categories.

Fodor’s account, essentially, makes use of just the predicate arm of DA-finsts, in that Mentalese words refer to categories in the same way the containers captured by the predicate arm do. And Barsalou’s account, essentially, makes use of just the subject arm of DA-finsts, in that perceptual symbols disclose what the object being thought of appears like in the same way that the information captured by the subject arm does. Both Mentalese symbols and perceptual symbols are taken to stand directly for categories in the world. In contrast, DA-finsts stand directly for recognisings (entities recognised as belonging to categories), and inherit their meanings from what the two arms of the finst point to. DA-finsts symbolise internal representations, which in turn stand for things in the world.

So the DA-finst account can be thought of as fusing Fodor’s and Barsalou’s accounts, in many ways. But the DA-finst account has the advantage over Fodor’s account that perceptual information is fully integrated into the propositional story. And the DA-finst account has the advantage over Barsalou’s account that, because disposable symbols are amodal (nothing about a DA-finst itself, as opposed to what it is pointing to, reveals what the
symbol is about), there is no problem dealing with categories that are not perceivable. In addition, there is not a problem in determining whether a perceptual symbol refers to, for example, the category *dog* or the category *pet*.
Chapter 7

Fodor’s LOT Account: Mentalese Symbols versus Disposable Symbols

7.1 Introduction

In the last chapter, I set out how the DA-finst account, Fodor’s LOT account, and Barsalou’s perceptual symbol account satisfy Machery’s propositional desideratum. I noted that the three accounts differ, primarily, in what they take the nature of the symbols involved to be. But the three accounts were in agreement that the expression of complex propositional thoughts involved juxtaposing discrete symbols into larger wholes.

In this chapter, I will examine in more detail Fodor’s LOT account. I will claim that what is right about the account is that propositional thoughts are formed by compositions of symbols, much in the way propositions are expressed in public languages. I will claim that what is wrong about the account is that the composed symbols must have persisting meanings, in the same way that public language words have persisting meanings. I claim that DA-finsts, which have disposable meanings, will do the job just as well.

Public language words are a convenient way to achieve publicity (Rey 1983, Peacocke 1992, Fodor 1998) – i.e., words let others understand what proposition a thinker is expressing because everyone uses the same set of symbols that have the same set of persisting meanings. But, in private thought, publicity is irrelevant, because only the thinker has to understand the thought being expressed. No-one else can directly perceive the thinker’s thoughts. What I argue, in this chapter, is that, in private thought, publicity is irrelevant, and, so, it does not matter if the composed symbols keep changing their meaning, provided the system has a way to tell what the current meaning is.

In Fodor’s (1975, 2008) LOT account, he considers Mentalese symbols to be just like public language words, in that the system knows what a symbol means from its ‘shape’. Each time the system encounters the same shape of symbol (e.g., the shape of a word on a page, or
the shape of a neuronal configuration in the brain) the system takes it to have the same meaning. This line of thought leads Fodor and Pylyshyn (1988) to propose the Compositionality Principle which states that:

\[
\text{insofar as a language is systematic, a lexical item must make approximately the same semantic contribution to each expression in which it occurs (Fodor & Pylyshyn, 1988, p. 42).}
\]

The idea is that each time the word ‘dog’ is encountered in English it will mean dog, and, similarly, each time the mental symbol M(dog) occurs in the language of thought it will mean dog. It seems to be widely accepted that the Compositionality Principle is essential for an atomic symbol manipulation system. But I will show, in this chapter, that thoughts composed entirely of DA-finsts, whose semantic contributions are constantly varying depending on what they are currently pointing to, can be fully understood by the system. And, indeed, as pointed out in the last chapter, compositions of DA-finsts will usually pick out more precisely just what proposition a thinker has in mind. This is because the two arms of the finsts can point both to the object and the category that the thinker has in mind. In the DA-finst account, the constituents of private thoughts (thoughts which only the thinker can perceive) are DA-finsts. And a DA-finst symbol does not get its meaning from the shape of the DA-finst itself, but from the percept and container that the two arms of the finst are pointing to.

What is retained from Fodor’s LOT account, then, is that complex propositional thoughts are composed from discrete symbols, just as complex propositions are composed from words in public language. What is discarded is that concepts are word-like Mentalese symbols. What is substituted for Mentalese symbols are DA-finsts, and it is claimed that concepts are containers rather than Mentalese symbols. Mentalese symbols, like words, have persisting meanings. In contrast, DA-finsts have disposable meanings – they inherit their meaning from the container that they are currently pointing to, and discard that meaning once they stop so pointing.

7.2 TV and AV set out

Thought, like language, seems to be productive and systematic. Language is productive because there seems to be no limit to the number of new propositions that we can express. It is systematic in that if one can form the sentence ‘John loves Mary’ then one can
form the sentence ‘Mary loves John’. Compositionality explains productivity and systematicity in public language: a finite stock of primitive atomic components (mostly words) can be composed into a (potentially) infinite number of larger wholes, according to systematic rules. The meanings of the wholes are dependent on the meanings of the primitive atomic components. So the meaning of ‘the dog chased the cat’ is derived from the meanings of the atoms (‘the’, ‘dog’, ‘chased’, and ‘cat’), together with the rule that determines it was the dog doing the chasing and not the cat. What is crucial is that the same atomic meanings can (systematically) recur in other sentences, and that one can always pick out the components that make up the whole. The atomic symbols (words) can simultaneously have meanings, and take on physical form, so as to have causal effects in the world. We also know that this kind of productive and systematic compositionality is very powerful when deployed in computer programs.

Fodor’s (1975) LOT hypothesis posits that thought works just the way that public language does. As he says,

Language (/thought) is productive and systematic because it is compositional... If there really is such a thing as Mentalese, then the productivity of Mentalese and English are strictly analogous (Fodor, 2004, p. 37).

And of course the Compositionality Principle will apply to Mentalese, just as it applies to English. The idea is that each time the word ‘dog’ is encountered in English it will mean dog and, similarly, each time the mental symbol M(dog) occurs in a composition it will mean dog.

What this means is that, for Fodor, the realisers of expressions of propositions in the brain are maximally real. You could, in principle (though it might be extremely hard in practice), identify certain neurons that are activated in working memory, and know from the pattern of them (their ‘shape’) that they mean, for example, dog. Dennett, in Real Patterns (1991), refers to a synactoscope – a hypothetical instrument for detecting the syntactical patterns in the brain that have meanings, the way words have meanings in public language. He notes that, for Fodor, those patterns will be as real as the patterns on a page containing written expressions of propositions.

As stated in the Compositionality Principle, Fodor also thinks that thought achieves its systematicity in the same way as English: by having the same physical shapes always mean the same thing. In English, e.g., the word ‘dog’ has a physical auditory ‘shape’ when
spoken, and a physical visual ‘shape’ when written. Whenever this same physical shape is encountered in a propositional expression someone who understands English knows that it means *dog*, because that ‘shape’ always means *dog*. So, according to Fodor, in thought, there is a particular physical ‘shape’ (— a neuronal pattern) that, whenever it is encountered, the system takes to mean *dog*. In effect, there is a neuronal pattern that physically implements the Mentalese symbol M(*dog*), and, each time it is encountered, it makes approximately the same semantic contribution (it means *dog*), in accordance with the Compositionality Principle.

This is the target view (TV) that I want to replace. It arises because it is assumed that thought is just like public language, and that compositionality is achieved in the same way. The target view is:

TV: any time there is a syntactic tokening of a type of symbol in a thought it systematically has the same semantics/meaning.

So, to give an example, using Fodor’s notation in LOT 2, every time a syntactic token of the Mentalese symbol type M(*bucket*) appears in a thought it means *bucket*.

Of course there are exceptions. As Crane (2003) points out it won’t apply, in public language, in cases of ‘metaphor, idioms, ambiguity and … [where] more than one person can share a name’ (p. 138). Fodor & Pylyshyn give an example of idiom: the meaning of ‘the man kicked the bucket’ cannot be derived from the meanings of the atoms. These kinds of exceptions must be learnt separately; using the normal compositional method will result in failure to understand.

But, aside from these kinds of exceptions, the principle of compositionality is supposed to apply. To express the proposition *the dog chases the cat* in LOT, the system will combine the Mentalese words M(*dog*), M(*chases*) and M(*cat*). Those Mentalese words can occur in other propositional thoughts, and, each time they occur, they will have roughly the same physical implementation, and will mean the same thing (*dog*, *chases*, and *cat*).

The alternative view (AV) with which I want to replace TV is

AV: each time there is a syntactic tokening of a DA-finst symbol in a thought in working memory, the DA-finst inherits its meaning from the meaning of what its arms are currently pointing to, and the meaning of the DA-finst can be changed by changing what the arms are pointing to.
Instead of instantiating, for example, a token of the type M(dog) which always has the same meaning (dog), an arbitrary token, a DA-finst, is instantiated, which has no pre-existing meaning. The meaning of the DA-finst is determined by the container that the predicate arm is pointing to (e.g., it acquires the meaning dog once the DOG container is captured by the predicate arm).

DA-finsts, in themselves, have no persisting meanings. They inherit their meaning, on the fly, from what the arms of the finst are currently pointing to. In LOT and public language, the symbol M(dog) and ‘dog’, respectively, refer to the category dog. In the DA-finst account the same DA-finst means dog if its predicate arm is pointing to the DOG container, but means cat if it is pointing to the CAT container.

To summarise, TV (the target view) considers atomic mental representations to be similar to words in a public language. There is a finite, persisting stock of primitive atom types, with fixed meanings (just like there is a vocabulary of English words with fixed meanings) which are composed into greater wholes. The same atom types retain the same meanings. AV (the alternative view), in contrast, says that the atoms used to compose have no fixed meanings; instead, they get their meanings on the fly, as and when they are used, and dispose of them afterwards.

7.3 Why TV is assumed

We are familiar with atomic symbols in public language: words typically play the role. Words have the following three features: (a) an atomic word cannot be broken down into parts that are semantically evaluable; (b) an atomic word represents arbitrarily (it doesn’t have to carry information about what it refers to) (c) each word is distinguishable from other words. Because anything can be used as atoms, it is efficient to make the atoms as compact as will do the job. Each different atom will have a different ‘shape’ or ‘syntax’, which can in turn have a different semantics or intentionality. The way semantic differences can have causal effects is via syntactic difference. What TV assumes is that, just as words do, each atomic symbol type, at some point, acquires an initial intentionality, and from then on retains that same intentionality.

That TV is assumed to be necessary for atomic symbol manipulation is hardly surprising given that the symbol systems we are familiar with comply with TV. Take for example public language, digital computers, and the genetic code used in cells.
It is hard to imagine how a public language could operate without TV. Users of a language have to ensure that they are using the same stock of primitive syntactical types that everyone else is, and that they have paired each syntactical type with approximately the same semantics/meaning as everyone else. Whenever I use or encounter the syntactical shape ‘dog’ in English sentences, it can be relied on to have the same meaning (barring exceptions such as idiom or slang) that it always had.

When writing a program for, let us say, a calculator to be run on a digital computer, the programmer will work out what ‘meanings’ are required (e.g., 0,1,2,...9,+,-,/,*,,=,...) and will arbitrarily assign these meanings to different ‘shapes’ of primitive atoms. Once this assignment has taken place, whenever a particular ‘shape’ occurs in the program, it will have the same ‘meaning’ paired with it.

Meaning, in the case of the calculator, was fixed by the programmer, but it could have been fixed by ‘evolution’ as is the case in the genetic code. In the nucleus of every living cell there are chromosomes that contain DNA, which is interpretable as a code. Ribosomes use the DNA code to systematically make proteins from the strings of amino acid building blocks that are coded for. The atoms in this code are triplets of DNA, and they always stand for the same amino acid. For example, the triplet ‘CAG’ always means glutamine (it will always cause the ribosome to produce glutamine).

So we have a pattern here. Each time the atomic syntactical symbol (e.g. ‘dog’) is encountered in a public language it means the same thing (dog). Each time the same atomic symbol (e.g. ‘101’) is encountered in a calculator program it means the same thing (the number 5), and each time the same atomic syntactical symbol (e.g. CAG) is encountered inside a living cell, it means the same thing (glutamine). It is only natural to assume that if thought involves manipulating atoms of symbols into greater complexes, then there will be a stock of primitive atomic symbol types in which the same syntax is persistently paired with the same meaning.

Another reason to consider TV essential is that symbols at some point in time have to be paired with their initial intentionality. It is easy enough to tell a story about how words get

---

8 It might be thought to be controversial as to whether the genetic code represents. I would argue there is computation going on in the cell, therefore representation (Sprevak, 2010), but be that as it may, the scientists who ‘cracked the code’ did so on the basis of each triplet having a fixed meaning that was set by evolution.
their initial intentionality (someone coins a new word for something, and, if enough people adopt it, then the word becomes an atom type in the language). In the case of a computer program, the programmer sets the initial intentionality. In the case of the genetic code, we can understand how natural selection could bring it about that each triplet acquired an initial intentionality. But how to do this for mental representations is, as we will see in Chapter 9, difficult to account for naturalistically. So, if it is hard to set the initial intentionality, then all the more reason to keep it fixed to the same syntactical shape from then on.

Also, if the same meaning persists once it is initially set, then we can see how Haugeland’s Formalists’ Motto can be effective. The motto is: “if you take care of the syntax, the semantics will take care of itself” (Haugeland, 1985, p. 106). The way semantics takes care of itself is by keeping itself paired to the same syntactical shapes all the time.

Another way of saying this is to note that one of the most important purposes of symbols is to stand for things when they are not present, and it is very hard to figure out how tokening a shape of a symbol type should be paired with a particular meaning, if there is no principle of compositionality keeping the meaning and shape in sync.

Similarly, when we retrieve a thought stored as a memory, the syntactical shapes that constitute the stored thought had better have the same paired meanings as they did when they were stored. For example, if I see smoke and think M(smoke), then if I retrieve from memory ‘If M(smoke) then M(fire)’, I will not be able to derive the conclusion M(fire), unless the M(smoke) token I retrieve from memory means the same as the M(smoke) I tokened when I saw smoke.

7.4 Technical counter-examples to TV

AV claims that having the same constituent syntactical vehicles is not the only way to achieve the same semantic contribution in thought, as TV claims. What follows are examples from public language and computation in which syntactical shapes vary greatly, while the semantics stays the same. These will all turn out to be pointless, or easily accommodated by advocates of TV, but they will open the way for AV.

The first way involves what can be thought of as one symbol shape passing the semantics it is paired with on to a new shape, with the system able to keep track of the change. For example, in 2002, the president of Turkmenistan arranged for the government to
pass a law changing the words to be used for the months of the year and the days of the week. We can easily cope with this kind of setting new fixed meanings for a small number of words. In principle, the dictator could order the changing of the words for a number of meanings every day, or even every hour. The populace would have to learn the new syntactical pairings with meanings, before they could communicate with others again (assuming most people complied with the law). So here we would have a clear case of there being meanings that are not paired with the same persisting syntactical shapes, as TV requires, albeit one that would seem to serve no purpose in thought.

A variation on this theme is a code (such as the ones used by 'Enigma' machines during World War 2) that changes every day, which is used to encrypt messages, so that only recipients with the right decoding means can decrypt the message. The 'shapes' that are paired with meanings, when sending messages, will be different every day. Of course, once decoding has taken place, the symbols that emerge are in fact fixed, and this is required for understanding; TV can claim that this is what matters.

The following case is trickier for TV to accommodate. Consider a calculator program: it will have internal symbols for each of the 'meanings' that will be displayed to the user (the numbers 0 to 9, plus, multiply, etc. Now, it is perfectly possible for a programmer to incorporate into the program the behaviour that, every time the calculator is turned on, the program randomly changes the arbitrary internal symbols it uses for the different meanings. The system retains no fixed atoms with fixed meanings, and this case differs in that when the calculator is off, there are no persisting syntactical shapes that retain any intentionality. Someone who had 'cracked the code' of the program would find it provided for arbitrary symbols but would be unable to say what semantics each discrete symbol had, while the calculator was off. The program randomly assigns to each symbol a brand new intentionality whenever the calculator is switched on: the symbol that stood for the meaning 3, the last time the calculator was on, will this time, perhaps, stand for plus.

These counter-examples are certainly no reason to abandon TV. Adopting these methods would serve no useful purpose in thought. However, what they do show is that all that matters is that the correct meaning is encountered each time it should be (that there is some way for the system to retrieve the same meaning even if the symbol shape keeps on varying). For example, every time I should be thinking of the first month of the year, I do so regardless of what shape of symbol is used. The most obvious and practical way to achieve...
this is TV: having the same syntactic shape paired constantly with the same meaning. But there is nothing necessary about this: if the meaning can be kept the same some other way, it doesn't matter if the syntactical shapes of the atoms are changing willy-nilly.

7.5 The alternative view

DA-finsts are precisely a way for the meaning of atoms to be kept the same, when they should be, that does not rely on the same meaning being persistently paired with the same 'shape' of atom. One can think of the contrast between the two views as follows. In TV the shape of the symbol determines what the meaning is (the one it is always paired with). In AV, on the other hand, the shape of the symbol (the DA-finst) itself is irrelevant; the meaning of the symbol is inherited from the meaning of the representations that the two arms of the DA-finst are pointing to. The meaning of a DA-finst changes completely when the container the predicate arm is pointing to changes.

In order to compare it with AV, I will set out how TV works in LOT. Using Fodor’s notation, the primitive Mentalese atomic shape for dog is M(dog). At some point M(dog) acquired its initial intentionality of standing for dogs and from then on continued to refer to dogs. The idea is that whenever M(dog) is tokened in my working memory, I think of a dog. Once the token fades from working memory, it still remains the case that the Mentalese atom type M(dog) persists, and continues to be paired with the same meaning, just the way the English word does, even if no-one is tokening it. The next time I encounter a dog, or want to think of one, I will take out a token of the same M(dog) type of symbol. And if I activate a memory that contains a M(dog) token, it makes me think of dogs, because that is what it always stands for. As in Haugeland’s Formalists’ Motto, the semantics ‘takes care of itself’ by each meaning rigidly sticking to the same syntactic primitive atom that it always does.

Under AV, in contrast, the syntax (shape) of a DA-finst symbol, itself, does not reveal what the meaning of the symbol is. Semantics cannot take care of itself by staying attached to the same persisting atomic shape. Instead, DA-finsts attach themselves to the required meanings in thought by having their predicate arms point to the concept/container that matches the meaning. How does the system know what meaning is required? In typical circumstances, the required meaning will be the category that the object pointed to with the subject arm of the DA-finst is recognised as belonging to. In this way, DA-finsts, which have no intrinsic meaning, are given intentionality on the fly, there and then. For example, a DA-
finst acquires the intentionality *that dog there*, where the ‘that’ and ‘there’ are given by the subject arm of the DA-finst, and the category *dog* is given by the DOG container. In effect, I am saying that, before compositional propositional thinking that involves the manipulation of atoms can take place, the meanings that are going to be used must be linked to the symbol (the DA-finst). Under TV, all that is looked after is the syntax, semantics has to take care of itself. Under AV meanings must be worked out (typically by the recognition process) before they can be assigned to disposable atomic symbols (DA-finsts).

In the DA-finst account, there are indeed entities with persisting meanings, namely, the containers of stored knowledge pertaining to single categories. But the vehicles (the DA-finsts) that are composed (arranged, juxtaposed) in working memory are not like words or Mentalese symbols, whose meanings are determined entirely by the ‘shape’ of the symbol itself (e.g. the word ‘dog’), and whose meanings persist from one use to another. A DA-finst disposes of its meaning as soon as its arms no longer capture percepts and containers, and a DA-finst acquires a new meaning when its arms capture new percepts and containers. How DA-finsts capture percepts and containers, as set out in Chapters 4 and 5, is typically a function of the recognition process. The cognitive system uses a DA-finst to keep track (with its subject arm) of what object in perception is to be recognised, and to keep track (with its predicate arm) of what category it is recognised as. Instead of a persisting symbol like the word ‘dog’ or the Mentalese symbol M(dog), whose meaning is always *dog*, the meaning of a DA-finst is worked out from what percept and container it is currently pointing to.

Note how much more informative compositions of DA-finsts can be, compared with public language words or Mentalese symbols. Say, for example, that there are three discrete DA-finsts. The first stands for *that dog there*, the second stands for *that cat there*, and the third stands for *that chasing there*. They can be juxtaposed together so as to express the proposition *that dog there chasing that cat there*. Recall, in the last chapter, the example of Mary and John out for a walk. The scene contains two instances of dogs chasing cats, and Mary says ‘the dog is chasing the cat’. Even though her statement is ambiguous, and could be misinterpreted by John, she has no doubt which dog and cat she has in mind. This is fully explained if the thought giving rise to her public language expression is underwritten by DA-finsts, whose subject arms pick out the dog and cat she is thinking of. But it is not explained if her thought is underwritten by Mentalese symbols which are wordlike. M(dog) M(chases) M(cat) does not disambiguate which dog and cat the thought refers to.
7.6 Containers/files

Fodor sees the need for containers of stored knowledge, and discusses them in the guise of files. The file, as an explanatory concept, is commonly used. Prinz (2005)⁹ and Fodor (2008) explicitly discuss it. It is either explicit or implicit in, for example, Margolis (1998), Barsalou (1999), Millikan (2000), Cummins (2002), Papineau (2006), and Recanati (2012).

As Fodor (2008) says, files should be thought of as quite like files stored in filing cabinets, in the real world. You could also think of them as like computer data files, but they more akin to computer folders (whose icons often look like real world files). A more modern way of thinking of them (occasionally used below) would be as interactive websites or webpages. They are to be thought of as containers whose purpose is to gather any kind of information and knowledge pertaining to a single ‘topic’ or ‘category’. As set out in Chapter 3, the idea is that for each primitive concept a subject possesses, they possess a separate file or container into which is placed knowledge as it is accrued.

Fodor (2008) proposes that, for example, a Mentalese token of M(John) is used to think about John (and means John), but is also the name of the file that stores information about him, which can be looked up. So M(John) does double duty as an atom that has meaning, and a file name. Fodor says, ‘That one thinks in file names is the best short summary I’m able to formulate of the version of RTM that I’m currently inclined to endorse (Fodor 2008, p. 95).’ However, he also specifies that the contents of files ‘are written in Mentalese’ (p94), i.e., in file names, which contrasts sharply with the view that I set out in Chapter 4 that the contents are largely perceptual and sensorimotor in nature, as Barsalou (1999) and Prinz (2002) have argued. But the most crucial difference between his Mentalese LOT account and the DA-finest account is that Fodor is adamant that Mentalese symbols (file names) directly refer to categories in the world entirely independently of the files they name, and their contents. He holds that e.g. M(dog) would mean dog, even if no DOG file existed. In contrast, a DA-finest cannot refer to dog if there is no DOG container for it to point to. The problem for Fodor is that he cannot give a satisfactory account of how Mentalese symbols get their initial intentionality. He suggests that the symbols have an innate intentionality but

⁹ Prinz more often uses the metaphor of a database. It is possible that he may not think of knowledge as being so determinately segmented into files as presented here. However discrete files are essential to the irenic position.
many find this suggestion very problematic. In Chapter 9, I offer an account of how containers, from which DA-finsts inherit their meaning, acquire their intentionality based on the information in a container.

So, Fodor acknowledges that something like files/containers are required to explain inference and interaction, but he does not consider files/containers to be concepts. He considers them to be incidental to concepts, which he identifies with word-like Mentalese symbols. Fodor claims that the contents of files are written in Mentalese, the way information is written in a book using words. And, in accordance with TV, Mentalese symbol types retain their meaning over time.

According to container theory, concepts are containers and what are juxtaposed in working are not the concepts/containers themselves, but DA-finsts, which inherit their intentionality from the container that they are pointing to. Under AV, there are no persisting atoms with fixed meanings, and so file contents cannot be written in Mentalese-like atomic symbols. Instead, according to container theory, the contents are either multimodal perceptual representations, or expressions of propositions that have meaning determined by the containers/concepts that are pointed to.

One can think of the file/container, itself, as referring to a type or category, and so having a meaning, but the contents of containers do not have meanings but are meanings (Jackendoff, 2006). This is the point that, e.g., an exemplar of a banana does not by itself have the meaning banana; instead it gives meaning/understanding to the subject depending on what container it is in. The BANANA container has the meaning banana, and the exemplar informs the subject what bananas are like. But the same exemplar, in the FRUIT container, informs the subject what fruits are like. Unless they are pointers to other containers, the multimodal representations in these containers directly reveal ‘what it is like’ to encounter and interact with instances of the category of the container. The contents of the DOG container disclose how to identify a dog, what the parts and features of dogs are, what different dogs look like, what behaviours dogs typically engage in, how to interact with dogs, and so on. The container contents constitute our understanding of dogs. One could think of containers as like interactive websites with movies and sounds, except that they also contain representations that are pertinent to the other senses, and to movement, actions and feelings.

Our long term memory is ‘content addressable’. This means any information about a topic can be used to retrieve the rest of the information about that topic. In addition to the
obvious case of a visual percept of a dog activating my DOG container, any of the following could also: a barking sound, paw prints on the kitchen floor, the smell of my dog, a scratching sound on the door, the sound of the word ‘dog’, and so on. The idea is that any content of the container can be used to activate the whole container. It is like the idea of a search engine in which you are not just limited to typing in words, but can use any sensory image (or part of one) as well. McLeod et al (1998) show how connectionist networks can have this feature, using the McClelland & Rumelhart’s (1986) ‘Jets and Sharks’ network as an example. They contrast this kind of content addressing with the system having to know how to get to the ‘head’ of the file (e.g. by using a file name) before being able to access the rest of it. So, for example one needs to know the word ‘dam’ to look it up in a dictionary. One can’t use any part of the definition to find out the word. The same goes for telephone books: one can’t use the phone number to find out the name. If a file is content addressable, then the file can be considered to have many entry points to the file (many exemplars, prototypes and theories that can be matched). An instance, for example a particular percept, just needs to exceed the threshold of any one of these access points in order for the whole of the information inside to become available.

I want to stress that from a computational point of view, DA-finsts that point to containers and percepts are just as effective as Mentalese atoms for composing in a symbol system. They are compact, arbitrary, and easily distinguishable from each other. The only difference is that they derive their meaning from the meaning of the container they are currently pointing to, rather than have their own fixed meanings. And they are distinguishable from each other, not by their intrinsic shape, but by what container and percept they are pointing to. In TV, semantics takes care of itself by the same meaning attaching itself to the same shape of symbol. In AV semantics takes care of itself by attaching a container to the DA-finst symbol.

7.7 Pointers and demonstratives

I have been saying that DA-finsts ‘point to’ containers and this can be understood computationally by analogy with computer pointers (cf. Ballard et al. 1997, p. 725). Rather than these vehicles storing data values, they store the location of the data. As set out in Chapter 5, this is what I claim DA-finsts in working memory (in pre-frontal cortex) achieve via recurrent loops: they keep track of the location of the data (I tentatively suggest is
located) in post-parietal cortex, sensory cortex, and infero-temporal cortex. The main reason for using pointers in computation is as a substitute or proxy for large data vehicles (such as files). This increases the speed of processing and reduces the amount of working memory required.

These pointers are very much like demonstratives in a public language (cf. Section 1.3.3. of Pylyshyn (2007)), in that the intentionality of the symbol cannot be determined from its 'shape'. Instead it is the 'shape' of the information at the location that is pointed to that determines the intentionality intended. But they differ from demonstratives in that, if one says 'look at that' (while pointing), the listener must perform a context dependent interpretation to pick out what is being referred to by attending to the location pointed to (Pylyshyn 2007, p16). By contrast in computation pointers point quite determinately to a definite internal memory location; no interpretation is needed. They also differ in that we cannot be aware of them at the personal level, in the way that we can be aware of what words like 'this', 'that', 'I' or 'now' mean independently of what they are picking out in any particular use. These pointers have a meaning only for the cognitive system operating on them (something like use the information here to activate a file). The person is only ever aware of the container pointed to, not the pointer.

On first hearing, AV may seem like quite a radical departure from TV, but, once all the details are in, it could be thought of more as a rearrangement of elements. From a compositional point of view, there is not much difference between an activation of a token of M(dog) and an activation of a DA-finst that points to the DOG container. There is, however, the major upshot that the entities that have persisting intentionality are not word-like atomic symbols but are atomic containers (where atomic means not made up of smaller units of the same type). And, because containers contain within themselves information pertaining to the category of the container, we will see in Chapter 9 that this allows us to explain how containers acquire their intentionality.

10 Another way to think of these demonstrative pointers is like hyperlinks to files. For example an email might say, '...which can be found here'. Clicking on 'here' will bring up the file or webpage pointed or linked to.
7.8 Conclusion

To sum up, I am agreeing with Fodor that agents can productively and systematically produce an unlimited number of expressions of single propositions by composing atomic symbols into larger wholes. But I disagree with him that those atomic symbols have fixed persistent meanings, the way words do. I am disagreeing with him that there is a vocabulary of Mentalese words that directly refer to categories in the world. Instead, I am proposing the following two things: (a) the symbols that are composed into larger wholes (into expressions of complex propositions) in working memory are disposable symbols (more specifically DA-finsts) that have no persisting meanings, but, rather, inherit their meaning from what the predicate arm of a DA-finst is pointing to; and (b) what the predicate arms point to are concepts i.e. containers of stored knowledge pertaining to single categories. Those containers are what have persisting meanings, and not the disposable symbols that are combined. The DA-finsts can constantly change their meaning simply by pointing to different things – the same DA-finst can mean dog one moment and then cat the next simply by changing what is captured by its predicate arm from the DOG container to the CAT container.

Fodor thinks that concepts, the entities with persisting meanings, are word-like Mentalese symbols, and I claim that concepts are containers, and the entities with persisting meanings are those containers. Fodor claims that what are composed into expressions of propositions are Mentalese symbol concepts, while I claim that what are composed are disposable symbols (DA-finsts) that point (via their predicate arms) to concepts/containers from which the disposable symbols inherit their meaning. According to container theory, DA-finsts can be substituted for Mentalese symbols and the rest of Fodor’s LOT account can be retained. Next we will see how DA-finsts can be substituted for perceptual symbols in Barsalou’s account.
8.1 Introduction

In this chapter, I will examine Barsalou’s (1999) account. We will see that it is similar in many respects to the DA-finst account. This is, of course, no accident; many aspects of the DA-finst account are inspired by Barsalou’s, and, in effect, this chapter sets out how I have interpreted him, and why the DA-finst account differs in certain respects.

Barsalou’s account centres on perceptual symbols; the main aim of the chapter is to argue that DA-finsts can be substituted for perceptual symbols, without any great detriment to the rest of Barsalou’s account.

I will focus on three main issues:

1. That perceptual symbols like DA-finsts inherit their meaning from the container they are linked to, and not from what they resemble.
2. That what Barsalou calls ‘frames’ are the equivalent of container theory’s discrete containers.
3. That propositions are expressed by sentential juxtapositions of symbols in the DA-finst account, and by iconic/perceptual juxtaposition of symbols in Barsalou’s account.

8.2 Container theory vs Barsalou and Fodor

8.2.1 The perceptual area, the conceptual area, and the working memory area

In order to better understand Barsalou’s account, and in particular his diagrams, it will be useful to highlight the three distinct functional areas that I have previously described. As I have said, the neural localisation claims are not essential to container theory.

1. The area (call it the ‘perceptual area’) where occurrent perceptions and simulations of perceptions (e.g. imaginings, dreams, episodic memories)
are recorded; broadly speaking this can be characterised as sensory cortex (e.g. visual perceptions are a result of activity in the visual sensory cortex).

2. The area (call it the ‘conceptual area’) where containers of stored knowledge pertaining to single categories are located; this can be characterised as associative cortex.

3. The area (call it the ‘working memory area’) where discrete symbols are composed/juxtaposed into expressions of complex propositions; working memory is regarded by many to be located in prefrontal cortex.

Bear in mind that these areas are not meant to be rigid segregated locations in the brain. Each area may well comprise many different regions distributed about the brain. What I am arguing for are functional areas. Areas that are united by their connections rather than by contiguity. Container theory claims that there are regions where information is represented iconically, regions where information is stored in containers, and regions where information is juxtaposed, combined and manipulated in real-time.

As I set out in Chapter 2, perceptions recorded in the perceptual area are iconic in nature, which means that they simultaneously plot many pieces of information against independent dimensions, and, as a consequence, there are always very many propositions that can be made about them.

As I set out in Chapters 3 and 4, the conceptual area contains concepts that are containers of stored knowledge; these containers contain representations that facilitate recognition, inference, and interaction with members of the category. The containers can be associated with parts of perceptions, thus conceptualising them.

As I set out in Chapters 5 and 6, I propose that the working memory area contains DA-finsts (disposable symbols), whose subject arms point to percepts/entities in the perceptual area, and whose predicate arms point to containers in the conceptual area. As I set out in Chapter 7, I take DA-finsts to be juxtaposed sententially, the way Mentalese symbols are juxtaposed in Fodor’s account, but with the difference that DA-finsts are disposable symbols (they do not have persisting meanings), while Mentalese symbols have persisting meanings the way words do.

In this chapter, we will see that the Barsalou account differs in two respects from both Fodor’s and the DA-finst account. First, he regards the juxtaposed symbols to be perceptual (neither word-like nor disposable). Second, in the expression of complex propositions, he
proposes that the symbols are juxtaposed iconically (and by extension perceptually) in working memory. In contrast, in Fodor's account and in the DA-finst account, symbols are juxtaposed sententially. I will go into greater detail on this later in this chapter.

8.2.2 Marking up perceptions

It can be useful to think of DA-finsts as, in effect, 'marking up' perceptions. Just as one could write the word 'dog' on a photograph with an arrow pointing to the dog image, so the predicate arm of a DA-finst, in effect, marks the object captured by the subject arm as belonging to the dog category, because the predicate arm points to the DOG container. Even as the object moves about in the perception, the DA-finst keeps track of the category it has been marked up as. One DA-finst marks the part of the perception that corresponds to, for example, the dog shape as belonging to the dog category, and another DA-finst marks the part of the perception that corresponds to, for example, the cat shape as belonging to the cat category. As we will see below, Barsalou also makes use of the idea of marking up.

8.2.3 How the three areas feature in the DA-finst account and Fodor's account

All three accounts acknowledge the perceptual area, the conceptual area, and the working memory area, but they differ in how they regard them to interact.

In the DA-finst account, all three areas are involved in determining the meanings of the composed symbols: the DA-finsts in the working memory area inherit their meaning from what the subject arm and the predicate arm are pointing to. The subject arm will be pointing to something in the perceptual area, and the predicate arm will be pointing to a container in the conceptual area. For the system to determine the meaning of a DA-finst in working memory, it must have regard to both the perceptual and conceptual areas.

This is in sharp contrast to Fodor's account. Fodor fully acknowledges all three areas. Chapter 6 of LOT2 (2008) argues that perceptions are iconic, and, as we saw in the last chapter, he embraces the idea of 'files' of stored knowledge. However, when it comes to determining the meanings of Mentalese symbols, he insists that the perceptual and conceptual areas are irrelevant. He claims that Mentalese symbols have direct reference, e.g. M(dog) directly refers to and means dog. Regardless of what information might be in the DOG file (it could be all pertaining to cats, for example) the M(dog) symbol refers to dog. And it doesn't
matter if the percepts in the perceptual area that activate the M(dog) symbol all happen to be percepts of, for example, cats, the M(dog) symbol still refers to dog. So the meanings of Mentalese symbols that are composed into propositions in working memory are independent of the other areas, in the sense that what proposition is being expressed is revealed entirely by the shapes of the Mentalese symbols, just as in public language. The symbol M(dog) always has the same shape, and directly refers to dogs; what the shape means is given entirely by this direct reference, and is not inherited from anything else in the system. Fodor accepts that the perceptual and conceptual areas must be included in any full account of cognition. For example, some account must be given of how parts of perceptions are tagged, and how creatures come to have inferential and interactional knowledge concerning what they are thinking of. But, in order to determine the meaning of a symbol in working memory, all the system has to do is examine the shape of the Mentalese symbol in working memory. The M(dog) shape can be relied on to always mean dog.

We will now see that Barsalou’s account is closer to the DA-finst account than to Fodor’s account in terms of interweaving all three areas.

8.3 Barsalou’s perceptual symbol systems

8.3.1 Summary of Barsalou’s account

I will start by setting out just the bare bones of how Barsalou says complex propositions are expressed. As was discussed in Chapter 4, with the example of a car, he takes the system to build up ‘frames’ (containers) of stored knowledge by selectively attending to parts of perceptions. He calls the pieces of information that get placed in frames ‘perceptual symbols’. Once perceptual symbols are stored in frames, they can then be utilised to recognise objects in perceptions. Perceptual symbols are very similar to what I have been calling perceptual exemplars and prototypes, and the recognition process he envisages is compatible with what I have set out in Chapter 4.

Barsalou claims that complex propositional expressions are expressed by combinations of perceptual symbols (PSs). A perceptual symbol that is involved in recognising, for example, a cloud is passed forward to working memory, and becomes available, as a discrete symbol, to be a constituent in the expression of a proposition. In other words, rather than the recognition of a cloud causing the activation of the Mentalese symbol M(cloud), as in Fodor’s account, the same perceptual symbol involved in the recognition of
the cloud is forwarded to working memory where it can be used as a constituent in compositions that express propositions. In Barsalou’s account, it is not the case that there are a fixed number of recognitional PSs in, for example, the CLOUD container, and one of these is passed forward. Instead the symbol used to stand for cloud is created on the fly, based not just on the closest stored PS to the cloud being perceived, but also adapted, as necessary, to take account of what the currently perceived cloud is like. As a result, the shape of PS(cloud) is very likely to be different each time it is used. The perceptual symbol used in working memory includes information about the appearance in some sensory modality of what is being thought about. Indeed, it can be multimodal (a PS(dog) might include information about what the dog looks and sounds like).

Barsalou gives an example, which we will return to below, where, in addition to a PS(cloud), there are also PSs for jet, balloon, and above/below passed forward to the working memory area. Barsalou claims that the proposition The balloon is above the cloud is formed by juxtaposing the corresponding three discrete PSs, and he claims that changing the proposition to The jet is above the cloud is a matter of substituting PS(jet) for PS(balloon). In other words PSs are discrete symbols, which can be manipulated in the same way one could substitute M(jet) for M(balloon) in Mentalese, or substitute the JET container for the BALLOON container, in the predicate arm of a DA-finst, in the DA-finst account, or substitute the word ‘jet’ for the word ‘balloon’ in English. However, an interesting difference, as discussed below, is that he regards the PSs to be juxtaposed iconically, i.e., plotted against independent dimensions (a reference frame), rather than sententially.

So all three areas are involved in Barsalou’s account: the perception of the cloud occurs in the perceptual area; the PS that recognises the cloud comes from a ‘frame’ in the conceptual area; and that PS is passed forward to the working memory area, where it can be combined with other PSs to form expressions of complex propositions. I now turn to how the meaning of PSs are determined. Can the system interpret the meanings of PSs in working memory just from their shapes, like Mentalese symbols, or are the meanings inherited from the conceptual area, like DA-finst symbols?

8.3.2 Perceptual symbols do not get their meaning from what they resemble

How can the meaning of perceptual symbols be determined, if they can be a different shape every time they are used? A tempting response is to say that resemblance determines
what they mean. But there are well known problems with using the notion of resemblance to determine meaning, and Barsalou is well aware of this. He says

the designation of a perceptual symbol determines whether it represents a specific individual or a kind – the resemblance of a symbol to its referent is not critical . . . Suffice it to say for now that the same perceptual symbol can represent a variety of referents, depending on how causal and contextual factors link it to referents in different contexts . . . (Barsalou, 1999, p. 584).

What does he mean by ‘the designation’? He never explicitly says, but how I have interpreted him is that he means the container/frame that the PS designates (the container/frame from which it came), and not anything about the PSs nature (e.g., what it resembles or what it can detect).

In his discussion, Barsalou uses the example of a PS that resembles the Empire State Building, and notes that it could stand for the building itself, or a skyscraper, or New York, or a clothing firm based in New York. As he says, the ‘designation’ of the symbol is what matters, and resemblance is not critical. So how would the system know which of the above options applies? The answer, I suggest, is that the system determines the meaning of the symbol by determining which container the PS is linked to, or which container it came from. It will mean skyscraper if the PS came from the SKYSCRAPER frame/container and it will mean New York if it came from the NEW YORK frame/container.

This echoes the point that I made previously, when I claimed that the same exemplar of a banana means banana if it is in the BANANA container, and fruit if it is in the FRUIT container. This notion of meaning being determined by which container a symbol is linked to is a much more concrete proposal than Barsalou saying it will depend on ‘how causal and contextual factors link it to referents in different contexts’, so Barsalou may not agree with my interpretation. Still, it is what I will be assuming going forward.

Another way of putting this is that the meaning of a PS is inherited from what it is pointing to, and not from any of the inherent properties of the PS itself. What the PS resembles is informative, but it does not determine what the PS refers to.

Barsalou says that resemblance is not critical in determining what a PS stands for, but I go further and say that it is entirely irrelevant. Say, for example, a PS that resembles a cat is linked to the DOG container/frame then what it would mean to the thinker is dog. Using
Barsalou’s terms, the PS designates the DOG container. If what the PS resembles is irrelevant, and what the PS designates is what does matter, it follows that DA-finsts, which don’t resemble anything, can be substituted for PSs, because they also designate/link to what does matter, namely, the frame/container. One might argue that PSs are superior because they contain information about what the object being thought of appears like, but DA-finsts, via their subject arms, also contain perceptual information about what the object appears like, with the additional benefit of keeping track of where the object is in the perception. DA-finsts are also superior in that they do not require the transmission of what is ultimately redundant data about what the object of the PS is like from the container system (associative cortex) to working memory (pre-frontal cortex).

In other words the system cannot tell the meaning of a PS just from its shape; to see what it stands for it is necessary to have information about what container it comes from or points to.

The main claim, then, that I am making in this chapter, is that, in the same way that I showed in the last chapter that we could substitute DA-finsts for Mentalese symbols, without detriment to the rest of Fodor’s LOT theory, so we can substitute DA-finsts for PSs, without detriment to the rest of Barsalou’s theory. Just as DA-finsts do everything that Mentalese symbols do, they also do everything that Barsalou’s discrete PSs do, in terms of the expression of propositional thoughts.

8.3.3 Barsalou opposes Fodor’s account

I have characterised the three accounts as all holding that expressions of complex propositions are composed out of discrete symbols in the working memory area, and that what differs between them is the nature of the symbols. Barsalou posits perceptual symbols in opposition to amodal symbols. Barsalou says that he is against the idea that perceptual systems pick up information from the environment and pass it on to separate systems that support the various cognitive functions, such as language, memory, and thought. I will argue that this view is fundamentally wrong. Instead, cognition is inherently perceptual, sharing systems with perception at both the cognitive and the neural levels. I will further suggest that the divergence between cognition and perception reflects the widespread
assumption that cognitive representations are inherently nonperceptual, or what I will call amodal (Barsalou, 1999, p. 577).

There are a number of issues here. The first is his objection to the idea of perceptual systems passing information on to separate systems that support language and thought. Obviously there must be some additional capacity over mere perception that allows creatures compose complex propositions because very few creatures (possibly only humans) can do so. That doesn’t mean there has to be separate systems that facilitate this additional capacity; it could be just an augmentation to existing systems. However, Barsalou does claim that PSs stored in the functional conceptual area (i.e., stored in containers/frames) are combined in working memory, on the fly. Working memory just is the name for an area where representations are operated on (e.g. juxtaposed, arranged, etc.). So I don’t think he can be objecting to the idea of information being passed on. What is distinctive about Barsalou’s account is that he claims symbols are juxtaposed against independent dimensions, the same way perceptions are, and this is why he thinks that cognition (which I’m interpreting here in the limited sense of being the expression of complex propositions) shares ‘systems with perception at both the cognitive and the neural levels’. He can contrast this iconic juxtaposition of symbols with systems (such as Fodor’s LOT and the DA-finst account) where symbols are concatenated sententially. We will come to this later in Sections 8.5.1. and 8.5.2.

The second point he makes is that he thinks it is wrong that the discrete symbols are inherently nonperceptual or amodal. But, as I have said, I claim that DA-finsts (which are amodal) can be substituted for perceptual symbols. The correct distinction, in my view, is the one between symbols that have fixed persisting meanings (like Mentalese symbols) and symbols that inherit their meaning from what they are pointing to (like perceptual symbols and DA-finsts).

In Fodor’s account, Mentalese symbols are amodal and word-like, and in Barsalou’s account the symbols are perceptual. That is, perceptual symbols pertain to some sensory modality—they are, at the very least, in a format that could underwrite perception. Perceptual symbols are ones that reveal something about their meaning by what they are similar to, or could detect in some sensory modality. Amodal symbols by contrast are arbitrary, in the sense that any symbol can arbitrarily stand for any meaning. There is no need for an amodal symbol to resemble or be capable of detecting what it refers to. Public language words are
usually amodal in this sense. There is nothing about the word ‘dog’ that resembles dogs. Any word/symbol could have been used, and that the symbol ‘dog’ is used is arbitrary. But note that because any symbol can stand for dogs in an amodal system, it is fine if the symbol does resemble its meaning in some modality. So the word ‘dog’ could be substituted in speech with ‘wuff-wuff’ which is like the sound dogs make, or, in writing, with a pictogram of a dog. The system would still be an amodal one, provided resemblance is optional and the matching of symbols with meanings can be arbitrary. It would still be amodal provided it is possible in the system for the sound ‘wuff-wuff’, or the pictogram of a dog to mean cat. The system can be considered perceptual or modal if, in the system, the symbol used must reveal its meaning through resemblance, or by virtue of what it can detect, in one or more sensory modalities. But we saw above that Barsalou acknowledges that resemblance is not critical. And I claim that resemblance is not necessary. Designation, i.e., the frame/container that the symbol points to or that the symbol originates from, is doing all of the work. Consequently amodal DA-finsts that also designate or point to the container from which they inherit their meaning can be substituted for PSs. This is to say that it is arbitrary whether a PS resembles the category it is designating/pointing to.

I agree that Barsalou’s PSs are different to Fodor’s Mentalese symbols, but not because PSs are modal/perceptual, and Mentalese symbols are amodal/nonperceptual, but because the meanings of PSs are, just like the meanings of DA-finsts, determined by the container that they point to, and not by the intrinsic features/shape of the symbol itself. The main claim of this chapter, then, is that DA-finsts can be substituted for PSs with no detriment to the rest of Barsalou’s account.

### 8.4 Simulations, simulators, frames and perceptual symbols

Before setting out how Barsalou claims propositional thoughts are expressed, I will discuss the most important elements in his account.

#### 8.4.1 Simulators

According to Barsalou, information pertaining to a category is gathered in the form of perceptual symbols (PSs) in frames. Frames are a kind of container, so, in container theory, they would equate to concepts. And, in container theory, the stored knowledge (discrete PSs
in Barsalou's account) contained in a frame constitute a conception for that category, as described in Chapter 3.

But that a frame is equivalent to a concept/container is not how Barsalou sees it. He claims that a concept is equivalent to 'a simulator' (Barsalou, 1999, p. 587), so for each concept a creature possesses (e.g., the concept of dog or of car) there is a corresponding simulator. What is a simulator? He says 'Together, a frame and the simulations it produces constitute a simulator.' (ibid., p. 590) And,

Thus, a simulator contains two levels of structure: (1) an underlying frame that integrates perceptual symbols across category instances, and (2) the potentially infinite set of simulations that can be constructed from the frame (ibid., p. 586).

I find this confusing on two counts. First there is a potential constitutes/contains confusion. In one place he says that a frame and the simulations the frame produces constitute a simulator, but then he says that a simulator contains a frame and the simulations that can be constructed from (the information contained in) the frame. Sometimes, of course, constitution and containment amount to the same thing. For example, a lamp both contains a shade and a stand, but one might also say that the lamp is constituted by the shade and the stand.

However, I have been arguing that we should view concepts as containers that are distinct from what they contain. Let us revert to the DVD example. A DVD is not constituted by the body of information it contains, and the movie/simulation (for example) that can be generated from that information. Using language carefully, a DVD is the particular medium used to store the body of information from which a movie can be constructed by a DVD player. The disk is not constituted by that information (and the simulations it generates), for many other mediums could store the exact same information, and they would not be DVDs. And a DVD still exists even if it contains no stored information. In other words, the container has an independent existence from what it contains. A simpler example is a box: a box is a container, but it is not constituted by what is put in it. The box is entirely distinct from its contents. Barsalou may claim that concepts are like lamps, and so are constituted by what they contain, but, in container theory, one must distinguish between the concept/container and the conception/contents of the container. Barsalou can mean by the term 'simulator' either (1) the entity constituted by the frame (plus its contents) and the simulations produced from that information or (2) the entity that contains the frame (plus contents) and simulations,
but has an independent existence. To comply with container theory he would have to mean the latter.

And there is a second confusion. We usually use the term ‘simulator’ to mean a mechanism that produces simulations, and that meaning fits neither of the above two possible meanings. A simulator is an entity that produces different output (simulations) depending on the information input to it (here this would be information contained in a frame). So, understood as a mechanism that produces simulations, a simulator is clearly not constituted by the stored information it makes use of, and the simulations it produces; nor does it seem right to say that a simulator contains the simulations it potentially generates. Rather, a simulator is constituted by whatever is the mechanism that takes information and transforms it into simulations. Again, using the DVD example, a simulator, in common parlance, would be equivalent to a DVD player, and not the DVD, or its contents; the DVD player produces simulations on a screen, in accordance with the information it reads on the DVD.

So, it seems that Barsalou is not using the term ‘simulator’ in its normal sense, but, rather, in the sense of, either, (a) the entity constituted by a frame and the simulations it produces, or (b) the container that contains a frame and the simulations it produces. This is unfortunate for it is easy for a reader of Barsalou’s paper to mistakenly interpret the term ‘simulator’ as meaning a simulation generator, given that simulations are so important in his account, and given that nowhere does Barsalou explain what is generating those simulations.

But perhaps, it is implicit, in what Barsalou says, that what he calls a ‘simulator’ is what generates simulations. Perhaps, Barsalou takes a ‘simulator’ to be a module (Fodor, 1983). The simulator/module would contain not just a frame and the simulations that could be generated, but also a simulation-generator. But recall that he says ‘a concept is equivalent to a simulator’. That means that, for every concept that we possess, we have the equivalent simulator. So, we have a DOG simulator, a CAR simulator, and so on, and each such simulator would include its own simulation-generator. This would certainly make the account massively modular (Carruthers, 2006). But it would seem to be much more efficient to have general purpose simulation generators (e.g. for creating dreams), which can take information from any frame (e.g. the DOG frame or the CAR frame), and produce simulations (e.g. rotate the dog/car image) of instances of the input category. And, second, even if we allow that there is a discrete simulation-generator for each category the subject possesses, it is still necessary to posit a separate general purpose complex simulation-generator that can produce
simulations that combine a number of categories (as we will see below, for example, a simulation of a balloon above a cloud).

I am not objecting to the various entities and phenomena that Barsalou posits in connection with simulators: there are, in my view, frames (containers that gather information according to single categories), perceptual symbols (stored knowledge contained in frames/containers), and simulation generators (mechanisms that produce simulations on the basis of the stored knowledge in frame/containers).

But, at least in container theory, the term ‘simulator’ does not describe a useful scientific kind, it is sufficient to make generalisations about just frames, perceptual symbols, and simulation generators. I will be emphasising Barsalou’s notion of a frame, and downplaying his notion of a simulator, in what follows.

8.4.2 Perceptual symbols

According to Barsalou, a perceptual symbol (PS) arises when selective attention focuses on part of a perception. He claims a PS consists does not consist of an exact copy of the percept, but rather a schematic of the percept, just as Locke (1690) proposed. A PS of a visual image might include information about vertices and edges, and what colours to fill in, rather than a point by point representation of the image. He also claims that selective attention may just focus on shape, and not colour or texture. An important point for him is that PSs are ‘dynamic not discrete’ (Barsalou 1999, p. 584), and by this he means that a PS can be used not just to reproduce the originally perceived entity, but can be altered by subsequent similar symbols (much like prototypes can alter when new instances are encountered), and by contextual information which strengthens some features and not others. He notes that PSs are componential not holistic. So a PS would have a shape component, an orientation component, a colour component, a texture component, and so on. A PS, then, is not like a conscious image, which holistically represents shape, orientation, colour and texture. For example, Barsalou claims that if you consciously imagine a triangle it will have a determinate orientation and shape. In contrast, the system can unconsciously represent just the shape of a triangle without specifying an orientation. The PS just includes information about the shape, and not the orientation. The system does not also have to ‘draw’ the triangle with a particular orientation.
Barsalou also claims that PSs do not have to represent specific individuals. The distinction between exemplars and prototypes is relevant here: exemplars are examples of specific individuals, while prototypes are summary averages of a number of examples of specific individuals; hence prototypes do not represent specific individuals.

Barsalou notes that PSs can be indeterminate. For example, a PS can represent a tiger without specifying a determinate number of stripes. The PS can just specify that a conscious image of the tiger should fill in a striped patterning, and not a specific number of stripes. Barsalou says that a system can represent a line as being present, without specifying the exact length, position or orientation of the line. So a PS for a triangle could be simply a specification to represent three lines, such that the first is joined to the second, the second is joined to the third, and the third is joined to the first. One might wonder how this is perceptual rather than amodal. Does it not require the system to have amodal symbols for categories like LINE, LENGTH, ORIENTATION, etc? If not, how does the system implement triangles that do have line length and orientation? Is this not exactly how amodal programming languages produce triangles? Barsalou says:

Consider the representation of triangle. Imagine that certain neurons represent the presence of lines independently of their length, position, and orientation. Further imagine that other neurons represent vertices between pairs of lines independently of the angle between them (Barsalou, 1999, p. 585).

Representing lines independently of length, position, and orientation seems to me to be exactly what would make a line representation amodal or nonperceptual. The same goes for neurons that represent vertices independently of angle. Barsalou goes on,

Three qualitative detectors for lines, coupled spatially with three qualitative detectors for vertices that join them, could represent a generic triangle. Because all of these detectors are qualitative, the lengths of the lines and the angles between them do not matter; they represent all instances of triangle simultaneously. In this manner, qualitatively specified neurons support perceptual representations that are not only indeterminate but also generic. (ibid., p. 585)

The idea seems to be that the neurons representing e.g. line do not arbitrarily stand for that category because they must be able to detect lines (regardless of length, position and
The representing neurons do not have to appear like lines, but must be capable of detecting or generating lines. For example, a connectionist network may be able to reliably detect lines but nothing in the network resembles lines with determinate lengths and orientations. It would be the LINE network because of what it could detect and because of what images it could simulate. Barsalou does say that attended to/captured percepts are matched to similarly represented stored knowledge (as opposed to amodal feature lists) to recognise them. Again, one might wonder whether the network is closer to an amodal feature list, rather than a perceptual ‘what-it-is-likeness’. This concern makes no difference to the DA-finst account; I raise the issue merely to show that Barsalou’s account may not be as purely perceptual as some think, and that, as a consequence, substituting amodal DA-finsts for perceptual symbols does not radically overhaul his account.

Roughly, Barsalou’s PSs are entities that can be used to detect what category things belong to, can be used to produce perceptual simulations, and can be used to stand for individuals and categories. According to container theory PSs are what comprise the stored knowledge in the container system. PSs are often perceptual exemplars and prototypes, but are also pieces of information that simulation generators can use to produce simulations (both sensory and motor).

That the container system is largely compatible with what Barsalou proposes can be seen from the following.

Perceptual symbols do not exist independently of one another in long-term memory. Instead, related symbols become organized into a simulator that allows the cognitive system to construct specific simulations of an entity or event in its absence (analogous to the simulations that underlie mental imagery) (Barsalou, 1999, p. 586).

Container theory emphasises that PSs are stored in a frame/container pertaining to a single category, rather than ‘organized into a simulator’. But otherwise, what Barsalou claims is compatible with container theory. Barsalou also claims, as we will see below, that PSs can be combined in working memory to create expressions of propositions, and container theory differs in that it claims that DA-finsts are combined in working memory.
8.4.3 Frames

Containers, in container theory, are a more abstract notion than Barsalou’s notion of frames. In particular, container theory is neutral on whether all of cognition has a perceptual basis. Nonetheless, Barsalou’s account of frames, is compatible with container theory. He says:

A frame is an integrated system of perceptual symbols that is used to construct specific simulations of a category (Barsalou, 1999, p. 590).

I have been putting it that a frame/container contains exemplars, prototypes and theories that are often in perceptual format (i.e., entities that are in the right format to directly compare percepts with). For Barsalou, the contents of frames are PSs that come about, for example, as follows,

On the perception of a first car ... the schematic symbol formation process in section 2.2 extracts perceptual symbols for the car’s overall shape and some of its components, and then integrates these symbols into an object-centered reference frame... (ibid., p. 590).

So all the PSs pertaining to the category car are gathered together in the CAR frame. If the window or handle of the front door, for example, was selectively attended to this will be reflected in the information in the frame. In addition, spatial information (e.g. where the windows are) is represented separately. Barsalou claims,

This distinction between levels of representation follows the work of Ungerleider and Mishkin (1982), who identified separate neural pathways for spatial/motor information and object features (also see Milner & Goodale (1995). Whereas the spatial representation establishes the frame’s skeleton, the content specializations flesh it out (ibid., p. 590).

When a second car is seen, it will cause a ‘reminding’ of the first car, he says. I prefer to put it that the stored information in the car container is capable of detecting/recognising the second car because it is sufficiently similar (it exceeds the threshold for belonging) to the first car. After this ‘reminding’ or recognition, information extracted from the perceiving of the second car gets added to the car frame/container, and so,
Following experiences with many cars, the *car* frame accrues a tremendous amount of information (ibid., p. 591).

Frames exist to gather together information pertaining to single categories (there is a frame for *car* and a frame for *eating* etc. For each concept a creature possesses, the creature will have a frame containing stored knowledge about that category that the concept pertains to, and that stored knowledge is used to recognise instances, draw inferences and engage in interaction.

**8.4.4 Simulations**

The simulations Barsalou has in mind are simulations of perceptual states, i.e., modellings or representings of perceptual states; when experienced consciously we call them imaginings, and, if experienced while sleeping, we call them dreams. In Barsalou’s theory, related [perceptual] symbols become organized into a simulator that allows the cognitive system to construct specific simulations of an entity or event in its absence (analogous to the simulations that underlie mental imagery) (Barsalou, 1999, p. 586).

Note the interesting wording: it is not the simulator that constructs simulations; rather the simulator allows the cognitive system to construct simulations. For Barsalou, a simulator is not a simulation generator, but the entity that either contains, or is constituted by, a frame that contains PSs, and the simulations that can be generated (by the cognitive system, not the simulator) on the basis of those PSs. In some manner, simulations get constructed, but Barsalou does not discuss the mechanism that constructs them. But it makes sense that the cognitive system does have general purpose simulation generators which can take PSs as input from any frame/container, and generate simulations as output. In what follows, I will use the term simulation-generator for the mechanism that constructs simulations, and it should be borne in mind that Barsalou means something completely different to a simulation-generator when he refers to a simulator.

According to Barsalou, information is stored when the system selectively attends to parts of a perception. Over time, a large amount of information pertaining, for example, to cars will come to be stored in the CAR frame. Using this stored information, the system can allow someone to make inferences about a particular car, as follows,
they can anticipate how the car would look from its side if they were to move around the car in the same direction as before; or they can anticipate how the car would look from the front if they were to go around the car in the opposite direction. Because they have integrated the perceptual information extracted earlier into an organized system, they can later simulate coherent experiences of the object (Barsalou, 1999, p. 586).

While it is not necessary for container theory, I do like Barsalou’s description of how stored information not only directly gives rise to inference, but also gives rise to simulations that go beyond what has been experienced in the past, and from which inferences can be drawn. Barsalou notes that this is by no means a new idea and that

Simulators have important similarities with other constructs. In the philosophical literature, Lockean (1690/1959) dispositions and Kantian (1787/1965) schemata are comparable concepts. Both assume that unconscious generative mechanisms produce specific images of entities and events that go beyond particular entities and events experienced in the past. Similar ideas exist in more recent literatures, including Russell (1919b), Price (1953), and Damasio (1994). In all cases, two levels of structure are proposed: a deep set of generating mechanisms produces an infinite set of surface images, with the former typically being unconscious and the latter conscious. Mental models are also related to simulators although they are not identical (K. Craik 1943; Gentner & Stevens 1983; Johnson-Laird 1983). Whereas a simulator includes two levels of structure, mental models are roughly equivalent to only the surface level, namely, simulations of specific entities and events. Mental models tend not to address underlying generative mechanisms that produce a family of related simulations (Barsalou, 1999, p. 586).

This generative aspect is emphasised when he discusses recognition:

Whereas many theories assume that relatively static, amodal structures determine category membership (e.g., definitions, prototypes, exemplars, theories), simulators suggest a more dynamic, embodied approach: if the simulator for a category can produce a satisfactory simulation of a perceived entity, the entity belongs in the category. If the simulator cannot produce a satisfactory simulation, the entity is not a category member (Ibid., p. 587).
While I have said that recognition takes place by matching stored information in the form of perceptual (as opposed to amodal) exemplars and prototypes, I do think it is likely that the matching process is more dynamic than just comparing stored information to perceived information. Instead, it is proposed that a generative process takes the stored information, and actively tries to produce a simulation that matches the perceived entity.

8.5 Complex propositions in Barsalou’s account

8.5.1 Perceived scenes

Now that I have detailed what Barsalou means by frames, perceptual symbols, and simulations, and why I consider frames, rather than simulators, to be the equivalent of concepts/containers, I can set out my interpretation of how Barsalou claims propositional expressions are implemented. His account differs, in some respects, between when the propositional expression is about a perceived scene, and when the propositional expression is about an imagined (simulated) scene. In this subsection, I will deal with perceived scenes.

Barsalou first describes how simple, rather than complex, propositions (ones that involve just a single category) are expressed, Figure 3 shows how the recognition process (which he calls type-token mapping) works. Recall the three areas that I discussed, at the start of this chapter. In Figure 3, the large rectangle, in the centre, corresponds to the perceptual area, i.e., where the scene, currently being visually perceived, is represented. The leftmost jet and the rightmost balloon correspond to frames/containers (or simulators in Barsalou’s terms) in the conceptual area. The more heavily outlined jet and balloon correspond to PSs (which Barsalou also calls simulations) in the working memory area. The idea is that the jet, in the perceptual area (the large central rectangle), has been recognised by the JET frame/container.
and this has caused a jet PS (a simulation of a jet) to be generated in working memory. Again, recall that I said DA-finsts mark-up perceptions. We can see marking-up occurs in Figure 3 also: note that the jet PS has two lines emanating from it, one linking it with the JET frame, and one linking it with the percept of the jet in the perceptual area. We could say that the jet PS (perceptual symbol) has two arms, just like DA-finsts have two arms. The key argument I am making in this chapter is that we could substitute a DA-finse for the middle jet PS, and nothing would be lost. This is because the meaning of the PS is inherited from the container that it points to, and what it resembles is irrelevant. The information that the PS gives about what the jet appears like is redundant, for that information is already available in the perceptual area which the PS points to.

Barsalou says that selective attention captures the jet shape in the perceptual area but does not specify any mechanism for selective attention. In the DA-finse account, selective attention is implemented by the subject arm of a DA-finse capturing the jet shape (this is, in effect, the system selectively attending to the jet shape). In Barsalou’s account, the jet percept is recognised by the JET container, and this causes there to be produced a jet PS, which is available for use in working memory. The DA-finse account is very similar. The subject arm captures the jet percept, which is then recognised by the JET container. Now, instead of producing a redundant jet PS, which is forwarded to working memory, the DA-finse, which is already in working memory, and whose subject arm has already captured the jet percept, simply links its predicate arm to the JET container.

I am not certain that Barsalou would agree with how I have interpreted him above. He says,

the line from the simulator [the JET container] to the simulation [the middle jet PS] stands for producing the simulation from the simulator. The line from the simulation to the perceived individual [the jet in the perception] stands for fusing the simulation with the individual in perception (Barsalou, 1999, p. 595).

This may very well be different from the picture I envisage, in which the central DA-finse keeps track of (keeps its fingers on), over time, with its two arms, both the JET container and the jet percept (even as it moves about). But if I have not interpreted him correctly, my account still draws inspiration from his account.
As Barsalou notices these type-token mappings implicitly constitute simple propositions. He says:

Most importantly, this type-token mapping implicitly constitutes a proposition, namely, the one that underlies ‘It is true that the perceived individual is a jet.’ In this manner, perceptual symbol systems establish simple propositions (Ibid., p. 596).

But Machery’s Propositional Desideratum requires concepts to be constituents of complex propositions. How does Barsalou propose to satisfy this desideratum?

The basic idea is that sequences of discrete simulations can be combined to form complex expressions. But the simulations/PSs are not standalone; the system cannot just examine the set of simulations, and determine what single complex proposition is being expressed. The linkings of the PSs to frames and percepts determines what complex proposition is being expressed.

Figure 4. The expression of the proposition *The balloon is above the cloud* (Reproduced from Barsalou, 1999, p. 595)

Figure 4 represents how the proposition *The balloon is above the cloud* is implemented for Barsalou. The rectangle on the left corresponds to the subject’s perception in the perceptual area. The three figures on the right correspond to the BALLOON frame/container, the ABOVE/BELOW frame/container, and the CLOUD frame/container, in the conceptual area. The rectangle in the middle corresponds to the working memory area, and contains juxtaposed PSs (simulations) produced by the information in the BALLOON and CLOUD containers/frames.

Here, Barsalou makes a very interesting proposal. Whereas, in Fodor’s account, and in the DA-finist account, the symbols in working memory are juxtaposed sentientially,
Barsalou, as I interpret him, juxtaposes the symbols iconically. Using Haugeland's principle set out in Chapter 2, this means that the symbols are plotted against independent dimensions (a reference frame). In Figure 4, the independent dimensions can be regarded as height and breadth, in the same way that visual perceptions plot information against these dimensions from the viewpoint of the subject. In the middle rectangle (which corresponds to the working memory area where PSs are juxtaposed) where the symbols are plotted matters: the fact that the balloon PS is plotted higher than the cloud PS is part of what makes it the case that the proposition being expressed is *the balloon is above the cloud*. If the proposition had been *the cloud is to the left of the balloon*, the cloud PS would be plotted to the left of the balloon PS.

How does Barsalou distinguish the proposition that *The balloon is above the cloud* from the proposition that *The cloud is below the balloon*? The plotting of the balloon PS and the cloud PS in Figure 4 is consistent with both propositions. According to Barsalou, it is because the ABOVE container (as represented by the two circles in the rightmost rectangle in Figure 4) is linked to the balloon PS (in Barsalou's depiction the balloon PS becomes circled with a thick outline) that the proposition being expressed is *The balloon is above the cloud*.

![Figure 4](image)

Figure 5. The expression of the proposition *The cloud is below the balloon*.
(Reproduced from Barsalou, 1999, p. 595)

Figure 5 shows how Barsalou proposes that the proposition that *The cloud is below the balloon* is expressed. Everything is the same as in Figure 4, except that the thick circle is now around the cloud PS, and is linked to the BELOW frame/container.

Once more, note that the PSs in the middle rectangle can be taken to be double-armed. Both PSs link to a frame/container that determines the category that is being thought of, and to a shape in the perceptual scene. We could substitute DA-finsts for the PSs, and no
information that is essential to the determination of the proposition being expressed. The predicate arm of a DA-finst points to the frame/container, and the subject arm of the DA-finst points to the percept of the object. The only difference is that the DA-finst, itself, tells you nothing about the entity being thought of. But that information is available from what the arms of the DA-finst are pointing to, in any event.

Barsalou is attempting here to make good his claim that thought is inherently perceptual, and his solution is ingenious. In the working memory area, the individual symbols are perceptual, in that they reveal what the entity in question appears like, and the juxtaposition of symbols preserves spatial relationships in a perceptual/iconic manner, because the symbols are plotted against the same kind of independent spatial dimensions (height, width, depth) that they would be in visual perceptions.

Barsalou is aware that perceptions, or iconic representations of scenes, represent many propositions simultaneously, as I set out in Chapter 2, and that it takes something extra to pick out just a single proposition. His account makes the expression of the proposition, in the working memory area, iconic (and by extension somewhat perceptual) in nature, but Barsalou’s account can pick out just one proposition because of how the symbols are ‘marked-up’ (e.g., the balloon PS is marked up as being in the above position; the balloon PS is marked up as being from the BALLOON container, and so on). For this reason, it is plausible that the system can interpret the working memory juxtaposition of symbols as expressing e.g. *The balloon is above the cloud* and distinguish that proposition from, e.g., *The cloud is below the balloon*. However, just as in the DA-finst account, the proposition that is being expressed cannot be determined just from the working memory area; the perceptual and conceptual areas must also be adverted to.

Barsalou’s proposal is ingenious. However, the perceptual/iconic nature of the juxtaposition of the PSs is not doing any work that a sentential juxtaposition of the PSs would not also achieve. All of the work is being done by the ‘marking up’. Compare Barsalou’s iconic juxtaposition with the DA-finst account, which is sentential. In sentential accounts, the order of the symbols is all that matters, and not their position relative to independent dimensions. The system will have rules about how the order of the symbols is to be interpreted. In English, the subject comes first, the relation second, and the object last (e.g., ‘the balloon above the cloud’). Order in speech is determined by earliest in time. Order in writing is usually determined from left to right, and from top to bottom. It is perfectly
possible for a language to adopt rules, so that the relation comes first or last (e.g., above (balloon, cloud) or (balloon, cloud) above). What matters, in determining the proposition being specified, is just the symbols’ position relative to each other, and not their position relative to independent dimensions, such as height or depth. There is no need, for example, for the word ‘balloon’ to be actually above the word ‘cloud’. Sentential accounts offer all the meaning that Barsalou’s iconic/perceptual account offers, but without the additional requirement of having to plot information against independent dimensions. Barsalou might argue that his account has the benefit of explicitly plotting what it means to be above or below, but note that, in Barsalou’s account, what determines if it is the above relation or the below relation is not the position of the symbols in working memory, but how the ABOVE/BELOW frame/container is linked. Also note the circling (marking up) of the cloud in the central rectangle in Figure 5, which indicates the below relation, is duplicated in the leftmost rectangle, i.e., in the perceptual area where information is already plotted perceptually.

The DA-finst account is perfectly compatible with Barsalou’s idea that symbols are juxtaposed iconically (that is isomorphically to how the objects that the symbols represent are juxtaposed in the world) in working memory. A system could arrange DA-finsts in above and below positions against a reference frame, for example. But arranging them sententially is simpler and the perceptual area already plots the objects spatially (e.g., plots the balloon above the cloud). Equally, Barsalou’s PSs could be juxtaposed sententially in working memory and be able to pick out a single proposition because of what the PSs are pointing to, or designate.

8.5.2 Imagined/simulated scenes

However, not all propositions are about perceived scenes. We often express propositions about imagined scenes, and Barsalou offers a different account of how these are expressed. Instead of the three areas (the perceptual area, the conceptual area, and the working memory described in Section 8.1.1) that were required for the expression of propositions about perceived scenes, Barsalou only envisages two of those areas to be necessary for the expression of propositions about imagined scenes. In effect, he proposes that the working memory area can also be used as the perceptual area.11

11 Or he can propose what amounts to the same thing: the perceptual area does double duty as the working memory area.
Imagined scenes (simulated perceptions) involve a process that is often the reverse of the process for perceived scenes. In perceived scenes, typically, the system will start with the perception (e.g., a scene in which there is, amongst other things, a balloon, a jet, and a cloud), then recognise (associate concepts with) objects in the perceptual scene, and finally combine a number of concepts into a propositional thought (e.g., the thought that There is a balloon above a jet to the left of a cloud). In imagined scenes, the system, often, starts with a propositional thought involving a number of concepts, then accesses the containers corresponding to those concepts, and, finally, generates the imagined scene using the stored knowledge about what things appear like that is contained in the container. For example, you may be asked to imagine A balloon that is above a jet to the left of a cloud. The cognitive system is then capable of using the perceptual stored information in the BALLOON, JET, and CLOUD containers to generate the imagined scene.

One of the distinctive features of predictive processing (Clark, 2013) is that it is envisaged that even in perceived scenes some (or even much) of what is perceived depends on top-down, generated simulations, rather than just bottom-up sensory recordings. Container theory can certainly accommodate such top down influences; instead of simply recognising what has been perceived with information in containers, the system can also generate simulations of what it predicts is being perceived based on information in containers.

Figure 6 shows how complex propositions are expressed about simulated (imagined) perceptions, rather than occurrent perceptions, in Barsalou's account.
C of Figure 6 shows the expression of three proposition, namely, from left to right *The balloon is above the cloud, The jet is above the cloud, and The cloud is above the jet*. D of Figure 6 expresses the proposition that *The balloon is above a jet that is to the left of a cloud*. The major difference between Figure 6, and Figures 4 and 5, is that the larger rectangle to the right of each propositional expression in Figure 6 represents the working memory area, and not the perceptual area as is the case in Figures 4 and 5. In D of Figure 6, the rectangles on the left represent the BALLOON, JET, CLOUD, ABOVE, and LEFT frames/containers. The idea is that PSs are produced from the frames/containers and are then plotted in the working memory area *iconically* (i.e. against the independent dimensions of height and breath in this case. In other cases, the independent dimension of depth can be added, if for example behind or in front of is part of the expressed proposition. Note how the independent dimensions can be very easily derived from, for example, an arrow on a map.

Figure 6 Simulations or imaginings that expresses proposition in Barsalou’s account
(Reproduced from Barsalou, 1999, p. 593)
indicating N, or on a box marked 'this way up'. The point is that, by arranging the PSs iconically, they are arranged isomorphically to how the objects they represent would be located in the actual environment. Whereas, if the PSs are arranged sententially, these isomorphisms are not preserved.

Barsalou's proposal is that the juxtaposition of the PSs in working memory, which expresses a single proposition, also amounts to a perceptual simulation of the scene. In contrast, the DA-finsts account juxtaposes DA-finsts sententially in working memory, and a simulation generator then, optionally, produces the imagining (the perceptual simulation) in the perceptual area.

In Barsalou's account, by arranging the PSs iconically, there arise, as was claimed in Chapter 2, very many propositions that can be expressed about the juxtaposed PSs. For example, D of Figure 6 picks out the single proposition _A balloon that is above a jet to the left of a cloud_, rather than _A cloud that is to the right of a jet below a balloon_, or any of the many other propositions that could be expressed about the simulation, because of the way the iconic working memory area is 'marked up'. It is because the ABOVE container/frame links the above symbol to the balloon PS that part of the expression is that _the balloon is above the jet_. And it is because the balloon PS is linked to the BALLOON container/frame, and not, for example, the LOLLIPOP container, that the thinker is thinking of a balloon, and not a lollipop. And so on, for the other objects and relations in the scene.

Compare C of Figure 6 to Figure 4. In Figure 4, there are three areas: the working memory area, where PSs are combined iconically; the conceptual area, which links the PSs to containers/frames that determine what category is being thought of; and the perceptual area, which links the PSs to entities in the perceived scene. In Figure 6 there is only the working memory area and the conceptual area. Because it is an imagined scene, and not a perceived scene, Barsalou does require the perceptual area. And, because the PSs are plotted iconically, the working memory area can be taken to be a simulation of a perception.

In Section 8.4.1, I said that, in the case of perceived scenes, it was redundant to juxtapose the symbols expressing the proposition in an iconic manner, because the spatial relations thereby revealed were already disclosed by the perceptual area. In the case of imagined simulations, there is not already a perception to refer to (and mark up). For example, there is not already a perception with a balloon plotted above a cloud. In Barsalou's account, the arranging of the PSs against the independent dimensions of height and width, in
working memory, results in the system simulating, and not just stating, the spatial relations. In Barsalou’s account, the method of expressing the proposition simultaneously generates an imagining of the spatial relations. And the fact that the PSs are, themselves, perceptual in nature (the balloon PS, if it is a visual PS, will look like a balloon) reinforces Barsalou’s claim that the expression of the proposition is inherently perceptual. On the face of it, one might even say that Barsalou’s account is more parsimonious than the DA-finst account, since it only requires the working memory and the conceptual areas, and not the perceptual area.

So, how does the DA-finst account deal with simulations (imaginings or episodic memories)? It takes propositions expressed about simulations to involve the same three areas, as propositions expressed about occurrent perceptions. DA-finsts are combined sententially, in the working memory area, with the predicate arms of the DA-finsts pointing to containers in the conceptual area. But the DA-finst account also takes imaginings to be simulated in the perceptual area, and the subject arms of the DA-finsts point to the entities in the simulation revealing where they are, and what they appear like.

In perceived scenes, the subject arms of DA-finsts, typically, first capture parts of the perception, and, then, once recognition has taken place, the predicate arm captures the container that contained the recognitional information. In imagined scenes it is the reverse: first the predicate arm captures a container (e.g., the BALLOON container), and, then, the subject arm is filled with the information about how the object (the balloon) is to appear in the simulation generated.

The idea is that the same perceptual area that already has the tools to record perceivings makes use of those tools to simulate perceivings. So, for example, when I imagine something visually, according to the DA-finst account, I use the same regions that underwrite visual perception. There is evidence that visual imaginings do indeed, often, activate the same regions in area V1 of the cortex as are activated in perception (Kosslyn 1994, Farah 2000, Kosslyn & Thompson 2003, Klein et al. 2004). What I am suggesting is that the independent dimensions of area V1 (how the dimensions of height, breadth, and depth are built into the system), against which sensory information is plotted, can be the same dimensions that simulated information is plotted against. This claim is by no means uncontroversial, but it is perhaps the dominant view in psychology. The ‘imagery debate’ is a major debate in psychology and philosophy, pitting Kosslyn against Pylyshyn (see Chapter 6
of Pylyshyn (2003) for his objections to the quasi-pictorial view). It has not been decisively resolved. It can be characterised as Kosslyn arguing in favour of the idea that visual imagery is quasi-pictorial, and, hence, iconic, whereas Pylyshyn argues imagery is sentential. I do not want to go into this debate in this thesis, and there is no need to, in this chapter, because both Barsalou and I side with Kosslyn. The difference between us is that Barsalou regards iconic imaginings to take place in the working memory area, while I regard iconic imaginings to take place in the perceptual area. But we both regard imaginings to be perceptual and iconic (visual perceptions, for example, are quasi-pictorial) in nature and not sentential. However, in the DA-finsts account, propositions are expressed sententially in working memory, and in Barsalou’s account they are expressed iconically in working memory (in conjunction with the PSs being ‘marked up’ by the container/frames they are linked to).

I claim that there are clear advantages in having propositions expressed by DA-finsts that are separate to the imagining of the proposition.

First, sometimes when we express a proposition that is not about what we are currently perceiving, there is no accompanying simulation or imagining. For example I might think that all dogs have tails, but I may not imagine or simulate any particular scene or dog. In the DA-finsts account, this is explained by there being DA-finsts in working memory linked to containers, but no additional imagining was generated in the perceptual area. But other times, when I express the proposition all dogs have tails I may go onto imagine some dogs with tails. According to the DA-finsts account, information from the DOG and TAIL containers will have been used to generate the imagery in the perceptual area. In Barsalou’s account, however, there will have to be some perceptual/iconic arrangement of the dog PS, and the tail PS, that can both account for the singling out of that particular proposition, and that gives rise to imagery. There will also have to be an explanation for why sometimes we experience thoughts without accompanying imagery.

Also consider what Clark (1998) calls second order cognitive dynamics, i.e. thoughts about thoughts. For example, I may wonder if my thought that The balloon is above the cloud is correct; or if my thought that Mary thinks that the balloon is above the cloud is correct. In the DA-finsts account, the expressed proposition that is thought about is very compact: it is the ‘sentence’ of DA-finsts juxtaposed in working memory. The object of the thought has a compact concrete existence. Whereas in Barsalou’s account the object of the thought is a whole imagining/simulation, together with the frames/containers that purport to pick out a
single proposition in the imagining. Moreover, the unwieldiness of simulations becomes worse with even higher order thoughts. Take, for example, the proposition that I think that John thinks that Mary thinks that the balloon is above the cloud. In sentential accounts it is simply a matter of juxtaposing additional symbols according to rules. In a perceptual/iconic account it is necessary to find some way of iconically juxtaposing the symbols to indicate that it is a thought about a thought about a balloon being above a cloud.

In addition, DA-finsts arranged sententially allows for easy expression of abstract propositions. For example, the proposition that It was a victory for democracy can be implemented by DA-finsts that link to the DEMOCRACY and VICTORY containers. In Barsalou’s account there would need to be PSs for the categories, and those PSs would somehow have to be plotted against independent dimensions to produce a simulation of the whole proposition. Barsalou (1999) and Prinz (2002) give examples of how abstract categories can be represented perceptually, but they don’t explain how they can be arranged iconically, or indeed what that would add.

8.6 Conclusion

In this chapter, I have set out how Barsalou’s perceptual symbols inherit their meanings from what they are linked to, or designate, and not from what they resemble or can detect. I claimed that DA-finsts, which also inherit their meanings from what they are linked to could be substituted for perceptual symbols, while the rest of Barsalou’s account could be retained.

I, then, set out how Barsalou proposes that complex propositions are expressed, and claimed that, once again, DA-finsts could be substituted for perceptual symbols, in those expressions. I noted that, in Barsalou’s account, symbols were juxtaposed iconically to produce simulations of propositional thoughts, and said that I preferred the DA-finст account, in which the symbols that constituted the propositional thought were juxtaposed sententially.

So the upshot of this chapter, and the previous one, is that I claim that DA-finsts (disposable symbols) can be substituted for both Fodor’s Mentalese symbols, and Barsalou’s perceptual symbols, in the expression of propositional thoughts, with the result that the DA-finст account, in effect, fuses those two accounts.
In the next chapter, I turn to the question of how container theory explains how new concepts/containers are acquired, and how they come to have intentionality.
Chapter 9
Containers and intentionality: private concepts

9.1 Introduction

This chapter deals with the meaning of concepts. I have been referring to, for example, the BANANA container and the DOG container. But, so far, I have not said much about how containers come to pertain to single categories. It has proven very difficult to account for how concepts come to be about determinate single categories, in a purely natural causal manner, that is, without the system already knowing a correct way to carve the world up into such determinate single categories. Take, for example, the categories mouse, vole, and shrew. The question arises how can it come about, purely causally, that a container gets to be about just mice and not, for example, mice and voles? And how does error get into the picture: what makes it wrong, for example, for a shrew to activate the MOUSE container?

In order to answer these questions, I will be distinguishing between three types of concepts: (a) ‘private’ mental concepts, which need not be shared with others, (b) ‘social’ mental concepts, which are shared with others in a community, and (c) ‘word’ concepts which are part of a public language. Interpreting meaning differs for each of these three types of concept.

In this chapter, I will focus mainly on private concepts. I will start with the example of an owl, a creature that, I hypothesise, lacks social or word concepts. I will set out how the owl acquires new concepts, how those concepts acquire intentionality, and how error or misrepresentation is accounted for.

The aim in this chapter, and the next, then, is to show that it is entirely possible to naturalise private and social concepts (but not word concepts). That is, it is possible to reduce private and social concepts to physical and computational entities and processes, without appeal to semantic notions. I claim that concepts are containers of stored information, and that the containers, and information inside them, ultimately have a physical basis that can, in
principle, be pointed to. However, I will be acknowledging that naturalising (reducing to scientifically describable, non-semantic properties) a single correct interpretation of what a concept means will, often, not be possible. But I will claim that there is no need for a single correct interpretation to be available in order for a concept to be naturalisable, or for the creature to be able to intelligently deploy the concept.

Container theory can be regarded as a vindication of the idea that Putnam calls 'mentalism'. He says

An ideal belief-desire psychology would be isomorphic to (a part of) the computational description of what goes on in the brain. Make that assumption and you have mentalism in its most recent form. Mentalism is just the latest form taken by a more general tendency in the history of thought, the tendency to think of concepts as scientifically describable ('psychologically real') entities in the mind or brain. And it is this entire tendency that, I shall argue, is misguided (Putnam, 1988, p. 7).

I do not think that it is misguided to think that concepts, conceived of as containers, are scientifically describable. We have to separate out two things here.

First, I claim that private and social mental concepts are constituted by discrete containers, which have a physical basis in the brain, and I claim that the information in those container/concepts also has a real physical basis. There is one container for each concept a creature possesses; they are 'psychologically real' entities in the sense that, in principle, one could point to the neurons that underwrite each discrete concept/container, and the information therein. Public language word concepts are not 'real', by contrast, in the sense that there is no single physical place that can be pointed to that contains the stored knowledge of that concept. I will discuss this in more detail below.

The second thing that has to be separated out is what the meaning of a container/category is. The reasons Putnam gives for thinking that mentalism is misguided centre on his claim that what a concept means is a matter of interpretation, which is a social phenomenon, and there may be no fact of the matter as to which interpretation is correct. Consequently, the interpretation is not reducible or naturalisable. I agree with him that, often, it will not be possible to determine a single interpretation of the meaning of a concept/container that all can agree upon. But I am claiming just that a concept, conceived of as a container, and the information within, can be reduced to physical vehicles. I will
demonstrate with examples below that there is no need for a private or social concept (i.e. container of stored information) to have a precise or unambiguous meaning that all can agree on for it to have an entirely scientifically describable physical basis, and for it to be deployable by a subject. For example, theorists might disagree on whether a concept/container pertains to mice, or to mice and voles. But, according to container theory, it will still be possible to give a scientific description of the container and the information contained in the container.

All that is required is to simply describe scientifically how meaning gets into the picture from non-semantic elements. So, for example, to describe how creatures come to recognise a percept as belonging to a category, and how this recognition gives them interactional and inferential knowledge, and to describe how error or misrepresentation can get into the picture. In container theory, it is possible to describe scientifically how a creature comes to recognise an entity it perceives by activating one container, rather than another, and how the meaning it gives to that entity depends on what information is in the container it has activated. If the creature had activated a different container, it would be entertaining a different meaning. The containers, and the information inside, that give the meaning have a real physical basis, according to container theory. What may, often, not be possible is to give a precise and unambiguous interpretation of what category that meaning pertains to in scientifically real terms. But, below, we will see that unambiguous interpretation is not necessary for a creature to experience that it is determinately thinking of this (because it activates container X) and not that (because it doesn’t activate container Y).

9.2 On interpretation

According to container theory, containers are what have meaning, and not the contents of containers. This is the point that I have made previously. The same exemplar of a banana can be in both the BANANA and FRUIT containers; the exemplar, by itself, does not have the meaning of banana, rather it will be an exemplar of a banana if it is the BANANA container, and it will be an exemplar of a fruit if it is in the FRUIT container. But what determines whether the container stands for banana or for fruit? There have been many approaches to answering this question for mental representations that could be adapted to container theory. For example, it is the BANANA concept/container if it contains a definition of bananas (Smith & Medin, 1981), or a description of bananas (Frege 1892, Russell 1911), or if its evolutionary (Millikan, 1989) or developmental (Dretske 1981, 1988) function is to
activate only in the presence of bananas, or if it is activated by bananas in normal conditions (Stampe, 1977); and correspondingly for the FRUIT concept. None of these solutions are able, at this time, to deal with all of the objections that they face. I do not propose to go into all of these approaches because I aim to show in this chapter that, in the case of private concepts, the issue can be simply sidestepped.

Private concepts are concepts that can be possessed by creatures that do not possess a public language, and/or do not have to share their concepts with conspecifics. According to container theory, the intentionality of private concepts is determined solely on the basis of the information that is contained in the container. If the container contains only information pertaining to bananas, then an outside interpreter (who already has a concept of what bananas are) may judge that it is the BANANA container. If the container contains information about other fruits, then the interpreter may judge that it is the fruit container. However, the creature that possesses the container will simply think, when the container becomes activated, that it is thinking about this-kind-of-thing, i.e. the kind of thing that the various pieces of information gathered in the container are about. The creature does not have to specify, precisely and unambiguously, what this kind of thing is in order for it to experience a determinate meaning. The determinateness is given by the fact that a particular body of information is contained in the activated container.

The account that I am putting forward is an informational theory (Dretske, 1981) rather than a strict causal or nomological theory (Fodor, 1990). The account shares many features with causal accounts, because the explanation of how the information gets to be in containers (see Chapter 4) is usually that encounters with instances of the category in the world caused it to get there. But it is not a problem, in container theory, if the information got placed in the container accidentally (e.g., a swampman situation (Davidson, 1987)) or by deceptive input (a brain in a vat (Putnam, 1981)), or by direct intervention (an evil scientist directly manipulating neurons). The actual cause of the information that comes to be in a container is not an essential part of the story; rather, what matters is simply what form the information takes. This is why, for example, there is no mystery about a container referring to unicorns, even though there are no such causes in the world.

In container theory, the key idea is that private concept/containers contain the information within themselves that allow for the interpretation of what they mean, what they are about, what they refer to, what intentionality they have, what category they pertain to. But
there is no need for the creature that is deploying its concept (bringing to bear the stored information in the activated container) to interpret what that meaning is, beyond it is 'this-kind-of-thing'. It does not have to be the case that there must be a single, unambiguous interpretation that must be settled on before the subject can mean things by its containers/concepts.

9.3 The owl and private concepts

In order to bring out how, according to container theory, meaning comes about for private concepts, I will use the example of an owl; a creature that I hypothesise can have private containers/concepts, but does not possess a public language or social concepts. I want to demonstrate how the meaning of a private container/concept will depend on the information that is contained in the container, and how the meaning can shift, as the information changes.

Imagine a particular owl encounters its first mouse. The visual percept of the mouse is captured by the subject arm of a DA-finst. That percept is input to the owl’s container system, but fails to find a matching exemplar/prototype in any existing container. That is, the percept is not sufficiently similar to recognitional information in any existing container. As a consequence, the cognitive system opens a new container (call it the ‘M’ container). The information from the percept in the subject arm of the DA-finst is used to form an exemplar (information that can be used, in the future, to recognise members of the category) that is placed in the container. In addition, perhaps an episodic memory of the mouse running away might be placed in the container (giving functional information about how members of the category can be predicted to behave); there may also be information about where the encounter took place. Now, let us imagine that the next five encounters that activate the container (exceed the similarity threshold – match sufficiently closely an exemplar in the container) are all with what are in fact shrews (cf. Cummins, 1989), and each of these shrew percepts gets stored as an exemplar. Again, further behavioural and locational information may be stored.

Then the next five encounters that activate the M container are all with what are, in fact, voles, and, again, these percepts are stored in the M container as exemplars, together with behavioural and locational information (including, by now, for example, information about how they can be caught, what they taste like, and so on).
What happens when a mouse is next encountered? Perhaps, it is sufficiently different from even the first exemplar that the threshold is not exceeded, so the cognitive system opens a new container, and stores the new mouse exemplar in it. However, more likely is that the percept of the mouse is still close enough to the first exemplar to exceed the threshold, and so this new exemplar is simply added to the M container.

Let us take stock of what the container is like now. The first thing to note is how thoroughly causal the account is, so far. There is nothing normative, no question of what ought to be the case, no question of the system knowing what categories there are in the world that it should mirror, and no need for the system to appeal to any semantic notions. The cognitive system simply causally individuates percepts, and matches them mechanically against stored recognitional templates. Upon recognition, the system can mechanically add further information to the container, based on the recognised instance. Whenever a percept fails to exceed the recognitional threshold of any existing container, the system mechanically opens a new container, and places in the container information for recognising future instances. The information added to containers is perceptual and sensorimotor in nature. That is, the system can plot the information against a reference frame, without having to know which categories were involved. As more encounters with instances occur, the containers come to be filled with more ways of recognising and interacting. The functional information might now include some information about how the creatures behave, how they can be caught, where they are commonly found, what they smell like, what they taste like, and so on.

What is the meaning attached to the private M container? It was first opened because of an encounter with a mouse, but since then only shrews and voles have been encountered. As far as the owl is concerned, it does not care (yet) about how Mother Nature carves up the world into natural kinds. It simply has a single container (and therefore a single primitive concept). As far as the owl is concerned, when it encounters something that activates the container, it will simply ‘think’: there is another one of those things that belong to this container, and which sound like this, smell like this, and taste like this, and will behave in these kinds of ways.

Insofar as the container’s recognitional information can recognise mice, shrews or voles, an outside observer might declare that the semantics attached to the container is mouse or shrew or vole. Another interpreter might declare that the container refers to mouse, because the first thing that caused the container to be opened was a mouse (Donnellan 1972, Putnam
1975, Prinz 2002). Another interpreter might argue that what mattered was the critical period when extra attention was given to what should and should not belong to the category/activate the container (Dretske (1981)). This interpreter might claim that, apart from one mouse, all the other activators of the container, in the critical period, were shrews, so the container refers to shrews.

Note that finding a principle that determines unambiguously what category the private container refers to is an entirely intellectual exercise that will have no impact on the owl's ability to deploy the information in its container/concept. As far as the owl is concerned, the container refers indexically to this-kind-of-thing, where 'this' is pointing to/focusing upon the exemplars in the container (the pre-activation recognitional information) and the behavioural (post-activation functional) information. The functional information, for example, will inform the owl how to successfully catch and eat the creatures that are recognised. The owl, if it could express complex thoughts, might be inclined to say to the observers/interpreters: 'Let me know when you reach a consensus on what my container refers to; in the meantime I will continue to recognise this kind of thing and bring to bear this kind of useful functional information.'

In container theory, the category that the container pertains to (what meaning it has) need not be static. It can change over time. Before discussing the shifting of meaning, and how prototypes enter the picture, let me discuss the idea that the cognitive system is a feedback system: the cognitive system has evolved to be constantly comparing what it predicts will happen, based on stored information in containers, with the incoming detected signals of what actually happens in the world. As I will discuss more fully below, this is how error gets into the picture: the generalisations in the container may result in predictions that do not match actual outcomes. If a percept of a mouse is recognised by the M container, the information in the container can be used to predict, for example, what the mouse will look like from another angle, or what it will taste like, or how it will attempt to evade capture. Having made its inference or prediction, the system then compares this information with what actually occurs. That is, the internal predicted signal can be matched against the incoming perceived signal (what the taste is actually like, or how it in fact ran away rather than froze, for example).

So the system is constantly mechanically comparing predicted signals with incoming feedback, and when error occurs, i.e., when the predicted and incoming signals do not match,
the system will endeavour to update and revise the information in containers, so as to reduce error. The system tries to maximise the extent to which the functional (post-activation) information in a container results in accurate predictions about the things detected by the recognitional (pre-activation) information in the container. A very basic way of achieving this mechanically is to strengthen the weight that is given to information that results in accurate predictions, and weaken the weight that is given to information that results in inaccurate predictions.

So far we have just talked about exemplars in the M container. Let us introduce prototypes, now, in order to see how containers might come to split. Prototype theorists hold that a summary representation (a kind of average) is formed out of all the exemplars so far encountered. So the prototype for the M container will have been formed out of the five shrew, five vole, and two mouse exemplars.

In addition to this matching of predicted outcome signals to sensed incoming signals, we can also imagine that the system engages in top down 'data mining' on the contents of its containers, and, in the M container, the system finds that five of the exemplars all share the feature of having a long snout. The prototype formed from just these exemplars turns out to be different from the prototype formed from the remaining exemplars. In addition, where these five long snouted exemplars were found was different from where the rest of the exemplars were found. In combination, this may be enough to trigger the opening of a new container (call it the M1 container), and all of the long snouted exemplars together with their associated functional information are placed in it. These will all, of course, have been shrew exemplars.

What intentionality will the M and M1 containers now have? An interpreter will base her judgement on what information is in the containers. As the M1 container only contains shrew exemplars and prototypes, it seems likely that an outside observer (already possessing the concept shrew) will interpret it as the SHREW container. As for the M container, an observer might interpret it as a MOUSE/VOLE container because both types have been recognised as belonging. Another observer might argue that the container refers only to voles, and mice are wrongly included. But, of course, for the owl, the outside interpreter's judgement makes no difference. For the owl, the M container determinately pertains to this-kind-of-thing, the kind of thing that looks like these (visual exemplars), tastes like this, behaves like this, and so on. And the M1 container determinately pertains to the kind of thing
that looks, tastes, and behaves like this. It is irrelevant whether outside interpreters can agree on just what this kind actually is. The owl’s cognitive system does not have to possess semantic knowledge of which determinate category the kind belongs to, in order to recognise entities as belonging, or to produce inferences and predictions, based on the stored knowledge, about the recognised entities.

Let us next suppose that the owl starts to encounter mice and voles equally often, and the percepts of the mice and voles continue to exceed the threshold of the M container. Many observers would agree that the container pertains to mice and voles. However, let us suppose that voles taste much nicer than mice to the owl. Let us suppose, for example, that voles are nutritionally better for owls, and, hence, owls that are genetically disposed to prefer the taste of voles to mice would be favoured by natural selection. In causal terms, we can say that the taste of voles delivered higher valued ‘positive reinforcement’. The cognitive system will seek to maximise positive reinforcement and minimise negative reinforcement. Data mining reveals that the visual appearance prototype (recognitional information) formed from just the exemplars that taste ‘nice’ (functional information) differs sufficiently from the visual appearance prototype formed from the remaining exemplars (perhaps in addition to ‘tasting nice’ these exemplars tend to have a shorter tail). Knowing what something tastes like is generally post-activation functional information, to be applied once a particular instance has been recognised (e.g., by sight, sound, or smell). The visual information on its own was not enough to trigger the opening of a new container, but when combined with the taste information it is enough. As a consequence, all of the ‘bad tasting’ exemplars are moved to a new container (the M2 container) and these are, as a matter of fact, all mouse exemplars.

So we now have a situation where the owl has separate containers for what are, in fact, shrews, mice and voles. The owl can now ‘think’ about these categories separately. By ‘think’, I mean, here, that the owl can, for example, recognise a percept with the M container and, hence, bring to bear the information in the container for prediction and inference. Again, all of this is achieved entirely causally, with no advertence to any semantic or intentional properties. Things only happen if a threshold or a similarity metric is exceeded, or by connections being strengthened/weakened on the basis of whether predicted signals matched incoming sensed signals. It is not the case that the system has separate containers for these three species because it ought to, nor because they are three different species; it is because of metrics that depend ultimately on predictions matching outcomes, and on maximising positive reinforcement and minimising negative reinforcement.
In fact, it is possible for the owl to come to have two containers that both pertain to mouse/vole. Say both short-tailed mice and voles taste nice, and long-tailed mice and voles taste bad, then it is to be expected that the owl will have come to have one container for nice tasting mice and voles (which give positive reinforcement) and another for bad tasting mice and voles (which give negative reinforcement). This-kind-of-thing will be good tasting mice and voles in the case of the M container and bad tasting mice and voles in the case of the M2 container. All that matters, for the owl’s cognitive system, is that the predictions and inferences (e.g. about expected reinforcement) made about the recognised entities match the incoming signals. What does not matter is that the containers pick out natural or non-disjunctive kinds, for example, or kinds that share an essence. Of course, it would be very unusual if tail length was a determinant of common taste, rather than common DNA. But if it was, that would be the way the owl’s cognitive system could be expected to organise the information in containers.

I claim that specifying what category a container pertains to, very often, may be a matter of interpretation, and depend on the criteria that the interpreter uses. And there may be no fact of the matter as to which interpretation is correct. In addition, it is possible for the intentionality of a container to shift over time, as new information is added to containers, and as containers are split into more precise or useful versions.

9.4 Recognition and functional information and how error occurs

I have described how the meaning of a concept/container for the owl itself is just the kind of thing that looks, tastes, behaves, etc. like these pieces of stored information. Now one objection that might be made to this formulation is that it makes error impossible. There is nothing that makes it wrong, for example, for mice to activate the M1 container (which as described above contains only shrew exemplars); either the mouse percept will exceed the system’s threshold for belonging, or it won’t. As I have set it out above, if mice exemplars continue to be added to the M1 container then this will make the M1 container pertain to mice and shrews (or so an interpreter might judge). In that case, why would it be an error for the mouse to activate the M1 container? Where there is representation, there must be the possibility of error (Dretske 1985, Cummins 1989, Crane 2003). The way out of this problem is to see that containers do not just contain recognitional information; they also contain functional (interactional and inferential) information. It is true that, under my formulation,
there cannot be errors in recognition — something either exceeds the threshold for belonging or it does not. But error can arise in terms of the functional information failing to apply to the entity that has been recognised.

Let me say a little more about the distinction between recognitional and functional information. Containers exist to gather information that pertains to single categories. This information can be divided into two distinguishable (albeit overlapping) kinds. There is recognitional information that is deployed pre-activation of the container. The container is activated if an entity is recognised as belonging. And there is functional information that is deployed post-activation of the container. Inferential, predictive, interactional, and behavioural information becomes available once the container has been activated. For example, a visual recognitional exemplar or prototype in container M recognises a percept of a vole, which causes the activation of the M container, making available the functional information in the M container for making predictions and inferences about the vole (for example, how to catch it, what evasive action it will take, what it tastes like and so on).¹²

There is evidence that the cognitive system is constantly matching predicted signals to incoming sensed signals (see, for example, Grush 2004, Frith 2007). The predicted signals are formed based on functional post-activation generalisations in an activated container. What caused the container to be activated was recognitional (pre-activation) generalisations in the container (typically exemplars and prototypes). The post-activation functional generalisations licence predictions about outcomes (e.g. if I eat it, it will taste like this; if I perform action x I will capture it; if I move here it will move there, and appear like this, etc.). The creature will also be in a position to choose what to do based on the values it attributes to the predicted outcomes (e.g. I will get reward X (tastes nice) if I eat it; I will get punishment Y (pain sensation) if it bites me). The cognitive system will test if the actual outcome matches the

¹² It is not that the information in the container is either functional or recognitional. It is that information in the container can be used either pre-activation or post-activation. For example, what might normally be considered functional information can be used to recognise. Say there is information in a container that this kind stings, and should therefore be avoided. If I see one of these things, then once I have recognized it, I can use the information that it stings as post-activation functional information and infer that I must avoid it. But, if I feel a sting, the information that this kind stings can be used as recognitional information to activate the container. If I activate the dog container because of perceiving a barking sound I may infer that that creature has four legs because my visual recognitional information (e.g. information that can be used to simulate what dogs look like) indicates that members of this category have four legs.
predicted outcome (does the actual taste match the predicted taste; does the perceived appearance match the predicted appearance after movement, and so on). If the predicted outcomes do not match the (perceived) actual outcomes then error has occurred. This is how error gets into the picture. It is not that container M stands for voles and this is a mouse so error has occurred. Rather it is that container M stands for the kind of thing that looks like *this* . . . , tastes like *this* . . . , and so on, and the mouse does not taste like *this*. This is entirely a matter of comparing physical signals: the signal of the predicted taste, and the signal of the detected perceived taste. Even though the mouse percept did activate the M container (the percept was sufficiently similar to one of the exemplars, for example), it was an error, because the mouse did not taste as predicted. In fact the mouse would taste as predicted by the information in the M2 container.

The owl doesn’t have to be able to state in one word (it doesn’t of course even have any words) what the type of the container is for error to occur. There doesn’t even have to be a single word for the type of this category. The owl doesn’t have to go on and give necessary and sufficient conditions, or a definition, or a foolproof test, for being of this type. Nor does the category have to be unambiguous and non-disjunctive. When, in the above example, the owl started off by including all of its information about shrews, mice, and voles in the one container, that was not an error. Say the generalisations are *<looks like these* exemplars, smells like *these* exemplars, can be caught by *these* kinds of actions, will try to escape by *these* kinds of behaviours, will taste like *this* or *this>*. If the post-activation generalisations apply equally well to all three types that are recognised, then the owl is not making any error in not being able to tell mice, shrews, and voles apart. That is because its predictions of what will occur (e.g. what taste it will perceive) will be matched by what does occur (e.g. what taste it does perceive). But if the generalisations cause predictions that do not come about, then that is the source of error.

Imagine that the owl perceives what is, in fact, a mouse, and, following recognition, the M container is activated. The owl, now, post-activation, has available all of the information in the M container, which it can use to predict outcomes, to make inferences, and interact with the object recognised. For example, the owl might predict that *this* token (which it perceives head on) has a short tail, because that is what the information in the active container indicates. However, when it gets to be in a position to see the tail, the prediction turns out to be an error, in the sense that the incoming signal does not match the predicted internal one. The container may contain an emulator (Grush 2004) that informs it how to
catch the mouse. This emulator may predict that the mouse will flee. The emulator can be run offline, as it were, to predict what will happen before it occurs. If, however, the mouse freezes, then the incoming perceptual signal will not match the emulated prediction. Once the mouse is caught, the owl may predict that it will taste like *this* (based on stored taste information in the container). If in fact it tastes differently, then the incoming perceptual signal will not match the predicted signal.

I do not mean by error that the outcome is bad for the creature. If the owl recognises a fox, and activates its FOX container, and uses the information in that container to predict that the fox will attack it, and the fox does in fact attack it, then this counts as a good prediction in the sense I mean (the predicted outcome does occur), even though the outcome is very bad for the owl.

Or imagine a creature that can recognise mushrooms with its Q container, and use the information therein to predict that they taste nice, that they will make hunger go away, that they can be found at location Y, and so on. Say the creature perceives a hunger sensation, which activates its HUNGER container, making all of the information in that container available. This includes the information that things recognised by *this* container (the Q container) will make hunger go away if eaten. The activated container includes information about where to find mushrooms. The creature goes to that location. On arrival, the cognitive system can compare what it predicted would be the case with what is perceived to be there. It has made no error. There are mushrooms there. The creature eats one predicting that it will taste like *this*. It does taste as predicted, so again there is no error.

But now imagine that one of the mushrooms is, in fact, poisonous. It has never before encountered a poisonous mushroom. Soon, the creature feels very ill, a feeling which is very negative for it. But all of its predictions, so far, have not resulted in error.

However, the creature has encountered a very unpleasant feeling, which it did not predict. Because the negative feeling is severe, the system may take steps to ensure, if possible, that it can recognise what causes this feeling before it occurs again. The system will be disposed to open a new container to gather together information about the source. On closer examination, it may be possible to distinguish this poisonous mushroom from the prototype of the ones it ate without problem before, and so be able to recognise them, and store functional information that gives rise to predictions that will be matched by outcomes.
All of this is learned by experience. It is a matter of organising containers so that predicted outcomes about what has been recognised match actual outcomes.

According to container theory, in the case of private concepts/containers, the container is not marked as being the X container. An outside observer can mark a container as being, for example, the MOUSE container, but for the owl this makes no difference. The owl’s cognitive system is only concerned that this functional information applies to the kind of thing detected by this recognitional information. What makes it an error for a piece of information to be in the container is if it results in false predictions. The owl’s container may contain recognitional information that detects mainly mice, but also some voles. Is it a MOUSE container, or a MOUSE/VOLE container? That doesn’t have to be settled, in order to account for error. What matters is whether the functional information applies to what is recognised. If the generalisations apply equally to mice and voles, then there is no error (no false predictions will occur). But if the functional generalisations apply only to mice, then there is error (false predictions will occur). For example, if, in fact, mice taste nice, and voles taste bad to the owl, but the functional information in the container indicates that this type of thing tastes nice, then a false prediction will occur (in the sense that the predicted signal will not match the detected signal) when a vole is recognised as belonging to the category. The source of error is not that the container was the MOUSE container all along; the source of error is the information that this type of (recognised) thing tastes like this. The cognitive system is in the business of reducing surprise, where surprise is to be taken as a measure of the difference between the predicted physical signal and the actual received/input physical signal. The system attempts to reduce surprise by updating either the functional information or the recognitional information, whenever error occurs.

9.5 Three sources of error for private concepts

If errors in prediction occur, there are three possible sources: (a) the recognitional information is at fault, (b) the functional information is at fault, and (c) the information itself is not at fault, but the recognising process has gone wrong (due to e.g. bad light).

Say, for example, the owl predicts that a mouse it perceives will behave like this (e.g. a particular evasive course of action), but that prediction turns out to be wrong (the mouse freezes rather than flees). Now, there are three things that might have gone wrong. The first is that the recognitional information is wrong (mice exemplars should be in a different container
because mice always freeze, and it is voles that run away). Second, the functional information might be wrong. So the entities recognised by this container do normally flee, but sometimes they also freeze. This information needs to be added to the container; the mouse was correctly recognised as belonging to the container, but the functional generalisations need to be augmented. By a combination of trial and error and data-mining, the cognitive system can determine whether to augment the existing container with the freezing information, or to change the recognitional information. As discussed above, opening a new container might occur where, e.g., exemplars with short tails coincide with exemplars that taste good, and exemplars with long tails coincide with exemplars that taste bad.

The third possible source of error occurs when there is nothing wrong with the stored information; rather the problem is with the recognising process. It is a feature of threshold exceeding mechanisms that the threshold can be adjusted by feedback, so as to achieve what is optimal for the system, in terms of allowing false positives and avoiding false negatives. In general, an intelligent system is better off allowing a certain amount of false positives, rather than insisting on one hundred per cent accuracy in detection rates, especially if the penalty for false recognition is small (cf. signal detection theory (Swets, 1964) – for a Bayesian approach see Schonhoff & Giordano, 2006). In other words it is worth it for an owl to chase the occasional leaf, if that allows it to catch more voles. By ruling out all ‘leaf’ causes, the system is also likely to fail to detect many causes that were actually voles.

In the above example, where voles taste nicer than mice, the owl will catch more voles if the vole threshold is set at a level that allows some mice to be detected, than if the threshold is set at a level that excludes all mice, for this latter will exclude many genuine voles as well. But this leads us back to the original problem: how do we establish a principled reason for saying it is wrong for mice to activate the M container, but correct for voles to activate it, while still allowing that it is possible for mice to activate it (or indeed shrews or baby rats).

Error (misrepresentation) can occur if the recognition process is less than ideal. Let us say that the actual threshold is set at 75%\(^1\) – i.e. be 75% similar to some recognitional information (e.g. an exemplar) in the container – and this lets in mice and shrews encountered

---

\(^1\) At night-time, for humans, the cognitive system is likely to have to reduce the threshold further (say to 60%) for there will be little colour information and foveal vision will be relatively weak. The same will go for things seen far away because of lack of detail. In other words the cognitive system is likely to work with an ever shifting threshold in order to get the best results.
in less than ‘optimal conditions’ (for example, bad light, far away, or unusual viewing angles). But if the threshold had been 100% or 95% then only (some) voles would exceed the threshold, and mice and shrews would not. And we now have a type of asymmetric dependence. Mice can only exceed the lower threshold because voles can exceed the higher. The mouse activation counts as a misrepresentation, because ‘an ideal’ mouse would not exceed the ‘ideal’ higher threshold of the vole container. In essence, if you want to interpret what semantics is attached to a container, you look at the best recognitional exemplifications of the category in the container, along with the functional information in the container.

The ‘best exemplifications’ will be those exemplars stored within the container that come closest to achieving a score of 100%, when matched against the prototype for the container. Assuming I have seen enough exemplars in ‘good’ or ‘optimal’ conditions, these ‘best’ exemplars are most likely to be the ones that score highest against the prototype. The prototype, and therefore the best exemplifications, will have to be separable into sub-prototypes for each of the modalities. If I can only see a vole, but not hear or smell it, then only its visual features should count towards passing the threshold. In addition, for containers like the DOG container, there would have to be a number of different visual sub-prototypes, because dogs can vary in appearance so much.

Why does the owl think it is chasing a vole, and not a vole/mouse, when in fact it is chasing a mouse in less than optimal conditions? According to container theory, the owl does not think in terms of vole or vole/mouse; it just thinks it is chasing one of those where those is pointing to the best recognitional exemplifications in the container and to functional information in the container such as tastes like this and behaves like this. An outside interpreter may judge that the ‘one of those’ is a vole, rather than a vole/mouse, on the basis of the information in the container. For example, if the owl catches the mouse, the owl will find that some of the predictions it makes, on the basis of the information in the container, about what the mouse will look like from another angle, or how it will behave, or what it will taste like, will not match the sensed outcomes. An outside interpreter would judge that ‘this kind of thing’ is in fact a vole because the information (the exemplars etc.) best recognise voles and the generalisations (particularly what the taste will be like) apply to voles. Because the owl’s cognitive system employs a threshold designed to maximise its success rate, it will suffer occasional disappointments, but in the long run the owl will catch more voles, than it would if it only chased instances that scored 100%. 


172
Note the interplay between the container contents, and what the container represents. What the container represents is given by the ‘best’ interpretation of what the container contents are about. But, at the same time, there is no guarantee that the container’s recognitional and functional information is in a finished state. The cognitive system is constantly performing checks on and updating the container system. The cognitive system’s aim is to recognise all and only entities that the same set of stored inductive knowledge can be applied to. What counts as the category of the container is always a work in progress. What count as the best exemplifications can always change due to updating, and we saw above how containers can split.

The cognitive system, in order to minimize surprise (have its predicted signals match as closely as possible the incoming sensed signals), will constantly be updating, adjusting and revising the contents of containers, in order to maximize the extent to which the functional information applies to the instances detected by the recognitional information. Registering that a prediction has failed will cause the system to mechanically adjust its recognition mechanisms or its stored generalisations. Data-mining may cause a new container to be opened if a threshold is exceeded. Two containers may be merged into one, if the system discovers that the information in both pertains to the same kind of thing. The contents of containers may be constantly changing, as the cognitive system revises, updates and adds new information. These changes may cause an outside observer to say that the meaning of the container has shifted. But from the perspective of the creature, if it is activating the same container over time, it just thinks ‘here is another one of this kind of thing’ regardless of what the outside interpreter judges. Because the recognition threshold may be set at a level that allows false positives to occur, from time to time, predictions made on the basis of functional information in a container may not be matched by outcomes.

9.6 What is a single category?

I have been saying that concepts are containers that pertain to single categories. We are accustomed to thinking of single categories as things like water, dogs, spatulas, and so on. Intuitively, they seem discrete and non-disjunctive to us. Perhaps they have an essence, something that can be used to determine what all and only Xs are.

But I am claiming that private containers do not have to be about such intuitive distinct kinds. The restrictions on what a container can pertain to does not depend on the
extension having some intrinsic discreteness. The restrictions come about solely from the fact that the functional information must apply to the entities detected by the recognitional information. According to container theory, a container pertains to a single category, even if most interpreters would that category to be disjunctive (e.g., mouse/vole).

In other words, the direction of fit is that if there is a discrete container, then there is a single category; not if there is a discrete category then there should be a single container. So what can count as a single category is just about anything (things to take on a picnic, dogs chasing cats, the conjunction of water and twin earth water). What determines, in container theory, if it is a single category (what makes it a single category for a creature) is simply whether there is a separate container, in which is gathered together recognitional and functional information.

Of course, it is generally the case that the kinds of things that good and useful generalisations can be made about will be the kinds of things that we intuitively think of as single categories. The set of things that look, sound, smell, taste, and behave like the stored information in this container will, most likely, coincide with a single species, if the information is sufficiently detailed. But there is no error being made if, in fact, the container pertains to two or more species, provided that the generalisations apply to the entities that are recognised. The question is do the inferential and interactional generalisations apply to the entities that the container recognises? Do the predictions made match the perceived outcomes?

In container theory, the container pertains to a single category. In the case of private containers, the contents of the container can be used to interpret what that single category is. As the information in the container changes, the aboutness may change too. It is not the case for private concepts that the container has an independent intentionality that holds steady regardless of the information in the container. That is, indeed, the way it is for social concepts, when the container is headed by a public language word, as we will see in the next chapter. But, for private concepts, the interpreted meaning of the container can change, depending on what information is currently in the container. If it originally contained only mouse information, but now contains an equal amount of mouse and vole information then the meaning will be interpreted as having shifted.

For private concepts, the ideal set of generalisations may arise where all the members of the kind have the same sensory appearance, the same constitution, and the same set of
behaviours. But it is rare that a category is this determinate, and, even if a category is, a cognitive system cannot be guaranteed to find all of the relevant generalisations. In our sophisticated scientific community, we have come to realise that same constitution is a better way of recognising things that behave the same way, than same appearance is. But for creatures like owls that do not have the benefit of communal knowledge (much less scientific communal knowledge), and do not have access to constitutional information, same appearance is a generally good way of detecting things that behave the same way, and that a common set of inferences may be made about.

The set of containers that a creature possesses will tend to emphasise either same appearance, and/or same function/behaviour, and/or same constitution. But this is not because containers must be about discrete non-disjunctive kinds, but because getting the (post-activation) functional information to provide accurate predictions about the entities detected by the (pre-activation) recognitional information is much more likely to occur when the kind has a common appearance or function or constitution.

The only requirement, then, for being a single category is that there is a single container that corresponds to it. If a container exists, then that fact alone means that the container pertains to a single category. It is irrelevant if an outside observer judges that that ought not to be a single category.

Equally, note that containers are atomic. What determines the number of atomic concepts that a creature possesses is the number of discrete containers that the creature possesses. BROWN COW is an atomic concept if there is a single container devoted to gathering recognitional and functional information about this category. And it is a complex concept if it is made out of two discrete containers, one for BROWN and one for COW.

9.7 Conclusion

In this chapter, I argued that it is possible to naturalise private concepts: containers, and their contents, can be reduced to physical entities. In principle, one could point to the neurons that realise the representational vehicles. I also described how containers will seem to have determinate meanings for a creature: the creature will experience thinking about the kind of thing that appears like this, behaves like this, that this can be done to, and so on, where in each case this refers to information contained in the same container That is, what category the concept/container represents depends on the information that is in the container.
In a sense, what the container pertains to depends on the inferential role (Field 1978, Block 1986) of the information in the container. Inferential role will be discussed in Chapter 11.

Specifying unambiguously and precisely just what the category of a container is, however, is a matter for an outside interpreter, and interpreters can bring different rules of interpretation to the table. There may be no way of settling disputes about which is the best way to interpret the category that the container pertains to. Consequently, it will not be possible, in many cases to reduce the meaning of a container/concept to something as precise as category X. But, I claim, this failure to specify an unambiguous and precise interpretation, when it occurs, makes no difference to the possessor of the concept. She will be able to deploy the information in the container to recognise, predict and infer entirely successfully.

I also set out how error or misrepresentation is possible. It does not occur because the container is about category X, and the recognised entity does not belong to category X. Error or misrepresentation occurs when a predicted outcome based on stored functional information does not match the actual outcome. There are three sources of error: (a) when the recognitional information recognises entities that the functional information does not apply to; (b) when the functional information licences predictions that do not apply to entities that are recognised; and (c) when the recognitional and functional information are in accord but the recognising conditions are less than ideal.

However, these private concepts that do not have to be communicated to others in a community are not the most common type of concept possessed by humans, and in the next chapter we turn to the intentionality of social and word concepts.
Chapter 10
Containers and intentionality: social and word concepts

10.1 Introduction
In the last chapter, I described how containers that are private concepts can be reduced to physical entities without recourse to semantic notions, and how those concepts come to have meaning for the possessor. However, we will see, in this chapter, that concepts that are shared with other members of a community, or that are part of a public language operate differently.

It is common to distinguish between public language word concepts and mental representation concepts. I propose that, in the case of social creatures, mental representations can be subdivided into private concepts and social concepts, making three kinds in total. I will now set out, for these three kinds of concept, how they are acquired, how their meanings are interpreted, and how error arises.

10.2 Private, social, and word concepts
10.2.1 Private concepts
In order to compare them with social concepts and words, I will say some more about private concepts. Private concepts are the only kind of concept that non-social creatures can have. Each private concept a creature possesses is realised by a container for gathering recognitional and functional information. A new private concept is acquired when a new container is opened. This occurs whenever information (e.g., a visual percept) in the subject arm of a DA-finst fails to activate any existing container, i.e., when perceptual information is not recognised by any of the creature's existing recognitional information. Note that, in effect, the creature's cognitive system hypothesises just that this subject matter belongs to
some new category. It does not hypothesise that the subject matter belongs to category X, which would entail that the system already knows that there is a category X. This is in contrast to an outside interpreter, who must already possess the category X in order to interpret that the concept represents X. The visual percept will form the basis of the first exemplar that is placed in the container. In addition, perhaps some functional information will be placed in the container (for example, an episodic memory of how the creature behaved). The exemplar can now be used as recognitional information that will cause the container to be activated, if a sufficiently similar percept is encountered. If the container is activated, the creature can bring to bear its functional information about how the entity will behave. Over time, more and more recognitional and functional information is gathered, and provided they result in good predictions the new concept will become firmly established.

The information that is gathered together can be entirely perceptual and procedural, and so does not depend on the creature having any other containers/concepts, or semantic notions, in order for it to be able to recognise and interact with instances. These concepts/containers have original intentionality (Haugeland 1981, Searle 1983, Dennett 1987, Fodor 1987). That is, the meaning is not derived or copied from someone or something that already has the meaning. What the concepts/containers stand for is given by ‘this-kind-of-thing’, which in turn is given by the kind of thing that the recognitional information is about and that the functional generalisations apply to. Just what that kind is, is a matter of interpretation. Perhaps there will be a number of rival interpretations, and no fact of the matter settles which interpretation is correct. But that makes no difference to the creature that possesses the container; the creature will consistently pick out a particular kind of thing to which the set of generalisations in its containers will apply. Sometimes the creature will get it wrong (make errors, misrepresent), but mostly it will apply its generalisations to entities to which they do apply. The container pertains to this-kind-of-thing where the category of container A will be experienced by the creature as being distinct to the category of containers B, even if containers A and B are co-referential.

14 Cf. Margolis and Laurence’s (2011) arguments against Fodor (2008) that you can learn a new category without already having a concept of that category. The example they give is of a complex concept made up of the different steps of a dance that is being learned. The learner does not have to already have the hypothesis that the steps to be combined are e.g. AABBAC in order to confirm that hypothesis. He must simply hypothesise that there is some combination of steps and try out this particular combination.
Private concepts do not need to satisfy the publicity desideratum (Fodor 1998, Prinz 2002). For example, two owls do not need to share the same set of containers. One owl can have a container that gathers information about what are, in fact, mice and shrews, with a separate one for voles. Another owl can have a container that gathers information about mice and voles, with a separate one for shrews. And a third owl can have a container that gathers information about voles and shrews, with a separate one for mice. Because they do not communicate or share their concepts, the owls do not have to ensure that they each have roughly the same set of concepts.

Error arises for private concepts in the three ways I have described in the previous chapter.

10.2.2 Social concepts

But once a creature is part of a community, and especially a linguistic community, things change substantially; in addition to private concepts such creatures can possess social concepts. Social concepts, just like private concepts, are containers of stored knowledge pertaining to a single category, but with the difference that the concept will be shared with others.

In what follows, I will be taking the paradigm of a social concept for humans to be a container that is headed by a public language word. For example, the MOUSE container would be headed by the word ‘mouse’ in English. Whereas, the owl could have a private MOUSE container independently of the concepts that any other owls possessed, a human’s MOUSE concept will be shared with many others in the community. In order for successful communication to take place, it is incumbent on each member of a community to possess roughly the same set of social concepts as other members possess.

Although what follows will be a discussion of social concepts in a human linguistic community, I do think that non-linguistic creatures can have social concepts. Consider, for example, the alarm calls of vervet monkeys. The experiments of Cheney and Seyfarth (1990, 1992) are consistent with the idea that they possess social concepts. Vervets share the same calls for different kinds of predators (snakes, eagles and leopards). On hearing the ‘leopard’ call, they respond by climbing high into a tree to branches that wouldn’t take the weight of the leopard. On hearing the ‘eagle’ call, they look up and then hide in the nearest bush.
Young vervets often overgeneralise calls, and adults will, often, ignore their calls. As they grow older, the vervets become much more accurate at recognising when they should sound an alarm. We can say that they come to share the same concepts with other members of the community, and so successfully communicate when danger is present. Through the sanction of being ignored when they make wrong calls, their stored recognitional and functional information will become ever closer to that possessed by other members of the community. For example, we can posit that the recognitional information in the container that contains the leopard alarm call may at the outset recognise any animal with four legs, but after sufficient encounters the exemplars and prototypes in the container come to be very similar to those in the corresponding container of other community members.

A vervet that spots a leopard knows which call to make, because (what outside interpreters would interpret as being) the LEOPARD container contains that information. Any vervets who hear the alarm call will activate their LEOPARD container, and can use the information in the container to inform them of how to respond. They can also use the information in the container to recognise the leopard, if it is in view. By recognising the leopard, a vervet monkey can ensure that it runs away from the leopard, and not from, say, the nearby gazelle.

What makes a social concept/container different to a private concept/container is that it is incumbent on each possessor of the same concept to activate it in roughly the same circumstances, and to possess roughly the same functional information about how to react. These calls are, of course, very similar to public language words, even if most theorists would argue that the monkeys do not possess a fully-fledged language.

But it may also be the case that, even without words or calls, social creatures may possess social concepts. For example, a pack of wolves may ensure that each member’s set of containers about prey should be largely the same, by sanctioning non-conforming members. Here the meaning of each container is set by the community’s practices, and not by the information that is currently in the container, as would be the case with private concepts/containers. In other words, there is a publicity requirement (Prinz, 2002) for social concepts: what a container should be about is determined by the community’s practices. We can imagine an interpreter saying something like:

The information in this young wolf’s container is about rabbits and snakes right now. But we will find that the container comes to be just about rabbits
over time, because whenever the wolf seeks, for example, to instigate a hunt of a snake, the rest of the pack will sanction the young wolf, by ignoring it. Over time the pack will ensure that all of its members share the same set of concepts, where sociality is a factor.

So it seems plausible that social concepts/containers can be possessed by non-linguistic creatures, but in what follows I will be discussing social concepts that are inherited from public language words.

A social concept inherited from a word is, in effect, a private concept/container ‘headed’ by a public language word. By ‘headed’ I mean that the container can be activated by the word. For example, hearing or seeing the word ‘mouse’ would activate the MOUSE container, just as perceptual encounters with actual members of the category would. Just like a private container, a social container gathers recognitional and functional information that goes together. But the container will also contain the means to recognise the word for the category.

These containers are opened on the same principle as private concepts, except that what no match can be found for is a word rather than a percept. For example, when a child hears the word ‘zebra’ for the first time, because the word fails to activate any existing container, a new container will be opened into which will be placed the means for recognising the word in the future. The container is now poised to gather information that pertains to the category zebra. Typically, ways of recognising zebras (e.g. visual exemplars) and generalisations about zebras (e.g., that they live in Africa) will come to be stored in the container, just as would be the case for a private concept container. Indeed, it may be the case that the child already possesses a private container for zebras (she saw pictures of them in books without encountering the word), and at some point the cognitive system will come to merge the private concept with the container headed by the word ‘zebra’, when it is realised that both containers pertain to the same category.

But, despite the only difference between a social and private concept being that the former is headed by a public language word, the consequences for intentionality are substantial. Once headed by a word, the concept no longer has an aboutness determined by the recognitional and functional information in the container, as is the case for private concepts. Error no longer occurs just in case the functional generalisations do not apply to
what is recognised. The aboutness of a social container, once it is headed by a word, is inherited from what the word means for the community.

Imagine I have a private concept/container that allows me to very reliably detect voles, and all of the generalisations in the container pertain to voles. As a private concept (that is, one not headed by a word) the intentionality of the container is simply the kind of thing that the recognitional and functional information pertains to. Perhaps all outside interpreters would agree that that kind is \textit{vole}.

But, say, I hear the word ‘mouse’ for the first time when someone points to a mouse. Now, if my percept of that mouse happens to activate my private vole container, then the cognitive system will head that container with the word ‘mouse’. Now, what happens to the intentionality of the container? It automatically pertains to mice even though all of the information in the container pertains to voles. Why? Because of the publicity desideratum for social concepts. Once I am a member of a linguistic community, if I want to avoid miscommunication and sanction from my community, I must ensure that my social concepts have the same intentionality as my community has for them. In private concepts, the requirement for a good concept is just internal consistency between the recognitional and functional information. In social concepts, not only must I strive for consistency between my functional and recognitional generalisations, but, in addition, those generalisations must be, broadly speaking, about the same kind of thing as the community takes the word to refer to.

Social concepts (containers of stored information which are headed by a word) have disquotational intentionality (Davidson, 1990). If the container is headed by the word ‘mouse’, then it refers/pertains to the category \textit{mouse}. This appears to give us the kind of unambiguousness of reference that so many seek for concepts. But that lack of ambiguity is largely illusory. For the word ‘mouse’ simply refers to what the community take to be mice, and we will see in the next subsection that that is, once again, a matter of interpretation.

Although the category that a person’s social concept/container pertains to is given by what the community means by the word that heads the container, nonetheless, what the person \textit{takes} the word to mean is given by the recognitional and functional information in the container. Whether the person is right about what it takes the word to refer to will depend on how closely the information in the container matches what the community takes the word to refer to. This will be evidenced by how successful the person is in communicating. If miscommunication arises, or the person is sanctioned for their use of a concept, the cognitive
system will be disposed to change the information in the container to reflect what the
community indicates it ought to be.

Also note that it is possible for an individual to discover that the community is wrong
about what the community takes the concept to pertain to. For example, we can imagine
Hubble thinking something along the lines of:

I know that my community thinks all of those points of light in the night sky
are stars and planets, but I can convince them that I have evidence that some of
those things they recognise as stars are in fact galaxies.

Here Hubble is able to shift his community’s meaning for the word to his meaning. But he
avoids miscommunication and sanction, by presenting his evidence and argument for why the
meaning should shift.

Note how easily the publicity desideratum is achieved with the notion of social
concepts in communities with a public language. What members of a community will share
are containers headed by the same word. The fact that two individuals both have containers
headed by the word ‘cat’ means that they share a concept, even if one of them has entirely the
wrong kind of recognitional and functional information in his ‘cat’ container (for example, he
thinks that ‘cat’ refers to mice). But even where no miscommunication occurs, two
individuals can have very different sets of recognitional and functional information about
cats: what ensures publicity is that both containers are headed by the same word.

10.2.3 Word concepts

Public language words fall under the third type of concept. I claim that words, just
like private and social concepts, should be regarded as (nominal) containers of recognitional
and functional information. However, unlike mental concepts, where the container and the
information within have a real physical realisation, the information gathered by words cannot
generally be pinpointed in any precise manner. There is no physical container in which all of
a community’s generalisations about, say, mice or water is to be found. Words do not gather
a community’s stored information in a physical location. Of course words themselves have a
real physical realisation, when, for example, they are spoken or written, but what they
(nominally) contain (the community’s generalisations) is just a matter of interpretation of the
community’s overall dispositions.
What does my community mean by the word ‘mouse’? It means the kind of thing it recognises as mice and the kind of thing all its mouse generalisations are about. And what are they? No one book, or no one brain, can be said to contain all of such generalisations. An interpreter of what the community takes the word to mean will have to resort to saying things like ‘they appear like this, behave like this, are constituted by this,’ where in each case ‘this’ is pointing to or focussing on some concrete example that she takes to be typical of what the community would say, or what its dispositions are. Of course some words, like ‘odd (number),’ will have definitions or tests for belonging that are not disputable by rival interpreters. But most words are not like that.

Note that, in container theory, the intentionality of words is just like the intentionality of private concepts insofar as it is a matter of the consistency of recognitional and functional information. What a word means is a matter of what kind of thing the community takes the generalisations to be about. As the information in the nominal container changes the meaning/aboutness of the word can shift (for example, the word ‘star’ might once have had a meaning that included planets and galaxies). The difference is that, in the case of a private concept, the recognitional and functional information, and the container that they are in, have a real physical implementation, whereas, for a word, not only is the information spread about in multiple locations (many minds and books, and in modern times the internet etc.), but also an interpreter has possibly to give much greater weight to the beliefs of experts and authorities (e.g., religious or political), and so on. For most words, no single interpreter can evaluate all of the ways of recognising members of the category, nor all of the generalisations that the community makes about the category. As Putnam points out, meaning is holistic and involves very intangible things, such as estimating simplicity (which itself is not a single factor, but different things in different situations), and weighing simplicity against our desire for successful prediction and also against our desire to preserve a certain amount of past doctrine. It involves having a nose for the ‘right’ trade-off between such values . . . (Putnam, 1988, p. 11).

I agree with Putnam’s claim that there is no prospect for reducing single correct interpretations of the meaning/intentionality of most words. Still, it will be possible for an interpreter to point to some subset of recognitional and functional information that will suffice for a new user of the word to use it correctly in the community (‘it is what these have in common’).
Where do new words come from? It seems to me to be a logical truth that there must be a first *coiner* of new words. It can’t be that new words are floating around in the community awaiting discovery before having been uttered by anyone. The principle for ‘opening’ (coining) a new word is much the same as for private concepts, except that an individual will notice that he possesses a private concept that he doesn’t think the community already has a word for. In that case, he can coin (open) a new word (which now functions as a nominal container), which he can communicate to others, so that they can gather the kind of information about the category that he already has in his private container. In effect, he transforms his private container into a social container, by heading it with a word. Of course, now, assuming that the word takes hold in the community, what that word means will no longer depend on what he takes it to mean, but on what the community takes the word to mean.

Where does error come from? Error comes about in the same way as for private concepts: when the functional information gathered by the word is not in accordance with the recognitional information, resulting in predictions that do not match outcomes. For example, thinking that the sun and the planets are the same kind of thing, while Earth is a different kind of thing, will result in predictions that are not matched by outcomes.

### 10.3 Mice shrews and voles, and social concepts and words

Let me return now to the owl, and its containers pertaining to mice, shrews and voles, as discussed in the previous chapter. But now I want to discuss the relation of private concepts with social concepts and word concepts. I surmised that owls only possess private concepts, but imagine a person who has grown up in an environment where mice, shrews and voles were unknown. This person is castaway on a remote island, where these animals are present. Imagine that he now encounters a mouse for the first time. Because it does not activate any of his existing containers, his cognitive system opens a new container. Imagine that his future encounters with mice, shrews and voles are roughly as was described in the previous chapter for the owl. But because he does not eat any of them, his cognitive system never finds a reason to open separate containers for shrews and mice. His functional information does not result in any false predictions, and so he just recognises mice, shrews and voles as the same kind of thing (they look like this, they are found here and here, they behave like this and this). Even though he is a language user, this container is a private
concept for it is not headed by a public language word, and he does not share his concept with other members of a community. There is no publicity requirement for the concept. He is making no error by including three different species within the one container. The only criterion for error is if the functional generalisations in the container do not apply to the things detected by the recognitional information. For now, because he doesn’t interact very closely with the creatures, the recognitional and functional information is consistent. This might change if, for example, he ate voles and mice and preferred the taste of one to the other. As with the owl, he might make false predictions about what a recognised instance will taste like.

But, now, imagine, he finds that he is not, after all, alone on the island; one day, he finds a way to descend into a valley, and discovers a community there who welcome him. The next time he encounters a vole, he is told that it is called ‘a vole’. He now heads his container with this word, and this transforms it from a private concept into a social concept. No longer does the container simply refer to the kind of thing his recognitional and functional information is about, now it refers to what the community take voles to be, because of the publicity desideratum. Sometime later, he encounters a mouse and says to his companion, ‘look a vole’. ‘No’, says his companion, ‘that is a mouse’. He is in error, even though the only change to the container was the addition of the public language word as a heading. Before that he was quite correct to recognise the mouse as belonging to this container. But now the meaning of the container is inherited/derived from what the community take the word ‘vole’ to refer to. In order to communicate successfully, he must endeavour to make the contents of his container correspond with what the community recognise as voles, and what generalisations the community make about voles. As a first step, his cognitive system will open a new container headed by the word ‘mouse’, and will seek to reorganise the stored information into containers, in a way that corresponds with the community’s meanings for the words.

But note that there is nothing special about words that mean they have to correspond with natural kinds or with different species (for example the word ‘jade’ refers to a disjunctive kind). It is perfectly possible that, in fact, in this community the word ‘vole’ does mean mouse, shrew, or vole. Words are just like private concepts in that what intentionality they have depends on what kind of thing is recognised, and what kind of thing the functional generalisations are about. The only difference is that, when it comes to interpret what a word means, an interpreter must take into account the whole of the community’s practices and
beliefs about the word, while giving extra weight to the views of experts and authorities. The word ‘vole’, in a particular community, could apply to mice, shrews and voles, if there was no practical need for the community to organise its functional generalisations more finely. As with private concepts, the only determinant of error is when false predictions occur. For example, a community may not value same constitution as being more important than same appearance or same behaviour. In our scientifically sophisticated society, it has been found that classifying natural things according to their constitution will result in fewer erroneous predictions/inferences about behaviour, than classifying things according to appearance. But this is an entirely pragmatic concern. If it turns out that some kinds are better grouped according to appearance rather than constitution, when it comes to making good inferences about them, then there is no reason to say that this is a worse way of classifying than by constitution.

It is plausible that I might have separate containers for mice, voles and shrews but be unable to distinguish them perceptually. I can re-identify (via discrete containers) more words than I can distinguish perceptual categories. Words track distinctions that the community has decided are worthwhile, and can be used to pass down stored knowledge built up over generations.

10.4 Words refer, and are usable, even if they are ambiguous

The meaning of word concepts and private concepts depends on the information contained in the concept/container and samples in the world. Both types of concept can be interpreted as being about this kind of thing, where this kind of thing is picking out both instances in the world recognised by recognitional information in the concept/container, and the kind of thing that the functional information in the container pertains to. There is no need specify unambiguously what this kind of thing is. A person can learn what water is just by having a few samples of it pointed out. The person doing the pointing out, is doing so by matching recognitional information in her WATER container (the container she possesses that is ‘headed’ by the word ‘water’) to her percept of water. Having pointed to the water, she may, for example, say ‘you drink it, when you are thirsty; it’s used to clean things’, and so on. She selects a few pieces of functional information from her WATER container. The person learning what water is can add these pieces of information to his newly formed WATER container (which is headed the word ‘water’). To possess the concept water, and to
be able to deploy it, a person does not have to know definitively and unambiguously either what it actually is, or what his community takes it to be. And, equally, a community does not have to know definitively and unambiguously what a word refers to, over and above this kind of thing, for it to be perfectly usable as a concept. Of course, a community may well aspire to making ever finer distinctions and generalisations, so as to reduce any ambiguity to an absolute minimum. It may aspire to have necessary and sufficient conditions for recognising what the functional information will apply to. And there will be many concepts that a community possesses where this definitiveness exists.

Still, it is widely agreed that we do not have definitions for most words. If it should become apparent at some future time that word ‘x’ does refer to a category that has necessary and sufficient conditions for membership, well and good. But the word remains usable if no one ever figures out what those necessary and sufficient conditions are. And the word ‘x’ is entirely usable and has intentionality even if, in fact, there are no necessary and sufficient conditions. The reason a word remains useable and manages to refer is simply that the kind of thing that looks like this and that behaves like this etc., carves the world into two without dispute for very many instances (a number of instances clearly either do appear and behave like this or do not). This will leave some boundary cases (maybe even many), where it may not be clear what belongs and what doesn’t. The fact errors may occur over boundary cases (predicted outcomes do not materialise for example) is no good reason to abandon the whole concept. Concepts work just fine, if they produce mostly good results. As Millikan (2000) says, we are fallible creatures; we are not surprised when our ways of recognising and ways of inferring/interacting result in some false positives and false negatives. It is simply incorrect that there has to be some independent perfect interpretation of what a concept refers to. All that is needed (and all that is experienced by concept users) is that concepts refer to this kind of thing where this kind of thing is given a large measure of determinacy by the contents of the container.

Indeed, it is often our experience that, when we are asked what a word means, the best we can do is resort to giving examples of what is recognised as belonging to the category and citing inferences that can be made about the category. Even though this is the best we can do, we succeed perfectly well in deploying the word. The same goes for private concepts.

Words and social concepts do however give the appearance of unambiguity. This is because of disquotation. The word ‘dog’ refers to all and only dogs. The word can serve as a
concrete entity that is a nominal as opposed to literal container. The word can be thought of as an ideal repository for gathering all recognitional and functional information that pertains all and only to that category. And this is so, even if it is impossible even for God to determine what are all and only dogs, and what are all and only the generalisations that can be made about them.

But even though there is the concrete entity that is the word ‘dog’, and it refers to *dogs*, just what dogs are will depend on what functional and recognitional information that the community is gathering together with the word, and that is always a matter of interpretation. Even though ‘dog’ unambiguously refers to the category *dog*, what that category actually is, remains a matter of interpretation for the community. The term still is perfectly usable, even if rival interpreters cannot agree on which is the best interpretation.

Interpretation is still very important. For example, an interpreter can change what a community takes a word to mean. An interpreter can point out that the functional generalisations are inconsistent, in some respects, with the recognitional information, and can suggest changes that make them consistent. An interpreter can even seek to eliminate a word, which is precisely what Machery argues for in the case of the word ‘concept’, as set out in Chapter 3. Machery’s argument was that the generalisations that were being applied to what are recognised as concepts would be better off broken up and applied to three distinct concepts, namely exemplars, prototypes and theories. That is, instead of there being one container (word) for gathering information in, there should be three containers.

10.5 Conclusion

I have said there are three kinds of concepts: private concepts, social concepts and word concepts. The meaning of private concepts is given by the kind of thing the functional and recognitional information in the container is about. Because these containers, and the information in them, are realised by physical entities they can be reduced to scientifically describable properties. What is not reducible in many cases is a single definitive, unambiguous interpretation of what that meaning is, over and above an indexical *this kind of thing*. Error (misrepresentation) occurs when functional information is not in accordance with the recognitional information in the container.
Social concepts have a publicity requirement; their meanings are to be shared with other members of the community. The meaning of social concepts is given by what the community take the word heading the container to refer to, and not by what information is present in the container. The presence of the word in the container allows for disquotational reference (e.g., the container headed by the word ‘dog’ refers to the category dog). However, what the person, at any moment in time, takes the word to refer to is given by the information in the container. That information, having a physical basis, is reducible. But, as with private concepts, it will often be the case that it is not possible to reduce to scientific properties either what meaning the community have for the concept or what the person takes the concept to refer to. This is because there will often be no definitive and unambiguous interpretation that all can agree on. Error, for social concepts, occurs when a person takes (as evidenced by the information in the container headed by the word) a word to mean something different to what the community actually takes the word to mean. This error will be evidenced by miscommunication and/or the sanction of the community.

The meaning of word concepts is given by the kind of thing that the community takes the functional and recognitional information, gathered by the word, to be about. As Putnam claims, this meaning is holistic, and depends on intangibles like simplicity. As such, these meanings cannot be reduced to scientifically describable properties. In many cases, where no universally agreed upon necessary and sufficient conditions are available, the best that can be done is to interpret what the meaning is, and different interpreters may fail to agree. Error occurs when the generalisations gathered by the word result in false predictions about what is recognised as belonging to the category that the word pertains to.

I agree with Putnam that the unambiguous intentionality of the three types of concept cannot be reduced, in many cases. One can reduce the meaning to a definition or infallible test, when such a definition or test is available, but, in most cases, such a definition or test is not available. I agree with Putnam that interpretation of meaning is usually an act of judgement that requires ‘a nose for trade-offs’ between what are the relevant values in any given case, and there will often be no way of deciding between rival interpretations.

But it does not follow from this that private and social concepts (as opposed to their meanings), conceived of as containers and the information within them, cannot be reduced. Container theory claims that mental concepts are containers of stored knowledge pertaining to single categories that can be described in scientific terms. The interpretation of whether a
container refers to mice, or to mice or shrews, may not be reducible (i.e., there may be no way of settling which is the correct interpretation), but the underlying facts which allow a human, or an owl to think quite determinately that they are thinking of ‘this-kind-of-thing’ are, I have argued, entirely scientifically describable.
Chapter 11
Sense and Reference and other issues

11.1 Introduction
Previously, I have discussed how DA-finsts can be symbols that are constituents of propositional statements, and how the information contained in concept/containers can determine what the container refers to. In this final chapter, I want to demonstrate how DA-finsts and containers that gather together information pertaining to single categories can explain why theorists have felt the need to distinguish between sense and reference, and between narrow and wide content. I will, then, briefly discuss the disjunction problem and why it does not arise for container theory. Finally I will note that what I take 'sense' and 'narrow content' to be is a kind of 'inferential role', and I will show why problems Fodor raises for inferential role theories do not arise for container theory.

11.2 Sense and reference
The terms 'sense' and 'reference' were introduced by Frege (1892) in connection with public language, but his claims are also relevant to private and social concepts, i.e., mental representations. The reference of a term or concept/container is the entity it refers to. The sense of the term or concept/container is what it means to the expresser: its cognitive significance, or its mode of presentation (the way it is thought about or understood). Peacocke (1992) puts it that two concepts are distinct if by substituting one for the other an uninformative thought becomes informative. So Lewis Carroll is Lewis Carroll is uninformative but Lewis Carroll is Charles Dodgson is informative. Both Carroll and Dodgson have the same referent, but they can have different cognitive significances or senses to a thinker.

A number of theorists hold that meaning is exhausted by reference (Kripke 1980, Putnam 1975, Fodor 1998), and some, such as Fodor, claim that 'it's very likely that
reference is the only semantic property that compose (Fodor, 2008, p. 18). I will set out below why I disagree with these claims, but as Prinz says,

Even if these reference-based accounts are correct, Frege is surely right to say that reference cannot exhaust our understanding of terms. When we consider the psychology of language, or when we consider the non-linguistic cases just described, the need for a kind of content that transcends reference is manifest. Even if one insists that such contents should play no part in a theory of linguistic semantics, they are indispensable for understanding the concepts we deploy in thought (Prinz, 2002, p. 7).

As Prinz puts it, we need an account of how ‘two coreferential representations, be they terms or concepts, [can come] to seem semantically distinct to a cognitive agent’ (Prinz, 2002, p. 7). Prinz also notes that we need to account for how ‘concepts that are not coreferential can seem alike.’ Putnam (1975) uses a thought experiment about Twin Earth as an example. In this case, two twins can have identical concepts or terms, but refer to different substances. I will deal with ‘coreferential’ cases in this section and I will deal with Twin Earth-type cases in the next section. The solutions to the two puzzles rest on the fact that, in container theory, DA-finsts have two arms.

The subject arm of a DA-finst has a referential content: for example a percept of what is in fact a dog captures a particular perceptual representation from the perceptual field, and what it actually stands for is given by the object in the world that caused that percept.

The second arm of the DA-finst, the predicate arm, ascribes a sense to the object captured by the subject arm. The sense ascribed depends on which container is captured by the predicate arm. If the container activated is the DOG container, then the understanding that is made available to the thinker is given by the information in that container. If the PET container is activated, then the generalisations about pets in that container make that be the sense in which the percept is understood. The activated container in the predicate arm might also be the ANIMAL container, or the MAMMAL container. So, in the DA-finst account, it is straightforward to explain how a representation (via its subject arm) with the same referent can be assigned many different senses (via its predicate arm).

In the DA-finst account, there is no mystery about how two co-referential terms can seem to have different meanings to a cognitive agent. One example of a Frege puzzle involves Hesperus and Phosphorus (or the Morning Star and the Evening Star), which the
Ancient Greeks thought were two distinct bodies, but, in fact, they were both the planet Venus. In the DA-finst account, an Ancient Greek’s percep of the brightest object in the nightsky, as picked out by the *subject arm* of a DA-finst just before morning, will, in fact, have the same referent as the brightest object picked out by the subject arm of a DA-finst in the evening. But, through the process of recognising, in the morning the percep will activate the HESPERUS container, and in the evening will activate the PHOSPHOROS container.

There is nothing mysterious about the community failing to realise that the brightest star in the evening is the same entity as the brightest star in morning, and thus maintaining different containers/words in which to place different generalisations. At some stage, someone could, for example, keep track of Phosphorus right through the night, and come to realise that it is the same entity as Hesperus. Once that happens, the two containers of stored knowledge can be merged.

In this kind of case both the individual and the community as a whole do not realise that the two objects are co-referential. But there are also cases where many in a community do know that two terms are co-referential. A common example is the one mentioned by Peacocke above: Lewis Carroll and Charles Dodgson are in fact the same person, so have the terms have the same referent. As detailed in the last chapter, I propose that a new (social) concept/container is opened when a new word is encountered. Let us assume a person already has a concept/container headed by the word ‘Lewis Carroll’, in which is gathered together information about this author. It is easy to imagine contexts in which this person later first encounters the term ‘Charles Dodgson’ without realising that it refers to the same person as the term ‘Lewis Carroll’. This will cause a new container/concept to be opened. The individual thus comes to have two distinct containers, one in which is gathered all of her information about Lewis Carroll, and the other in which is gathered all of her information about Charles Dodgson. In such circumstances, it is informative to discover that the container pertaining to Lewis Carroll refers to the same person as the container pertaining to Charles Dodgson; whereas, it is uninformative to realise that the container pertaining to Lewis Carroll pertains to Lewis Carroll.

As I discussed in the last chapter, the subject arms of DA-finsts generally capture percepts of objects, but they also capture percepts of words. So, when the individual next hears ‘Charles Dodgson’ uttered, the percep of these words can be grabbed by a DA-finst subject arm. The reference of this term is a particular person out there in the world (the same one that is referenced by the term ‘Lewis Carroll’), but the sense of the term for the
individual will depend on what information is in the container activated by the percep of the
words, namely, the one headed by the term 'Charles Dodgson'. The subject arm of the DA-
finst would have the same referent regardless of whether the LEWIS CARROLL or
CHARLES DODGSON container was captured by the predicate arm of the DA-finst. But,
because she has two separate containers for Lewis Carroll and Charles Dodgson, filled with
differing information, she will attribute a different sense to the object captured by the subject
arm of the DA-finst, depending on which container is captured by the predicate arm of the
DA-finst. Of course, for individuals who know that Lewis Carroll and Charles Dodgson are
the same person, the one container can be headed both by the term 'Lewis Carroll' and
'Charles Dodgson'. In other words, both terms will activate the container. We can also
imagine that the individual who does have separate containers will at some stage, through the
sanction or help of the community, come to realise that the terms refer to the same person. At
that stage, the cognitive system can merge the two containers.

So far we have been discussing cases of social concepts/containers that are headed by
words, but the same distinction between sense and reference can also apply to private
concepts/containers. Imagine a non-linguistic creature that has a container for things that look
like this, taste like this, are found in this location, etc. And it has a discrete container for
things that look like this, sting like this, are found in this (different) location etc. These things
are, in fact, the same, but the creature has been stung by them at location 1 but has never
eaten them there, and has eaten them at location 2 but never been stung by them there.
Because the creature has two discrete containers, it can think of these things in two different
senses. The creature can have two modes of presentation for them. Imagine, then, on two
separate occasions, at location 3, the creature encounters one of these objects. Both times it
captures a percep of the thing with the subject arm of a DA-finst, and both times the subject
arm will have the same referent. The first time it activates the container with the stinging
information (and assigns that container to the predicate arm of the DA-finst). As a
consequence it flees the thing. However, the second time it activates the other container with
the taste information, and as a consequence the creature approaches the thing to eat it. So, we
can say that the percep in the subject arm had the same reference on both occasions, but
because different containers were activated there were different senses (cognitive
significances, modes of presentation, understandings) involved. And, of course, it is always
open to the cognitive system to discover that the two containers pertain to the same kind of
thing, to discover that the generalisations in the second container apply to the things
recognised by the first container, and vice versa, and, as a consequence, the system will merge the two containers. When that happens there will no longer be two different senses that the same referent can be understood by.

All of this ties in with what I have claimed before about how the same exemplar or prototype can have different meanings in different containers. So a visual image exemplar of a banana can refer to *bananas* if it is contained in the BANANA container, but can refer to *fruit* if it is contained in the FRUIT container. We can say that the exemplar itself has the same referent (it refers to the thing it visually resembles – in perception this is the object that caused the exemplar to be formed), but the sense of the exemplar is given by what container it is contained in or associated with.

### 11.3 Narrow and wide content

The second puzzle concerns how the same term or concept can have two different referents, and this puzzle gives rise to the distinction between narrow and wide content. Content here is roughly synonymous with meaning. A mental concept's content is how it represents the world to be. Narrow content is meaning that depends entirely on what is going on in 'the head', as Putnam (1975, p. 227) puts it, i.e., on the intrinsic properties of the thinker. Wide content, in contrast, is meaning that depends to some extent on environmental factors, on what is out there in the world. Once again, the fact that a DA-finst has two arms will provide the solution to the puzzle. I will claim that the subject arm generally has wide content, but the information or sense assigned by activating a container is narrow. However, in the Twin Earth example below, we will see that it is now the subject arm that has different references, while the predicate arm bestows the same 'sense' (that is, captures the same container), rather than, as in the previous section, where the subject arm had the same referent while the predicate arm captured different containers.

The most discussed reason for making the distinction between narrow and wide content is Putnam's (1975) Twin Earth thought experiment. He asks us to imagine that Twin Earth is identical in all respects to Earth, except that on Twin Earth water is constituted by XYZ rather than H2O. Putnam claims that in 1750 (when nothing was known about the constitution of water) an individual on Earth (call him Hilary) will be identical in all respects to Twin Hilary on Twin Earth, including his mental representations to do with water. However what Hilary on Earth calls water is a different substance to what Twin Hilary on
Twin Earth calls water. So the meaning of Hilary and Twin Hilary’s identical mental representations to do with water is determined in part by factors external to them, namely, the constitution of the substances their representations pick out.

As with Frege cases, the fact that DA-finsts have two arms allows us to solve the puzzle. The subject arms of DA-finsts have referents and wide content, and the sense bestowed by the container captured by the predicate arm of the DA-finst has narrow content. On Earth, the subject arm of Hilary’s DA-finst captures a percept of water constituted by H2O. What the subject arm refers to depends on what is out there in the world, and not on what is in Hilary’s head. If Hilary was transported to Twin Earth, his subject arm capturing the same percept would now refer to twater constituted by XYZ. So the subject arm of the DA-finst has wide content. But the predicate arm, which captures the container for the category that Hilary recognises the percept as, has narrow content, in the sense that the inferences and predictions that Hilary can make about what is recognised depends entirely on the contents of the container, if the information is not in the container, which is ‘in the head’, it cannot be brought to bear.

By hypothesis, Hilary and Twin Hilary have containers pertaining to water filled with identical information. When one of Hilary’s DA-finsts captures a percept of water (or the percept of the word ‘water’) with its subject arm, the subject arm’s reference is given by what is out there in the world (namely H2O). The percept activates the WATER container and this container is captured by the predicate arm of the DA-finst. The information that is in the container determines the inferences and interactions that Hilary can make about the object being thought of. For Twin Hilary everything is identical, except that the referent of the DA-finst’s subject arm is now XYZ. But because the information in the container is identical to what is in Hilary’s container, he will make the same inferences and interactions: that is entirely a narrow matter, a matter of what is in the head.

As with Frege cases, the subject arm of a DA-finst picks out percepts of objects in the world. The subject arm refers to whatever is actually picked out, so it has wide content. But what an individual can think about that object, what cognitive significance or sense he attributes to the object, is given by the contents of the container, which has narrow content. That is, the contents of the container are ‘in the head’. It can also be argued that the contents of a container can be realised, in part, by things outside the head, such as Otto’s notebook (Clark & Chalmers, 1998). In my view, such contents would also be narrow.
If Hilary is teleported to Twin Earth, when he next sees water the subject arm of his DA-finst will, in fact, pick out something that is constituted by XYZ, and not by H₂O, but nothing will change in his behaviour, for this is entirely determined by what inferences and interactions are licensed by the information in the container he activates. The actual constitution of water on Twin Earth is irrelevant to which container is activated, because by hypothesis, Hilary's water container contains no information about the constitution. Of course, if one of the contents of Hilary's water container is that water is constituted by H₂O, then some of the outcomes he predicts will not be matched by what actually occurs. But then there would be no puzzle, for Twin Hilary's container would no longer be identical to Hilary's container, as it would contain the information that water is constituted by XYZ.

11.4 Meaning of containers: wide or narrow?

11.4.1 Meaning of containers distinguished from the sense attributed

Note that I said, above, that the percept captured by a subject arm has reference and wide content, but I did not say that the container captured by the predicate arm has sense and narrow content. What I said was that the container attributed a sense with narrow content to the object captured by the subject arm. This is to say that the process of recognising bestows a sense or cognitive significance (with narrow content) on the object to be recognised. But there still remains a question of whether the meaning of a container/concept, in its standing state, has narrow or wide content: does the category that the container pertains to have wide content (depend to some extent on what is in the environment) or have narrow content (depend entirely on what is in the head).

When I discussed Pylyshyn's notion of a finst, a key element was that it tracked an object regardless of its properties, appearance or location; those could change substantially but the finst would 'keep its finger' on the object over time. And this, of course, is what the subject arm of a DA-finst does also. So given that the reference of a representation is the object it refers to, it is no surprise that finsts that keep track of objects have a reference, and so have wide content (what is actually out there matters). And the whole point of the predicate arm of a DA-finst is to make 'sense' of what is captured by the subject arm; to provide a set of generalisations that can be used to guide interaction and inference with the
DA-finsts have the purpose of matching a container/concept/sense with a representation (typically a percept) that picks out an object/referent. But an interesting question is whether containers/concepts, in themselves, have wide or narrow content when considered independently of the recognition process. Obviously the information that is stored/contained in the container is narrow (entirely in the head), but I have been arguing all along that the entity that is the concept is not the information in the container but the container itself. So does the container, itself, have wide or narrow content? Does the meaning of a container depend entirely on what is in the head, or can it depend to some extent on what is in the environment? In 1750 did an individual’s WATER container have wide content because it referred to H2O, and not XYZ, which are factors ‘outside the head’ of the individual?

Note that this issue only concerns the interpretation of the meanings of containers (of what category a container pertains to). The question is whether an interpreter has to take account of external factors in determining what a container pertains to, or can she advert solely to matters that are intrinsic to the subject (that are in the subject’s head). According to container theory, a subject will be able to successfully deploy concepts whether or not interpreters can settle on what the meaning of a container is.

11.4.2 Social concepts have wide content

In the previous chapter, I distinguished between social concepts and private concepts. Social concepts are the same as private concepts except that there is a publicity requirement: each individual’s social concept X should be sufficiently similar to everyone else’s, in order to facilitate communication and to avoid sanction. In the examples I discussed in the last chapter, social concepts were typically ‘headed’ by a public language word. I claimed that the meaning of an individual’s social concept did not depend on what information was in the container, but on what the community took the word ‘heading’ the container to refer to. It

---

15 I don’t mean to rule out that an individual can bring to bear general intelligence to guide interaction and inference, but this will be a matter of doing so on the fly, i.e., working out new generalisations. Containers gather previously stored information that pertain to the same category that do not need to be freshly figured out.
follows that social concepts have wide content (the meaning does not depend just on what is in the individual’s head, but also on what is in the social environment). We can see this by examining Burge’s (1979) classic ‘arthritis’ case.

Burge asks us to imagine a person (call him John) who does not realise that arthritis is a disease of the joints. John believes that he has arthritis in the thigh. Burge argues that in John’s actual community that belief is false. But he claims there could be another community where ‘arthritis’ is the term for rheumatoid ailments in general, and if John was in that community his belief would be correct. So, what ‘arthritis’ means does not depend just on what is in John’s head (on John’s intrinsic properties), but also on what is in the (social) environment. Consequently, the term has wide content.

Putting this in container terms, we can imagine the following. John feels a nasty pain in his thigh, and his cognitive system opens a new container X to gather information about the pain. Over the years, he experiences similar pains in other parts of his body (including some in his joints), and this information is placed in the X container. An outside observer might interpret the meaning of the container to be the RHEUMATISM container. At some point John hears people discussing arthritis, and because the symptoms are very similar to the information gathered in the X container, his cognitive system causes the X container to be headed by the word ‘arthritis’. According to container theory, this causes the X container to change from being a private concept to being a social concept. As a private concept, the category the container pertained to was the kind of thing the generalisations in the container was about. As I discussed in the previous chapter, if the information in a container changed sufficiently, the meaning would shift (e.g. from mouse, shrew, or vole to just mouse). What the container meant depended entirely on what information was in the container, i.e., on what was ‘in the head’ and so it had narrow content.

But as soon as the container is headed by the word ‘arthritis’, the container’s meaning is deferred to what the community take the meaning of the word to be. This is because John wants to be able to communicate successfully, to avoid the ill-effects of misunderstanding, and to avoid the sanctions a community may impose for using the term ‘arthritis’ incorrectly. It is incumbent on John to change the information in his X container, so that it conforms to what the community takes the term ‘arthritis’ to refer to.

Imagine John feels a pain in his thigh. The percept of that pain is captured by the subject arm of a DA-finst, and because it is recognised by the X container (which is now
headed by the word 'arthritis') that container is captured by the predicate arm of the DA-finist. The sense or cognitive significance that John attributes to the pain percept captured by the subject arm of the DA-finist is narrow content, i.e., the sense is determined by the information that is in the X container. However, what the X container, itself, means depends not on what information is in the container, but on what the community takes the word 'arthritis' to mean, and, so, the container has wide content.

Social concepts/containers, themselves, always have wide content; also percepts captured by the subject arms of DA-finsts always have wide content. But it is important to see that, even if a social concept/container with wide content is captured by the predicate arm of a DA-finst, still, the sense or cognitive significance that is attributed to the object of the subject arm is narrow, because it depends entirely on what is in the container. This is just to say that even though the X container, headed by the word 'arthritis', means whatever the community takes the word to mean, and so has wide content, what John thinks the container means (the sense he attributes), when he applies it to a particular percept of a pain, is given entirely by the information in the container, and so this attribution of sense has narrow content.

11.4.3 Private concepts and words can have wide or narrow content

Social concepts/containers, then, always have wide content, but what about private concepts/containers and word concepts? Do they always have narrow content? That is, does the category a private concept or a word concept pertains to always depend entirely on what information is in the container (narrow content), or can it also depend on matters in the environment (wide content)?

I claimed, in the previous chapter, that public language word concepts and private concepts refer in the same way: what a word concept or private concept means is given by what kind of thing the recognitional and functional generalisations gathered by (contained in) the concept are about. And I said that this was a matter of interpretation: interpreters might fail to agree on which interpretation is best. Nonetheless, interpreters would base their interpretation on the information contained in containers/concepts. The question is whether there are circumstances where matters in the environment would also play a role in the interpretation.
How might environmental factors play a role, in addition to information in a container? Sometimes containers should be capable of keeping track of a particular individual, and what information is known about the individual does not matter. If my system opens a container to gather information about my newly moved in next door neighbour, that container should keep track of that particular individual regardless of what information gets put in the container. Perhaps, I mistakenly assume that a visitor to the house is my new neighbour, and I come to store recognitional information about that visitor in the newly opened container. But that recognitional information should not make the reference of the container shift, in the way described in the previous chapter. The container would continue to pertain to my next door neighbour, despite the information in the container being about the visitor. Therefore, this container has wide content; what it refers to depends not just on what information is in the container, but on what individual in the world it was opened to keep track of. The same case can be made for an owl that does not have any social or word concepts. The owl could, in principle, open a container for gathering together information about the owl in the adjacent territory to the north, for example. If this container comes to have information pertaining to a non-neighbouring owl, the meaning should not shift; what the container is about is a particular individual in a particular location in the world.

But how about containers that track kinds and substances? Can these have wide content? In the previous chapter, I described how an owl might come to have a container that tracked mice, shrews, and voles, in the sense that it stored recognitional and functional information about these three species, all in the one container. And I described how the category that the container pertained to could shift, depending on what information was in the container, so that the owl came to have three discrete containers, one for mice, one for shrews, and one for voles. The upshot of that discussion was that, when an outside observer interpreted what the container referred to, what was taken into consideration was just what the recognitional and functional generalisations pertained to, and external factors such as the DNA constitution of the creatures did not matter. This would make those kinds of containers have narrow content.

But could it be argued that, even granting the above, in addition to the information in the container, environmental factors can make a difference, and at least some private containers have wide content? Consider Putnam’s Twin Earth argument adapted to owls. Ollie the owl on earth and Twin Ollie the owl on Twin Earth both have identical containers. Based on the recognitional and functional information in the containers, would not an
interpreter judge that Ollie’s container pertained to water and Twin Ollie’s container pertained to twater (Twin Earth water)? If this is correct, then both containers would contain identical recognitional and functional generalisations, but they are interpreted to refer to different categories, depending on what is in the environment; therefore meaning is not in the head. The containers have wide content.

That would be Putnam’s interpretation. However, container theory offers a rival interpretation, namely, that both of Ollie’s and Twin Ollie’s containers refer to $H_2O/XYZ$. This is because the generalisations in Ollie’s (and Twin Ollie’s) container (call it the ‘W container’) are perfectly consistent with them being about XYZ, as well as $H_2O$. The information in Ollie’s W container will result in no predictions that do not match outcomes, whenever he encounters XYZ, and the same goes for Twin Ollie, whenever he encounters $H_2O$. Of course it is true that Ollie is never going to encounter XYZ on Earth, but that is just a contingency. If it was the case that some of the water on Earth was constituted by $H_2O$ and some by XYZ, it would be a perfectly good interpretation that the container referred to $H_2O$ and XYZ, given that the same set of generalisations applied to both substances, and given that Ollie does not have access to the constitutional information about $H_2O$ and XYZ.

Here we have two different interpretations of what the W container refers to, and one can argue over which is better. Here is Putnam’s argument:

If you had asked a person living in 1750 the hypothetical question, ‘Suppose that I gave you a glass containing 50 percent normal water and 50 percent some substance which is not found as a constituent of normal water [i.e. XYZ], but you couldn’t tell this by the appearance or taste or aftereffects, or by washing clothes in it, or anything like that (apart from using a still); would that mixture then simply be water?’ I think that even in 1750 a typical person would have answered, ‘No, I wouldn’t say it was water, I would say it was a mixture of water and something else.’ Of course, if it had turned out that normal water was itself a mixture, and that it contained an indefinite number of different ‘pure’ constituents, then the answer might have been different. But we might say that our intention, even in 1750, was somewhat as follows: On the assumption that normal water is in fact a pure substance, then we do not intend the description ‘water’ to be true tout court of anything which consists
to a significant extent (say, 20 percent or more) of any other substance (Putnam, 1988 p. 31).

But this paragraph just highlights how much the issue is dependent on interpretation. Putnam’s hypothetical question includes the information that there are, in fact, two substances with different constitutions in the glass. But it would require a time-traveller with knowledge of the existence of XYZ to impart that fact to an individual living in 1750. If the question had been along the lines of ‘there is this stuff here that is water and this stuff over there which appears and behaves exactly the same; would you call them both water?’ the answer would surely have been yes.

I believe the issue boils down to the fact that there are three criteria that can be used for determining when a distinct container or word is required; a criterion based on: (a) appearance, (b) behaviour, and (c) constitution. That is you can group information according to same appearance, same behaviour, same constitution, or some combination of these. Some kinds emphasise just one of these criteria:

a) Artifact kinds emphasise behaviour or functionality, and appearance and constitution is less important in determining what counts as a single category; it doesn’t matter what carburettors or mouse-traps are made of, or appear like, it just matters what they do;
b) Natural kinds emphasise constitution, and behaviour and appearance is less important; it doesn’t matter if other things appear or behave like gold or mice, if they do not have the right constitution;
c) Appearance kinds emphasise appearance, and arguably constitution and behaviour do not matter at all to, for example, the colour blue or the feeling of arthritic pain.

In the previous chapter, I stressed that concepts/containers contained information that could broadly be broken down into recognitional and functional strands. Information was needed to determine whether the container should be activated (whether an instance be recognised as belonging), and then, once activated, information was needed to guide inference and interaction with the instance. The general principle that the cognitive systems of animals like owls can be taken to follow is ‘same appearance will result in same behaviour’. But, in our scientific age, we know that for many things appearances can be deceptive, and the scientific way of thinking has come to realise that same constitution is a
much better guarantor of same behaviour. Whereas an owl cannot, generally, be expected to recognise/classify according to constitution, but, instead, can be expected to classify according to observed appearance and behaviour, humans, deferring to scientific experts, can classify according to constitution. Constitution often cannot be directly observed. Whether water is made of H\textsubscript{2}O or XYZ cannot be directly perceived. The constitution of a cat’s DNA cannot be directly perceived. But indirect methods can establish constitution, thus enabling more accurate generalisations about behaviour to be made.

Same appearance/same behaviour is a good principle for gathering together recognitional and functional information, but, particularly for natural kinds, same constitution/same behaviour is, often, a better one. Recognising the same kind of thing by appearance is something any perceiver can do, but recognising the same kind of thing by (hidden) constitution is generally only open to communities that practice science. Indeed, it might be argued that science is the practice of gathering together behavioural information according to constitutional information.

We can then say that appearance based recognition is narrow, while constitution based recognition is wide. Why? Because judgements about similarity of appearance to stored information is entirely a matter of what goes on ‘in the head’, but what something is constituted by is entirely a matter of what is in the environment. In a scientific community that emphasises same constitution, some private concepts/containers can have wide content.

However, the issue is not straightforward. If my container for water contains the information that water is constituted by H\textsubscript{2}O, then this makes my container have narrow content, because what water means is given entirely by information in the container. The same goes for word concepts. If the community’s generalisations about water include that it is constituted by H\textsubscript{2}O, then environmental factors don’t matter as well. That is why Putnam’s example of wide content, meaning that is not given by the intrinsic properties of the representational system alone, invoked people living in 1750, who did not know that water is constituted by H\textsubscript{2}O.

To return to the hypothetical question that Putnam poses to a person (call him Jim) living in 1750, what is being emphasised is that the glass contains ‘50 percent normal water’ and 50 percent a substance constituted by something else (namely XYZ). Jim is being asked whether, even if there is nothing about the appearance or behaviour that can distinguish the two substances, the knowledge that the substances have different constitutions is sufficient
grounds for there to be two different concepts. In a community that values distinguishing things by constitution, the answer may well be yes. Let us say Jim already has a container headed by the word ‘water’ which contains all of his recognitional and functional information that pertains to that category. Now he opens a new container headed by, say, the word ‘twater’, and all of the appearance and interaction information from the WATER container is copied into the new TWATER container. In addition Jim can add a statement to the container to the effect that he was informed that this substance has a different constitution than the substance that the WATER container is about, but there is no known method of telling the two substances apart. Now, both of Jim’s WATER and TWATER containers have wide content: the category that they refer to depends on factors in the environment (H2O or XYZ constitution), but Jim has no way of knowing what those factors are, nor of knowing whenever he encounters water in future whether the WATER or TWATER container should be activated. Luckily, because the same functional information is in both containers it makes no difference which one he activates. Both allow him to interact intelligently. I am being somewhat facetious here, but I do wish to stress that the new TWATER container really adds nothing useful to the individual’s knowledge of the world, until someone has actually discovered XYZ, and the means to distinguish it from H2O. And once that occurs, then, as I argued above, the constitution information can be placed in the respective containers, giving them narrow content.

In contrast, in a community that has no experience of the benefits of classifying things according to constitution, and has no means of doing so, we can imagine a different answer would be given, especially if the question is posed less tendentiously. If, for example, a person (call him Josh) in a non-scientific community was simply asked whether the substance in this glass, which appears exactly the same as the water in that glass, and cannot be told apart by any known test, also counts as water he would surely say yes. His WATER container would have narrow content: the category of the container depends entirely on the information in the container. And Josh would be in no way worse off than Jim, who has a WATER and TWATER container with wide content. Arguably he would be slightly better off, in that he would not have a redundant TWATER concept/container, whose contents are a duplicate of the WATER concept.
11.5 The disjunction problem

I now want to make some comments on the disjunction problem for mental representations, because what I have just been discussing is relevant to this issue. Prinz says the problem can be easily stated for information theories: ‘if a concept refers to whatever has the power to cause it, then there can be no error’ (Prinz, 2002, p. 243). He says that his bush pig concept can be tokened not just because of perceptual encounters with bush pigs, but also by encounters with a number of ‘illicit’ causes, e.g., pregnant warthogs and wild boars, because he cannot distinguish them visually from bush pigs. In addition, he notes that twin bush pigs from Twin Earth would cause the concept to be tokened. So there are three kinds of illicit causes according to Prinz: ‘wild-cause cases, Twin Earth cases, and Earth-bound twin cases’ (Prinz, 2002, p. 249).

We have already stated a solution to this problem in Chapter 9. Prinz says ‘if a concept refers to whatever has the power to cause it, then there can be no error’, but I have made clear that, in container theory, what a container refers to is not just given by the recognitional information in a container, but also by the functional information in the container. This means that the mere fact that something has the power to activate a container, by matching recognitional information, does not mean that the container refers to that cause. For the cause to be non-illicit the functional information must apply to the cause also. For example, if a shrew activated the owl’s MOUSE container, and the information in that container is used to predict what the shrew will taste like, this will not match the information of what the shrew actually tastes like. This non-matching of outcome to prediction is what makes it an error for the shrew to have activated the MOUSE container. If it had activated the SHREW container, the predicted and actual taste would have matched. This is an example of what Prinz calls a ‘wild-cause’: a case where the recognising conditions are less than optimal. If the conditions were better the shrew would have activated the SHREW container. And, because exemplar and prototype recognition is a matter of exceeding a threshold, that threshold can be set at a level that allows some false positives in order to eliminate more false negatives (e.g., setting the threshold at level x% causes some shrews to be recognised as mice but allows many more mice to be recognised as mice than if the threshold was very high). What makes it wrong to recognise a cow in bad light as a horse, in container theory, is that the functional information in the container will result in bad predictions (e.g. if you try to ride the cow recognised as a horse the outcome will not be as predicted).
But what about Twin Earth and Earth-bound twin cases? Prinz proposes a solution that applies to these cases as well as wild causes. He discusses the example of monarch and viceroy butterflies:

One can find genetically distinct species that appear indistinguishable because one has evolved to look like the other in order to fool predators. For example, viceroy mimic monarch butterflies. Imagine that I form a concept by observing a monarch, and after that, tokens of this concept are caused by encounters with other monarchs but also by encounters with viceroy butterflies.

Intuitively, the fact that viceroy butterflies happen to cause this concept is irrelevant. It is a monarch concept because it was created to detect monarchs. Viceroy butterflies only cause it to token because they happen to be monarch mimics. Compare: if I create a bear trap that also happens to catch caribou, it is not a caribou-and-bear trap. According to Fodor's etiological constraint, a concept refers to any category whose instances have actually caused (and would reliably cause) it to be tokened. If my monarch concept is sometimes reliably caused by viceroy butterflies, Fodor is forced to say the concept refers to both monarchs and viceroy butterflies.

Prinz simply bites the bullet in such cases, claiming that concepts whose tokens are actually caused by members of indistinguishable kinds are disjunctive. I think this is a hard bullet to bite. We take our natural-kind concepts to pick out unique natural kinds, not disjunctive sets of natural kinds.

I believe that my monarch concept refers to one kind of butterfly even if I suspect that I am frequently duped by mimics (Prinz, 2002, pp. 247-248).

Prinz offers a single solution to specifying when a cause is illicit that applies to all three types of cases. He adapts a proposal of Dretske's (1981). According to Dretske there is a critical period when a concept is first acquired when special attention is given to what is recognised, and what the concept refers to is what would be recognised in this period. Later the concept may be applied more loosely and so errors occur. Prinz notes that this proposal will only work for wild causes, and cannot handle twin causes, for even with special attention viceroy butterflies might have been recognised as the same kind as monarchs during the learning period. But Prinz believes that Dretske is on the right track, and proposes to remove the subjunctive element in Dretske's proposal. Instead of what would be recognised during the learning period, it is what is recognised during the period. He says, 'If I form a concept as
a result of a perceptual encounter with monarch butterflies, that concept refers to monarchs alone, even though it is also tokened when I see viceroys' (Prinz, 2002, p. 250).

Prinz deals with a number of objections to this proposal but I just want to focus on one which is central to what I have been arguing about how containers refer. Prinz says,

Suppose that Sally’s mother takes her to the zoo and shows her some alligators. As a result of that encounter she acquires a concept that would be reliably caused by further alligator encounters. But as it happens, she never sees an alligator again. What she does see are lots of crocodiles, and these reliably cause the alligator-derived concept to be tokened. I do not think that these tokenings qualify as correct applications of the concept. My intuition is that Sally is making some kind of mistake when she applies her alligator-derived concept to crocodiles (Prinz, 2002, p. 253).

Now it is important to bear in mind here that Prinz’s concern is to avoid the disjunction problem, and establish a principle that makes it an error for crocodiles to activate the ALLIGATOR concept or, to use the example in Chapter 9, for mice to activate the VOLE concept. The principle Prinz utilises, when translated into container theory terminology, is that when the container/concept is first opened whatever it refers to then (or during a critical early period) is what it always refers to.

This is contrary to what I have been claiming in Chapter 9 about the owl’s containers for mice, shrew, and voles. I argued that the meaning of a container could shift, as the information in the container changed. In contrast, in Prinz’s formulation, once the container first comes to have a meaning – which depends on what caused the container to be opened (or what caused it to be activated during a critical period) – then that is what the container always refers to, no matter what future encounters are with. Error occurs when causes that are not of the same type as the initial cause activate the container.

In the example in Chapter 9, the container started off as a mouse container, then later became a mouse/vole/shrew container, then a mouse/vole container, and, finally, became a vole container, while new containers for mice and shrews were opened. I stressed that what the container pertained to was given by what generalisations were in the container. If the functional and recognitional information changed sufficiently, then what the container referred to changed also. If, as in Prinz’s example, a container came to be opened by an alligator cause but subsequently all the information in it was about crocodiles, then, in
container theory, it would become the CROCODILE container/concept, as all of the functional generalisations would be about crocodiles. If, after many encounter with crocodiles, an alligator activated the container, some of the inferences and predictions made about the alligator would not be matched by actual outcomes. Of course, if the functional generalisations did, in fact, apply equally well to alligators, then, according to container theory, the container would refer to alligators/crocodiles.

But what is interesting is that this is the case just for private concepts. Prinz’s principle does apply to social concepts. In Prinz’s alligator example, we can imagine that Sally’s mother would have given Sally the word ‘alligator’, and by heading the container with this word, then, according to container theory, the container refers to what the community take to be alligators. And this is so even if Sally only encounters crocodiles from then on and comes to fill her ‘alligator’ concept with crocodile information.

However imagine that Sally is cast away on a desert island, and here encounters alligators and crocodiles for the first time. The first one she sees is an alligator, and she opens a new container. This is a private concept. It is not headed by a word, and is not shared (yet) with other members of a community (there is no-one else on the island). If from now on Sally only encounters crocodiles and gathers together plentiful generalisations that pertain to crocodiles and not alligators, then the container pertains to crocodiles.

It is also important to note that for private concepts there is no need for concepts to track non-disjunctive natural kinds. Note that in the quote above from Prinz, he say ‘We take our natural-kind concepts to pick out unique natural kinds, not disjunctive sets of natural kinds.’ This is no doubt true of natural kind words in our scientifically informed society (jade is probably the only commonly cited counter-example). But that is because countless experts have performed experiments, and engaged in substantive disputes with each other, to arrive at generalisations that apply to non-disjunctive natural kinds. However, in the case of private concepts that have to be acquired by an individual on their own, without the benefit of generations of passed down knowledge that is gathered together in words, there is no need to think that only ‘unique natural kinds’ are picked out. If the cast away Sally happens to see equal numbers of alligators and crocodiles, she is not making any great mistake if she happens to have just one container for them. If the information she gathers is just that they appear like this and this, are found here, here and here, behave like this and this, and should
be avoided at all costs, there is no reason for her to have two distinct containers just because they have different genetic constitutions (which of course she has no way of knowing).

The point is that when it comes to private concepts (containers opened generally as a result of perceptual encounters, and which are not shared with other members of a community) what will be used to recognise members will be appearance and functional/behavioural criteria, and not constitution criteria.

To return to monarch and viceroy butterflies, the reason that Prinz came to have two distinct concepts for the two categories is almost certainly because he heard or read the two different terms. When he encountered the term ‘monarch butterfly’ he opened a container to gather information about that category, and when he encountered the term ‘viceroy butterfly’ he opened a different container. His expectation is that experts in his community came up with two different terms because there are two different kinds (with differing functional information), even though they appear similar (the same recognitional information applies). It may well have occurred that he encountered both terms together. He may even have been on a field trip and actually perceived a monarch butterfly, which caused a new container to be opened. But, presumably, the way he came to have a separate viceroy concept/container will have been because he encountered the word from an expert. He now has two social concepts, one headed by the word ‘monarch’, and the other headed by the word ‘viceroy’. What those concepts refer to is inherited from what his community mean by the terms, and not from what perceptual exemplars are in the containers. His system may have wrongly put the monarch percept as an exemplar into the container headed by the word ‘viceroy’. According to container theory, that doesn’t make it the MONARCH container; because it is a social concept, what it refers to is given by the word heading the container.

But when it comes to birds like owls, or castaway humans, it is likely that they will not have distinct containers/concepts that distinguish between monarchs and viceroys (on the assumption that they cannot be easily told apart), for the simple reason that owls and castaway humans have no access to constitution information. Indeed, this is what mimicry in nature relies upon. The mimicking species comes to mimic the dangerous species because predators cannot tell them apart. It could be the case that predators think there are two kinds but cannot tell which are which. Sometimes they eat monarchs and are ill and sometimes they eat viceroys and are not. This differing functional information warrants two containers, but since the predator cannot tell which is which, they ‘play safe’ and adopt the same behaviour.
of not eating either monarchs or viceroys. Or it could be the case that the predator just has one container because the first one it ate was a monarch, and it was poisoned, and so learned never to eat any butterflies with that appearance.

The point is that we should not think that classifying things by constitution is to be taken as the norm for private mental concepts for the reason that non-scientists cannot come by the necessary constitution information. These concepts do not have to be shared with other members of a community, and, so, anything that works well enough will do for the individual.

11.6 Container theory and inferential/conceptual/functional role semantics

In this section, I deal with the idea, noted at the start of this chapter, that many theorists hold that meaning is exhausted by reference. But Frege cases, prima facie, seem to make this impossible. Fodor notes that friends of RTM and naturalism have two options for dealing with Frege cases: (a) accept their implications and posit sense in addition to reference or (b) somehow deflate Frege cases. He opts for the latter course. This is because he thinks that ‘these days senses are widely supposed to be something like inferential roles’ (Fodor, 2008, p. 53), and he thinks that the resulting view (inferential/conceptual/functional role semantics) is untenable for two reasons.

The first reason is,

... arguably, positing senses would require some kind of IRS [inferential role semantics]; and IRS would require some kind of semantic holism; and semantic holism would require a not-very-plausible account of the metaphysics of words and languages (Fodor, 2008, p. 54).

The second reason is that he does not believe that senses, as inferential roles, can be compositional. Indeed he believes ‘that reference is the only semantic property that composes’ (Fodor, 2008, p. 18).

The account that I have been putting forward in this chapter and the previous one is certainly not the same as traditional IRS accounts, but, nonetheless, it can be argued that I am claiming that inferential or conceptual or functional roles are determinants of the meanings of most concepts/containers. This is because I argue that the information in a container is what is used to interpret the meaning of a container, and that information is what gives rise to the
inferences a subject makes, and determines what actions the subject will take. That is, the information in the container determines the inferential/functional/conceptual role of the container/concept. In my account what a container pertains to (at least for private concepts) is generally determined by the recognitional and inferential information in a container, and not by its reference. Rather, the inferential role determines the reference. If the information in a container changes, the referent can change also. As we have seen the MOUSE container can become the VOLE container. So do Fodor’s objections to IRS threaten my account? I claim that they clearly do not.

Let me deal with the holism problem first. I do not wish to go into the merits of traditional IRS accounts; I merely wish to show that Fodor’s objections to them do not apply to container theory. Fodor puts it that, in IRS,

the inferential role of an expression is constituted by its inferential relations to every other expression that belongs to the same language. In consequence, according to IRS, the individuation of whole languages is metaphysically prior to the individuation of the expressions that belong to them. Likewise, whole conceptual systems are prior to the concepts by which they are constituted (Fodor, 2008, p. 54).

Prinz puts it this way,

The most widely discussed objection to functional-role theories is that it is difficult to determine which inferences matter for concept individuation (Fodor (1987), Fodor and Lepore (1992)). Should the narrow content of a concept be identified with its total functional role (consisting of a Ramsified description of every inference into which it enters) or a partial functional role? If total roles are used, then no two people have concepts with the same narrow contents, because no two people draw exactly the same inferences. If partial roles are used, then one must say which inferences contribute to narrow content and which do not. Fodor and Lepore (1992) argue that this can only be done if one can come up with something like an analytic/synthetic distinction. Quine’s arguments against such a distinction convince Fodor and Lepore that individuation by partial roles is untenable. This leads them to reject functional role theories (Prinz, 2002, p. 264).
In container theory only the information in a container contributes to the inferential role of a concept (container). There is no need to have recourse to all of the subject’s possible inferences (total functional/conceptual role) and so the problems associated with semantic holism do not arise. An interpreter can interpret what the container pertains to just by examining the contents of the container. This means that in container theory there is a way of specifying just which recognitional and inferential/functional information contributes to the interpretation of the meaning of a concept/container. There is no need for an analytic/synthetic distinction. If the information got into the container then it contributes. If it did not, then it is not part of the inferential role information.

I would stress here the work that containers are doing. In traditional IRS accounts either all possible inferences matter for the individuation of a concept, in which case you have the problem that whole conceptual systems seem to be prior to the concepts that constitute them, and the problem that no two people can have the same concept because no two people draw the same inferences. Or only some inferences matter for the individuation of a concept, in which case a principle is needed to determine which inferences those are. Fodor and Lepore claim that this amounts to finding something like an analytic/synthetic distinction. And most theorists agree that Quine (1953) has shown that an analytic/synthetic distinction cannot be found. But in container theory, what matters for interpreting what a container pertains to are only the inferences licenced by the recognitional and functional information in the container. That information got placed in the container purely mechanically, and not because it was, for example, analytic that it belonged in the container. As we saw in Chapter 4, when a subject selectively attends to something in perception (captures it with the subject arm of a DA-finst), the cognitive system will add any novel information to the container that recognised the object of the percept. The system does not have to judge if it is analytic that this new information belongs to category X. The system simply adds it to the container if a threshold was exceeded. If adding new information over time causes the meaning of the container to shift (as interpreted), then so be it. As we saw with the mouse/shrew/vole example, what matters is being able to successfully predict outcomes, and the system will alter information in the container, and potentially alter the meaning of the container, purely so as to maximise accurate predictions.

Fodor argues that
it's also common ground that you need more than one concept to draw an inference, so if IRS is true, conceptual atomism isn't (Fodor, 1998, p. 14).

It is true that expressions of propositional inferences will require more than one concept. But the kind of inferential/functional information that is contained in containers, as set out in Chapter 4, can be entirely perceptual and sensorimotor in nature. I have claimed that it does not require the subject to have any concepts to plot information against a reference frame.

Fodor also claims that,

by contrast [to sense understood as IRS] it is plausible prima facie that reference is atomistic; whether the expression ‘a’ refers to the individual a is prima facie independent of the reference of any other symbol to any other individual. Indeed, it’s plausible prima facie that ‘a’ might refer to a even if there are no other symbols. The whole truth about a language might be that its only well-formed expression is ‘John’ and that ‘John’ refers to John. I do think that uncorrupted intuition supports this sort of view; the fact that ‘John’ refers to John doesn’t seem to depend on, as it might be, such facts as that ‘dog’ refers to dogs (Fodor, 1998, pp. 54-55).

The problem he is getting at here is that if the meaning of a concept depends on inferences that can be made about the category, then the subject will have to already possess other concepts that give rise to the inference. But again this is answered by the point that the inferential information in containers can be perceptual and sensorimotor in nature, and so not require the subject to possess any concepts. For example, a subject will be able to predict what the recognised object will appear like (from other angles, or in other sensory modalities) and what the object will behave like (as emulated by, for example, a Grush (2004) style emulator). All of these inferences can be non-conceptualised. Of course, a creature that is able to express propositions using other concepts, in the manner set out in Chapter 6, can add propositional information to a container, in addition to the perceptual and sensorimotor information, and for this other concepts will have to be possessed.

In container theory, in principle, a subject could have just one container/concept pertaining to, say, mice and have no other containers at all. The container refers to mice because all of the information/generalisations in the container pertains to mice, and this would be so even if the subject had no other concepts.
Fodor's final objection to IRS is that it is surprisingly difficult to devise a version of IRS that is compositional, and that exhibits the productivity and systematicity that languages and minds exhibit. But, as I have set out in Chapters 6 and 7, compositions of DA-finsts will do all that Fodor's Mentalese symbols can do in terms of productivity and systematicity.

Recall that containers are atomistic: for each primitive concept that a subject possesses, there will be a single container, and containers are not made up of smaller containers. That is, they are the smallest units in the conceptual system; they are atoms. Even though what a (private) container pertains to depends on the stored information in the container, which in turn gives the sense in which things are recognised, it is not the stored information which is composed, in propositional expressions, it is the atomic containers (via DA-finsts) that are composed. The contents of containers are not concepts, rather the containers themselves are concepts. The inferential role information is not what is composed, rather concepts/containers contain the inferential role information.

Clearly there would be problems if what was composed directly was the inferential information itself. To use an example of Fodor's, if one combined all of the inferences one could make about pets and fish, one would not arrive at the set of inferences one would make about pet fish. Or if a prototype of a typical fish was combined with a prototype of a typical pet, one would not come up with a prototype of a typical pet fish (Fodor & Lepore, 1996). But when it is the containers that are composed (via DA-finsts), rather than the inferential information that is composed, one achieves the same effect as one would if one composed symbols whose meaning is determined by reference. In container theory, the meaning of the PET and FISH containers is given by the inferential information in those respective containers. And the meaning of pet fish can be derived from the subset of information in the former that can be applied to the latter (or vice versa).

In conclusion concepts whose meanings are interpreted on the basis of their inferential role, that is, on the basis of the recognitional and functional information in a container, are composable. In container theory, what a container pertains to is a matter of interpreting what the information in the container pertains to.
Conclusion

What I have sought to do with container theory is put forward a notion of concept that is capable of accounting for all that cognitive scientists and philosophers expect of concepts. It satisfies both Machery’s Judgement Desideratum and his Propositional Desideratum. In Prinz & Clark’s and Fodor’s terms, concepts are both for doing and for thinking. Many disputes can be traced to the fact that theorists have either been confusing the contents of concepts for concepts themselves, or have been confusing symbols of concepts for concepts themselves. Once it is realised that a concept is the container, and not the container’s contents or its symbol, many disputes just vanish.

In container theory, DA-finsts keep track, with their two arms, of a percept of an object and the container for the category that it has been recognised as. These DA-finsts, I claim, are capable of being composed into expressions of propositional thoughts. These thoughts are meaningful for a thinker, in that the system knows exactly what the components of the thoughts are. The subject arm of each DA-finst picks out which entity is being thought of, while each predicate arm picks out the category that the entity is being thought of as falling under. The containers captured by the predicate arms provide the kind of inferential and functional information about the entities that give the thinker understanding. DA-finsts also explain why theorists have been tempted by both referential and sense accounts of content: the subject arm of a DA-finst has a reference, while the predicate arm assigns a sense.

Perhaps the most crucial pay-off from the claim that a concept is a container is that it allows concepts to be individuated by the container structure itself, rather than by meaning or content (whether determined by sense or reference). If there is a discrete container, then there is a discrete concept; there does not have to be a discrete, determinate meaning as well. According to container theory, each container/concept pertains to a single category X, but what that category X is, is a matter of interpretation for an outside observer and not something that the system need determine itself. The system merely gathers recognitional and functional information that go together. The result is a container that pertains to a category. But the system doesn’t have to be able to specify which X that category is, in order for the
information in the container to be useful and predictive. It is possible for outside observers who possess social, as opposed to merely private, concepts to interpret which (social) concept the subject’s (private) concept corresponds to. But it may be that interpreters cannot agree on this: for example, one theorist might claim that the container pertains to mouse, while another claims that it pertains to mouse/vole. Which interpretation one makes depends on competing sets of criteria, and there will often be no way of settling which set of criteria is better. This failure to agree on what the container pertains to need not matter to the concept possessor at all. The subject will have available a determinate set of inferential information (the information in the container), regardless of the interpretation of which category X the information pertains to. For the subject, it refers to a distinct category merely because it is in a discrete container/concept.

Let us return to the questions I posed in the Introduction and consider the answers container theory offers:

A. What is the difference between perception and (propositional) cognition and how do concepts fit in?

Right the way through the thesis, I noted a distinction between perception, which provides information in a holistic iconic manner and which represents many propositions at once, and propositional cognition, which picks out single propositions that can be evaluated as true or false. Simple propositions (i.e. ones involving just one concept) can be expressed by a single DA-finst, with the subject arm effectively expressing ‘that’ and the predicate arm expressing ‘is a X’ where X is the category of the container pointed to. Complex propositions involve the juxtaposition of more than one DA-finst.

B. Is there a single notion of concept or is the term applied to different things?

One of the greatest advantages of container theory is that once it is realised that a concept is the container, and not the information contained in the container or the symbol of the container, theorists who had previously disagreed can find that they agree on the substantive issues at hand. For example, the contents of a concept/container are what facilitate interaction and prediction (doing), while the symbols of a concept/container are what facilitate propositional expression (thought). And one should not think that concepts are
meanings, but concepts/containers contain information that bestows understanding and that can be used, by an interpreter, to assign meanings.

C. How do concepts allow us to recognise, infer, predict, and interact?

Concepts/containers contain recognitional and functional information pertaining to a single category. When the subject arm of a DA-finst captures a part of a perception, that part of the perception becomes available for recognition by information (e.g. exemplars and prototypes) in containers. Whichever container contains the closest matching recognitional information will be activated (provided a threshold is exceeded), thus making available to the system the stored functional information in the container that allows the subject to interact, infer, and predict successfully with respect to the entity that has been recognised.

D. How do concepts allow propositions to be expressed?

According to container theory, propositions are expressed by DA-finsts. Single DA-finsts, because they have subject and predicate arms, can be evaluated for truth. Combinations of DA-finsts can be evaluated for truth that depends on the truth of each of the constituent DA-finsts, and on the system's composition rules. In effect, container theory adopts the spirit of Fodor's LOT account but does away with the Mentalese symbols to which many theorists object. Container theory also adopts much of what is best in Barsalou's (1999) account, but substitutes DA-finsts for Barsalou's perceptual symbols.

E. How are concepts and their meanings acquired?

As I have mentioned, one of the most important contributions container theory makes is that it claims that concepts are individuated not by their meaning or reference but by their discrete container. The number of concepts a subject possesses is the number of separate containers the subject has. A subject acquires a new concept whenever the cognitive system opens a new container. The cognitive system opens a new container whenever it encounters something that it cannot recognise with any of its existing containers. There is no need for the system to first determine the category of the new container and then gather information that pertains to that category. Instead, the system will mechanically gather information that goes
together based on what the system attends to. If a part of a perception activates a container by exceeding a threshold, then the system will add new recognitional and/or functional information to the container based on what is perceived and attended. It is not the case that the perceived information must pass some semantic test as to whether it belongs to a certain category or not; if it activates the container, then the system will treat it as belonging to that container. The system examines whether what it predicts on the basis of stored information is matched by incoming signals, and may alter the information in the container if the signals do not match. What I set out in chapter 9 is that in, the case of private concepts, additions and alterations of information could change the external interpretation of which category a container pertains to over time.

F. Can concepts and their meanings be naturalised?

It is straightforward to naturalise private and social concepts because they do not have any semantic structure; they are just containment vehicles. It is also possible, in principle, to naturalise information contained in concepts/containers. It seems plausible that the realisers of the recognitional and functional information that permits subjects to make recognitional and inferential judgements has a neural basis. However, I acknowledge that in many cases it will not be possible to naturalise an unambiguous and precise meaning of the contents of concepts/containers. This is because the category X a container pertains to, according to container theory, is a matter of interpretation by outside interpreters, and in many cases there may be no way of settling the correct manner to carry out this interpretation. Nevertheless, what I have argued is that there is no need to naturalise the intentionality of concepts/containers, because that intentionality is not how concepts are individuated, according to container theory. Concepts are individuated non-semantically by whether they are a discrete container or not. This does not mean that concepts do not have meaning. I demonstrated with the example of the owl that the owl would experience determinate meaning (in the sense of experiencing that it is thinking of this and not that because it has activated this container and not that container, and has made available this set of inferential information and not that set of information). There is no need for the owl or its cognitive system to also know just which category X this is. And this should be no surprise, for that is also the way the nominal containers that are words operate. For example, what the word 'mind' means depends on which recognitional and functional information it gathers together,
and different theorists may interpret this differently. The fact that it may not be possible to specify what a mind is does not prevent members of a community using the concept successfully. This is so even if those members have rival interpretations. It is also the case that interpreters can engage in useful and interesting substantive debate about what the concept refers to. This might even cause the meaning of 'mind' to shift in the community over time. The point is that concepts can be individuated by the container vehicle, and do not have to be individuated by an unambiguous and precise meaning. Consequently, there is no significant cost if the meanings of concepts cannot be naturalised.

G. Do concepts have senses, narrow content, inferential roles?

I have set out how concepts contain recognitional and functional information, and that this information can be regarded as sense, as narrow content (content that supervenes on the brain), or as inferential role. I have stressed that one should not think that concepts are constituted by sense, narrow content, or inferential role. Rather, concepts contain information that constitute such entities. I have also pointed out the role DA-finsts play: typically the subject arm of a DA-finst has reference (what it represents is determined by the entity it picks out), while the predicate arm of a DA-finst assigns a sense to the entity picked out. What the subject arm picks out typically has wide content (it depends on what is in the environment); but the understanding or inferential role that is provided by the information in the container/concept as picked out by the predicate arm has narrow content (understanding or inferential role depends entirely on what is in the head).

To sum up, I claim that a concept is a container of stored knowledge pertaining to a single category. The contents (the stored knowledge) of a concept facilitate inference and guided interaction. Symbols of concepts (DA-finsts in container theory) facilitate propositional thought. This notion of a container of stored knowledge allows concepts to be individuated by non-semantic structure (the structure of the container), and not by, for example, sense or reference, thus avoiding the problem that other theories of concepts have so far been unable to solve. I believe that container theory satisfies the desiderata for concepts better than any other theories.
References


228


Rensink, R., O'Regan, K., & Clark, J. (1997). To see or not to see: the need for attention to perceive changes in scenes. *Psychological Science, 8*, 368-373.


