Hypothesis:

principles + substantial computation

⇓

inductive biases ⇒ High-level cognitive

inductive priors in DL

⇓

preferences, priors, or assumptions

in-distribution generalization

out-of-distribution (OOD) generalization

transfer learning to new tasks with low sample complexity

high-level to talk about variables that are manipulated at the conscious level of processing and are thus generally verbalizable.

low-level or intermediate-level features, e.g. by describing an odd-coloured pixel, not just very abstract concepts like objects or social situations

Implicit and explicit knowledge

An important question for us is how knowledge can be represented in these two forms, the implicit—intuitive and difficult to verbalize—and the explicit—which allows humans to share part of their thinking process through natural language.

Table 1. Examples of current inductive biases in deep learning. Some have to do with the architecture while the last one influences the training framework and objective.

inductive bias	corresponding property
distributed representations	patterns of features
convolution	group equivariance (usually over space)
deep architectures	complicated functions = composition of simpler ones
graph neural networks	equivariance over entities and relations
recurrent nets	equivariance over time
soft attention	equivariance over permutations
self-supervised pre-training	$P(X)$ is informative about $P(Y\|X)$

inductive biases → encourage the learning algorithm to prioritize solutions with certain properties

neural architecture

how the networks are trained

Table 2. Proposed additional inductive biases for deep learning: much progress has been made in learning representation of high-level variables (entities or objects). Much more progress is needed on other inductive biases such as the ones listed above. It would also be useful to think about integrating these inductive biases into a unified architecture.

inductive bias	corresponding property	relevant references
high-level variables play a causal role	learning representations of latent entities/attributes	[90–108]
changes in distribution are due to causal interventions	changes in distribution are localized in the appropriate semantic space	[103,109–113]
knowledge is generic, defined over abstract variables, and can be applied on different instances	factorizing knowledge in terms of abstract variables and functions that encapsulate how these variables interact with each other	[99,114,115]
sparsity of the factor graph	learned functions operate on a sparse set of variables (like arguments in typed-programming languages)	[99,114]
relevant causal chains tend to be very short (in time)	causal chains used to perform learning or inference (to obtain explanations or plans for achieving some goal) are broken down into short causal chains of events that may be far in time from each other	[116–121]
context-dependent processing involving goals, top-down influence and bottom-up competition	top-down contextual information is dynamically combined with bottom-up sensory signals at every level of the hierarchy of computations relating low-level and high-level representations	[122–124]

Attention: dynamic information flow

(dynamic connection between different blocks)

Attention is about sequentially selecting what computation to perform on what quantities.

Soft attention:

convex combination of the values of the

elements at the previous level.

- Convex weights are coming from a softmax that is conditioned on how each of the elements' key vector matches some query vector.

Stochastic hard attention:

one samples from a distribution over elements to choose the attended content.

Soft attention:

one mixes these contents with different positive convex weights.

Attention: process sets of key/value pairs query (read key) $Q \in \mathbb{R}^{N_r \times d}$

key (write key) $K \in \mathbb{R}^{N_o \times d}$

(d: dimension of each key)

Values (write values)

$$\text{Attention}(Q, K, V) =$$

$$\text{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V$$

Declarative knowledge of causal structure