

Introduction to Language - Lecture Notes 2A

Language and Thought

Goal: A popular view (the Sapir-Whorf hypothesis) suggests that our thoughts are determined by the systems of classification of the particular language we speak. According to this hypothesis, the mental universe of an English speaker may be completely different from that of a Chinese speaker because they happen to speak different languages. This view has not received scientific support - quite the opposite. Furthermore, the cases of dissociation studied at the end of Lecture Notes 1A suggest that we do not literally 'think in words': if we did, patients with a language deficit should automatically have a deficit in thought as well, which does not appear to be the case. Thus verbal language and thought should in principle be taken to be distinct. This does not mean that thought is not a system that manipulates symbols; in fact a widespread contemporary model, the 'computational model of the mind', suggests that the mind should be analyzed by analogy with a computer, which manipulates abstract symbols. On this view, thought is just *symbol manipulation*. But the symbols in question need not be part of verbal language; they may be part of what Pinker calls 'Mentalese', which is just another term for 'language of thought'.

Thus on closer inspection, the Sapir-Whorf hypothesis in its unrefined form ('language determines thought') is not particularly plausible. Still, there are considerably more refined contemporary attempts to show that *some* aspects of language might determine *some* aspects of thought. Two such attempts are reviewed in the last section of these Lecture Notes. The first one tries to establish that *variable* (i.e. non-universal) aspects of language affects *spatial reasoning*. The second attempt tries to establish that some *universal* aspect of language affects *reasoning about other minds* (specifically, the ability to represent false beliefs). Both attempts are still the object of heated -and exciting- debates.

1 The Sapir-Whorf Hypothesis

Sapir-Whorf Hypothesis (=linguistic determinism): People's thoughts are determined by the categories (=systems of classification) made available by their language (cf. Pinker's *Language Instinct* p. 46).

(The intuitive motivation for such a hypothesis is that we think in words - a claim that appears to be false, at least in such a strong, unqualified form).

1.1 *Eskimo Vocabulary: 'snow'*

Whorf ('Science and Linguistics'): "We have the same word for falling snow, snow on the ground, snow packed hard like ice, slushy snow, wind-driven flying snow - whatever the situation may be. To an Eskimo, this all-inclusive word would be almost unthinkable; he would say that falling snow, slushy snow, and so on, are sensuously and operationally different, different things to contend with; he uses different words for them and for other kinds of snow."

Pinker's reply (*Language Instinct*): (i) Eskimo is no different from English in this respect! English has lots of terms for snow too: *snow*, *sleet* (=rain that is partly frozen), *slush* (=snow that is partly melted), blizzard, avalanche, hail (=frozen raindrops), hardpack, powder, flurry (=sudden, light fall of rain or snow), dusting...

(ii) Even if Eskimo *were* had more terms for snow than English, one should say: so what? As Geoffrey Pullum wrote (*The Great Eskimo Vocabulary Hoax*) [see Pinker p. 55]:

Imagine reading: 'It is quite obvious that in the culture of printers... fonts are of great enough importance to split up the conceptual sphere that corresponds to one word and one thought among non-printers into several distinct classes...' Utterly boring, even if true.

1.2 Hopi conception of time

Whorf ('Science and Linguistics'): 'Hopi may be called a timeless language'. [...] It 'does not distinguish between present, past, and future of the event itself (...)'.

Pinker's reply (*Language Instinct* p. 53): 'What, then, are we to make of the following sentence translated from Hopi?

Then indeed, the following day, quite early in the morning at the hour when people pray to the sun, around that time then he woke up the girl again.

Perhaps the Hopi are not as oblivious to time as Whorf made them out to be. In his extensive study of the Hopi, the anthropologist Ekkehart Malotki, who reported this sentence, also showed that Hopi speech contains tense, metaphors for time, units of time (including days, numbers of days, parts of the day, yesterday and tomorrow, days of the week, weeks, months, lunar phases, seasons, and the year), ways to quantify units of time, and words like 'ancient', 'quick', 'long time', and 'finished'."

The Fallacy: The fact that the *grammar* of a given language (e.g. the inventory of its pronouns or its rules of agreement) does not distinguish between two objects entails **neither**

(a) that the language does not provide other means to draw the distinction, **nor**

(b) that the speakers of the language treat these objects as belonging to the same class for non-linguistic purposes.

Examples:

(i) In English a baby may be referred to using the pronoun *it*. This does not entail that English speakers treat babies as inanimate creatures.

(ii) In English a ship may be referred to with the feminine pronoun *she*. This does not entail that English speakers cannot tell the difference between females and ships.

(iii) In German the term for 'young woman' (*Mädchen*) is neuter. This does not entail that German speakers treat young women as inanimate objects.

(iv) In Mandarin Chinese there is no difference in the pronunciation between feminine and masculine pronouns - e.g. both *he* and *she* are: *ta*¹ (here ¹ indicates that the word is pronounced with a tone [tone #1 in a list of 4 different tones]). This does not entail that Chinese speakers treat males and females in the same way.

(v) In Chinese there is no tense. But there are all sorts of temporal adverbs, such as *earlier*, *later*, *yesterday*, *tomorrow*, etc. Of course this does not entail that Chinese speakers do not conceive time as we do.

1.3 *Dissociations between Language and Thought*

If we simply 'thought in words', impairments of language should systematically lead to disruptions of thought. But as was discussed in Lecture Notes 1A, there are cases of dissociation (e.g. Selective Language Impairment, Broca's Aphasia) in which language is impaired but other cognitive abilities are not. (Refer to Lecture Notes 1A for more information)

Conclusions: (i) There has been no clear proof that any important aspect of thought is determined by the particular language that the subject speaks. (This does not necessarily mean that such a proof could not be found).

(ii) Still, one could attempt to show that *some* aspects of language determine *some* aspects of thought. This could be done in two ways:

A. One could try to show that some *variable* (i.e. non-universal) aspects of language determine some aspects of thought. This would be a weakened form of the Sapir-Whorf hypothesis: two individuals might think about some aspect of reality differently because they speak different languages

B. If one accepts the hypothesis that there exists a Universal Grammar, which is innate and is common to all languages, one could try to show that some *universal* aspect of language determines some aspect of thought. This would not entail any cognitive difference between individuals that speak different languages, and would thus be a rather different claim from the Sapir-Whorf hypothesis.

See Section 3 for some discussion of A. and B.

2 Thought Without Language

2.1 *Thought Without Language: An Example*

Pinker discusses several examples of thoughts that do not appear to be represented in the mind in anything like verbal language (*Language Instinct* pp. 57-63). Several examples are somewhat speculative. But one is particularly striking: when asked to determine whether different shapes are a stilted or a mirror-reversed version of a given letter (e.g. the letter F), subjects take longer to reply when the angle at which the letter is stilted is greater. For instance an answer comes faster for a letter that is stilted at a 45 degree angle than for one that was at a 90 degree angle. This would be entirely surprising if subjects compared the relevant shapes through some sort of verbal description; on the other hand the result is expected if subjects perform a mental rotation of the objects in question, to determine whether their shapes match. See Pinker's *Language Instinct* p. 62-63 for details.

Note: Some important aspects of this discussion, covered in class, are not reproduced here.

2.2 The Computational Model of the Mind: Thought as Symbol Manipulation

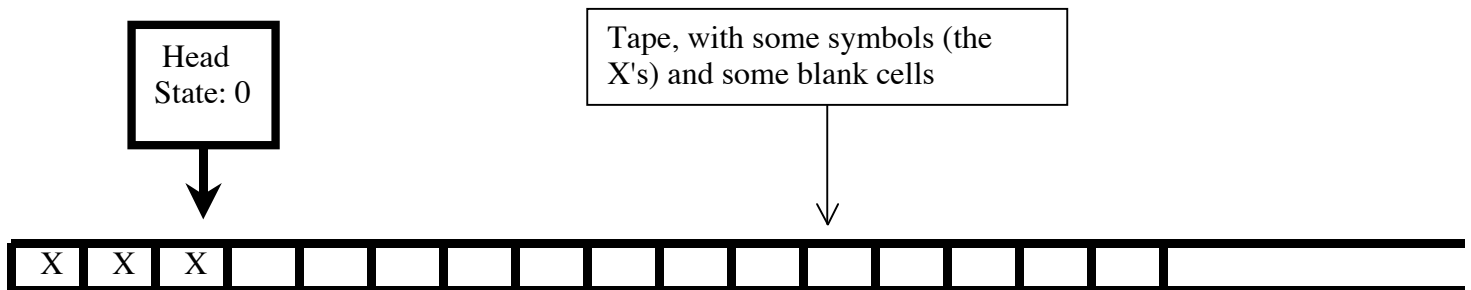
(This section is intended to supplement pp. 64-69 of Pinker's *Language Instinct*, which are somewhat allusive, esp. with respect to the notion of a 'Turing Machine').

How can we think without words? By performing mental computations with non-linguistic symbols - in a kind of 'language' (not a verbal language) that Pinker calls 'Mentalese', or Language of Thought. On this model, the mind functions like a computer, manipulating simple symbols to produce complex results. The underlying theoretical model is called a 'Turing Machine', which is *the most powerful model of computation that is known*. Now one would not want to say that the mind *is* a Turing Machine - if so it could compute *everything* that can in principle be computed, which isn't the case (can you compute everything?). But the idea is that there are no thoughts that are so complex as to be in principle beyond the reach of any machine.

What, then, is a Turing Machine? (Pinker does not make this very explicit). Surprisingly, it is a very simple device, which comprises:

- (i) an infinitely long tape, on which symbols may be written or erased
- (ii) a head, which may write and erase symbols (one symbol at a time), and move left or right on the tape (one step at a time)
- (iii) a list of states which the machines may be in (e.g. state 0, state 1, etc.)
- (iv) a list of instructions to the machine, of the form:

when you are in state n, reading symbol X, write instead symbol Y and move left/right, entering state n'



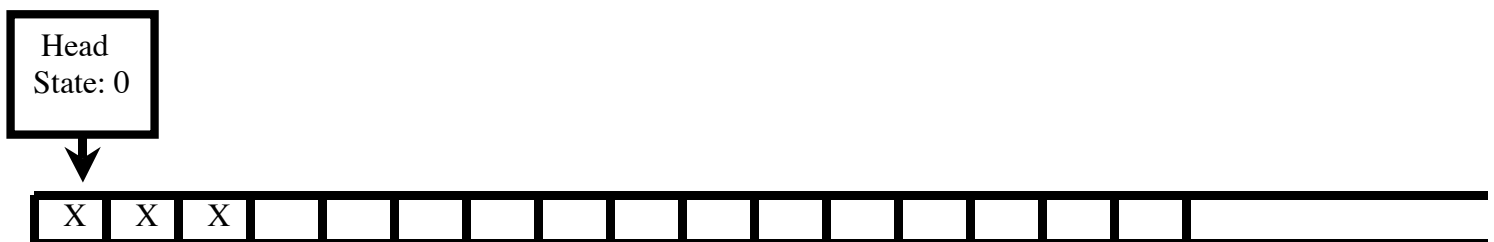
Example: The following program tells a Turing Machine to add 1 X at the end of the string of X's found at the beginning of its tape. (This can be thought of as a program that adds 1 to a number entered in unary notation, i.e. a notation in which, say, 10 is represented as a string of 10 X's, etc.)

-Program (the machine starts on the leftmost cell in state 0):

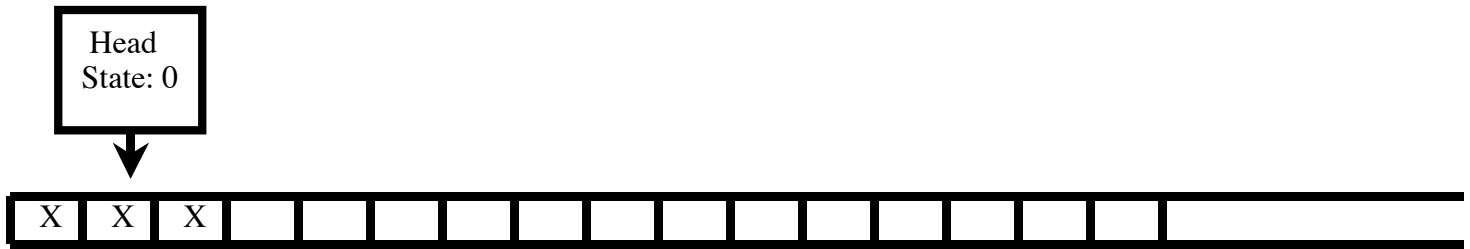
- (i) When you are in state 0, reading X, leave X as it is and move right, remaining in state 0
- (ii) When you are in state 0, reading (i.e. a blank), write X instead and enter state S (for STOP).

-How the Program Works

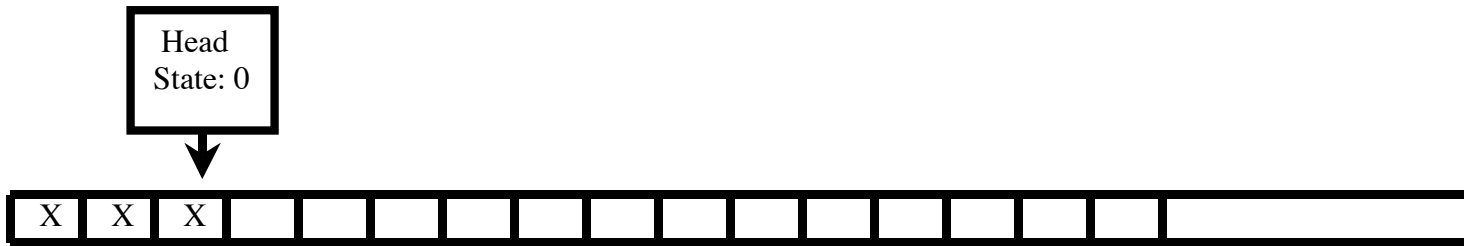
Step 1



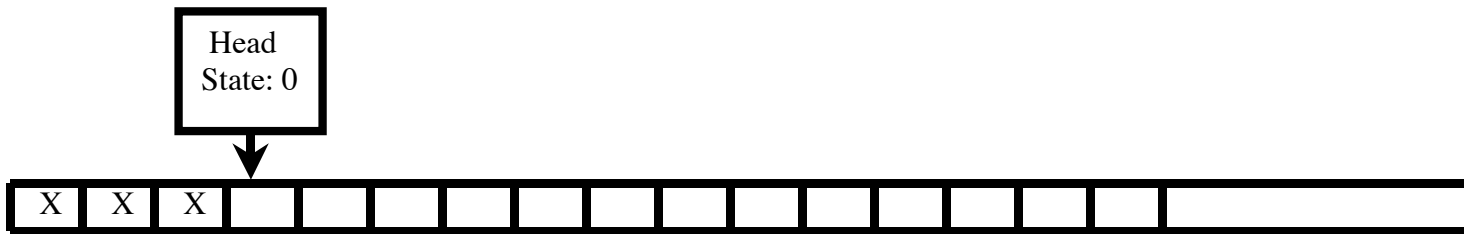
Step 2



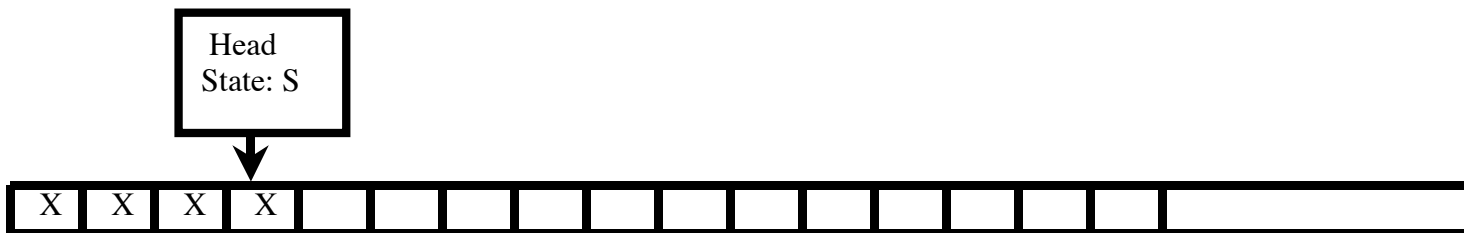
Step 3



Step 4



Step 5



Thus we see that at the end of the computation, when the machine has reached State S (for 'Stop'), there are 4 X's on the tape, i.e. one more than was originally the case. The machine has succeeded in 'adding one' to the number that was initially represented.

Simple though they are, it appears that *Turing Machines can compute everything that can in principle be computed (by any mechanical means whatsoever)*. This claim is known as Church's Thesis, or sometimes as the 'Church-Turing thesis' [Alonzo Church was for many years a professor at UCLA]:

Church's Thesis: Everything that can be computed at all can be computed by a Turing Machine.

What is the relevance of Turing Machines for our purposes? Church's Thesis suggests that we are unlikely to find symbolic computations that are beyond the theoretical reach of a machine (since Turing Machines can compute everything that can be computed at all); and therefore that we can in principle hope to analyze the mind as a mechanism that manipulates symbols as well (i.e. we are unlikely to find thoughts that are 'beyond the reach of any machine' because of their sheer complexity). Needless to say, this certainly does not mean that the mind *is* a Turing machine - among others, because Turing machines have an infinitely long tape, whereas the mind (/brain) certainly doesn't contain anything which is infinitely long (it wouldn't fit!). But the theoretical existence of Turing Machines makes it more plausible that the mind can be analyzed as machine that manipulates symbols (in effect, as a computer).

3 The Sapir-Whorf Hypothesis Revisited (and Modified!): Recent Debates

One might get from Pinker's discussion the impression that the question of the relation between language and thought has been settled once and for all. The result, one might think, is that language and thought are just different things, with little direct interaction. But in fact many aspects of the question remain open.

-First, there are attempts to defend a weakened version of the Sapir-Whorf hypothesis; their goal is to show that speakers of different languages think in different ways *in a particular domain*, and that furthermore this difference in thought is *caused* by a difference in language. For instance, according to a theory discussed below, individuals that speak different languages may use different frames of reference to situate objects in tasks of spatial reasoning [Section 3.1]

-Second, there are also attempts to show that some *universal* property of language determines some universal property of thought. For instance, according to another theory discussed below, children manage to represent false beliefs only when (and only *because*) they have mastered the linguistic form of the embedded proposition (a proposition is *embedded* if it is contained in another proposition. Thus in the bare sentence 'It is raining', *it is raining* is not embedded; but in 'John thinks that it is raining', *it is raining* is embedded). [Section 3.2].

Both debates should be considered to be open at this point.

Note: Some important aspects of this discussion, covered in class, are not reproduced here.

3.1 Do variable properties of language determine some aspects of thought? Language and Spatial Reasoning

The Argument for Linguistic Determinism

A *frame of reference* is a system that serves to classify the position of objects. Some frames of reference (called 'absolute') classify objects in the same way no matter what the position of the observer is. For instance, *New York is north of Washington* is true no matter where you or I stand. Other frames of reference (which we will call 'non-absolute') do not have this property. *Washington is to my left* may be true or false depending on where I stand; and similarly for: *Washington is to the left of New York* (true if you are looking from the Atlantic Ocean; false if you are looking from California).

Some languages may, in day to day life, display a preference for one kind of frame of reference over another. While it would be in principle possible to say in English: 'Give me the spoon that's northeast of your tea cup', we are more likely to say '... to the left of your teacup'. Thus to localize objects that are in our immediate environment we generally use a non-absolute frame of reference (of course if we wish to make general geographical statements, this is no longer true, as in the examples above involving New York and Washington). By contrast, some languages simply lack a non-absolute frame of reference, and classify objects in terms of absolute coordinates, even for objects that are in the speaker's immediate environment. Stephen C. Levinson and his co-workers claim that this *linguistic* difference across social groups is responsible for differences in *spatial reasoning*. They state their main point as follows:

'There are human populations scattered around the world who speak languages which have no conventional way to encode *left, right, front, and back* notions, as in *turn left, behind the tree, and to the right of the rock*. Instead, these peoples express all direction in terms of cardinal directions, a bit like our *East, West, etc.* Careful investigation of their non-linguistic coding for recall, recognition, and inference, together with investigations of their dead-reckoning abilities and their on-line gesture during talk, shows that these people think the way they speak, that is, they code for memory, inference, way-finding, gesture and so on in 'absolute' fixed coordinates', rather than in non-absolute coordinates.

[Levinson et al., 'Returning the tables: language affects spatial reasoning', *Cognition* 84 p. 157]

Here is a type of experiment used by Levinson and his co-workers to prove their point (the figure and discussion are from Li & Gleitman's 'Turning the tables: language and spatial reasoning', *Cognition* 83, 2002)

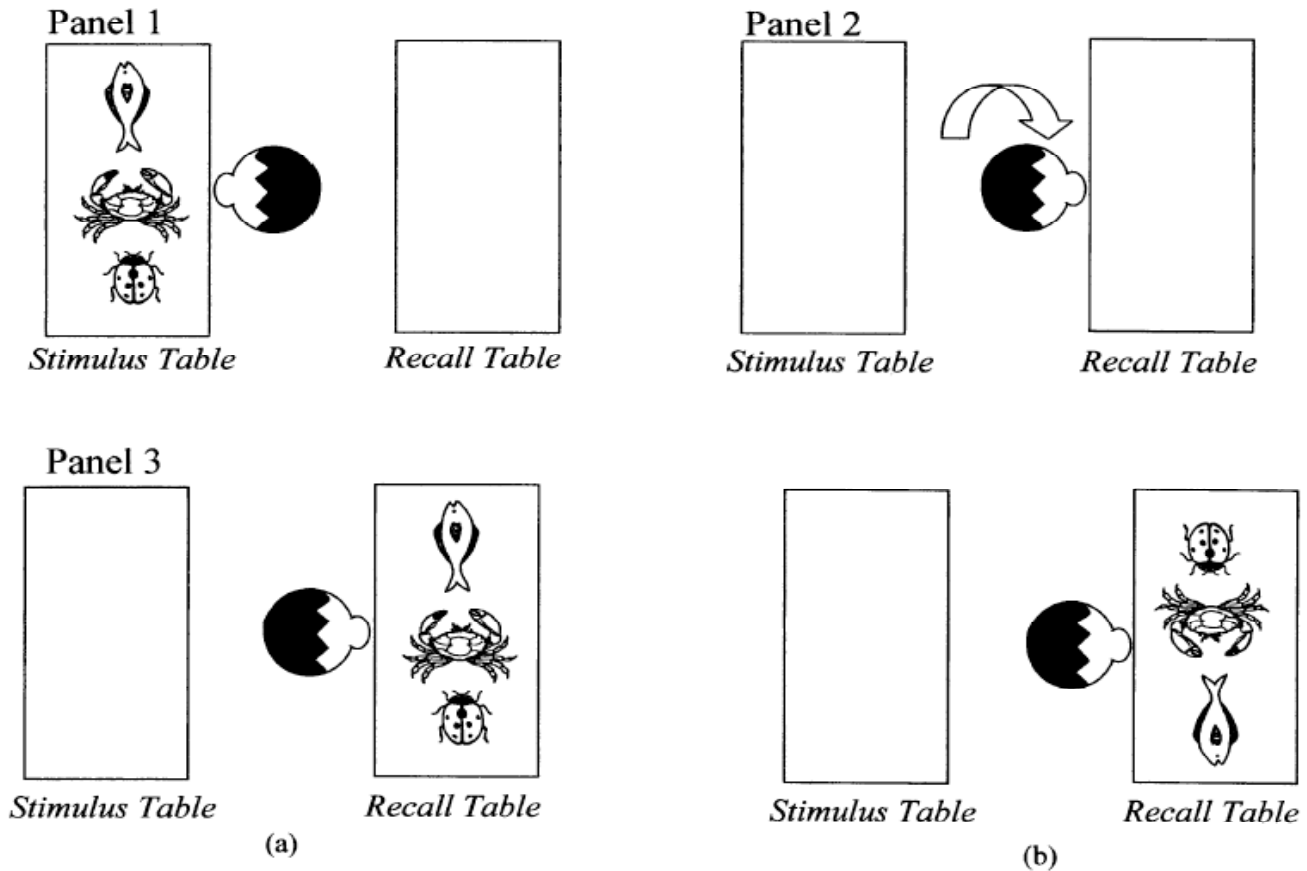


Fig. 1. The Animals in a Row task. Panel 1: Stimulus Table with array of toy animals; Panel 2: rotated 180 degrees at Recall Table; Panel 3a: the “absolute” rearrangement; Panel 3b: the “relative” rearrangement (adapted from Pederson et al., 1998).

The experiment proceeds as follows:

- First, the experimental subjects memorize the positions of three animals, positioned in a line in front of them on a tabletop, the 'Stimulus Table' (Panel 1).
- Second, the animals are removed from view, and after a brief delay, the subjects are turned around (usually by 180 degrees, as in the 'Recall Table' of Panel 2).
- Third, the subjects are handed the three original animals in random order and asked to position them 'in the same way as before'.

Suppose that the animals were originally displayed with their noses facing north, which happens to be to the subject's right as he faces the Stimulus Table. There are two possible outcomes:

- A. If the subject places the animals on the Recall Table still facing north, he has used an *absolute frame of reference*, as in Panel 3a.
- B. If the subject places the animals on the Recall Table still facing right, he has used a *non-absolute frame of reference*, as in Panel 3b.

In detailed studies, Levinson and co-workers have shown that speakers of Tenejapan Mayan, a language which only has absolute frames of reference, choose Solution A. By contrast, speakers of Dutch, which prefers non-absolute frames of reference, go for Solution B. Levinson and co-workers argue that spatial reasoning is thus influenced by language.

A Critique

The above argument has not gone unchallenged, however. In a recent critique, P. Li and L. Gleitman claim that in Levinson's studies two parameters were varied at once: (a) the particular language that the subjects spoke, and (b) the environments in which they were tested. This is because, for instance, the Tenejapan population was tested 'on its hills, out of doors, near a largish rectangular house', while the Dutch population was tested 'indoors in a laboratory room'. However, when Li and Gleitman held factor (a) constant (testing only English speakers) while varying (b) (the experiments were performed in different settings), they were able to *induce the English speakers to use an absolute frame of reference* when a sufficiently salient fixed point of reference was provided in the experimental situation (see the discussion in class).

Still, this leaves a question open. Why are there differences in preferred modes of spatial orientation across social groups? According to Li and Gleitman, there are two factors: whether the geography provides natural points of absolute reference, and whether the society is sufficiently coherent to include mostly individuals who know them. They reason as follows:

'There seems to be no consensual 'uphill' that can serve as a reference point in the very large and shifting communities in which linguistically interacting English, Dutch, or Japanese speakers generally find themselves. 'You'll find the railroad station just northeast of the Drexel University parking lot' is not too useful a direction to give the British tourist who has just arrived in Philadelphia. Body-orientation is the obvious alternative (or auxiliary) in establishing momentary spatial coordinates ("Go down to the corner, turn left and walk three blocks"). In contrast (...), people who live in a small, mutually familiar, geographical area typically use its local landmarks to devise a spatial coordinate system that makes reference to its stable features ('uphill', 'inland', etc.) This is so even when the traditional populations have the formal linguistic resources for encoding both absolute and relative spatial terminology (...) Of course the present authors do not know too much about traditional unschooled cultural groups who live in faraway places. (...) Luckily one does not have to go all the way to Chiapas or Papua-New Guinea to find communities that favor landmark-based spatial terminology: one of us is a native of a highly urbanized culture whose members live and work all crammed together on a skinny little island, about 16 miles long, at the mouth of the Hudson river; namely, Manhattan Island. Culturally diverse (some would even say 'literate') as this community is, its residents share a small, geographical landscape, rich in mutually known landmarks. Likely this is why their terminology for locations in the

community is absolute and - like that of the Tenejapans - typically makes do with only three terms in habitual use: *uptown, downtown, crosstown*.¹ (...)

Li and Gleitman conclude:

'In sum, the causal engine both for the engrained spatial reasoning styles and the fashions of speech that we find in different communities may well be a derivative of their ambient spatial circumstances. Whatever these circumstances are, communities of humans will develop terminology to fit. (...) Linguistic systems are *merely* the expressive medium that speakers devise to describe their mental representations and manipulations of their reference world. Depending on the local circumstances in which human beings find themselves, they select accordingly from this linguistically available pool of resources for describing regions and directions in space. (...) In the end, it's the thought that counts'.

[Li and Gleitman's position appears in *Cognition* 83 (2002). A reply by Levinson and co-workers appears in *Cognition* 84 (2002). The journal *Cognition* is available online from UCLA at: www.sciencedirect.com/science/journal/00100277]

3.2 ***Do universal properties of language determine some aspects of thought? Language and the Representation of Beliefs***

As can be seen from the preceding section, there is certainly no consensus about the question whether some *variable* (=non-universal) properties of language determine some aspects of thought. There is also recent research that attempts to show that some *universal* properties of language might determine some aspects of thought. One example is offered by recent work on young children's theory of mind, i.e. on their ability to *represent other people's beliefs*. A hypothesis due to J. de Villiers and co-workers is that *children acquire the ability to represent the fact that other people hold (potentially false) beliefs only when (and only because) they have mastered the linguistic form of the embedded proposition*. Here is a summary of the argument, which is still highly controversial.

A standard observation is that young children have difficulties analyzing situations in which another person holds false beliefs. A description of a typical experiment is reproduced below:

"The child was presented with a three-dimensional dolls house, props and dolls, The experimenter (E) moved the characters and objects and simultaneously told the child (C) a story:

This is a story about a boy named Johnny and his Dad. This is John and this is his Dad, and this is the kitchen in their house. Johnny's Dad made a delicious chocolate cake for their tea and gave Johnny a piece. But Johnny wanted to go out to play. So he put the cake away in the cupboard and went outside. Later the Dad said to himself: 'Hmm... I better put the cake in the refrigerator so the frosting doesn't melt'. So he took the cake out of the cupboard and put it in the refrigerator. Then he went out to the store to buy something for their dinner.

¹ One could object here that the Manhattan population is really wildly transient and therefore does not 'really' have stable landmarks that could be used by all its residents and massive numbers of visitors. That is true. But the trick is not caring. We have the following story from a Swedish tourist entering Manhattan via the George Washington Bridge (which hits the island's west flank at approximately 175th street): "We saw the signs, once labeled 'Uptown', the other 'Downtown'. We knew we were expected in 'Midtown' but this did no good at all and we were lost in the Bronx for two hours'. The moral here is that natives may rely on cues that are unusable by visitors to their island home.'

To make sure C remembered the essential facts, E asked: 'Where did Johnny put the cake before he went out to play' and 'Where is the cake now' (the order of these questions was varied across children). The story then resumed:

Now Johnny has finished playing and he comes back to have his cake. But he hasn't gone inside the kitchen yet. (Doll pauses at the door). Now C was asked the critical false belief question: 'Where will Johnny look for the cake?'"

As it turns out, three-year olds typically fail this task, whereas four-year-old typically pass it.

And now to the interesting experiment. Some researchers² at Smith College have tested deaf children at an oral school for the deaf. They had received no formal exposure to sign language and had for this reason significant delays in their language development, although they had a normal IQ. A similar experiment as the one above was then performed. The result was that *deaf children showed significant delays in the above false belief task*. The researchers' suggestion is that a child's ability to represent somebody else's false beliefs is determined by his/her level of linguistic development - specifically by his/her ability to use complex sentences with embedding such as 'Johnny believes that the cake is in the cupboard'. None of this should be taken as a definite result yet, but this experiment illustrates the fact that the relation between language and thought is still an open question, even if crude versions of the Sapir-Whorf hypothesis have been disproved.

[see Gale, de Villiers, de Villiers & Pyers, 'Language and Theory of Mind in Oral Deaf Children', *Proceedings of the 20th Annual Boston University Conference on Language Development*].

Appendix. Contents of Chapters 3 of Pinker's Language Instinct
(page numbers are indicated in parentheses)

3. Mentalese (=Language and Thought)

I. The (failed) attempt to identify Language and Thought

Introduction (44)

The Sapir-Whorf Hypothesis (48)

- a. Color (51)
- b. Hopi concept of time (52)
- c. Eskimo words for 'snow' (53)
- d. Counterfactual conditionals in Chinese (56)

II. Thought without Language: Examples

- a. Ildefonso (58)
- b. Experiments with babies: numbers (59)
- c. Experiments with monkeys: kinship relations (60)
- d. Experiments with adult humans: images and mental imagery (61)

III. Thought Without Language: Theory

Thought as symbol manipulation: the Turing Machine (64)

Whorf Revisited: why we don't think in language (69)

² See for instance Gale, de Villiers, de Villiers & Pyers, 'Language and Theory of Mind in Oral Deaf Children', *Proceedings of the 20th Annual Boston University Conference on Language Development*, Cascadilla Press 1996.