Cours 2013:

Le bébé statisticien

Stanislas Dehaene
Chaire de Psychologie Cognitive Expérimentale

Cours n°1

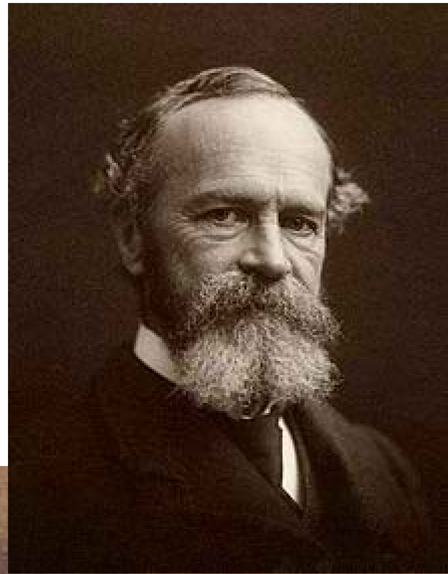
L’apprentissage bayésien:
Sommes-nous des scientifiques dès le berceau?

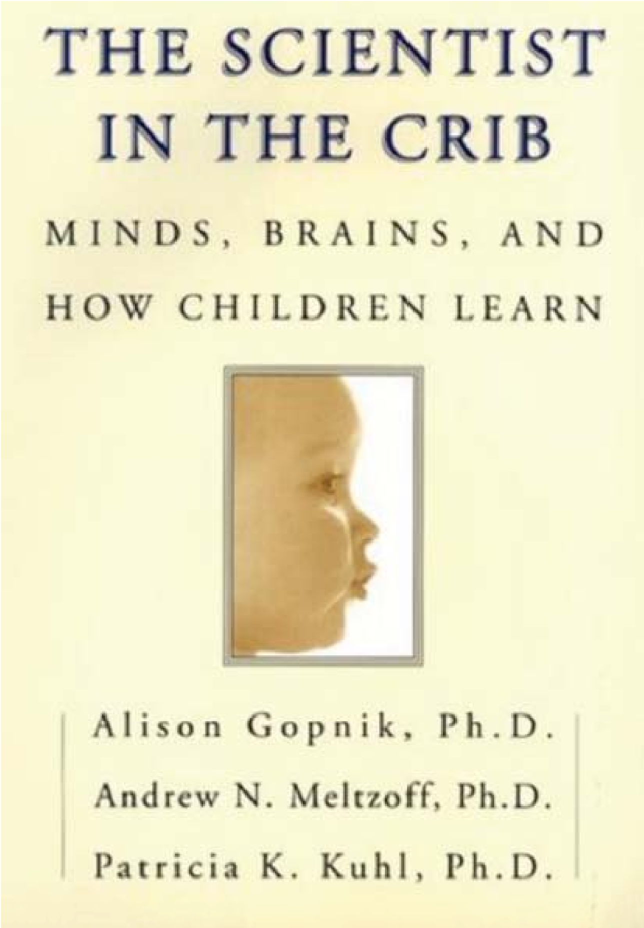
Deux visions du cerveau du bébé

Une confusion totale?

“The law is that all things fuse that *can* fuse, and nothing separates except what must. (…) The baby, assailed by eyes, ears, nose, skin, and entrails at once, feels it all as one great blooming, buzzing confusion.”

Williams James, *Principles of Psychology*





Des scientifiques au berceau?

“Over the past 30 years we have discovered an enormous amount about what children know and when they know it. In particular, young children, and even infants, seem to have intuitive theories of the physical, biological and psychological world. These theories, like scientific theories, are complex, coherent, abstract representations of the causal structure of the world” (Gopnik &Schulz, TICS 2004)


“Playing Doctors” by Frederick Daniel Hardy (1827–1911)

Deux étapes dans la révision des connaissances sur le bébé

1. Les multiples compétences du bébé

Le bébé dispose, dès la première année de vie, de « connaissances fondamentales » (*core knowledge*)

- concepts d'objet (Spelke, 1983), de nombre (Mehler & Bever, 1967; Wynn, 1992), d'espace (Landau et al., 1981)…
- grammaire universelle (Chomsky, Gleitman, Mehler)

Ces systèmes seraient innés et guideraient les apprentissages ultérieurs.

2. La métaphore du bébé comme un scientifique… ou un détective

- Le bébé élabore des **modèles mentaux du monde extérieur**
- Il **évalue leur plausibilité** en liaison avec les observations qu'il effectue.

→ il faut le supposer doté de compétences précoces pour

→ la manipulation des probabilités

→ l'évaluation, en parallèle, de très nombreux modèles du monde

→ la sélection des variables pertinentes, l'élimination des variables de non-intérêt, le repérage des ambiguïtés ou des interprétations multiples

Cette vision n'est pas incompatible avec la perspective nativiste. Cependant, certains principes, même très abstraits, peuvent être sélectionnés par apprentissage bayésien. Il n'est plus nécessaire de supposer qu'ils soient innés.

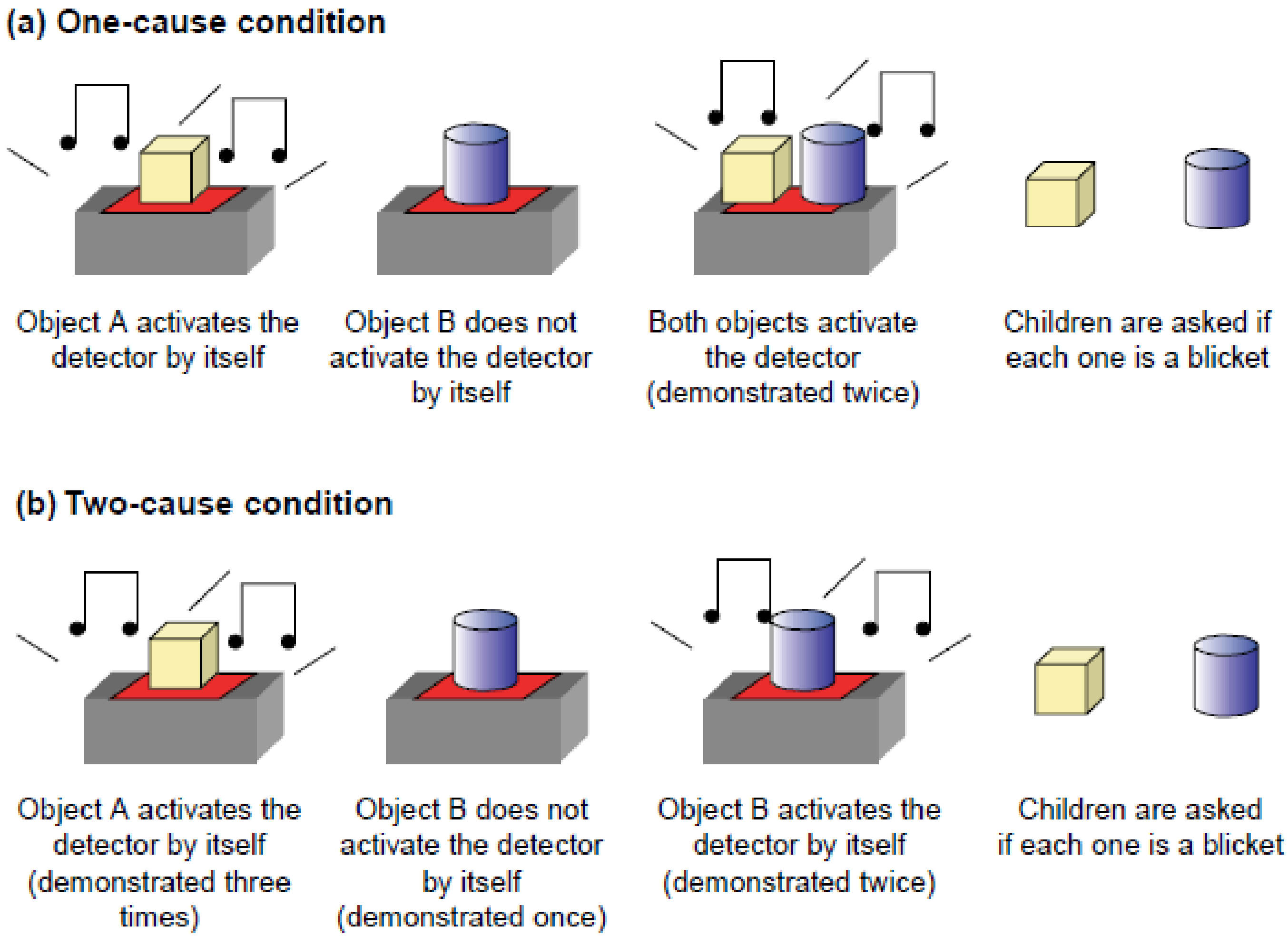
Un raisonnement quasi-scientifique chez l'enfant

Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Dev Psychol, 37(5), 620-629.*

Les enfants de deux ans et demi sont capables d'inférences rapides et sophistiquées.

-dans la condition (a), le détecteur ne s'allume qu'en présence du cube. --> les enfants disent que le cube est le blicket

- dans la condition (b), le détecteur s'allume tout autant, et autant de fois en présence de chaque objet, mais la seule explication causale est que les deux objets sont des blickets → c'est la réponse des enfants.



Les idées essentielles de la théorie bayésienne (cours 2012)

L'**inférence bayésienne** est une théorie mathématique simple qui caractérise le **raisonnement plausible** en présence d'incertitudes.

L'inférence bayésienne rend bien compte des processus de **perception**: étant donné des entrées ambigües, notre cerveau en reconstruit l'interprétation la plus probable.

L'inférence peut être **hiérarchique** et donner accès à des connaissances très abstraites.

Nos **décisions** combinent un calcul bayésien des probabilités avec une estimation de la valeur probable et des conséquences de nos choix.

L'architecture du cortex pourrait avoir évolué pour réaliser, à très grande vitesse et de façon massivement parallèle, des inférences bayésiennes.

L'algorithme utilisé pourrait expliquer la manière dont notre cerveau anticipe sur le monde extérieur (**codage prédictif**) et dont il répond à la nouveauté **(propagation des signaux d'erreur).**

Le cours de cette année

Le **bébé** semble dôté, dès la naissance, de compétences pour le **raisonnement probabiliste**.

Le raisonnement bayésien fournit un **puissant algorithme d'apprentissage** de régularités statistiques.

L'apprentissage du **langage** , la reconnaissance des mots, et la **théorie de l'esprit** pourraient également être modélisés comme des inférences bayésiennes.

Plan du cours

Aujourd'hui: Exposition des grands principes de la théorie bayésienne:
sommes-nous des scientifiques dès le berceau?

Mardi 15 Janvier: Le sens des probabilités chez l'enfant

De Piaget à Xu, Bonatti, Teglas

Mardi 22 Janvier : L'apprentissage de régularités statistiques

Le bébé internalise les régularités du monde extérieur et anticipe sur les entrées qu'il va recevoir

Mardi 29 Janvier: La découverte et l'apprentissage des mots

Segmentation des mots et apprentissage de leur sens

Mardi 5 Février : Vers une théorie Bayésienne du lexique

Mardi 12 Février : La théorie de l'esprit : une inférence Bayésienne ?

Principe 1. Le raisonnement bayésien permet des inférences « en avant » mais surtout « en arrière »

La théorie des probabilités, telle qu'enseignée à l'école, est surtout utilisée pour calculer la probabilité d'une observation, étant donné certaines hypothèses sur l'état du monde.

Par exemple, soit une urne contenant 3 boules noires et 7 boules blanches. Quelle est la probabilité, lors de deux tirages sans remplacement, de tirer deux boules noires?

$$p(H \,\&\, N_1 \,\&\, N_2) = p(N_1|H)\, p(N_2|N_1 \,\&\, H) = (3/10) \times (2/9) = 1/15$$

Mais, les données d'observation D et les hypothèses H jouent des rôles strictement symétriques. Rien n'empêche d'utiliser les équations pour inverser le procédé:

Etant donnée l'observation D, quelle est la probabilité de l'hypothèse H?

Application de la règle fondamentale: $p(H \,\&\, D) = p(D|H)\, p(H) = p(H|D)\, p(D)$

D'où $p(H|D) = p(D|H)\, p(H) / p(D)$ ou $\boxed{p(H|D) \;\alpha\; p(D|H)\, p(H)}$

p(H) = *probabilité « a priori » de H* (*prior* en anglais)

p(H|D) = *probabilité « a posteriori » de H*

p(D|H), considéré comme une fonction de H, est la *vraisemblance de H*

C'est cette équation fondamentale qui va permettre de modéliser comment l'enfant révise ses croyances en fonction des observations qu'il effectue.

Une inférence bayésienne chez l'enfant: p(urne|échantillon) =?

Dans les expériences 1 et 2:
- On familiarise d'abord des enfants de 8 mois avec des balles rouges et blanches, et avec des boites qui contiennent, soit une majorité de balles blanches, soit une majorité de balles rouges.
-Puis, à chaque essai, l'expérimentateur montre une urne à l'enfant, mais son contenu n'est pas visible.
- Les yeux fermés, l'expérimentateur retire, une à une, cinq balles (4 d'une couleur, une de l'autre couleur).
- le contenu de l'urne est révélé

Résultat: l'enfant regarde plus longtemps l'urne dont le contenu ne correspond pas à l'échantillon observé.

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proc Natl Acad Sci U S A, 105(13), 5012-5015.*

Une inférence bayésienne chez l'enfant: p(urne|échantillon)=?

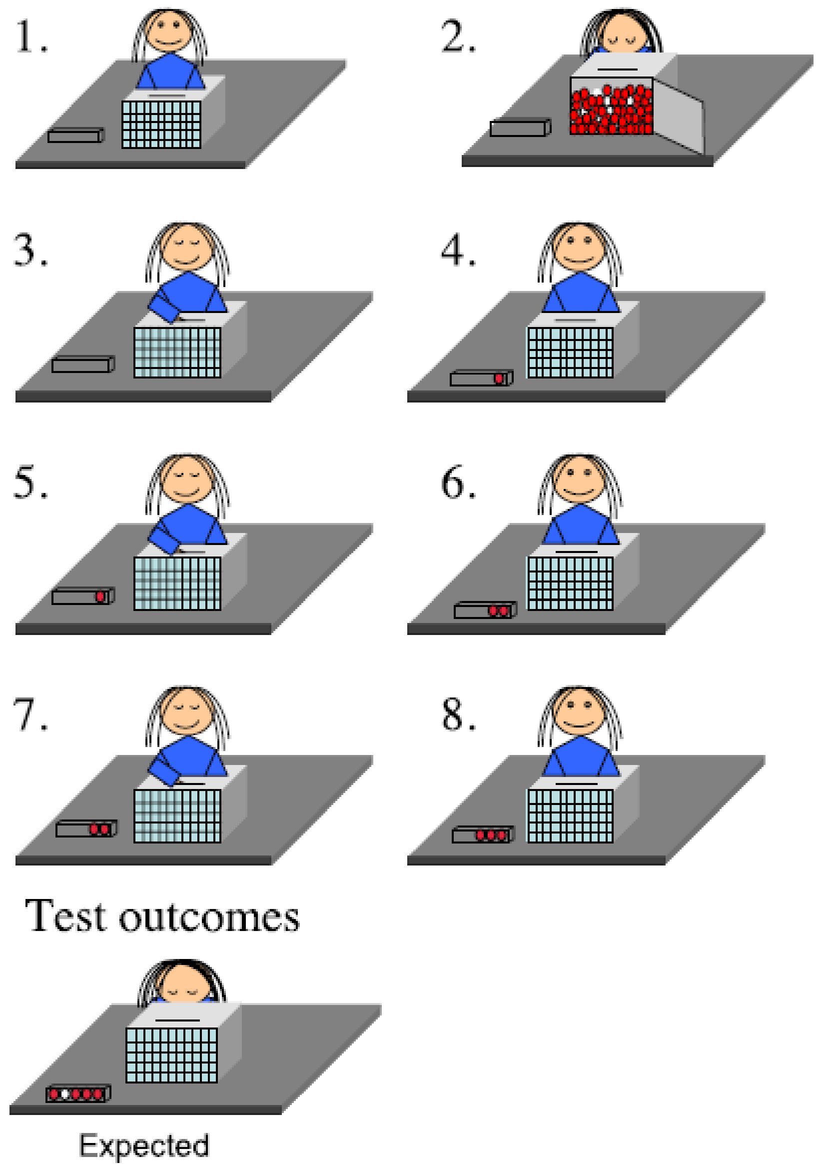
L'enfant est-il seulement sensible à la congruence des représentations sensorielles?
Expérience de contrôle: (exp. 3): tout est identique, sauf que l'expérimentateur retire visiblement les balles de sa poche et non de la boîte.
Résultat: Aucune préférence visuelle n'est observée.

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proc Natl Acad Sci U S A, 105(13), 5012-5015.*

La capacité d'inférence inverse: p(échantillon|urne)=?

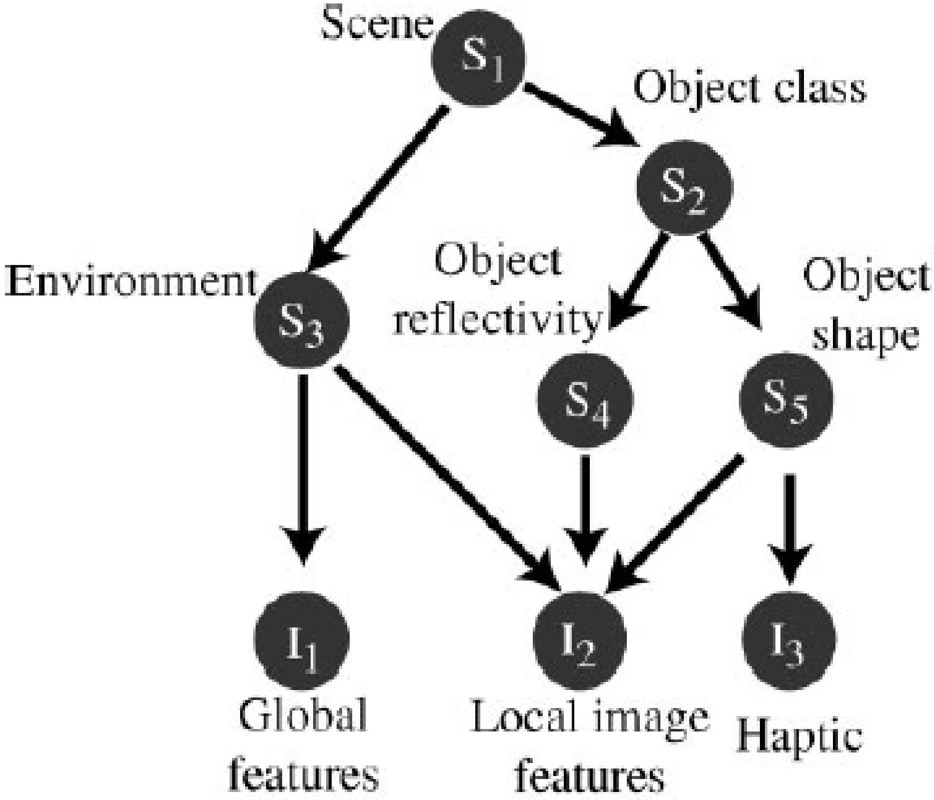
L'enfant peut-il faire l'inférence inverse: utiliser une information sur la population pour prédire les échantillons?
Expériences 4 et 5: le contenu de l'urne est visible dès le départ.
Résultat: les enfants regardent plus longtemps les essais où le contenu de l'échantillon est improbable.
De nouveau, aucune préférence n'est observée lorsque l'expérimentateur retire les balles de sa poche.

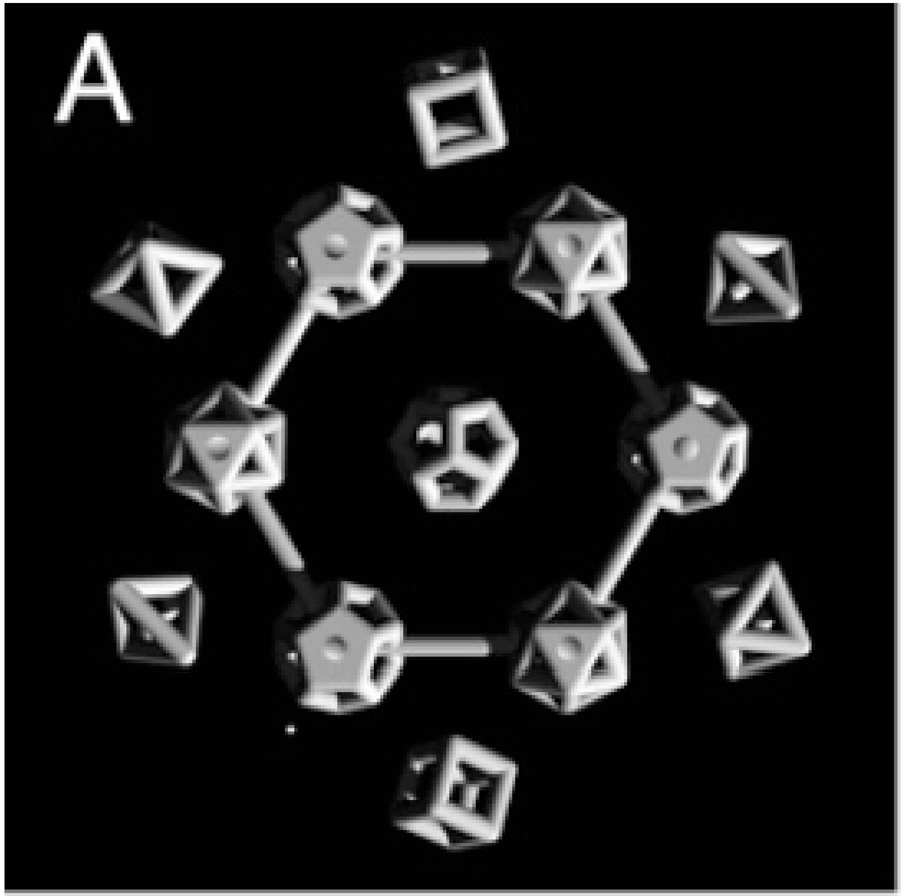
Conclusion: l'enfant de huit mois possède déjà des capacités d'inférence bidirectionnelles, fondées sur l'observation de la numérosité d'ensembles d'objets.



Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proc Natl Acad Sci U S A, 105(13), 5012-5015.*

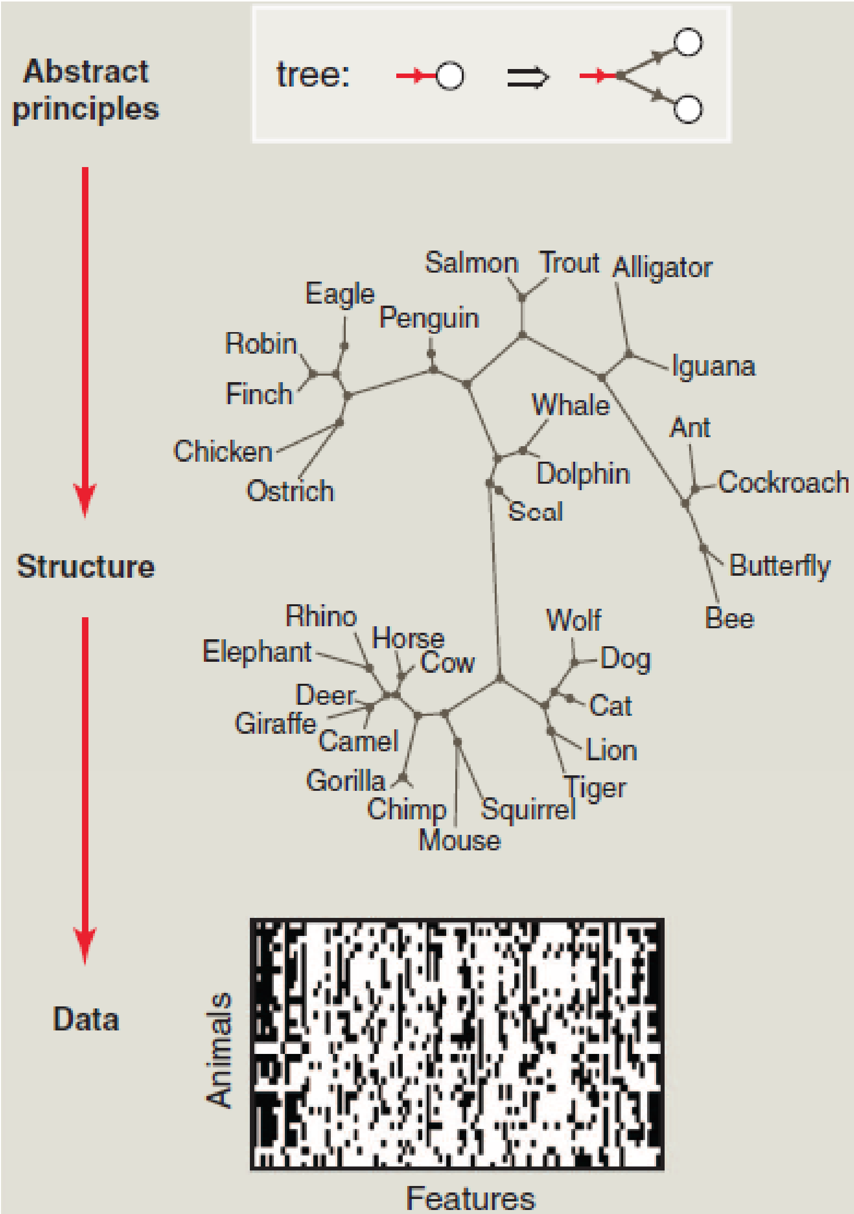
Principe 2. Le raisonnement bayésien permet des inférences **hiérarchiques, en cascade**

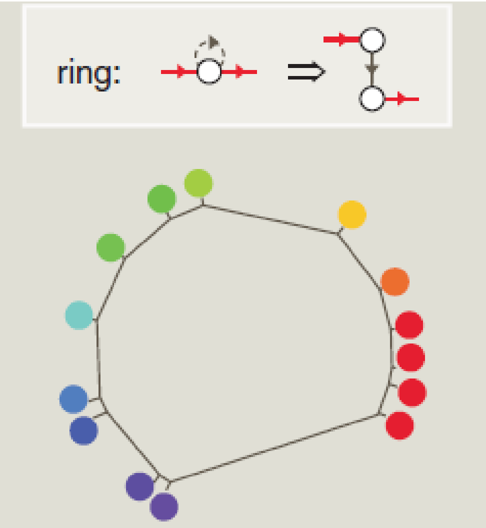




Le *modèle génératif* qui engendre les observables peut être organisé de façon hiérarchique, avec des *hyper-paramètres* de niveau supérieur qui contraignent les hypothèses de niveau inférieur.
La théorie bayésienne montre comment inverser le modèle et, partant des observations, remonter à des inférences de niveau très abstrait.

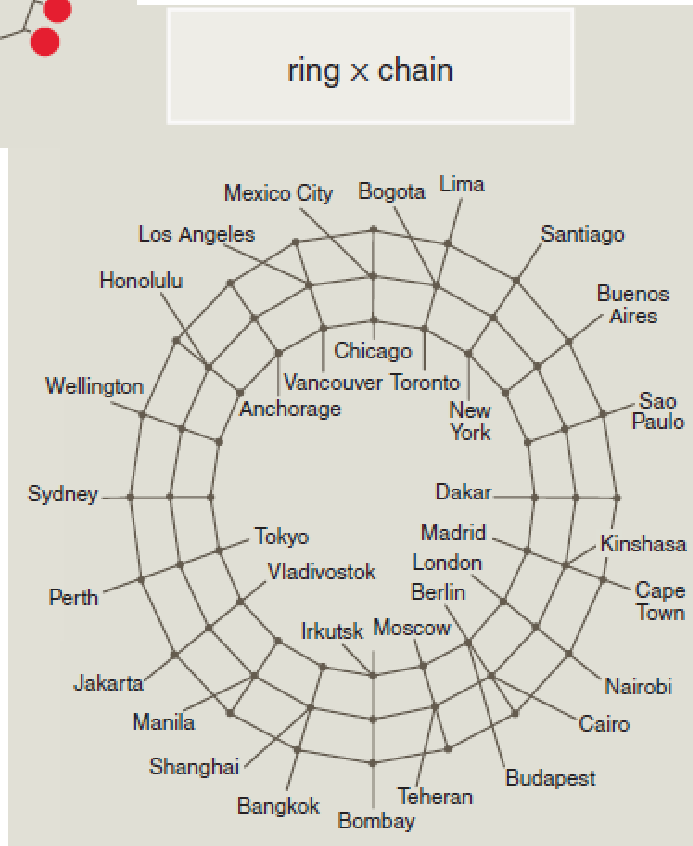
Principe 2. Le raisonnement bayésien permet des inférences **hiérarchiques, en cascade**





Kemp et Tenenbaum 2008: Découverte automatique que les animaux forment des arbres, les couleurs un anneau, et le monde une sorte de globe.

Blessing of abstraction: L'apprentissage peut être plus rapide au niveau le plus élevé: principes abstraits *avant* modèle particulier.



Une cascade d'inférences chez l'enfant de 16 mois

Gweon, H., & Schulz, L. (2011). 16-month-olds rationally infer causes of failed actions. *Science, 332(6037), 1524.*

Les enfants savent-ils inférer l'origine d'une anomalie: le monde extérieur, ou eux-mêmes?
Si un objet ne fonctionne pas: Est-ce le jouet qui est en panne? Ou bien moi qui n'y arrive pas?

Une cascade d'inférences chez l'enfant de 16 mois

Gweon, H., & Schulz, L. (2011). 16-month-olds rationally infer causes of failed actions. *Science, 332(6037), 1524.*

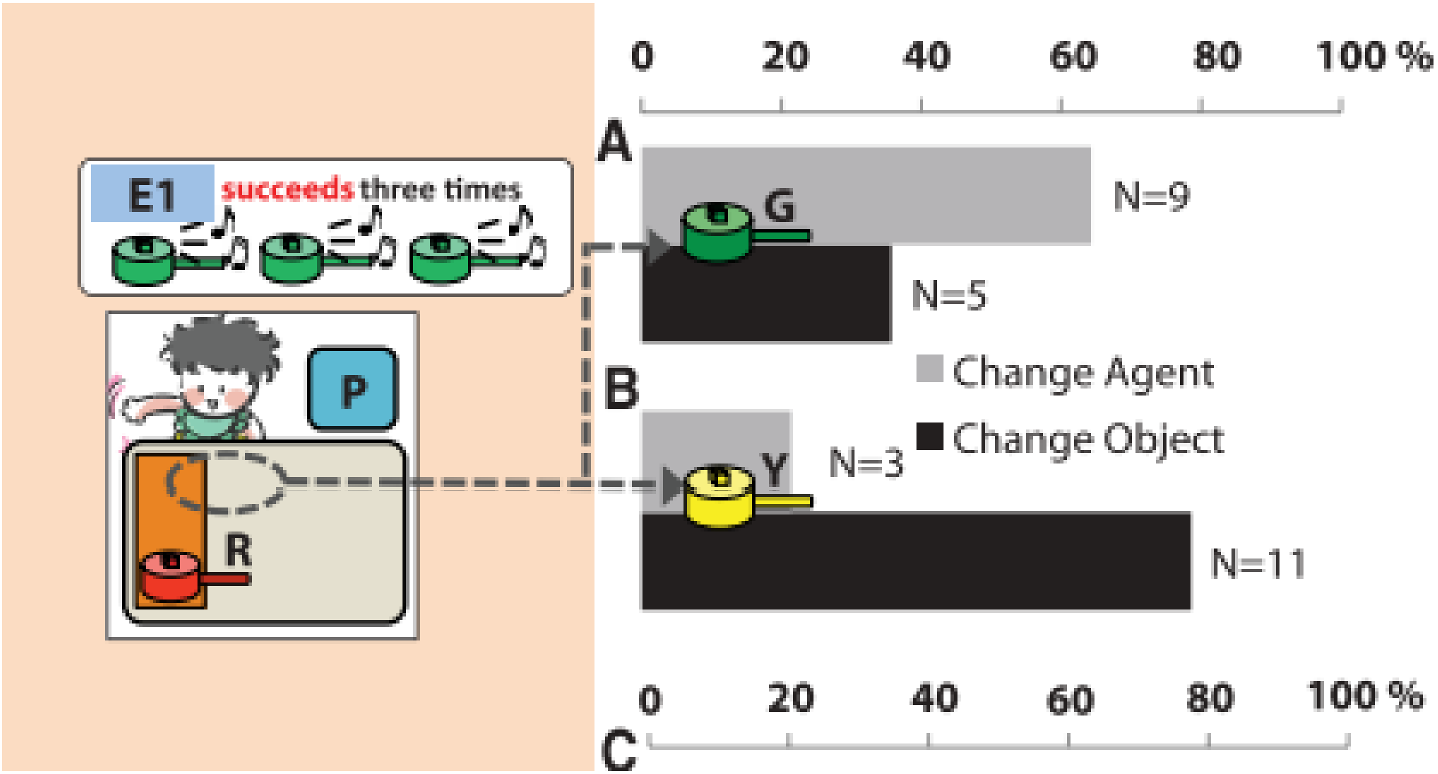
L'enfant réalise une inférence rationnelle:

-si l'objet n'a pas changé, alors c'est sans doute moi qui fais mal

→ demander à un autre agent.

-si l'objet a changé, il est plus probable que ce soit l'objet qui ne fonctionne pas

→ changer d'objet

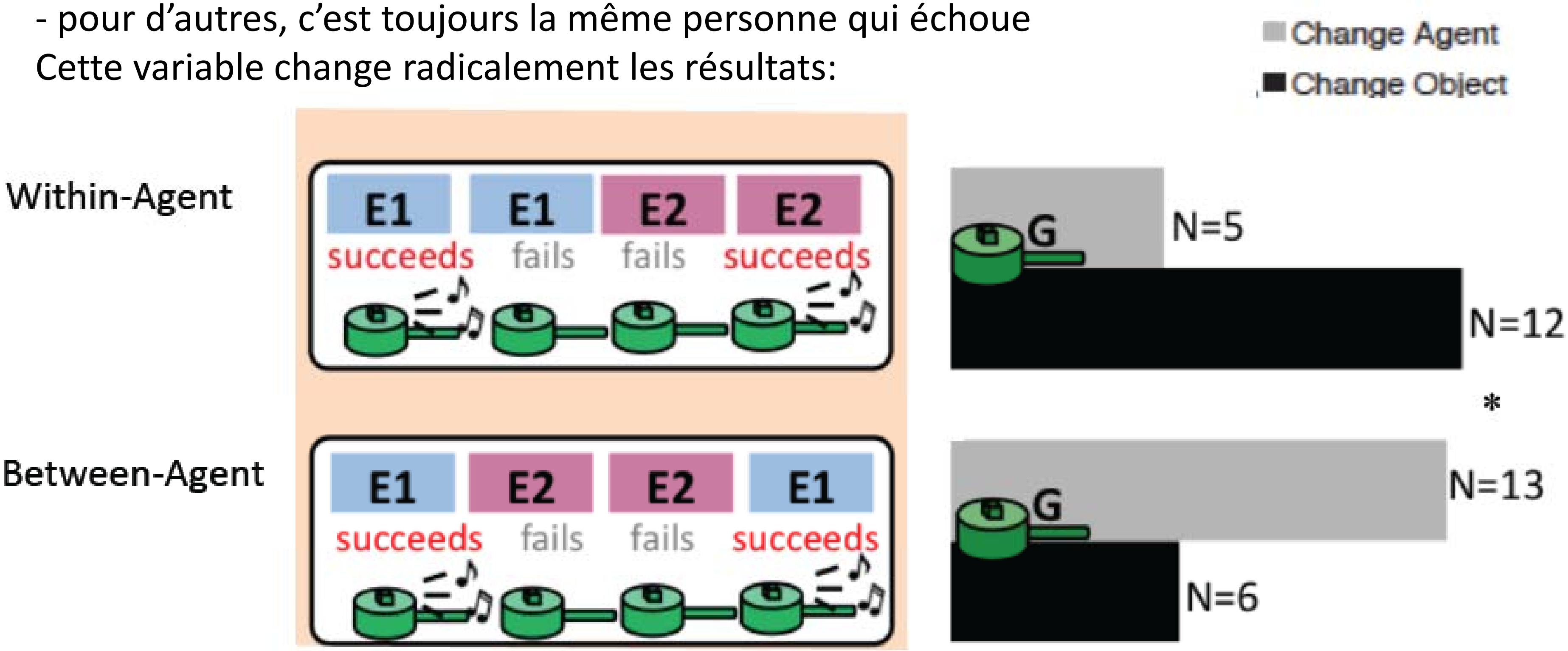


Une cascade d'inférences chez l'enfant de 16 mois

Gweon, H., & Schulz, L. (2011). 16-month-olds rationally infer causes of failed actions. *Science, 332(6037), 1524.*

Expérience 2: L'objet ne change pas, mais c'est l'agent qui change.
L'objet fonctionne une fois sur deux, mais
- pour certains enfants, la réussite et l'échec surviennent avec la même personne
- pour d'autres, c'est toujours la même personne qui échoue

Cette variable change radicalement les résultats:

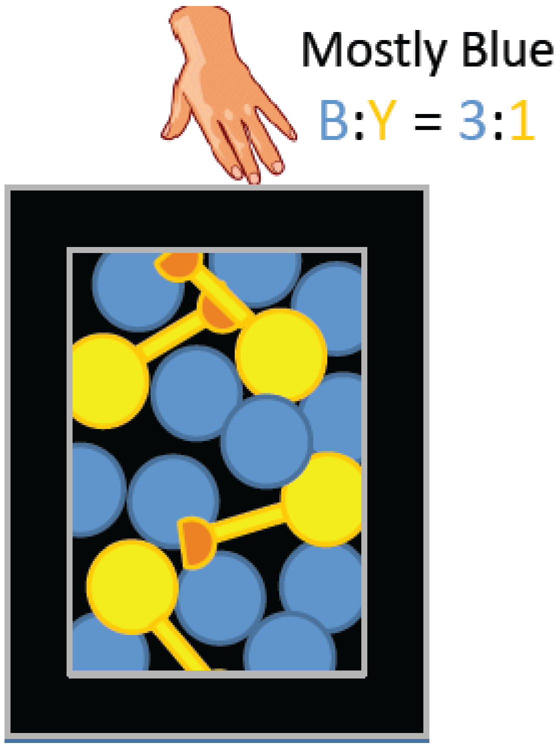


Conclusion:
il parait très difficile d'expliquer ces résultats sans supposer que l'enfant est capable, dès cet âge, de raisonner sur les **causes** des événements.

Qu'est-ce que les enfants connaissent de l'échantillonnage?

Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proc Natl Acad Sci U S A, 107(20), 9066-9071*.

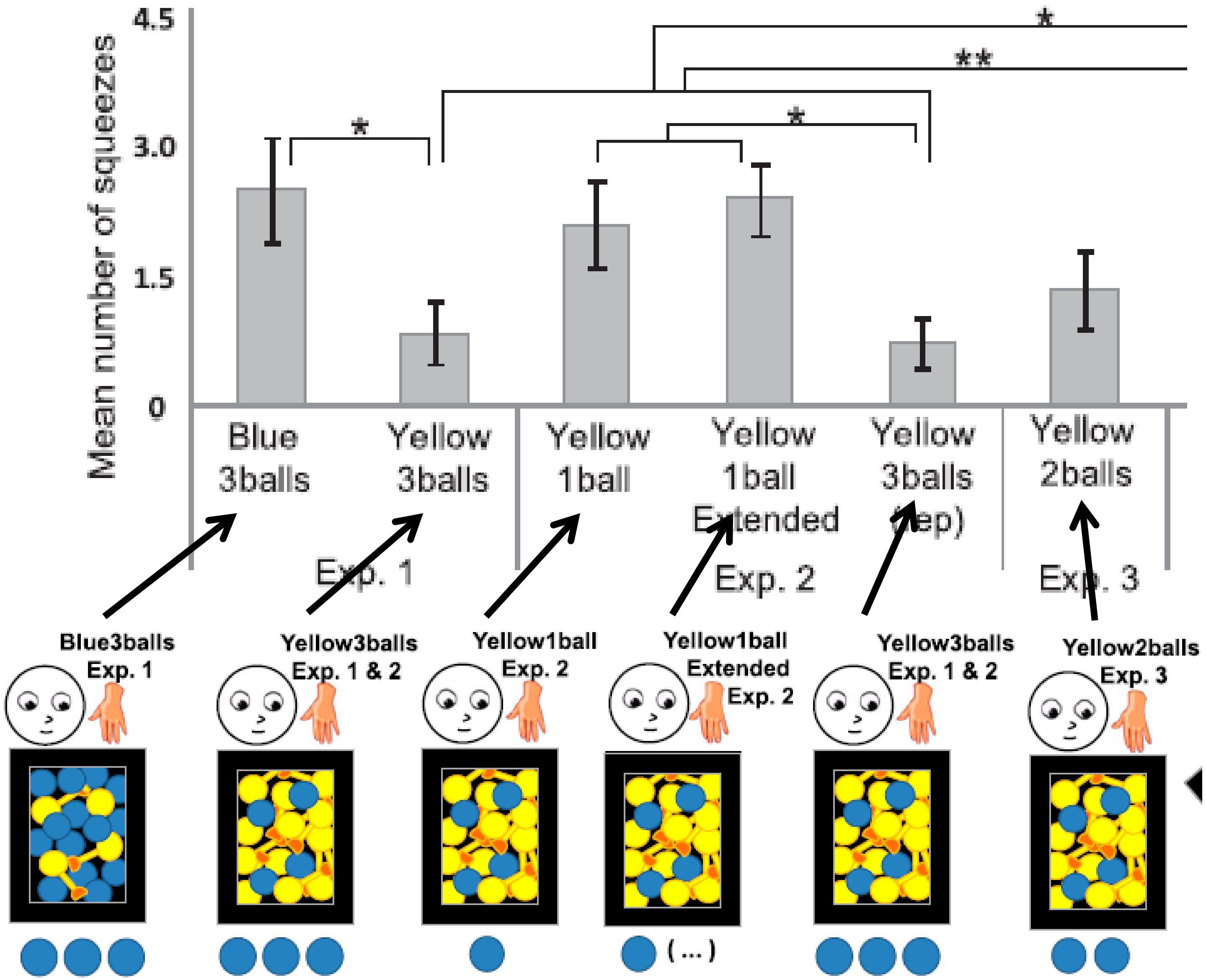
L'induction ne peut être correcte que si l'échantillonnage n'est pas biaisé, en sorte que l'échantillon est représentatif de la population.
La détection des biais est un processus essentiel de l'enquête scientifique (par exemple, si une météorite martienne contient de la silice, est-ce le cas de toute la surface de Mars?)
Les enfants savent-ils détecter les biais d'échantillonnage?



Une main retire trois objets bleus d'une urne: tous émettent un joli bruit quand on leur appuie dessus.

On en retire un objet jaune: l'enfant va-t-il généraliser et se mettre à appuyer?

Les enfants de 13 à 18 mois généralisent sur la base d'une solide connaissance des processus d'échantillonnage

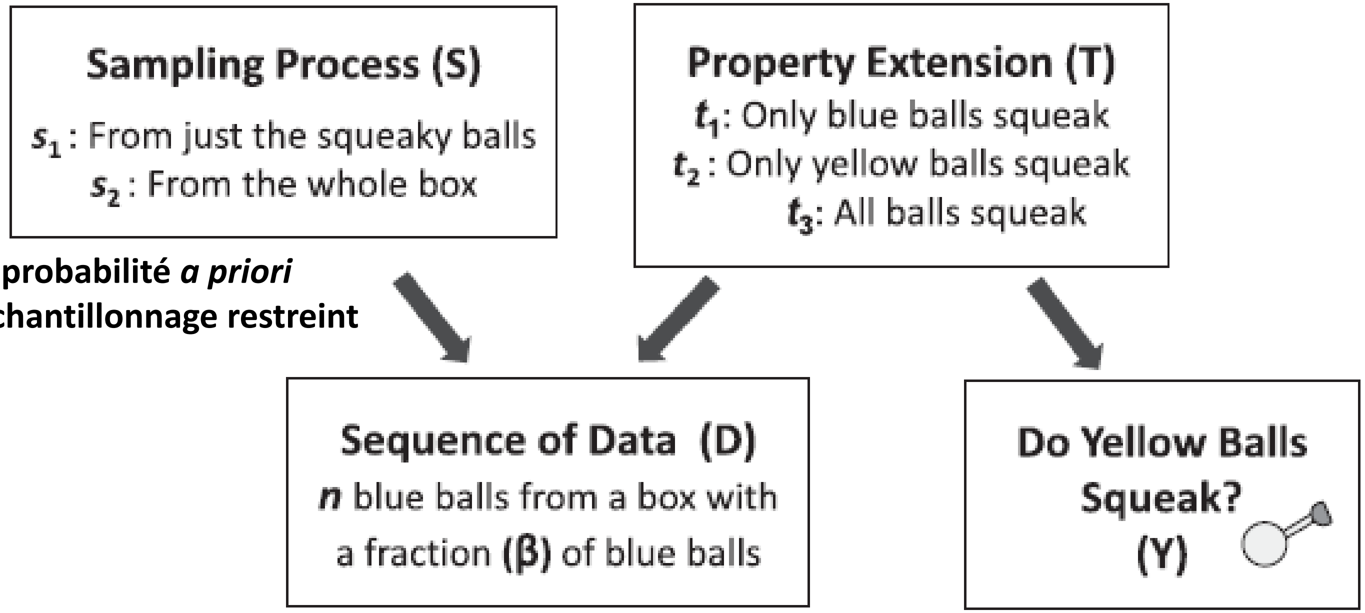


Conclusion: les réponses des enfants dépendent:
- de la proportion de balles bleues
- mais aussi, semble-t-il, du fait que les balles aient pu être échantillonnées au hasard.

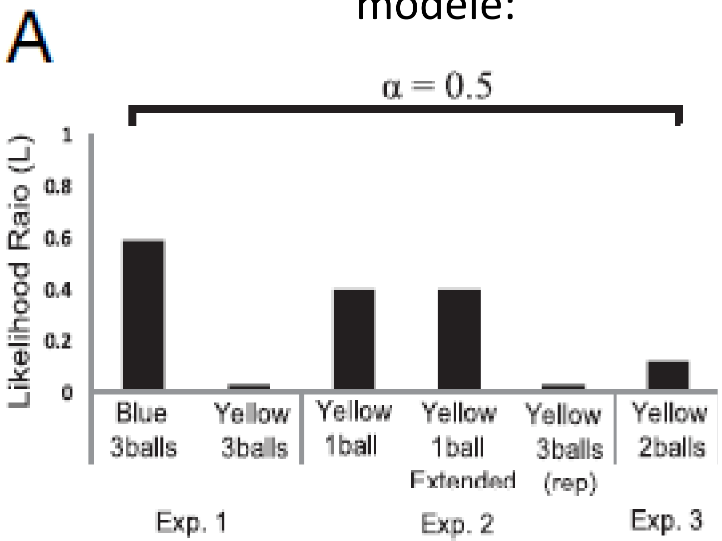
L'enfant semble inférer que, si l'échantillonnage n'est pas au hasard, c'est que l'adulte suit une stratégie, ce qui fournit des informations supplémentaires

Modélisation bayésienne de ces données

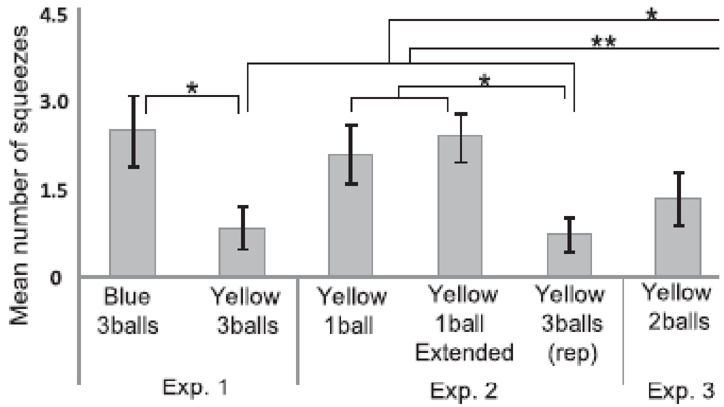
α = probabilité *a priori* d'échantillonnage restreint



modèle:



données:



Modélisation bayésienne de ces données

Sampling Process (S)
s_1 : From just the squeaky balls
s_2 : From the whole box

Property Extension (T)
t_1: Only blue balls squeak
t_2 : Only yellow balls squeak
t_3: All balls squeak

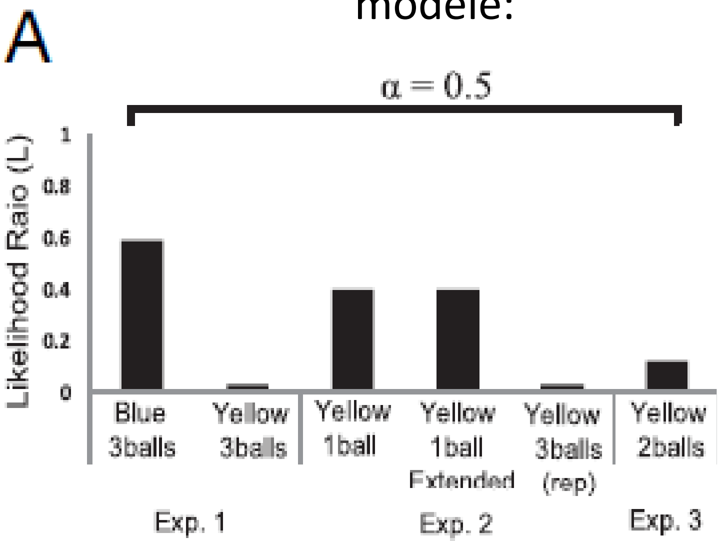
α = probabilité *a priori* d'échantillonnage restreint

Sequence of Data (D)
n blue balls from a box with a fraction (β) of blue balls

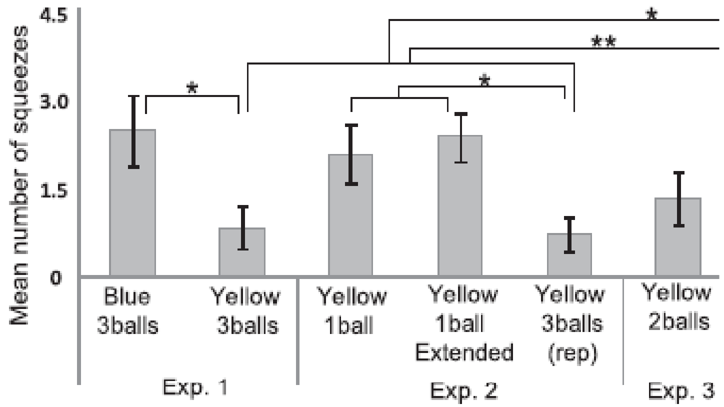
Do Yellow Balls Squeak? (Y)

Que se passerait-il si l'enfant disposait d'informations supplémentaires sur la nature de l'échantillonage?

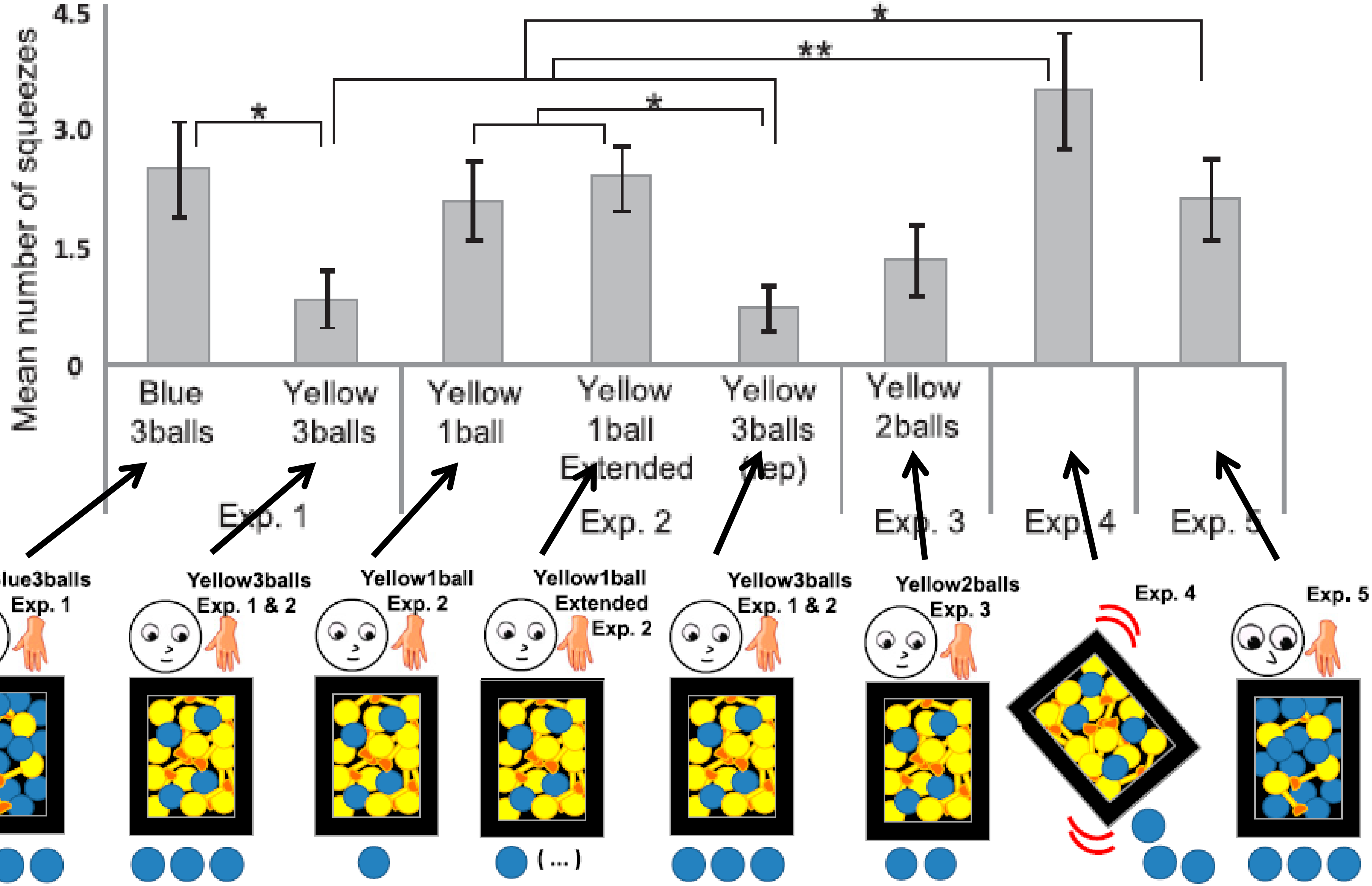
modèle:



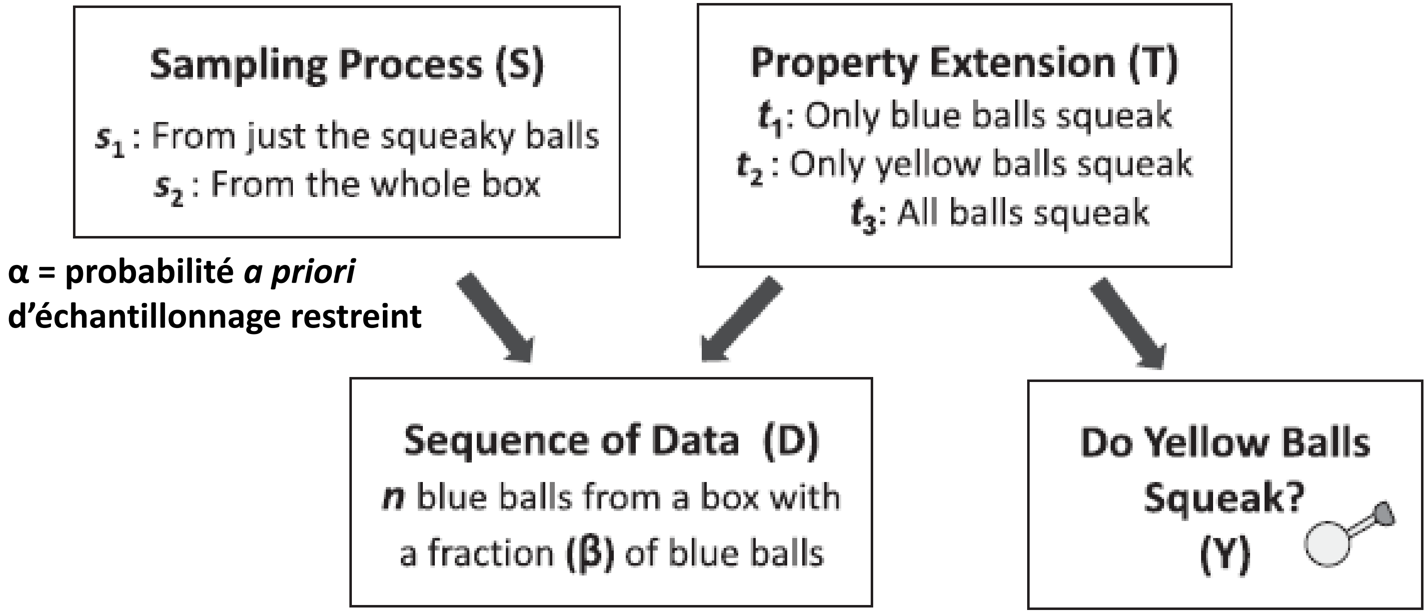
données:



Les enfants de 13 à 18 mois généralisent sur la base d'une solide connaissance des processus d'échantillonnage



Conclusion



Dans la seconde année de vie, les enfants seraient capables de
- formuler des **modèles génératifs complexes** pour expliquer ce qu'ils observent
- tenir compte, dans ces modèles, des **intentions** des personnes qui agissent autour d'elles
- utiliser des inférences bayésiennes pour remonter aux **causes cachées de leurs actions**.

Oui mais…
-d'où vient le modèle causal particulier qu'utilise l'enfant?
- ne faut-il pas supposer que la notion de causalité elle-même est innée? Ou bien, peut-elle apprise?

Apprendre une théorie de la causalité

Goodman, N. D., Ullman, T. D., & Tenenbaum, J. B. (2011). Learning a theory of causality. *Psychol Rev, 118(1), 110-119.*

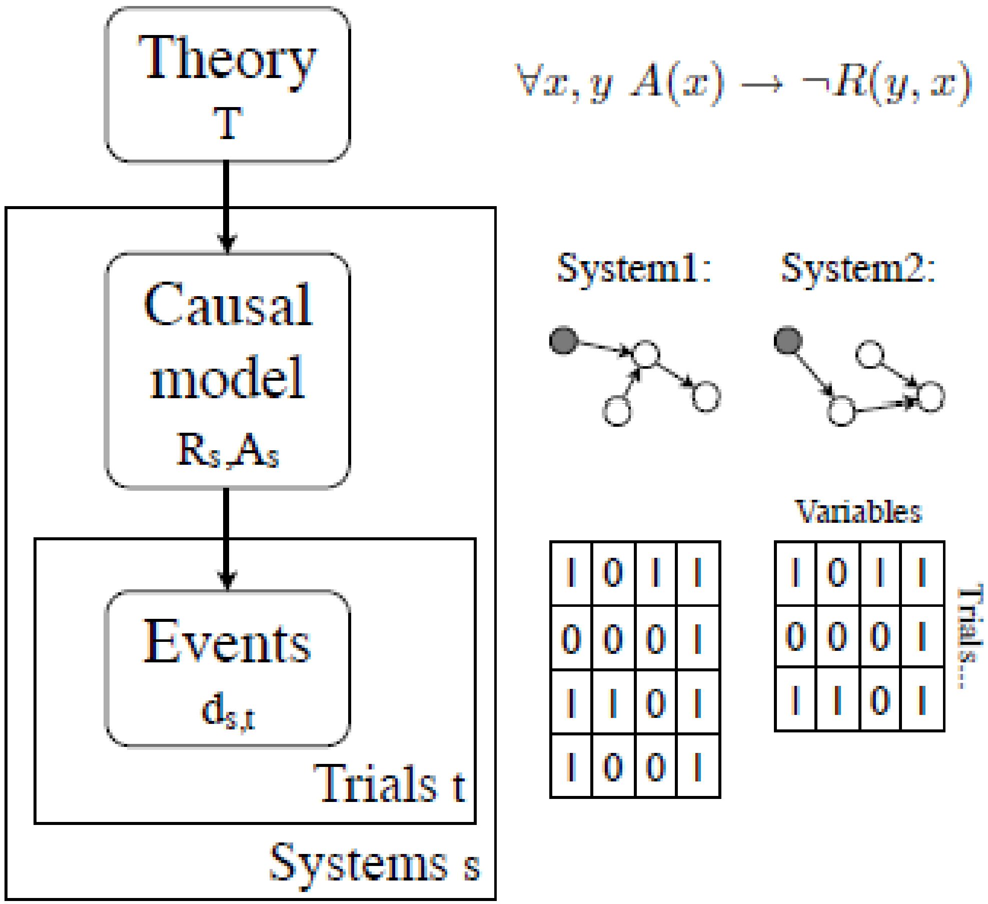
Law #1:	$\forall x\, \forall y\, A(x) \rightarrow \neg R(y,x)$	Interventions are exogenous.
Law #8:	$\forall x\, \exists y\, R(y,x)$	Variables have at most one parent.
Law #9:	$\forall x\, \forall y\, \forall z\, R(x,y) \wedge R(y,z) \rightarrow R(x,z)$	Dependence graph is transitive.

Il est possible de concevoir un système d'axiomes ('langage de la pensée') dans lequel on peut exprimer le principe même de causalité:

« Dans un graphe directionnel et acyclique liant diverses propriétés entre elles, il existe un sous-ensemble de propriétés dont toutes les autres dépendent. »

La théorie de la causalité n'est qu'une des théories relationnelles que ce langage peut exprimer.

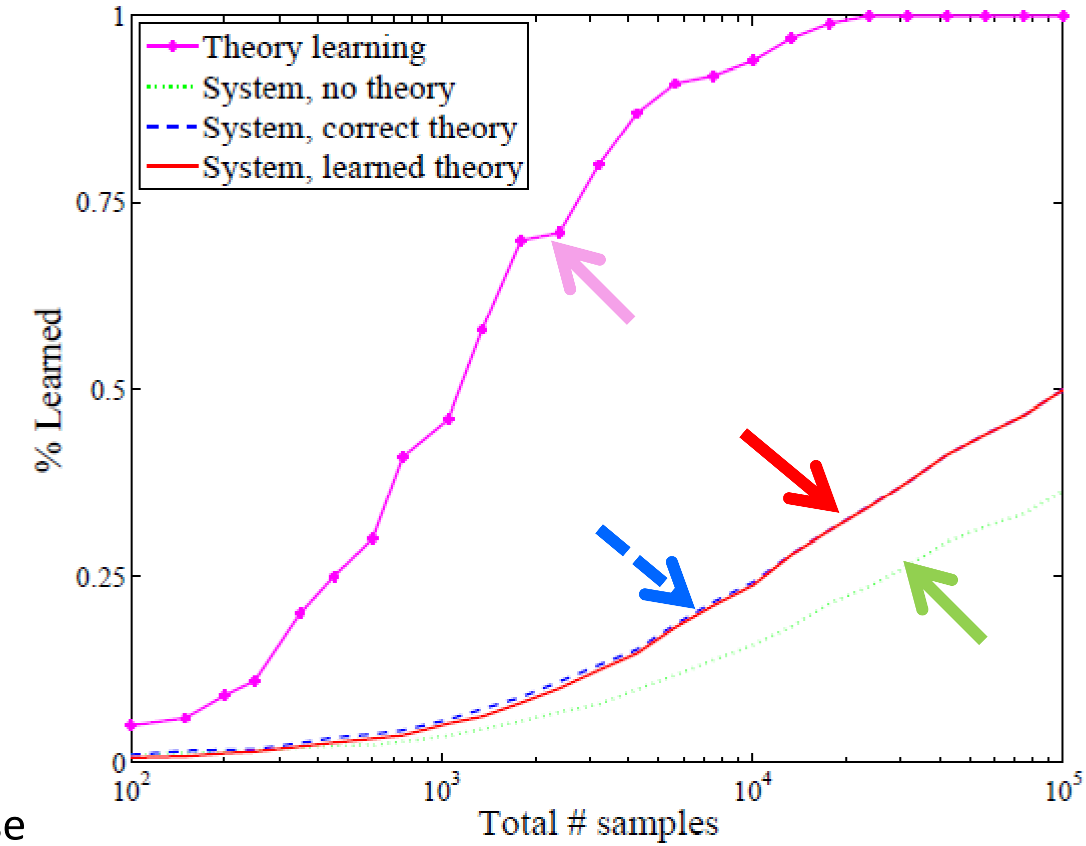
Goodman et al. (2011) simulent la capacité d'un tel système à converger rapidement vers le graphe qui décrit un ensemble d'entrées.



Apprendre une théorie de la causalité

Goodman, N. D., Ullman, T. D., & Tenenbaum, J. B. (2011). Learning a theory of causality. *Psychol Rev, 118(1), 110-119.*

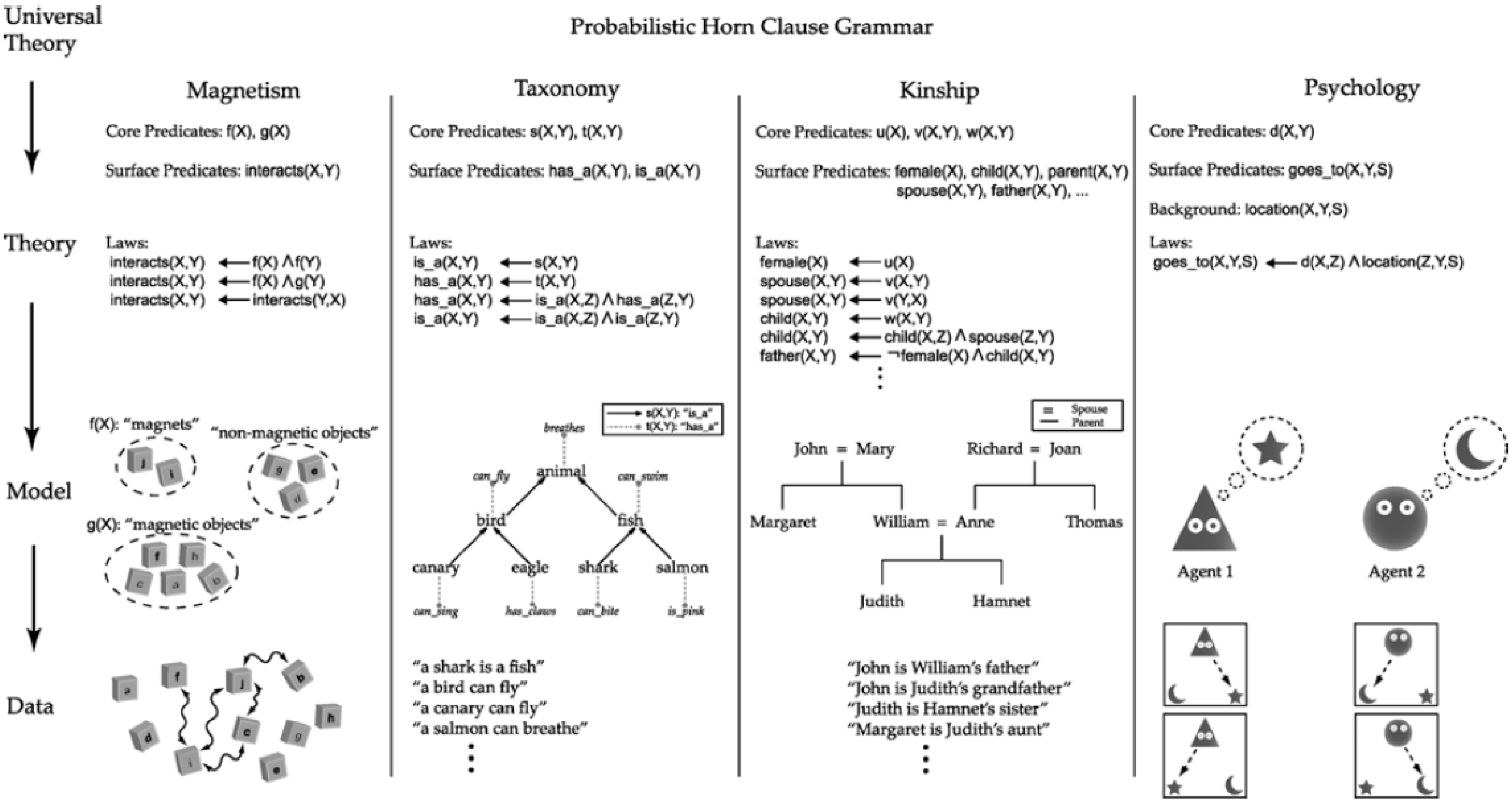
Les simulations montrent que
-le système converge très rapidement au plus haut niveau: les principes abstraits sont appris d'abord (*blessing of abstraction*)
-le modèle causal particulier converge ensuite
-et ce, beaucoup plus vite lorsqu'il est guidé par le principe approprié
- il n'est donc pas nécessaire de supposer que le principe de causalité soit « inné » -- l'apprentissage n'en est pas plus rapide.



Conclusion: le modèle bayésien réalise
un compromis entre nativisme et constructivisme/empirisme (***minimal nativism***):
-il ne requiert pas de présupposer des connaissances très particulières, propres à chaque domaine du savoir (contra Chomsky)
- mais son puissant algorithme d'apprentissage présuppose une compétence **statistique** (nombres, probabilités, inférences) et **expressive** (formules logiques, récursivité…) (contra Piaget)

Comment l'enfant construit-il des théories du monde?

Ullman, T. D., Goodman, N. D., & Tenenbaum, A. (2012). Theory learning as stochastic search in the language of thought. *Cognitive Development, 27, 455-480.*



Comment l'enfant construit-il des théories du monde?

Ullman, T. D., Goodman, N. D., & Tenenbaum, A. (2012). Theory learning as stochastic search in the language of thought. *Cognitive Development, 27, 455-480.*

Ullman et al. (2012) simulent l'apprentissage d'une théorie naïve du magnétisme chez l'enfant. Implémentation: *Monte-Carlo Markov Chain (MCMC)* combinée à un algorithme de recruit simulé.

Theory A

Rule 1: interacts(X,Y) ⟵ f(X) ∧ f(Y)

Theory B

Rule 1: interacts(X,Y) ⟵ f(X) ∧ g(Y)

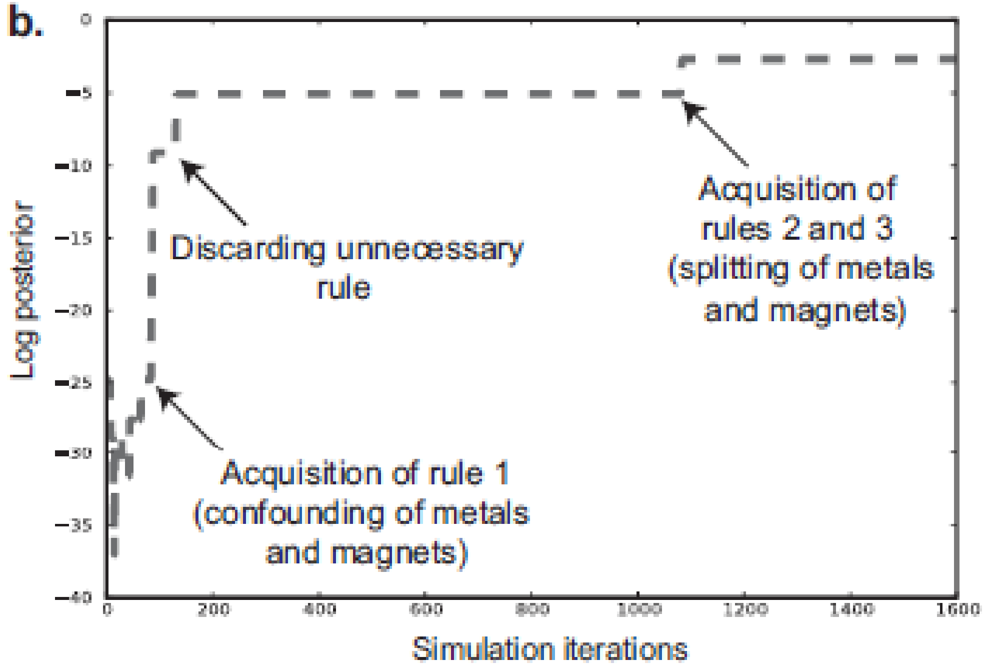
Rule 2: interacts(X,Y) ⟵ interacts(Y,X)

Theory C

Rule 1: interacts(X,Y) ⟵ f(X) ∧ f(Y)

Rule 2: interacts(X,Y) ⟵ f(X) ∧ g(Y)

Rule 3: interacts(X,Y) ⟵ interacts(Y,X)



Le réseau découvre progressivement une théorie raisonnable:

-il existe une propriété de certains objets (« aimants »)

- il existe une autre propriété (« métaux »)

- les aimants interagissent entre eux et avec les métaux. etc…


Allégorie de la nature forgeant un enfant

Conclusion: l'hypothèse du « bébé statisticien »

Selon le modèle bayésien, l'enfant serait capable, très tôt:

- de se représenter des distributions de probabilité
- de mettre à jour ces probabilités, en appliquant la règle de Bayes
- et ce, à de multiples niveaux hiérarchiques
- de représenter mentalement de très vastes espaces d'hypothèses, hiérarchisées, chacune avec sa probabilité a priori et sa vraisemblance
- d'évaluer simultanément la plausibilité de plusieurs hypothèses:
  - non seulement pour des hypothèses « statiques »
  - mais également pour des hypothèses « dynamiques » : des programmes récursifs en nombre potentiellement infini
- de les utiliser pour générer des prédictions et les comparer aux données reçues du monde extérieur (erreur de prédiction, signal de surprise)

Bibliographie succincte

Un livre accessible: Gopnik, A. (2000). *The Scientist in the Crib: What Early Learning Tells Us About the Mind New York: William Morrow. (traduction française: Comment pensent les bébés)*

Goodman, N. D., Ullman, T. D., & Tenenbaum, J. B. (2011). Learning a theory of causality. *Psychol Rev, 118(1), 110-119.*

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: causal maps and Bayes nets. *Psychol Rev, 111(1), 3-32.*

Gopnik, A. (2012). Scientific thinking in young children: theoretical advances, empirical research, and policy implications. *Science, 337(6102), 1623-1627.*

Gweon, H., & Schulz, L. (2011). 16-month-olds rationally infer causes of failed actions. Science, 332(6037), 1524.

Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. Proc Natl Acad Sci U S A, 107(20), 9066-9071.

Perfors, A., Tenenbaum, J. B., Griffiths, T. L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition, 120(3), 302-321.*

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: statistics, structure, and abstraction. *Science, 331(6022), 1279-1285.*

Ullman, T. D., Goodman, N. D., & Tenenbaum, A. (2012). Theory learning as stochastic search in the language of thought. *Cognitive Development, 27, 455-480.*

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proc Natl Acad Sci U S A, 105(13), 5012-5015.*

Xu, F., & Denison, S. (2009). Statistical inference and sensitivity to sampling in 11-month-old infants. *Cognition, 112(1), 97-104.*