ISSUES IN PRAGMATICS (PLIN 3001) 2006-7

LEXICAL PRAGMATICS

5. Fodor, concept acquisition and innateness

1. Introduction

As we saw in Lecture 2, Fodor sees his attack on the classical theory of concepts, and his conclusion that most word meanings are simple, unanalysable concepts, as having important consequences for innateness and concept acquisition. In particular, he has a simple argument to show that if his view of word meanings is right, then most lexical concepts (i.e. concepts encoded by words like *telephone, horse, tulip*), are innate:

Premise 1: Most lexical concepts (e.g. GIRAFFE, ELECTRON) are simple and unanalysable.
Premise 2: Simple, unanalysable concepts can't be learned.
Conclusion 1: Hence, most lexical concepts (e.g. GIRAFFE, ELECTRON) can't be learned.
Premise 3: If a concept can't be learned, it must be innate.
Conclusion 2: Hence, most lexical concepts (e.g. GIRAFFE, ELECTRON) are innate.

As we also saw, many cognitive scientists feel this conclusion is outrageous, and therefore reject Fodor's account of word meanings in favour of alternative accounts such as prototype theory, which are more compatible with the view that some aspects of concepts can be learned. Today, I want to look at Fodor's argument for innateness more closely and argue that there is a slightly weaker conclusion which is compatible both with his account of word meaning and with what is known about how concepts are acquired (and which still allows for a substantial element of innateness in concept acquisition). We'll then use Fodor's account of word meanings and concepts as the basis for our discussion of lexical pragmatics in the second half of term.

2. What is learning?

Generally, psychologists have not been very good at examining the content of notions such as 'learned' and 'innate'. If you doubt this, you might look up the entry on 'Learning' in David McFarland (ed.) (1981) *The Oxford Companion to Animal Behaviour* (an interesting survey of the varieties of ways in which animals might be thought of as 'learning'). This starts:

"Learning is a familiar enough phenomenon, but, as is often the way, one not easily captured by the scientist's definition". (McFarland 1981: 336)

and goes on:

"It may be foolish, therefore, to waste too much time attempting to provide a precise, all-embracing definition of learning at this point. For the time being, it will be more useful to provide particular descriptions of some of the particular situations in which learning has been studied". (337)

Notice that this presupposes an ability to identify situations in which 'learning' is taking place. The article ends by 'defining' learning simply as what takes place in these particular 'learning situations'. In other words, although it is full of interesting facts, it fails to say what learning is, or why the psychologists think that learning is taking place in these particular situations. Yet if we want to think seriously about whether concepts are learned or innate, we do need to provide theoretical accounts of both these terms.

Fodor's characterisation of learning ('The present status.of the innateness controversy', pp 275 ff.) at least improves on this. He starts by noting that learning obviously involves experience. For learning to take place, the environment must impinge in some way on the organism, and indeed it must affect the mental representations of the organism. If someone drops a brick on my foot, the environment is impinging on me all right, but it will not result in learning unless I form some thought – say, that my foot hurts because someone has dropped a brick on it – as a result of that experience. Learning, then, involves a modification of the individual's **cognitive** system as a result of some change in the **environment**.

Fodor claims that this revised definition is not specific enough. Not all modifications to our cognitive system that result from experience should count as learning. For example, we can imagine being hit on the head, or undergoing brain surgery, and waking up knowing Russian, or the theory of relativity. Although this would be a modification of our cognitive system as a result of some change in our environment, we would not call it **learning**. And here is the central point. For Fodor, what is crucial to learning is that it is a **rational** process: that is, a process of hypothesis formation and confirmation in which the change in our environment provides **evidence** for the hypothesis we eventually adopt. The reason we don't **learn** Russian or the theory of relativity by being hit on the head is that the relation between the experience and the resulting change in our cognitive system is **arbitrary**.

For Fodor, then, learning is defined as involving the construction and evaluation of hypotheses on the basis of evidence. If hypothesis formation and evaluation is not possible, then according to Fodor, learning is not possible. Given this definition is there any sense in which

new concepts can be learned? Fodor argues that, if learning involves the construction and evaluation of hypotheses, only definable concepts (e.g. BACHELOR) can be learned. And since simple unanalysable concepts (e.g. HORSE, TULIP, ELECTRON, TELEPHONE, etc.) can't be learned, they must be innate. Notice that the whole argument for concept innateness depends on this particular definition of learning, plus the one-step argument from 'This is not learned' to 'This is innate'. We'll consider some alternative accounts of learning later on.

3. <u>Why Fodor thinks simple, unanalysable concepts can't be learned.</u>

In 'The present status of the innateness controversy', pp 265-73), Fodor looks at what psychologists call 'concept learning' experiments, which typically involve the acquisition of complex, definable concepts, and do involve hypothesis formation and confirmation. Here's an example. The experimenter makes up a word, say *grack*, to which she assigns an arbitrary meaning, say [BLUE AND CIRCULAR]. The subject is given a set of cards with various coloured shapes on them, and has to decide which ones are GRACK. He does this by a process of hypothesis formation and confirmation. For example, he may start with the hypothesis that GRACK means BLUE, and point to a blue circle and ask the experimenter 'Is this grack?' The experimenter will say 'yes', and the hypothesis will be confirmed. On the other hand, if he points to a blue square and asks 'Is this grack?', the experimenter will say 'No' and the hypothesis will be disconfirmed. After forming and testing various hypotheses, the subject should eventually arrive at the correct hypothesis that GRACK means [BLUE AND CIRCULAR].

Fodor is reluctant to say that a new **concept** has been learned even in this case: what happens is merely that the subject learns to attach a new word 'grack' to a combination of concepts (BLUE, CIRCULAR) which he already had. However, he is prepared to concede that we might be said to 'learn' definable concepts such as GRACK, or BACHELOR by constructing and confirming a hypothesis such as (2) or (3):

(2) An object is GRACK if and only if it is GRACK.
(3) An object is a BACHELOR iff it is UNMARRIED, ADULT, HUMAN AND MALE.

So for Fodor, there is a sense in which complex, definable concepts can be learned.

Now let's consider how the child might learn simple, unanalysable concepts such as BIRD or ELEPHANT by the same method of hypothesis formation and confirmation. To learn the meaning of the word *bird*, he would have to be able to construct a hypothesis like (4):

(4) *Bird* means BIRD.

But to construct this hypothesis, he would already have to have the concept BIRD. Question: could the concept BIRD be learned by hypothesis formation and confirmation, as with GRACK or BACHELOR in (2) or (3) above? Answer: only by constructing a hypothesis like (5):

(5) An object is a BIRD if and only if it is X, Y, Z.

But, of course, simple, unanalysable concepts can't be defined in this way. So, according to Fodor, simple unanalysable concepts can't be learned; and according to Fodor, they must therefore be innate.

Fodor supports this claim that simple, undefinable concepts can't be learned by citing an impressive range of quotations from empiricist philosophers and psychologists who are prepared to concede that simple concepts which can't be decomposed into a definition must be innate (p. 272-7). For example, Hume, a committed empiricist, seems to have said quite clearly that simple, sensory concepts (e.g. RED) are innate. The question we need to consider now is, what is it for a concept, any concept, to be innate? (For arguments that simple concepts *can* be learned, see Margolis 1998; Laurence & Margolis 2002; for weakening of Fodor's views, see Fodor 1998, chap 5; we'll return to this below).

4. <u>What is innateness?</u>

Notice, first, that concepts –whether we see them as learned or innate – are not acquired independently of experience. For example, if you've never seen a bird, or a picture of a bird, or heard birds spoken of, it is most unlikely that you will acquire the concept BIRD. What follows from this is that, if the concept BIRD is innate, **innate** cannot mean 'acquired independently of experience'. And indeed, for Fodor, Chomsky, and most others currently working in cognitive science, it does not.

For Fodor, the crucial difference between learned and innate concepts is not that learned concepts are acquired from experience while innate concepts are acquired independently of experience. Experience is required in both cases, and what distinguishes the two types of case is the role that experience plays. As we have seen, for Fodor, learning is a rational process of hypothesis formation and confirmation, in which experience provides **evidence** for the hypothesis that is learned. By contrast, he claims that innate concepts are **triggered** by experience, and that triggering is not a rational process: the triggering experience does not provide **evidence** for the resulting hypothesis. For Fodor, the relation between triggering

experience and resulting concept is arbitrary rather than evidential: in principle, as arbitrary as the relation between being hit on the head and waking up knowing Russian.

Fodor's examples of triggering experiences are generally taken from ethology (the study of animal behaviour in natural environments). Thus, he notes that a baby goose becomes imprinted on the first moving object it sees, and treats it as its mother. Clearly, in such cases the gap between triggering experience (moving object) and acquired concept (MOTHER) can be quite large, and this is what makes the difference between triggering and learning: with triggering, the acquired concept goes well beyond the evidence provided. The notion of triggering, and the gap between experience and the resulting state of the cognitive system, also appears in Chomsky's arguments that language is not learned but grows. In the last 10 or 15 years, this approach has been reinforced by a mass of results in evolutionary psychology, which apply Chomsky's **poverty of the stimulus** argument in a variety of other conceptual domains. The claim is that the concepts we acquire are picked up so quickly, and go so far beyond the evidence, that it is more plausible to think of them as having evolved over millennia rather than being learned from scratch in a few months or years. (See for example Tooby & Cosmides 1997; Plotkin 1997, chap 4). We'll look more closely at some of these arguments towards the end of term.

There have been many attempts to show that Fodor's arguments for the innateness of concepts are unsound, or to develop alternative accounts of concept acquisition that would avoid them. One line of investigation we considered in Lecture 3 was to concede Fodor's argument that concepts do not have the classical structure of definitions, but deny his conclusion that they therefore have no internal structure at all, and claim instead that concepts have the internal structure of a prototype. We saw that this was inadequate as an account of word meaning. A second line of investigation is to look more carefully at Fodor's one-step argument from 'This concept is not learned' to 'This concept is innate', and to find a third position which avoids the more extreme consequences of Fodor's position – say by expanding our notion of learning or our notion of innateness. Here is how this might be done.

Let's agree (for the moment) with Fodor that most concepts are **triggered** rather than **learned.** Still, there are two possible ways of interpreting what **triggering** means:

6a. The **concept itself** is innate and merely **activated** by experience (e.g. BIRD by birds).
6b. We have **innately-determined** concept construction **mechanisms** which are **activated** by experience and construct concepts that go well beyond the evidence.

Fodor himself seems to have interpreted the notions of innateness and triggering pretty strongly in his early work (e.g. 'The present status…'), so that **innate** meant that the concept itself was fully formed and merely awaiting activation by experience, and **triggering** was simply the activation of a pre-formed concept, as in (6a). But there is room for alternative interpretations, and alternative accounts of how simple concepts are acquired. Much recent research on concept acquisition suggests a rather weaker interpretation, where **triggering** means something more like 'growing', as in Chomsky's claim that language is not learned, but grows in the mind. On this approach, the idea is that concepts are not pre-formed and awaiting activation, but constructed by **innately-determined concept-construction mechanisms** which may differ substantially across cognitive domains.

In the next section, we'll look briefly at the idea that there are innately-determined **kinds** of concepts, each with its own type of **innately-determined construction mechanism**. If this were on the right lines, then we could combine Fodor's view of word meanings with a much less outrageous view of innateness and triggering, along the lines in (6b) above.

5. Concepts and conceptual domains

Many philosophers and linguists have suggested that concepts come in several different **kinds**: e.g. LIVING THINGS, PHYSICAL OBJECTS, BELIEFS, ARTEFACTS, etc**.** The recent psychological literature on concept acquisition and cognitive development points in the same direction, suggesting a quite rich and differentiated set of conceptual kinds or domains.

Let's illustrate by looking at the work of Frank Keil (1992) on the development of several kinds of concept:

(a) moral terms (e.g. 'lie', 'tease', 'steal');
(b) meal terms (e.g. 'breakfast', 'dinner');
(c) tool terms (e.g. 'hammer', 'saw');
(d) kinship terms (e.g. 'grandfather', 'aunt');
(e) cooking terms (e.g. 'bake', 'boil').

Keil noticed that in all these domains, children start by categorising objects using representations based on clusters of superficial perceptual features (i.e. stereotypes), and then shift to more abstract descriptions or definitions. Take the concept ROBBER from the moral domain. At the first stage, looking *like* a robber is decisive: a kind, friendly woman who doesn't *look* like a robber, but takes something from you and doesn't give it back, would not be categorised as a robber. Take the concepts BREAKFAST and SUPPER. At the first stage, a meal of

orange juice, bacon and eggs, which *looks* like breakfast but is eaten at night, would be categorised as breakfast rather than dinner. Later, children shift to using more abstract definitions, and recognise that despite appearances the kind, friendly woman is a robber, and the meal eaten at night time is dinner or supper. The shift generally takes place at the same time for all the concepts within a domain, but different domains make the shift at different times. Interestingly, the shifts are made in the order listed in (a)-(e) above: i.e. moral terms are the first to make the shift from concrete to abstract, and cooking terms are last.

These basic developmental facts are compatible with many different stories about concept acquisition, both empiricist and innatist. A full-fledged empiricist story might go as follows: at the first stage, concepts are constructed by observing statistical associations among perceivable features of the world (‘stereotypes’); as the child matures, it develops theories to explain the associations and categorises objects in terms of the theories. The child is a little scientist, tabulating regularities and constructing explanatory generalisations by constructing and evaluating hypotheses based on the available evidence. The fact that different domains shift at different times does not reflect any internal differences in cognitive structure, but merely the accessibility of the theoretical vocabulary, which may be more concrete (closer to perceptual features) in some domains than in others. Here, the only difference from classical empiricist philosophers is in the idea that the 'mental operations' involved in concept construction are inductive inference processes of hypothesis formation and confirmation rather then logical procedures like addition, negation, disjunction, etc. More recently, psychologists (including Keil himself) have generally rejected this view in favour of more innatist accounts.

A full-fledged innatist story might say that *no* inductive inference processes are involved in concept construction: it's triggering all the way down. The mind consists of a set of innately determined conceptual modules, with their own internal concepts and mechanisms, which develop in their own time according to a genetic programme. The fact that different domains shift at different times may reflect (a) maturation of a later module, which takes over from an earlier one; or (b) maturation of the initial module, with more sophisticated techniques. On this approach, the idea that the child is a little scientist, tabulating regularities and developing inductive generalisations to explain them, has no place at all.

The empiricist account broke down mainly because children turned out to be born knowing much more about kinds of concepts than would be explicit on classical empiricst accounts. We’ll look more closely at studies of concept acquisition in Lecture 9. For the moment, I’ll simply summarise some of the arguments of Gabriel Segal (1996) against the empiricist view that the child is a little scientist, making observations and constructing theories

by inductive learning mechanisms along the lines suggested above, and suggest a more innatist approach:

1. On the empiricist approach, the child has to be not just a little scientist but a brilliant one, and this is puzzling because their knowledge is rather patchy (e.g. children seem to know the principles of contact mechanics but not the equally basic principle of inertia). This suggests special-purpose abilities rather than general-purpose ones.

2. The patterns and results of development are very similar across individuals and cultures. If children are little scientists, how do they find the same hypotheses so quickly? This suggests special-purpose rather than general-purpose abilities. ('Poverty of the stimulus' argument.)

3. One of the main forces in genuine scientific change is conscious reflection: the scientist is explicitly aware of counter-evidence and consciously tries to come up with an alternative account. This is clearly not what happens with children.

4. Perhaps the most compelling argument is the existence of Williams Syndrome. This is a rare genetic disorder resulting in average IQ of around 50, but combined with a high level of linguistic ability and, crucially, relatively high social skills. Williams Syndrome children pass the False-Belief (Sally-Anne) task with ease (Theory of Mind system 2), and also pass tests requiring the ability to attribute goals (Theory of Mind system 1). However, their general ability to acquire theoretical, explanatory knowledge is very poor. Similarly, as is well known, there are highly intelligent Asperger Syndrome individuals with good general explanatory skills but very deficient Theory of Mind skills. If anything, these people show what it is like to acquire Theory of Mind by reflection rather than by instinct, and their differences from others confirm the modularity of Theory of Mind (Sperber & Wilson 2002)

5. Despite what the empiricists say, it's hard to see how richer concepts can emerge by induction from more impoverished theories. This was Fodor's original argument (in *Language of Thought*), which started the whole recent innatist programme in developmental psychology.

On the whole, these arguments suggest a generally innatist direction for concept acquisition, which we'll look at more closely towards the end of term.

6. Fodor's recent views on innateness

Let's now return to Fodor's original argument that most word meanings are atomic concepts, which can't be learned by hypothesis formation and confirmation, and must therefore be innate. The recent developmental literature suggests an alternative picture. What is innate is not individual concepts but the **formats** and **mechanisms** for constructing concepts in different domains. In the natural-kind domain, for example, the child has (a) a propensity to pick out natural kinds by their stereotypical properties (shape, colour, movement, etc.), and (b) the assumption that natural kinds have a hidden causal structure (we'll discuss this further in lecture 5). As we've seen in earlier lectures, we could imagine that natural-kind concepts contain a **detector**, or **identification procedure**, based on the stereotypical properties, a format for constructing **logical entries** containing the information that a CAT, for example, is an ANIMAL OF A CERTAIN KIND, a TROUT is a FISH OF A CERTAIN KINE, and so on. Acquisition of particular natural-kind concepts could be triggered by encounters with the actual animal (say), or with a picture or film, and would involve filling the slots in a pre-determined format. There will be different kinds of format for different kinds of concept, with formats and construction mechanisms, but not the individual details, being innate.

This sounds a quite plausible picture, which preserves the idea that concepts are not learned by hypothesis formation and confirmation, but does not commit us to radical concept nativism (the assumption that all our individual concepts are pre-formed and innate). Recall that Fodor's original argument for innateness took only one step: concepts are not learned; hence, concepts are innate. But as noted above, the claim that concepts are not learned is compatible with two rather different pictures of how they are acquired: (a) the innate **mechanisms** for concept construction are not inductive learning mechanisms; (b) the concepts themselves are innate. Fodor's one-step argument showed only that **one** of these two conclusions must be correct. The conclusion that all concepts are pre-formed and innate was not warranted.

Interestingly, Fodor himself seems to have come round to this position more recently (Fodor 1998: 141-3). His model for concept acquisition is the sensory concept RED, and he considers whether, given the general agreement that we have an innate sensorium with mechanisms which trigger the concept RED on exposure to red things, it follows that the concept RED is innate. His answer is 'no'. We must distinguish innate **mechanisms** for concept construction from innate **concepts***:*

'Since the sensorium isn't an idea, it is a fortiori not an *innate* idea. So, strictly speaking, the innate sensorium model for the acquisition of RED doesn't require that it, or any other concept, be innate'. (1998: 142).

He then generalises the argument to other concepts, such as DOORKNOB, or BROCCOLI, and argues that here too, we only need innate concept-construction mechanisms, not innate concepts:

'The kind of nativism about DOORKNOB that an informational atomist has to put up with is perhaps not one of *concepts* but of *mechanisms.* (1998: 142).

This fits well with the sort of modular, domain-specific arguments that evolutionary and developmental psychologists have been favouring. Fodor's conclusion: 'Maybe there aren't any innate *ideas* after all'. (1998: 143). (See Margolis 1998; Laurence & Margolis forthcoming for further discussion.)

Reading

Fodor, J. 1981 The present status of the innateness controversy, pp. 257-292 and pp 298-end, in J. Fodor *Representations*.

Margolis, E. 1998 How to acquire a concept. *Mind & Language* 13: 347-69 (for arguments that Fodor's atomistic concepts can be learned).

Background references

McFarland, D. 1981 *Oxford Companion to Animal Behaviour*. OUP, Oxford.

Bloom, Paul 2000 *The Acquisition of Meaning.* CUP, Cambridge.

Tooby, J. & L. Cosmides 1997 'Foreword' to S. Baron-Cohen (ed.) *Mindblindness: An Essay on Autism and Theory of Mind.* (a quick introduction to evolutionary psychology).

Laurence, S. & Margolis, E. 2002 Radical concept nativism. *Cognition* 86: 25-55.

Fodor, J. 1998 *Concepts*. chaps 6 and 7. (for arguments that maybe it's not the concepts but the triggering mechanisms that are innate).

Keil, F. 1994 The birth and nurturance of concepts by domains: The origins of concepts of living things. In L. Hirschfeld & S. Gelman (eds) *Mapping the Mind: Domain Specificity in Communication and Culture*. CUP.

Gopnik, A. 1996 Theories and modules: creation myths, developmental realities, and Neurath's boat. In P. Carruthers & P. Smith (eds) *Theories of Theories of Mind*. CUP.

Spelke, E. 1990 Principles of object perception. *Cognitive Science* 14: 29-56

Leslie, A. 1987 Pretense and metarepresentation: The origins of 'theory of mind'. *Psychological Review* 94: 412-26.

Leslie, A. 1994 ToMM, ToBY and Agency: Core architecture and domain specificity. In L. Hirschfeld & S. Gelman (eds) *Mapping the Mind: Domain Specificity in Communication and Culture*. CUP.

Segal, G. 1996 The modularity of theory of mind. In P. Carruthers & P. Smith (eds) *Theories of Theories of Mind*. CUP.

Premack, D. 1990 The infant's theory of self-propelled objects. *Cognition* 36: 1-16.

Premack, D. & A. Premack 1994 Moral belief: Form versus content. In L. Hirschfeld & S. Gelman (eds) *Mapping the Mind: Domain specificity in Communication and Culture*. CUP.