Sandiway Fong
696 A
Syntax and Computation

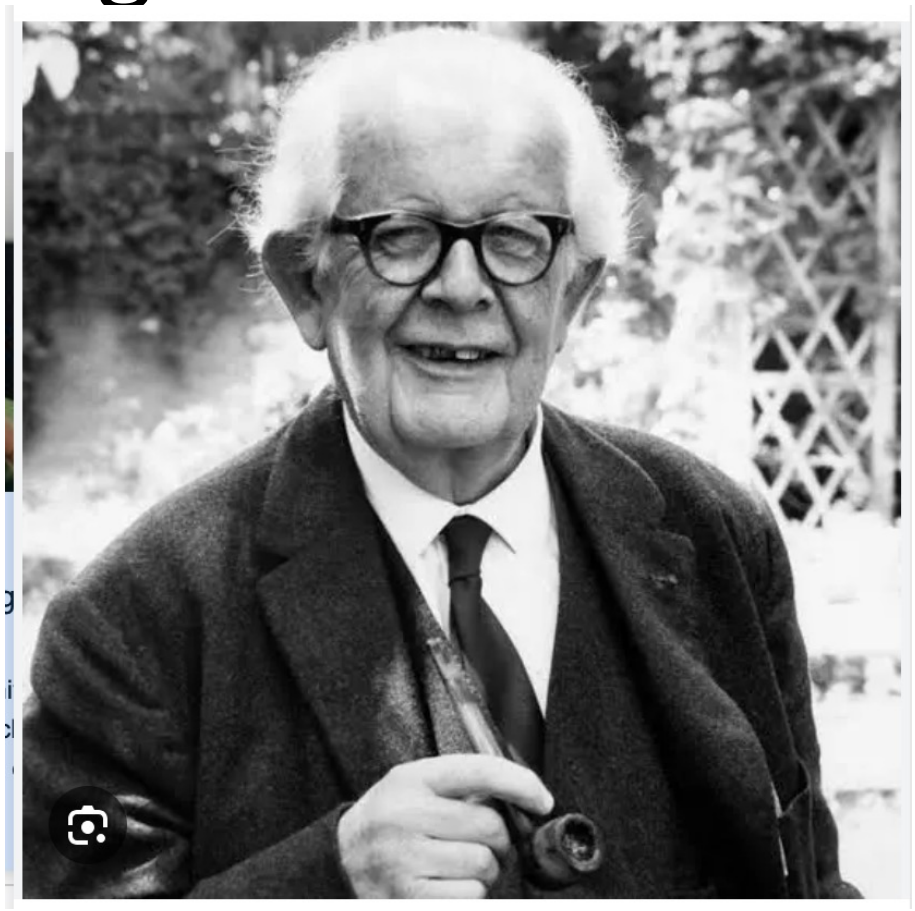
Massimo Piattelli-Palmarini
Guest lecturer

50 Years after the Chomsky-Piaget Debate

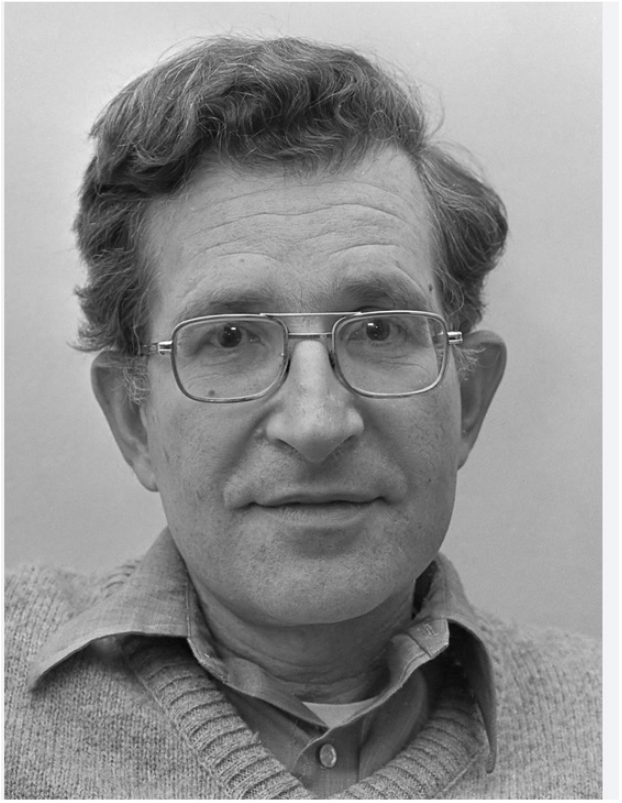
Opening

- October 10 2026 marks fifty years since the Chomsky-Piaget debate.
- Royaumont Abbey (35 kilometers north of Paris)

Jean Piaget



He was 79 years old at the time of the debate



Noam Chomsky

He was 47 years old at the time of the debate

Who participated in the debate?

- Jean Piaget and Noam Chomsky (of course)
- **Cognitive science**: Jerry Fodor, David Premack, Bärbel Inhelder, Jacques Mehler
- **Artificial Intelligence**: Guy Cellérier, Seymour Papert
- **Anthropology**: Gregory Bateson, Dan Sperber, Maurice Godelier, (Claude Lévi-Strauss), Thomas de Zengotita, Anthony Wilden
- **Ethology**: Norbert Bischof, Dieter Dütting
- **Neurobiology**: Jean-Pierre Changeux
- **Molecular Biology**: Jacques Monod, François Jacob
- **Philosophy**: Stephen Toulmin
- **Organizer**: MPP

About Piaget:

- **Jean William Fritz Piaget** (9 August 1896 – 16 September 1980)
- **Honorary doctorates**
- 1936 Harvard
- 1946 Sorbonne
- 1949 University of Brazil
- 1949 Bruxelles
- 1953 Chicago
- 1954 McGill
- 1958 Warsaw
- 1959 Manchester
- 1960 Oslo
- 1960 Cambridge
- 1962 Brandeis
- 1964 Montreal
- 1964 Aix-Marseille
- 1966 Pennsylvania
- 1966? Barcelona
- 1970 Yale
- 1971 Temple

Piaget's syllogism

- Life is, essentially, progressive auto-regulation and stepwise self-equilibration
- Cognition is an aspect of life
- Therefore
- Cognition is, essentially, progressive auto-regulation and stepwise self-equilibration

- The name of his game is: Genetic Epistemology
- The "psychogenesis of knowledge"

- Great article on Piaget in Wikipedia

How he describes his own work:

- *"My central aim has always been the search for the mechanisms of biological adaptation and the analysis and epistemological interpretation of that higher form of adaptation which manifests itself as scientific thought."*

- (Piaget, 1977 Foreword to H.E.Gruber & J.Voné (Eds),*The essential Piaget*. London: Routledge and Kegan Paul, p.XI)

How he describes his own work:

- *"The essential functions of the mind consist in understanding and in inventing, in other words, in building up structures by structuring reality."*

- (Piaget, 1971, Science of education and the psychology of the child. New York: Viking Press p.27)

- A *"dynamic Kantism"* (verbatim, in the debate)
- *"There is no genesis without structures, there are no structures without genesis"* (in the debate)

In essence:

- Cognition is not a static and passive process: It's a dynamic and creative process.
- Between the cognitive development of the child and the process of discovery in the sciences there are deep and illuminating commonalities.
- A central thesis of Piaget's: There can be necessity without innateness.
- The source of this necessity is to be found in logic and in "the structure of mathematics"

Piaget stages

- Piaget's four stage are:
- (1) The Sensorimotor Period (birth to 2 years).
- (2) Pre-Operational Thought (2 to 6/7 years).
- (3) Concrete Operations (6/7 to 11/12 years).
- (4) Formal Operations (11/12 to adult).
- These stages are central to Piaget's work
- Each stage is subdivided into sub-stages.

Some detail

- Piaget's ladder includes the following moments:
- the roots of logico-mathematical structures; reflective abstraction (*abstraction réfléchissante*); reflected abstraction (*abstraction réfléchie*); and the progressive regrouping, thematization, and reorganization of lower-level operations at a more abstract level, first implicitly and then explicitly..
- Abstraction and generalization are interdependent. Insufficiencies, gaps and unsolved problems cry out for completions, re-equilibrations, filling-in at higher level. Language is an application of the "semiotic function", a specific domain of this function.

Piaget's central hypothesis:

- There **is** a central representational capacity (the "semiotic function"). It develops at around 2 years of age.
- It applies to different domains (language being one of them)
- There is a dynamic "psychogenesis", whose working is the focal object of our concern.
- *"The functioning of intelligence **alone** is hereditary and creates structures **only** through an organization of successive actions performed on objects"*. (In his introductory paper. My emphasis)
- Neither empiricism, nor preformationism, but rather a dynamic constructivism.

Piaget's Stage Theory of Development

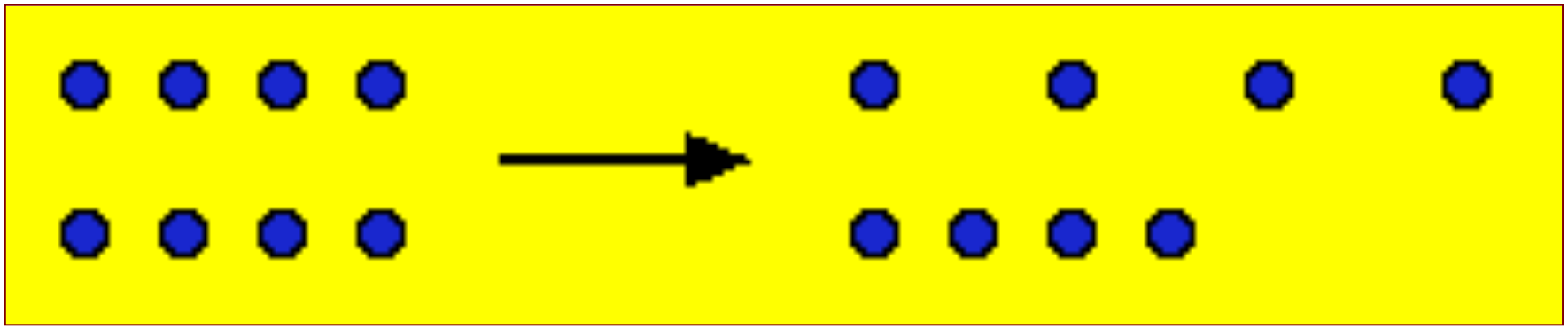
- The Sensorimotor Period (birth to 2 years)
- During this time, a child's cognitive system is limited to motor reflexes at birth, but the child builds on these reflexes to develop more sophisticated procedures. They learn to generalize their activities to a wider range of situations and coordinate them into increasingly lengthy chains of behavior.

- Pre-Operational Thought (2 to 6/7 years)
- At this age, children acquire representational skills in the areas of mental imagery, and especially language. They are very self-oriented, and have an egocentric view; that is, preoperational children can use these representational skills only to view the world from their own perspective.

Piaget's Stage Theory of Development

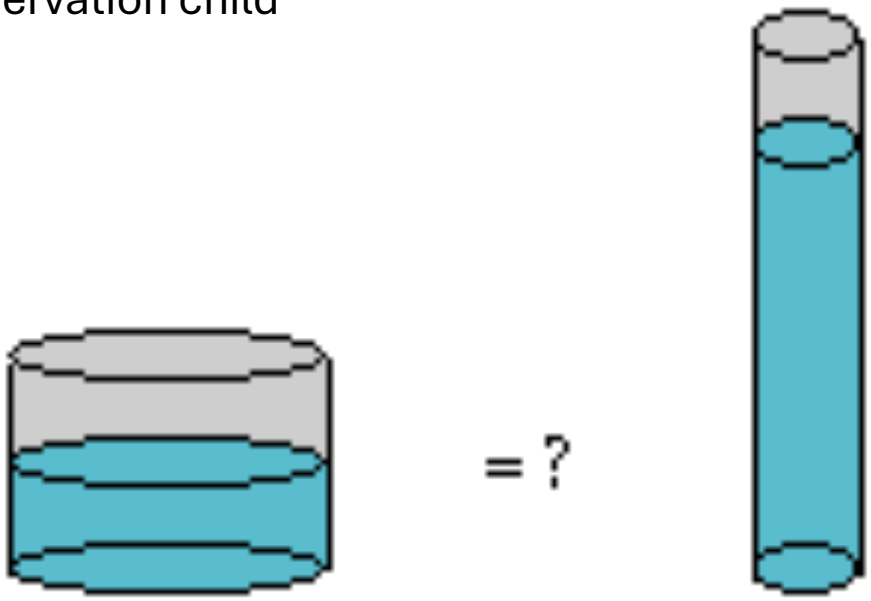
- Concrete Operations (6/7 to 11/12 years)
- As opposed to Preoperational children, children in the concrete operations stage are able to take another's point of view and take into account more than one perspective simultaneously. They can also represent transformations as well as static situations. Although they can understand concrete problems, Piaget would argue that they cannot yet perform on abstract problems, and that they do not consider all of the logically possible outcomes.
- Formal Operations (11/12 to adult)
- Children who attain the formal operation stage are capable of thinking logically and abstractly. They can also reason theoretically. Piaget considered this the ultimate stage of development, and stated that although the children would still have to revise their knowledge base, their way of thinking was as powerful as it would get.

The pre-conservation child



If you show a child four marbles in a row, then spread them out, the pre-operational child will focus on the spread, and tend to believe that there are now more marbles than before.

The pre-conservation child



If you pour the milk from the short, fat glass into the tall, skinny glass, the pre-operational child will tell you that there is more milk than before, despite having seen you pour it without spilling any.

By seven or eight years old, children develop conservation of substance:

If I take a ball of clay and roll it into a long thin rod, or even split it into ten little pieces, the child knows that there is still the same amount of clay. And he/she will know that, if you rolled it all back into a single ball, it would look quite the same as it did -- a feature known as reversibility.

http://www.ship.edu/~cgboeree/piaget.html

Piaget's ladder

- The roots of logico-mathematical structures:
- Reflective abstraction (*abstraction réfléchissante)*
- Reflected abstraction (*abstraction réfléchie*)
- Operations of a lower level are grouped, thematized, reorganized at a more abstract level
- First implicitly, then explicitly
- Abstraction and generalization are interdependent
- Insufficiencies, gaps, unsolved problems, cry out for completions, re-equilibrations, filling-in at a higher level.
- The "semiotic function" is an application, a specific domain. **And language an application of it.**

Forms of equilibration:

- (1) Assimilation: Linear velocity and angular velocity (common space-time relations in spite of superficial differences)
- (2) Sub-systems interlock into super-systems: Compensatory adjustments between partial negations and partial affirmations, between direct and inverse operations, with reciprocities.
- (3) Builds on (2), but constructs novel global systems. Compensatory steps of integration into a new totality, with enrichment. Cognitive equilibration is "accretive" (*majorante*), with a passage from the exogenous to the endogenous.
- The biological equivalent is the "phenocopy"
- Exact parallels in the history of science.

Piaget's curious (implicit) belief

- No one could possibly disagree with him.
- If someone disagreed with him, that meant that he/she **had not understood him**.
- Why so?
- 1 his many years of experimentation with children of different ages
- 2 the logico-mathematical foundation of his theories
- 3 Cogent parallels in the history of science and mathematics.

On the nature of reflection:

- One of Piaget's favorite quotes from John Locke.
- Locke said about reflection:

- *"the ideas it affords being such only as the mind gets by reflecting on its own operations within itself."*

- (Locke, Essay Concerning Human Understanding, 1690, Book II, 4)

Piaget challenges Chomsky

- If one fails to perceive the roots of these capacities in the vast process of adaptation and self-organization of life in general (and even before life), then one resorts to innatism, and then one has to go all the way down to protozoa. But this is a hopeless task. It makes the evolution of language "inexplicable". For Piaget, generality takes precedence over species-specificity.. Mere chance cannot explain necessity. Innatism is, in turn, based on a more general process: auto-equilibration.
- The forms of equilibration, according to Piaget, are:
- (1) Assimilation: linear velocity and angular velocity are assimilated (common space-time relations in spite of superficial differences).
- (2) Sub-systems interlock into super-systems: compensatory adjustments between partial negations and partial affirmations, between direct and inverse operations, with reciprocities.
- (3) Builds on (2) but constructs novel global systems. Compensatory steps of integration into a new totality, with enrichment. Cognitive equilibration is "accretive" (*majorante*), with a passage from the exogenous to the endogenous. The biological equivalent is the "phenocopy". There are, according to Piaget, exact parallels in the history of science.
- François Jacob reacted critically to this notion, offering a rather reductive interpretation from biology proper. If you swamp a female embryo with testosterone, then the adult female will resemble a male, you have a phenocopy.
-

Piaget challenges Chomsky

- If one fails to perceive the roots of these capacities in the vast process of adaptation and self-organization of life in general (and even before life), then one resorts to innatism, and then one **has to go all the way down to protozoa**. But this is a hopeless task. It makes the evolution of language "inexplicable". For Piaget, generality takes precedence over species-specificity.. Mere chance cannot explain necessity. Innatism is, in turn, based on a more general process: auto-equilibration.
- The forms of equilibration, according to Piaget, are:
- (1) Assimilation: linear velocity and angular velocity are assimilated (common space-time relations in spite of superficial differences).
- (2) Sub-systems interlock into super-systems: compensatory adjustments between partial negations and partial affirmations, between direct and inverse operations, with reciprocities.
- (3) Builds on (2) but constructs novel global systems. Compensatory steps of integration into a new totality, with enrichment. Cognitive equilibration is "accretive" (*majorante*), with a passage from the exogenous to the endogenous. The biological equivalent is the "phenocopy". There are, according to Piaget, exact parallels in the history of science.
- François Jacob reacted critically to this notion, offering a rather reductive interpretation from biology proper. If you swamp a female embryo with testosterone, then the adult female will resemble a male, you have a phenocopy.
-

Chomsky replies

- «Investigation of the human language has led me to believe that a genetically determined language faculty, one component of the human mind, specifies a certain class of “humanly accessible grammars”. Within a given speech-community children with varying experience acquire comparable grammars, vastly underdetermined by the available evidence».
- This is a formulation of the Poverty of the Stimulus (POS), a notion introduced by Chomsky in the Fifties and already fiercely resisted by philosophers like Nelson Goodman, Willard Quine and Hilary Putnam Quine’s quality space and other extremely simple and organism-independent elements.

Chomsky replies

- In later years, POS will be unpersuasively criticized by several authors. Chomsky then added that there are also “performance systems” for putting this knowledge to use.
- Very little is known about these. «My guess would be that, as in the case of grammars, a fixed, genetically determined system of some sort narrowly constrains the form that they [i.e. the performance systems] can assume»

Chomsky replies

- On Piaget's argument "to the protozoa", Chomsky says that the evolutionary development of language is, no doubt, yet unexplained. But it is not, contrary to Piaget's claim, "inexplicable".
- Genetic endowment is responsible for the cerebral cortex, the eye, the arms. Protozoa don't have these either, but this does not stop us from attributing these biological traits to genetic determination. Hypotheses about the genetic components of bodily organs are refutable. So are those concerning language. No circularity, no "question-begging".

Chomsky replies

- On “sensorimotor intelligence”, Chomsky says that no substantive proposals «offer any hope of accounting for the phenomena of language that demand explanation. Nor is there any initial plausibility to the suggestion, as far as I [Chomsky] can see»
- He then adds that we want to attempt to delimit certain cognitive domains, each governed by an integrated system of principles of some sort. «It is, surely, a legitimate move to take language to be one such domain, though its exact boundaries and relations to other domains remain to be determined» We do that exactly as we would for some organ of the body.

Chomsky replies

- On the development of language, Chomsky adds: «If E is the linguistically relevant experience available to a child and Ss the linguistic steady state of the adult of that community, we construct hypotheses on the «function mapping E into Ss»
- This function is LT(H, L), the Learning Theory for Humans in the domain of Language. Abstracting away from individual differences, it is «a genetically determined species character». In general, for a species O and a domain D, we have some LT(O, D) (the recognition of human faces, for instance).

Chomsky replies

- Contra “general learning”, these LT(O, D) will be rather different one from the other: «we would hardly expect to discover that there exists something that might be called “general learning theory”».
- The prospects of such a theory are «no brighter than for a “growth theory” intermediate in level between cellular biology and the study of particular organs and concerned with the principles that govern the growth of arbitrary organs for arbitrary organisms».

In a nutshell: The central theses

- **Piaget**
- Language is a special application of general cognition
- The reflex of an abstract, logical necessity
- Deeply rooted in formal structures and the continuous construction of cognitive equilibria
- Self-organization and self-equilibration are the central forces.
- Learning is a progressive, step-wise construction

- **Chomsky**
- Language is a specialized mental organ, a module
- The product of our genetic endowment
- Its formal structures are language-specific and cannot be derived from experience
- The core process is one of projection of internal structures onto the surface linguistic expressions
- Learning is essentially "growth" and selection

Sequels

- **From Piaget**
- Artificial intelligence
- Connectionism
- Evolutionary theories deriving language from action and interiorization of motor schemata
- Commonalities between language and perception and action
- A basis for the improvement of education

- **From Chomsky**
- Computational linguistics and the Chomsky hierarchy
- A new comparative linguistics
- A new historical linguistics
- Neonate cognition
- New approaches to language acquisition
- The theory of principles and parameters
- Emphasis on universality, contra relativism (contra the Sapir-Whorf hypothesis)
- Pathologies and "experiments in nature"
- Biology of language

On Hill Climbing

- Guy Cellérier, a close collaborator of Piaget and an expert of formalisms in artificial intelligence, offered a “compromise” (sic).
- Let’s postulate a hill-climbing, problem-solving system, with a generate-and-test method in the problem-space of grammars.
- There will be “substantive universals” (types of actions), and “formal universals” (general criteria for legal moves).
- The hill-climber possesses a minimal structure independent of the environment. There are neighborhood relations (vicinity between states), as suggested by Marvin Minsky in his 1961 essay *Steps towards Artificial Intelligence*.
- One must then specify how constrained the search has to be, and what is the source of these constraints.

Glitches

- However, there are several glitches with “hill-climbing”, some were stressed during the debate, but others are worth stressing here:
- (1) The system may always be trapped into a local maximum, unable to proceed any further (see Gibson and Wexler, 1994, for an insightful analysis of this predicament in the domain of parametric language acquisition).
- (2) A cognitive system must be capable of “sensing” (that is, understanding) how high it has climbed, what constitutes a satisfactory “height” in the be sheer “luck”, but no real “learning”).
- These abilities cannot be “given” by the hill-climbing process itself, they must be independently derived.
- (3) Most importantly, it is hard to conceive of any such process as devoid of the capability to monitor (again, in the cognitive domain, understand) how high it would have climbed, had it followed a different path. (4) Therefore, the power of handling counterfactuals must be attributed to the system. It must, once again, be independently granted. It cannot possibly be the outcome of hill-climbing.

Two strategies

- Cellérier also proposed two possible strategies, arguably complementary:
- (1) Maximize the initial structure of the learning device, and minimize the complexity of the search procedure.
- (2) Maximize the complexity of the search procedure, and minimize the initial internal structure of the learning device.
- We will need the combined action of both these strategies, in order to understand cognition. Whence the “compromise”. Cellérier then added: Gravity is universal, but not innate.
- Sickle cell anemia is innate, but not universal. None of Chomsky’s hypotheses has received any support from genetics, the way sickle-cell anemia has. Chomsky’s reply was: It makes no sense at all to minimize or maximize this or that. Nothing of the sort applies in physics, chemistry or biology.

Chomsky's reply:

- It makes no sense at at all to minimize or maximize this or that.
- Nothing of the sort applies in physics, chemistry or biology.
- The task is to understand which structures are there, for real.
- Hill-climbing is a doubtful metaphor.
- Better speak of *"successive maturation of a specialized hardware"* (analogous to sexual maturation, in kind)
- The very idea that cognitive structures are vastly more complex than those of any physical organ implicitly presupposes
- *"a metric of complexity which I [Chomsky] don't perceive or share"*.

A doubtful metric of complexity

- The task is to understand which structures are there, for real. Hill-climbing is a doubtful metaphor. Better speak of "successive maturation of a specialized hardware" (analogous to sexual maturation, in kind).
- The very idea that cognitive structures are vastly more complex than those of any physical organ implicitly presupposes "a metric of complexity which I [Chomsky] don't perceive or share".
- In essence, Chomsky defended the autonomy of syntax and the poverty of the stimulus, theses he had cogently defended already and that he has continued defending in the many following years. What was excluded from the process of language acquisition by the child was learning (in the traditional sense) and induction. (Contra Poverty of the Stimulus see Perfors et al. 2006, 2011).
- For a rebuttal see (Berwick, Chomsky, Piattelli-Palmarini 2012)

Structure-dependence

- Why are structure-dependent transformations tried out first by the child? An operation is structure-independent if it applies to the manifest properties of the words and the word-sequences. For instance, add the suffix -am to all words ending in a, o, or um. Or move the fourth word of the sentence to the front of the sentence.
- An operation is structure-dependent if it applies conditionally, depending on some internal, hierarchical property of the architecture of the sentence. For instance, it may act on the sister node of an embedded NP. The "simplest" structure independent hypothesis is not even tried out: Why? Take the formation of interrogatives:
- *The man is here* forming the interrogative *Is the man here?*
- Simplest hypothesis: Spot the first appearance of 'is' in the sentence and front it, to form the interrogative.
- For a neuro support of this radical difference, see (Musso, Moro et al, 2003). A paper that Chomsky, in later years, has frequently and approvingly cited. The neuro data show that the special status of structure-dependent operations in syntax and the difference with structure in-dependent operations has been validated by the sharp difference in their respective brain activations.
- In the domain of language acquisition, Neil Smith and Stephen Crain have shown that the privileged status of structure-dependent operations is known by the child as early as testing is possible, namely as early as 3 years-old. (Crain et al. 2006)
-

Counter-example

- Counter-example:
- *The man who is here is tall* Interrogative: *Is the man who is here tall?*
- *The man who is tall will leave* Interrogative: *Will the man who is tall leave?*
- Not
- **Is the man who here is tall?*
- **Is the man who tall will leave?*
- The formation of interrogatives is a structure dependent rule. How does the child know? The child does not even try the simplest rule first. No such errors are made by any child, the world over. Communicative efficiency is not the explanation. Such data suggest that these rules, the more complex ones (the structure-dependent ones), are part of the initial state.
- Other examples:
- *Each of the men likes the others*.
- *The men like each other*.
- These are near-synonyms. But this is not the case with.
- *Each of the men expects John to like the others*
- **The men expect John to like each other*.
- Chomsky then concluded that the explanation is strictly language-specific (free versus bound anaphora) and has nothing to do with sensorimotor schemata, communicative efficiency, or anything of the sort.

Fodor's buster

- In essence, Jerry Fodor wanted to show, cogently, that, if there are, in the cognitive development of the child, genuinely "more powerful" concepts, these cannot possibly be the result of learning.
- He squarely concluded, after his demo, there and then that **there must be some notion of learning that is so incredibly different from the one we have imagined that we don't even know what it would be like, as things now stand.** This is his "demo" (or proof, if one prefers): Fodor's buster: All concepts are innate.

Fodor's buster

- Based on three converging, but distinct, lines of evidence and reasoning:
- (1) No induction (no learning) is possible without severe *a priori* constraints on the kinds of hypotheses (concepts) that the learner is going to try out.
- (2) The failure of Locke's project: Derive all concepts from a Boolean combinatorial algebra of elementary sensory primitives.
- But you cannot even acquire concepts such as *buy* and *sell*, *automobile* and *typewriter* on the basis of Locke's schema. That has been shown.
- (3) Richer (more powerful) concepts cannot be developed out of poorer ones by means of learning (in any of the models of learning that have been proposed so far).

Fodor's buster

- General learning "situation": you present items to the learner and tell him/her/it (if a machine):
- These are "instances" {+ + + ……}
- These are not. {- - - ……}
- Instances and non-instances of what?
- Some property (concept, predicate) X.
- On the basis of the evidence presented, and partitioned into those two mutually exclusive subsets (is an instance, is not an instance, it's OK it's not OK, satisfies, does not satisfy, etc.), your task is to guess (discover) what X is.

Fodor's buster

- These are the possibilities:
- Schema 1: The learner already has a repertoire of relevant concepts (predicates, hypotheses), X1, X2, … Xn. He/she/it tries them out in some order of decreasing plausibility, and selects the best guess compatible with all the evidence seen so far. Inductive logic will tell you (not an easy task) which hypotheses will be tried out first, second, third, and so on., and what constitutes "sufficient" confirmation. It's totally silent on the **origins** of the repertoire. This is the innatist schema. We have some understanding of how it works.
- Schema 2: The learner has a repertoire of vaguely relevant, but weaker concepts (properties, predicates, hypotheses), x1, x2, … xn. He/she/it must find the means to develop (acquire, generate, compute) a "more powerful" concept X. Thesis: The methods for learning concepts also tell us how the more powerful concept is generated (see Piaget's theory).
- Call this: Feed-back, variational re-computation, abstraction, representational re-description, whatever. Fodor shows that no such schema could possibly work. Why?

Fodor's buster

- Sub-Option 1: The learner generates X by sheer luck, and X fits the available evidence by sheer luck. Otherwise, he/she/it dies, along with all the descendants. In fact, most of the time, X is wrong. Only rarely do such guesses work (extreme Darwinism). No learning has taken place, just lucky blind guesses.
- Special case: The target X and the guessed Y happen to be **co-extensive**.
- They mean quite different things (they are semantically distinct). (evening star, morning star)
- Y (not X) has been "fixed" by natural selection.
- This may (just may) apply to genuine triggers.
- Guidance by truth.
- Sub-Option 2: The learner, somehow, "tracks" the content (the meaning) of X, and why it is adequate with respect to (true of) the available evidence. The process is, somehow, guided.

Fodor's buster

- The content of X, and some sensitivity of the process to the truth/falsehood of X, supply the required "guidance" (tracking). Nothing else could supply it. Evolution's criterion can only be truth-value. There is learning (inductive fixation/rejection) of the **meaning** of X, but X cannot be fixed/rejected unless it is **actually available** to the learner and it is exploited in the process by the learner.
- A suggestion doomed to fail: The learner "works on" the previously available, weaker (primitive) concepts x1, x2, … xn by means of combinations, re-descriptions, thematizations, whatever, and thereby generates a genuinely new concept Y.
- One possibility: Y is literally a composite concept, composed out of the xs (brown cow) and what it means is that way of composing them. No less, no more. It's not "more powerful" in any interesting way. Obviously not all concepts can be composite. Some must be primitive
- The gap between co-extensionality and intensional characterization is a perennial glitch for behaviorism and for Darwinian explanations (see my book with Fodor for an extensive treatment). (Fodor and Piattelli-Palmarini 2010, 2011)

Fodor's buster

- Another possibility: one is acquiring genuinely "more powerful" primitive concepts. They are not decomposable (ex hypothesis). They are not "definable" out of more elementary concepts.
- Genuinely more powerful concepts cannot be exhaustively "defined" in terms of less powerful ones (Fodor's notorious "plus-X" problem).
- (See Fodor, Garrett, Walker & Parkes, 1980). Truth conditions on formulas containing more powerful concepts cannot be characterized with formulas containing genuinely less powerful concepts.
- Evidence, suitably labeled, can "activate" them, but not "engender" them, for all the above reasons. In other words: The manipulation of primitive concepts can (in fact, it typically does) produce "brown cow" from "brown" and "cow", and the syntax of the composition.
- But no repetition of "This A is a B", and "This A is a B", and "This A is a B" … can generate "All As are Bs", unless you have the universal quantifier ("every", "all") **already** in your conceptual repertoire.
- This point is also made very cogently by the American logician and philosopher Hilary Putnam (Putnam, 1975). You must have a record of past observations of As and Bs involving some general uniform way of representing "All As are B". Otherwise, you cannot do that, no matter how many As and Bs you observe.

Fodor's buster

- To put it differently:
- Learning a concept is learning its **meaning** (its unique semantic properties). At some stage you must entertain the following formula in mentalese:
- (A) For every x, P is true of x, if and only if Q(x).
- Q is a concept of mentalese. The one you have (allegedly) "learned".
- P is some concept you had already.
- P must be coextensive with Q, if (A) is correct. But this is not enough: P must be coextensive with Q **in virtue of what P means**. Otherwise (A) is not a correct semantic formula. So, Q is synonymous with P. So, you had Q already in your "language of thought", because you had P. So, Q is not "learned". Iterate this for every primitive concept, keeping in mind the failure of Locke's program.
- Conclusion: All primitive concepts are innate. And (due to the failure of Locke's program) they are not all mere constructs from sensory impressions. It is a shocking conclusion, but it is also unavoidable. Followed, there and then, a long discussion on why such "extreme" innatism is so strongly resisted.

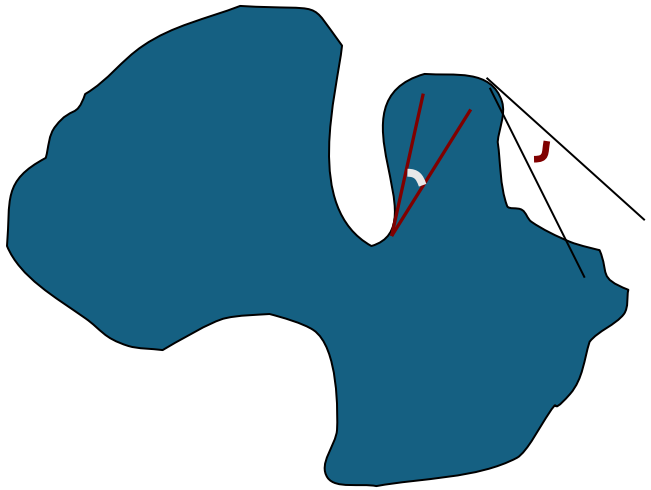
The perceptron

- Marvin Minsky and Seymour Papert had built an apparatus (the perceptron: Minsky and Papert, 1972) that had a kind of artificial retina, with interconnected local mechanisms, none of which covers the whole retina. None has any “global knowledge”.
- Weighted sums of the local decisions are reached by each sub-machine. As a result, we have emergent global decisions, not localized in any sub-part. There is a “learning function” sensitive to positive and negative feedback supplied from the outside.
- What can it learn? What is “innate” (in some sense)? The answer is far, far from obvious. Only by accumulating detailed analyses, case after case, can we answer that: It can easily learn to discriminate, say, between “triangular” and “square”, by looking at local angles in the retinal image. What about the predicate (the property, the concept) “connected”? Can it learn to decide whether the image is made up of one single connected piece, or several distinct connected pieces? The answer (far from obvious) is that it can.

The perceptron

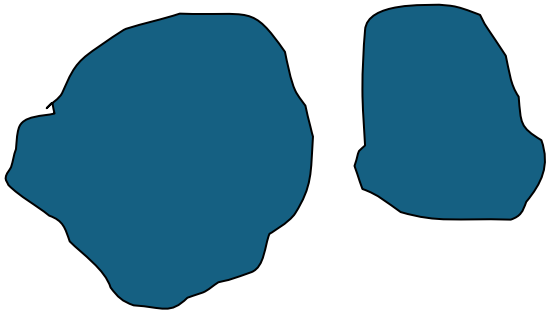
- Now, imagine an investigator (a Fodorian) who, therefore, concludes that “connectedness” is innate (prewired in the machine).
- But the wiring diagram cannot reveal anything that corresponds to “connectedness”. Big surprise! One has to be very careful.
- The key is a deep theorem, demonstrated by Euler, that links the total sum of external angles to connectedness, in a universal, necessary way. No piece of the machinery “possesses” (is sensitive to) the concept of connectedness, nor does it “contain” Euler’s theorem. The pieces just detect angles, and measure them. And the machine as a whole then sums up all the angles.
- It does not have to “know” (keep track of) whether two non-successive observations of angles are concerned or not with the same blob. As a result (not of innateness, but of the process itself), the machine is sensitive to connectedness. In exactly the right way. In virtue of Euler’s Theorem, if the total algebraic sum of the angles of the tangents is 2π then we have one connected blob. If n2π, then we have n connected blobs.

One connected blob



In virtue of Euler's Theorem:

Sum total = 2π



Two blobs

Sum total = 4π

The perceptron

- Suppose we did not know about Euler's theorem: we would have concluded that the perceptron had "learned" the concept of connectedness.
- That's wrong. It's neither innate (prewired) nor learned. It is the inevitable consequence of the deep property discovered by Euler.
- If the algebraic sum total of all the curvatures along the borders of a blob is 2π, then the blob is connected, whatever its shape. Otherwise, it's not. If it is n2π. Then we have *n* distinct objects. It's all so precise, and deep, and surprising. Terms like "concept", "notion", "predicate" are generic and misleading. We need better ones.
- The cognitive capacities of the adult may well be neither innate, nor learned. They have a **developmental history**. They emerge from other, different, components.

The perceptron

- Whatever is innate will not resemble in the least what you find in the adult's mind. The real search will have to track precursors, intermediate entities and constructions. Without the help of mathematicians and genetic epistemologists, we will never be able to track them, and understand how they interlock, to give the full-blown "concepts" of the adult mind. Artificial intelligence is a powerful tool in understanding these processes. It is a «long, arduous and technical path»

A Piagetian lesson

- Papert's Piagetian "lesson" was that, for instance, numbers are neither learned, nor innate.
- They are built through successive stages, by means of factorization (of a whole into sub-parts), the successor function (à la Peano), joint exhaustibility of sets by one-to-one pairing of their elements (à la Bourbaki).
- Different (though mathematically equivalent) definitions of numbers are more or less easy to learn, notably by the child.
- Most learnable is Bourbaki/Piaget, less so Russell/Whitehead, least of all Peano. All of them splendid mathematicians, but not all of them equally good "developmentalists".

Chomsky and Fodor strike back

- The machine has the concept “connected”, since it necessarily (not by sheer luck) applies the concept correctly to all and only the connected blobs.
- You would not have noticed that it had this concept, and why, if you were not as clever as Euler.
- But it does have the concept “connected”, exactly for the reasons explained by Papert, based on Euler’s theorem.
- A cognitive system does not have only the concepts that it’s easy for us, cognitivists, to ascertain that it has.
- Papert’s reply was that the perceptron, indeed, has the concept “connected”, but it’s precisely and exhaustively defined on the basis of other predicates the machine is sensitive to (local angles of curvature and their algebraic sum).
- So, contra Fodor’s thesis of the innateness of all concepts, this global concept is genuinely constructed from strictly local ones. If you had searched the “genome” of this machine to find where “connected” was encoded, the answer would have been: Nowhere! Yet, the machine has it, as also Fodor and Chomsky admit. Chomsky’s counter was: This is totally irrelevant. The set of actually available predicates has many interconnections, and it may be hard to discover some of these interconnections (it may take an Euler to discover some of them).

Chomsky and Fodor strike back

- Fodor's counter was: Predicates come in clusters, and we have to keep three problems separate:
- (1) What's the evidence that definitively compels the experimentalist to attribute possession of a concept (of a predicate) to the device?
- (2) Which predicates the device actually has?
- (3) What possessing a predicate amounts to?
- We have established that the concept is there, and that the device **actually exploits** it in carrying out its tasks. That's all that is needed.
- The problem of the definability of that concept in terms of other concepts is a separate one. One thing we now know for sure is that Locke's program was bound to fail. There is no buildup of such concepts from basic perceptual primitives.
-

Chomsky on syntax

- In essence, Chomsky wrote on the blackboard a few good English sentences and a few, superficially similar, but ungrammatical English sentences. The same for a few sentences of French
- The sentences offered by Chomsky were:
- **We expect John to like each other* (ungrammatical)
- *b) John seems to the men to like Bill* (ok).
- *c) *John seems to the men to like each other* (ungrammatical)
- Why could c) not mean: *John seems to each of the men to like the others?*
- That interpretation would seem perfectly fine. Why, then, is it excluded? The explanation offered there and then by Chomsky was the Specified Subject Condition (SSC).

Chomsky 1973

- The Specified Subject Condition (SSC) is a linguistic constraint that stops a grammatical rule from connecting two elements if an overt, contentful subject sits between them.
- **Definition:** A rule cannot link elements X and Y across a specified subject Z inside a phrase α (structured as … X … [α … Z … Y …] …).
- **Specified Subject:** A subject with actual semantic content, such as a proper noun, a full noun phrase (DP), or an overt pronoun.
- **Historical Context:** It originally restricted movement transformations and semantic interpretation rules, acting alongside the Tensed-S Condition. Later linguistic frameworks subsumed it into modern Binding Theory.

Chomsky on syntax

- In the following years, the SSC has been replaced by more abstract and unifying syntactic principles. Notably by Binding Theory. Originally, it was: Specified Subject Condition (SSC) “No rule can involve X, Y in the structure … X … [α… Z … - WYV …] … where Z is the specified subject of WYV in α.” (Chomsky 1973: 239)

The Tensed S Condition (Chomsky 1973)

- The **Tensed-S condition** (where S stands for "Sentence") is a condition proposed in Chomsky (1973) which essentially stipulates that certain classes of syntactic transformational rules cannot apply across clause boundaries. The condition is formalized as follows:
- **Tensed-S condition (TSC)** "No rule can involve X, Y in the structure … X … [α… Y …] … where α is a tensed sentence." (Chomsky 1973: 238)
- The rule accounts for such phenomena as the lack of passivization (a process which turns an active voice sentence into a passive voice in sentence (2) below:
- (1) The footballers are believed [to be talented]
- (2) *The footballers are believed [are talented]
- Based on the assumption that "The footballers" originates inside the square brackets in both active voice sentences (as the thematic subject of the predicate "be talented"), the TSC prohibits its raising (via A-bar movement) out of the finite clause in (2), but not the non-finite clause in (1).
- The TSC (along with the Specified Subject Condition (SSC) also has implications for later binding theory in conjunction with a simple rule of disjoint reference (which stipulated that any pronoun following a noun phrase (NP) antecedent in the same sentence, has disjoint reference with it; the rule applies anywhere unless it is blocked). The disjoint reference rule can apply in examples like (3) and (4) (which is an ECM verb structure) below, but is blocked from applying by the TSC in sentences (5) and (6) – where there are clause boundaries – thereby allowing the pronoun to refer back to the antecedent.
- (3) *John$_i$ likes him$_i$
- (4) *John$_i$ believes him$_i$ to like Mary
- (5) John$_i$ said that Mary likes him$_i$
- (6) John$_i$ said that he$_i$ likes Mary
- The way the TSC accounted for binding as well as movement phenomena (such as the passivization examples above), was influential for much subsequent research which tried to reduce binding and movement to the same set of principles (see Kayne (2002) for a recent implementation). The subsequent binding conditions A and B of Chomsky (1981) essentially replaced the TSC (as well as the SSC), and it is no longer a part of the toolkit of current researcher.

Chomsky on syntax

- His examples from French were:
- *J'ai laissé Jean manger X*.
- *J'ai laissé manger X à Jean*.
- If you have a quantifier, for example *tout* instead of X, then there is a rule that says that *tout* can be moved, which ought to give:
- d)* *J'ai tout laissé Jean manger*.
- *e) J'ai tout laissé manger à Jean*
- But e) is correct, while d) is not. Why?
- Because in d) there is a subject in the embedded sentence, therefore, by the SSC, only the subject is accessible to a rule that moves an element, and thus you can't move the word *tout*, just as you could not interpret the reciprocal phrase *each other*,
- in a) when there was a subject. In the second French case there is no subject, but rather something that functions as a subject in a prepositional phrase (*à Jean*), since there is no subject, the SSC does not apply, therefore *tout* is accessible to the rule, and therefore it moves. The cases look rather different, but the explanation is the same,

Chomsky on syntax

- Chomsky added that
- “there is nothing logically necessary about that principle”.
- And he also added: “English would be a perfectly fine language if it didn’t have that principle [SSC], only in English you would then say *John seems to the men to like each other*. There would be **no problem about communication**; all the properties would be acceptable. *It just wouldn’t be a human language* [italicized in the original]; it could be the language of another organism, namely, an organism whose initial state would not include this principle.”

Chomsky on syntax

- Then Chomsky added:
- “One could conceive of an organism exactly like humans, but minus the Specified Subject Condition, and it would talk with a fine language which it could use for all possible purposes.
- In fact, we could observe that organism for a very long time and not even know that it is not using the SSC, because these questions generally don’t even arise.”
- And reintroduces the Poverty of the Stimulus. No relevant experience for the child.

Chomsky on syntax

- Chomsky added (rightly, I thought then and still think today) that we might not have this syntactic condition in our head and we would communicate equally well.
- The German ethologist Dieter Dütting immediately raised his hand and said that he could not accept Chomsky's explanation.
- Chomsky invited him to offer an alternative explanation. Dütting said: "I have no idea, but cannot accept your explanation, because it is anti-Darwinian."
- This was, there and then, a first crack in my neo-Darwinian persuasions.
- To be developed, many years later, in my book with Fodor.

Abstraction in Piaget, and in Chomsky

- In Piaget it's a central cognitive "engine", acquired progressively, in a stepwise fashion, reflecting the "necessity" of mathematical constructions.
- Justified by these mathematical constructions.
- In Chomsky it's primitive, instantaneous, natural to the system, automatic.
- It's system-specific, not general.
- It's "justified" by the simplicity, elegance and predictive power of the linguist's theory.
- Just as in physics, chemistry, biology, no other "justification" is needed, or possible.
- The relentless search for such abstractions has propelled the theory ever after.

Fodor's triggers

- Monod: If sensorimotor schemas are so crucial for language learning, then a child born paraplegic should not be able to develop a normal language. But that's not the case.
- Inhelder: Very little movement is required, even just moving your eyes.
- Fodor: Then all we have is a **trigger**, not a structured precursor. This trivializes the doctrine that intelligence arises out of sensorimotor activity. Nothing is left in it.

Chomsky and Fodor strike back:

- The machine **has** the concept "connected", since it **necessarily** (not by sheer luck) applies the concept correctly to **all and only** the connected blobs.
- You would not have **noticed** that it had this concept, and why, if you were not as clever as Euler.
- But it **does** have the concept "connected", exactly for the reasons explained by Papert, based on Euler's theorem.
- A cognitive system does not have **only** the concepts that it's **easy** for us, cognitivists, to ascertain that it has.

Papert's reply:

- The point is: It, indeed, has the concept "connected", but it's **precisely and exhaustively** defined on the basis of **other** predicates the machine is sensitive to (local angles of curvature and their algebraic sum).
- So, contra Fodor's thesis of the innateness of all concepts, this global concept is genuinely **constructed** from strictly local ones.
- If you had searched the "genome" of this machine to find where "connected" was encoded, the answer would have been: Nowhere!
- Yet, the machine has it, as also Fodor and Chomsky admit.

Replies there and then:

- Chomsky: The set of actually available predicates has many interconnections, and it may be hard to discover some of these interconnections (it may take an Euler to discover some of them).
- Fodor: Predicates come in clusters, and we have the keep three problems separate:
- (1) What's the evidence that definitively compels the experimentalist to attribute possession of a concept (of a predicate) to the device;
- (2) Which predicates the device actually has.
- (3) What possessing a predicate amounts to.

Replies there and then:

- Fodor: We have established that the concept is there, and that the device **actually exploits** it in carrying out its tasks. That's all I needed.
- The problem of the definability of that concept in terms of other concepts is a **separate one**.
- One thing we now know for sure: That Locke's program was bound to fail.
- Locke's program: Derive **all** concepts from a Boolean combinatorial algebra of elementary sensory primitives.
- You cannot even acquire concepts such as *buy and sell, automobile* and *typewriter* on the basis of Locke's schema. That has been shown.

Why?

- Sub-Option 1: The learner generates X by sheer luck, and X fits the available evidence by sheer luck. Otherwise he/she/it dies, along with all the descendants.
- In fact, most of the time, X is wrong. Only rarely do such guesses work (extreme Darwinism)
- No learning has taken place, just lucky blind guesses.
- Special case: The target X and the guessed Y happen to be co-extensive. They mean quite different things (they are semantically distinct).
- Y (not X) has been "fixed" by natural selection.
- This may (just may) apply to genuine **triggers**.

Guidance by truth

- Sub-Option 2: The learner, somehow, "tracks" the content (the meaning) of X, and **why** it is adequate with respect to (**true of**) the available evidence. The process is, somehow, **guided**.
- The content of X, and some sensitivity of the process to the truth/falsehood of X, supply the required "guidance" (tracking).
- Nothing else **could** supply it.
- Evolution's criterion can only be truth-value.
- There is learning (inductive fixation/rejection) **of the meaning of X**
- But X cannot be fixed/rejected unless it is actually available to the learner **and it is exploited in the process** by the learner.

Fodor's conclusion:

- So, where do new concepts come from?
- Three possibilities:
- God whispers them to you on Tuesdays;
- You acquire them by falling on your head;
- They are innate.
- OR
- *"there must be some notion of learning that is so incredibly different from the one we have imagined that we don't even know what it would be like, as things now stand."*

My update on GG 50 years later

- I selected two main developments:

- Principles and Parameters
- Minimalism