Psych 156A/ Ling 150: Acquisition of Language II

Lecture 2

Mechanisms

Announcements

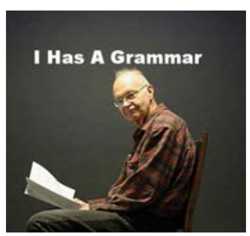
Galia office hours: M 1-2pm, W 10am-12pm

Be working on HW1 (due: 4/14/16)

Be looking over the review questions for introduction

What's being learned:

Patterns or "rules" of language = grammar



A distinction: Prescriptive vs. descriptive grammar rules

Prescriptive: what you have to be taught in school, what is prescribed by some higher "authority". You don't learn this just by listening to native speakers talk.

"Don't end a sentence with a preposition."
" 'Ain't' is not a word."

A distinction:
Prescriptive vs. descriptive grammar rules

Descriptive: what you pick up from being a native speaker of the language, how people actually speak in their day-to-day interactions. You don't have to be explicitly taught to follow these rules.

The dwarf is who Sarah first talked ***with***.

"You're horrible!" "No, I ***ain't*** - I'm Hoggle!"

A distinction:
prescriptive vs. descriptive grammar rules

The LingSpace: Word Crimes & Misdemeanors
~0.26 up through ~8:26

http://www.thelingspace.com/episode-3 (+ commentary)
https://www.youtube.com/watch?t=85&v=eFlBwBwL_iU

In a nutshell:
prescriptive vs. descriptive grammar rules

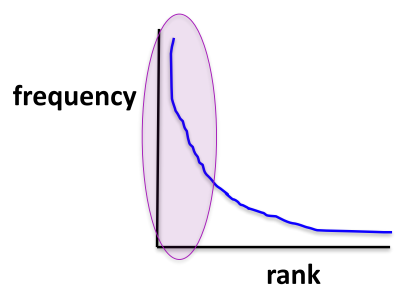
"You can't say that!" vs. "Can you say that!?"

http://specgram.com/CLIV.3/04.phlogiston.cartoon.xi.html

Learning grammars

One reason learning the rules of language is so difficult is that language follows a Zipfian distribution:

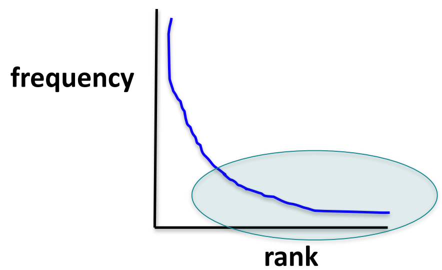
A few things are said very frequently…



Learning grammars

One reason learning the rules of language is so difficult is that language follows a Zipfian distribution:

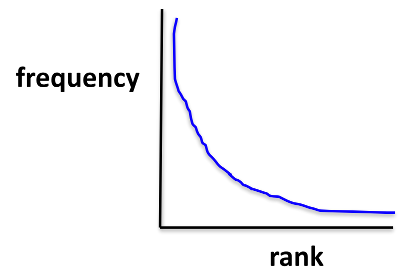
A few things are said very frequently and most things are said very infrequently.



Learning grammars

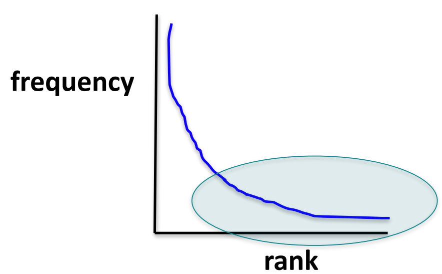
This means that children may only get a very few examples of any one linguistic structure spread out across years and years of input.

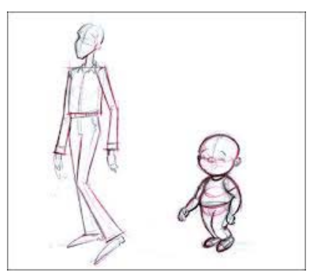
This makes acquisition particularly challenging.



Learning grammars

However, we do know that children are better at learning the mental rules of language than adults. One reason may be that they generalize from sparse and noisy data differently than adults do (Hudson Kam & Newport 2005, 2009).





Some evidence that adults and children differ

Hudson Kam & Newport (2005): Adults and 5- to 7-year-old children differ in their willingness to make generalizations.

Adults and children were presented with an artificial language that used determiners (words like “the” and “a” in English) inconsistently in noun phrases. Sometimes, the determiner would appear (maybe 60% of the time) and sometimes it wouldn’t.

Example of inconsistent use in English (rather than an artificial language):

“I want the pirate to win.”	(60%)
“I want pirate to win.”	(40%)

Some evidence that adults and children differ

Hudson Kam & Newport (2005): Adults and 5- to 7-year-old children differ in their willingness to make generalizations.

When presented with inconsistent input, adult learners matched the input and did not generalize determiner usage to all noun phrases. So, if they heard a determiner 60% of the time, they used a determiner 60% of the time when they produced sentences in this language.

Adult production:

"I want the pirate to win." (60%)

"I want pirate to win." (40%)

Some evidence that adults and children differ

Hudson Kam & Newport (2005): Adults and 5- to 7-year-old children differ in their willingness to make generalizations.

When presented with inconsistent input, child learners often generalized determiner usage to all noun phrases. So, if they heard a determiner 60% of the time, they used a determiner either 100% of the time when they produced sentences in this language - or 0% of the time (they didn't generalize the right way necessarily).

Child production:

"I want the pirate to win." (100%)

"I want pirate to win." (0%)

…but maybe not as much as we think

Hudson Kam & Newport (2009): Adults can be made to generalize too, when given inconsistent input.

When presented with inconsistent input but with one determiner being dominant (used 60% of the time as compared to others used 20% or less of the time)…

Example input:

"I want the pirate to win." (60%)

"I want pirate to win." (20%)

"I want two pirate to win." (20%)

…but maybe not as much as we think

Hudson Kam & Newport (2009): Adults can be made to generalize too, when given inconsistent input.

When presented with inconsistent input but with one determiner being dominant (used 60% of the time as compared to others used 20% or less of the time), adult learners often generalized only the dominant determiner and used it nearly all the time (90%).

Adult production:

"I want the pirate to win." (90%)

"I want pirate to win." (5%)

"I want two pirate to win." (5%)

…but maybe not as much as we think

Hudson Kam & Newport (2009): Children still differ from adults in *what* they generalize.

When presented with inconsistent input but with one determiner being dominant (used 60% of the time as compared to others used 20% or less of the time), child learners often generalized one determiner (even if it wasn't the dominant one) and used it nearly all the time (ex: 90%).

Child production:

"I want the pirate to win."	(10%)
"I want pirate to win."	(90%)
"I want two pirate to win."	(0%)

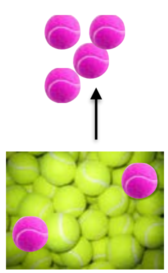
Children's learning abilities

11-month-olds don't probability-match in non-linguistic domains either, unlike adults (visual task: Yurovsky, Boyer, Smith, & Yu 2013)

Children's learning abilities

6-month-olds create probabilistic expectations about their environment, based on their observations of their environment. For example, after seeing that a box is mostly filled with yellow balls, they are surprised when someone pulls four pink balls in a row out of the box.

(Denison, Reed, & Xu 2011)



Children's learning abilities

Children selectively use their input: Children prefer to look at stimuli that are neither too boring nor too surprising, but are instead "just right" for learning, given the child's current knowledge state. This has been called the "Goldilocks Effect".

(Kidd, Piantadosi, & Aslin 2010, 2012)

Main points

Language acquisition is a process that involves inferring a structured system of rules from the available input, even if we're not consciously aware of this system when we use language.

Because of the Zipfian nature of linguistic data, the acquisition task may be very hard indeed, since many structures appear very rarely in children's input. However children may use the data very effectively, based on different helpful learning biases they have.

How do we explain how this process works?

Levels of representation (Marr 1982)

Describing vs. explaining in vision

"…it gradually became clear that something important was missing …neurophysiology and psychophysics have as their business to *describe* the behavior of cells or of subjects but not to *explain* such behavior….What are the problems in doing it that need explaining, and what level of description should such explanations be sought?" - Marr (1982)

Describing vs. explaining

"This is a common trick of psychologists, to pretend they solved a riddle of the human mind by giving it a name, when all they've done is invented an agreed upon name for the mystery rather than solved it." - Tom Stafford, "The Psychology of Tetris"

http://www.bbc.com/future/story/20121022-the-psychology-of-tetris/1

On explaining (Marr 1982)

"But the important point is that if the notion of different types of understanding is taken very seriously, it allows the study of the information-processing basis of perception to be made *rigorous*. It becomes possible, by separating explanations into different levels, to make explicit statements about what is being computed and why…"

On explaining (Marr 1982)

"But the important point is that if the notion of different types of understanding is taken very seriously, it allows the study of the information-processing basis of perception to be made *rigorous*. It becomes possible, by separating explanations into different levels, to make explicit statements about what is being computed and why…"

Our goal: Substitute "language learning" for "perception"

The three levels

Computational

What is the goal of the computation?

Algorithmic

What is the representation for the input and output, and what is the algorithm for the transformation?

Implementational

How can the representation and algorithm be realized physically?

The three levels: An example with the cash register

Computational

What does this device do?

Arithmetic (ex: addition).

Addition: Mapping a pair of numbers to another number.

(3,4) ⟶ 7 [often written (3+4=7)]

Properties:

(3+4) = (4+3) [commutative]

(3+4)+5 = 3+(4+5) [associative]

(3+0) = 3 [identity element]

(3+ -3) = 0 [inverse element]

True no matter how numbers are represented: this is what is being computed

The three levels:
An example with the cash register

Computational

What does this device do?

Arithmetic (ex: addition).

Addition: Mapping a pair of numbers to another number.

Algorithmic

What is the input, output, and method of transformation?

Input: arabic numerals (0,1,2,3,4…)

Output: arabic numerals (0,1,2,3,4…)

Method of transformation: rules of addition, where least significant digits are added first and sums over 9 have their next digit carried over to the next column

$$\begin{array}{r} 99 \\ +\ \ 5 \\ \hline \end{array}$$

The three levels:
An example with the cash register

Computational

What does this device do?

Arithmetic (ex: addition).

Addition: Mapping a pair of numbers to another number.

Algorithmic

What is the input, output, and method of transformation?

Input: arabic numerals (0,1,2,3,4…)

Output: arabic numerals (0,1,2,3,4…)

Method of transformation: rules of addition, where least significant digits are added first and sums over 9 have their next digit carried over to the next column

$$\begin{array}{r} 99 \\ +\ \ 5 \\ \hline 14 \end{array}$$

The three levels:
An example with the cash register

Computational

What does this device do?

Arithmetic (ex: addition).

Addition: Mapping a pair of numbers to another number.

Algorithmic

What is the input, output, and method of transformation?

Input: arabic numerals (0,1,2,3,4…)

Output: arabic numerals (0,1,2,3,4…)

Method of transformation: rules of addition, where least significant digits are added first and sums over 9 have their next digit carried over to the next column

$$\begin{array}{r} {\scriptstyle 1\ } \\ 99 \\ +\ \ 5 \\ \hline 4 \end{array}$$

The three levels:
An example with the cash register

Computational

What does this device do?

Arithmetic (ex: addition).

Addition: Mapping a pair of numbers to another number.

Algorithmic

What is the input, output, and method of transformation?

Input: arabic numerals (0,1,2,3,4…)

Output: arabic numerals (0,1,2,3,4…)

Method of transformation: rules of addition, where least significant digits are added first and sums over 9 have their next digit carried over to the next column

$$\begin{array}{r} {\scriptstyle 1\ } \\ 99 \\ +\ \ 5 \\ \hline 104 \end{array}$$

The three levels: An example with the cash register

Computational
What does this device do?
Arithmetic (ex: addition).
Addition: Mapping a pair of numbers to another number.

Algorithmic
What is the input, output, and method of transformation?
Input: arabic numerals (0,1,2,3,4…)
Output: arabic numerals (0,1,2,3,4…)
Method of transformation: rules of addition

Implementational
How can the representation and algorithm be realized physically?
A series of electrical and mechanical components inside the cash register.

The three levels: An example with a sandwich

Computational
What is the goal?

Make a peanutbutter and jelly sandwich.

Properties:
- slices of bread containing both peanutbutter and jelly
- number of bread slices: 2
- sandwich is sliced in half
- crusts are left on
- jelly type: grape
- peanutbutter type: crunchy

etc.

The three levels: An example with a sandwich

Computational
What is the goal?

Make a peanutbutter and jelly sandwich.

Algorithmic
What is the input, output, and method of transformation?
Input: ingredients (peanutbutter, jelly, bread slices), tools (knife, spoon)
Output: completed, edible sandwich with the required properties
Method: Use the spoon to put jelly on one slice & spread it with the knife. Use the spoon to put peanutbutter on the other slice & spread it with the knife. Put the two slices of bread together, with the spread sides facing each other. Cut the joined slices in half with the knife.

The three levels: An example with a sandwich

Computational
What is the goal?

Make a peanutbutter and jelly sandwich.

Algorithmic
What is the input, output, and method of transformation?
Input: ingredients (peanutbutter, jelly, bread slices), tools (knife, spoon)
Output: completed, edible sandwich with the required properties
Method: PBJ-making steps.

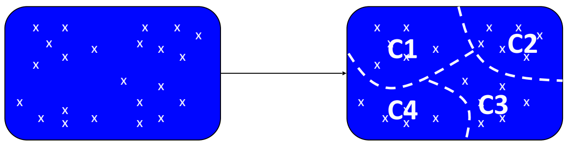
Implementational
How can the representation and algorithm be realized physically?
Directing your younger sibling to follow the steps above to make you a sandwich.

Mapping the framework

Goal: Understanding the “how” of language learning

First, we need a computational-level description of the learning problem.

Computational Problem: Divide sounds into contrastive categories



Mapping the framework

Goal: Understanding the “how” of language learning

First, we need a computational-level description of the learning problem.

Computational Problem: Divide spoken speech into words

húwzəfɹéjdəvðəbígbǽdwʌ́lf

↓

húwz əfɹéjd əv ðə bíg bǽd wʌ́lf

who‘s afraid of the big bad wolf

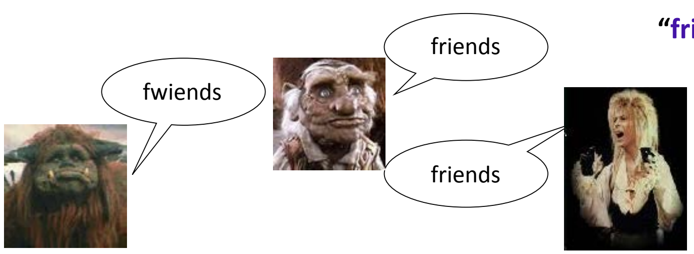
Mapping the framework

Goal: Understanding the “how” of language learning

First, we need a computational-level description of the learning problem.

Computational Problem: Map word forms to speaker-invariant forms

“friends”



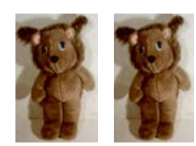
Mapping the framework

Goal: Understanding the “how” of language learning

First, we need a computational-level description of the learning problem.

Computational Problem: Identify the concept a word is associated with (Word-meaning mapping)

“I love my daxes.”



Dax = that specific toy, teddy bear, stuffed animal, toy, object, …?

Mapping the framework

Goal: Understanding the “how” of language learning

First, we need a computational-level description of the learning problem.

Computational Problem: Identify what a speaker means by using a specific expression.

“I love some of my daxes.”



Does the speaker not love all of them?

Mapping the framework

Goal: Understanding the “how” of language learning

First, we need a computational-level description of the learning problem.

Computational Problem: Identify grammatical categories

“This is a DAX.”

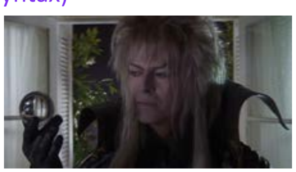
DAX = noun

Mapping the framework

Goal: Understanding the “how” of language learning

First, we need a computational-level description of the learning problem.

Computational Problem: Identify the rules of word order for sentences. (Syntax)



Jareth juggles crystals

Subject Verb Object

English

Subject Verb Object

Kannada

Subject t_{Object} Verb Object

German

Subject Verb $t_{Subject}$ Object t_{Verb}

Mapping the framework

Goal: Understanding the “how” of language learning

Second, we need to be able to identify the algorithmic-level description:

Input = sounds, syllables, words, phrases, …

Output = sound categories, words, words with affixes, grammatical categories, sentences, …

Method = statistical learning, prior knowledge about how human languages work, …

Recap: Levels of representation

Language acquisition can be viewed as an information-processing task where the child takes the native language input encountered and uses it to construct the adult rule system (grammar) for the language.

Main idea: The point is not just to describe what children know about their native language and when they know it, but also how they learned it.

Three levels:

computational: what is the problem to be solved

algorithmic: what procedure will solve the problem, transforming input to desired output form

implementational: how is that procedure implemented/instantiated in the available medium

Computational modeling: Understanding the mechanism

Computational Level:

Theoretical linguistic studies can often tell us what needs to be learned about language. Experimental studies can often tell us about when children seem to know different kinds of language knowledge. This defines the goal of language acquisition:

Learn the appropriate what by the appropriate when.

Computational modeling: Understanding the mechanism

Algorithmic Level:

But how do we know what the input is, what the output ought to look like, and what method(s) children use to get from the input to the output?

Computational modeling: Understanding the mechanism

Algorithmic Level:

Input: The CHILDES database has a wealth of child-directed speech transcripts and videos from a number of different languages. This can help us figure out what children's input looks like.

CHILDES Child Language Data Exchange System

http://childes.psy.cmu.edu

Video/audio recordings of spontaneous speech samples, along with transcriptions and some structural annotation. Extremely valuable resource to the language acquisition community.

```
@Loc:	Eng-NA-MOR/Rollins/al12.cha
@PID:	11312/c-00017262-1
@Begin
@Languages:	eng
@Participants:	CHI Target_Child , MOT Mother
@ID:	eng|rollins|CHI|||||Target_Child|||
@ID:	eng|rollins|MOT|||||Mother|||
@Media:	al12, video
@Activities:	Free Play
*MOT:	you haven't seen this .
%mor:	pro|you aux|have-neg|not part|see&PASTP pro:dem|this .
%gra:	1|4|SUBJ 2|4|AUX 3|2|NEG 4|0|ROOT 5|4|OBJ 6|4|PUNCT
*MOT:	that looks pretty cool .
%mor:	det|that n|look-PL adv:int|pretty adj|cool .
%gra:	1|2|DET 2|0|INCROOT 3|4|JCT 4|2|XMOD 5|2|PUNCT
*MOT:	do you know how to work that .
%mor:	mod|do pro|you v|know adv:wh|how inf|to v|work pro:dem|that .
%gra:	1|3|AUX 2|3|SUBJ 3|0|ROOT 4|3|OBJ 5|6|INF 6|4|XCOMP 7|6|OBJ 8|3|PUNCT
*MOT:	yes you do .
%mor:	co|yes pro|you v|do .
%gra:	1|3|COM 2|3|SUBJ 3|0|ROOT 4|3|PUNCT
```

Computational modeling: Understanding the mechanism

Algorithmic Level:

Output: Theoretical linguistics and experimental studies can tell us what the output should look like by observing adult and child knowledge of various linguistic phenomena.

Example problem: word segmentation

húwzəfɹéjdəvðəbígbǽdwʌ́lf — input

↓

húwz əfɹéjd əv ðə bíg bǽd wʌ́lf — output

who's afraid of the big bad wolf

Computational modeling: Understanding the mechanism

Algorithmic Level:

Method: Computational modeling can often help us figure out how children are getting from the input to the output.

húwzəfɹéjdəvðəbígbǽdwʌ́lf

↓ What goes here?

húwz əfɹéjd əv ðə bíg bǽd wʌ́lf

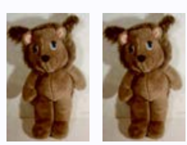
who's afraid of the big bad wolf

Computational modeling: What a "digital" child can tell us

We can construct a model where we have precise control over these:

- The hypotheses the child is considering at any given point [hypothesis space]

"I love my daxes."



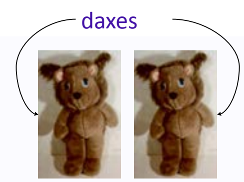
Dax = that specific toy, teddy bear, stuffed animal, toy, object, …?

Computational modeling: What a "digital" child can tell us

We can construct a model where we have precise control over these:

- The hypotheses the child is considering at any given point [hypothesis space]
- How the child represents the data & which data the child uses [data intake]

"I love my daxes."



Dax = that specific toy, teddy bear, stuffed animal, toy, object, …?

Computational modeling: What a "digital" child can tell us

We can construct a model where we have precise control over these:

- The hypotheses the child is considering at any given point
  [hypothesis space]
- How the child represents the data & which data the child uses
  [data intake]
- How the child changes belief based on those data
  [update procedure]

dax = that specific toy more probable

dax = any object less probable

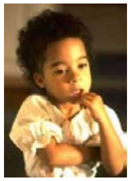
Computational modeling: What a "digital" child can tell us

Models are most informative when they're grounded empirically.

This is why most models make use of the child-directed speech data available through databases like CHILDES.

Many models will try to make cognitively plausible assumptions about how the child is representing and processing input data:

- Processing data points as they are encountered
- Assuming children have memory limitations (ex: memory of data points may decay over time)



Computational Methods

"Computational modeling can be used to examine a variety of questions about the language acquisition process, because a model is meant to be a simulation of the relevant parts of a child's acquisition mechanism. In a model, we can precisely manipulate some part of the mechanism and see the results on acquisition….Importantly, some manipulations we can do within a model are difficult to do with children…modeling data are thus particularly useful because of the difficulty of getting those same data through experimental means."

- Pearl 2010

General modeling process

(1) Decide what kind of learner the model represents (ex: normally developing 6- to 8-month-old child learning first language)

General modeling process

(1) Decide what kind of learner the model represents (ex: normally developing 6- to 8-month-old child learning first language)

(2) Decide what data the child learns from (ex: Pearl-Brent corpus from CHILDES) and how the child processes that data (ex: divide speech stream into syllables)

General modeling process

(1) Decide what kind of learner the model represents (ex: normally developing 6- to 8-month-old child learning first language)

(2) Decide what data the child learns from (ex: Pearl-Brent corpus from CHILDES) and how the child processes that data (ex: divide speech stream into syllables)

(3) Decide what hypotheses the child has (ex: what the words are) and what information is being tracked in the input (ex: transitional probability between syllables)

General modeling process

(1) Decide what kind of learner the model represents (ex: normally developing 6- to 8-month-old child learning first language)

(2) Decide what data the child learns from (ex: Pearl-Brent corpus from CHILDES) and how the child processes that data (ex: divide speech stream into syllables)

(3) Decide what hypotheses the child has (ex: what the words are) and what information is being tracked in the input (ex: transitional probability between syllables)

(4) Decide how belief in different hypotheses is updated (ex: based on transitional probability between syllables)

General modeling process

(5) Decide what the measure of success is

ex: making correct generalizations

- Knowing that *dax* refers to all teddy bears, even ones the child hasn't seen before

ex: achieving a certain knowledge state by the end of the learning period

- Recognizing useful units (such as words) in a fluent speech stream

The goal of modeling

Remember: the goal is generally to see if a particular learning strategy (as described by the hypothesis space, data intake, and update procedure) will allow the child to go from the input to the output. This then tells us about the process of language acquisition (the algorithmic level of explanation).

húwzəfɹéjdəvðəbígbǽdwʌ́lf

What goes here?

húwz əfɹéjd əv ðə bíg bǽd wʌ́lf

who's afraid of the big bad wolf

Recap: Mechanism of acquisition

One of the main goals of the study of language acquisition is to explain it, rather than just describe it.

There are three different levels of explanation, according to Marr: the computational level, the algorithmic level, and the implementational level.

The algorithmic level focuses on the process (the "how") of acquisition, and computational modeling is a technique that can be used to investigate different strategies a child might use to learn language.

Questions?

You should be able to do all the introductory review questions and up through question 4 on HW1.

Extra Material

Possible objections to a mental rule set

"Why should I believe I store a set of rules unconsciously in my mind? I just understand sentences because they make sense."

Possible objections to a mental rule set

"Why should I believe I store a set of rules unconsciously in my mind? I just understand sentences because they make sense."

But why do some sentences make sense and others don't?

Hoggle has two jewels.
*Two Hoggle jewels has.

Possible objections to a mental rule set

Why can we recognize patterns even when some of the words are unknown?

'Twas brillig, and the slithy toves
did gyre and gimble in the wabe...

Possible objections to an unconscious rule set

"When I talk, the talk just comes out - I'm not consulting any rule set."

Possible objections to an unconscious rule set

"When I talk, the talk just comes out - I'm not consulting any rule set."

Analogy: wiggling your fingers

When you want to wiggle your fingers, you "just wiggle them".

But your finger-wiggling intention was turned into commands sent by your brain to your muscles, and you're never conscious of the process unless something interferes with it. Nonetheless, there *is* a process, even if you're not aware of it.