

M2R Philosophie

Grenoble, 11/09/12

Des collaborations possibles entre philosophie et Intelligence Artificielle

Robin Lamarche-Perrin

Directeur **Denis Perrin**

Examineur **Denis Vernant**

UNIVERSITÉ PIERRE-MENDÈS-FRANCE

Stratégie générale du mémoire

Problème général
de l'intelligence artificielle

*Problèmes
scientifiques*

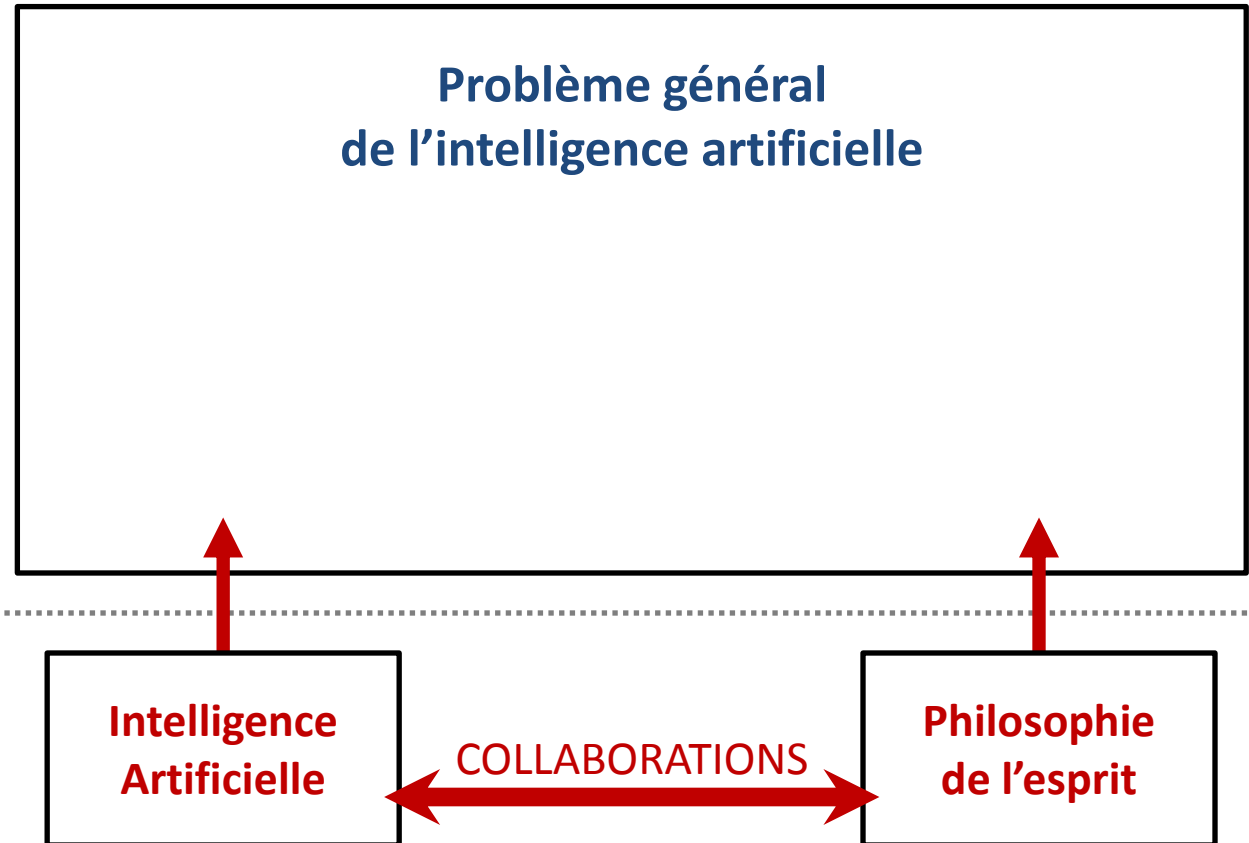
Disciplines

Intelligence
Artificielle

COLLABORATIONS

Philosophie
de l'esprit

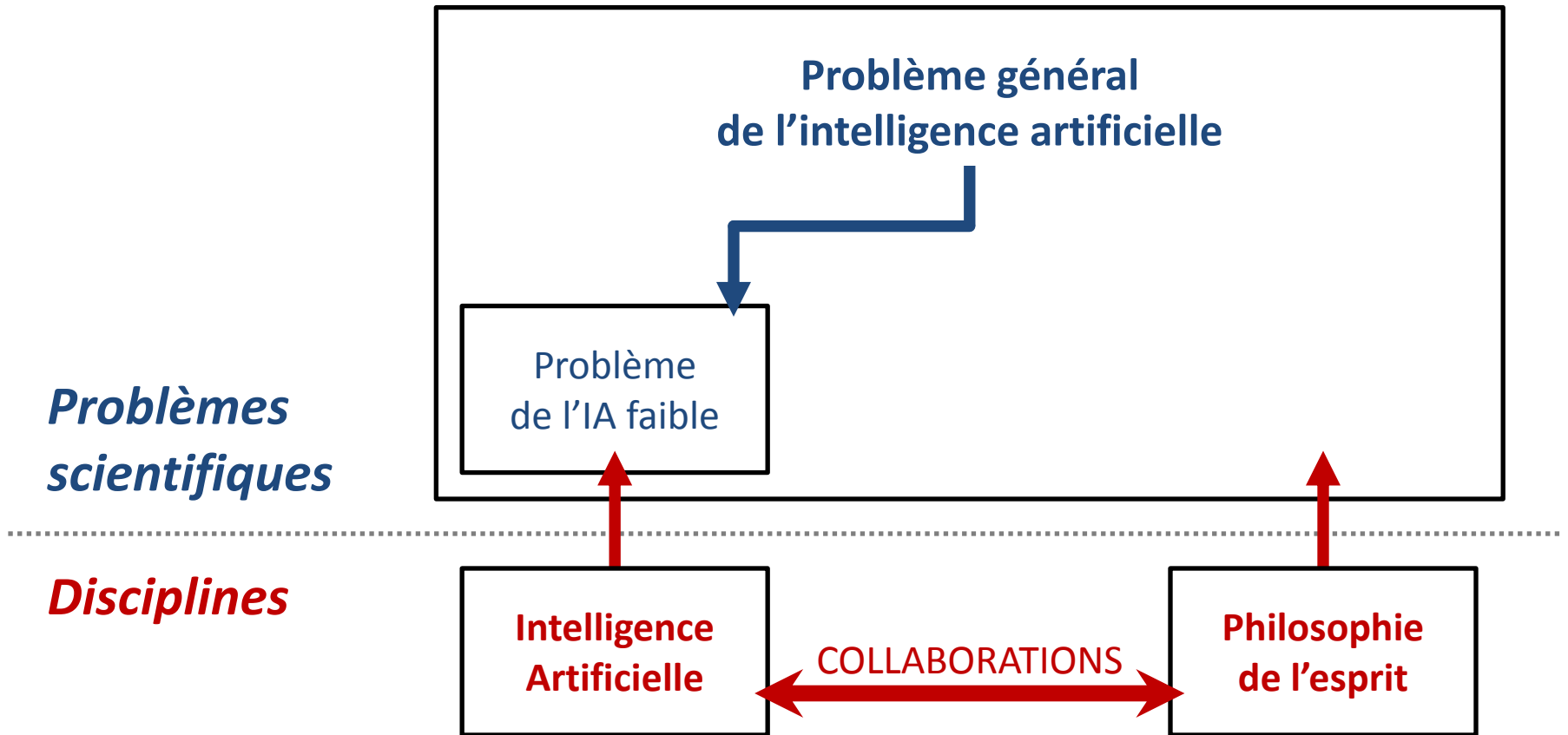
Stratégie générale du mémoire



*Problèmes
scientifiques*

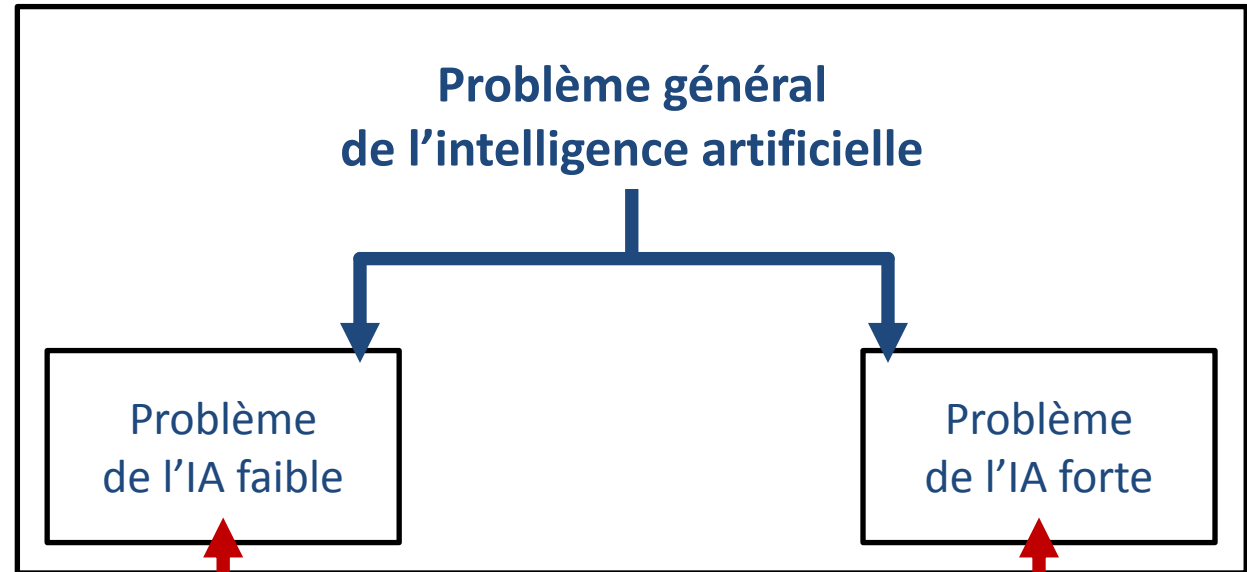
Disciplines

Stratégie générale du mémoire



Stratégie générale du mémoire

*Problèmes
scientifiques*

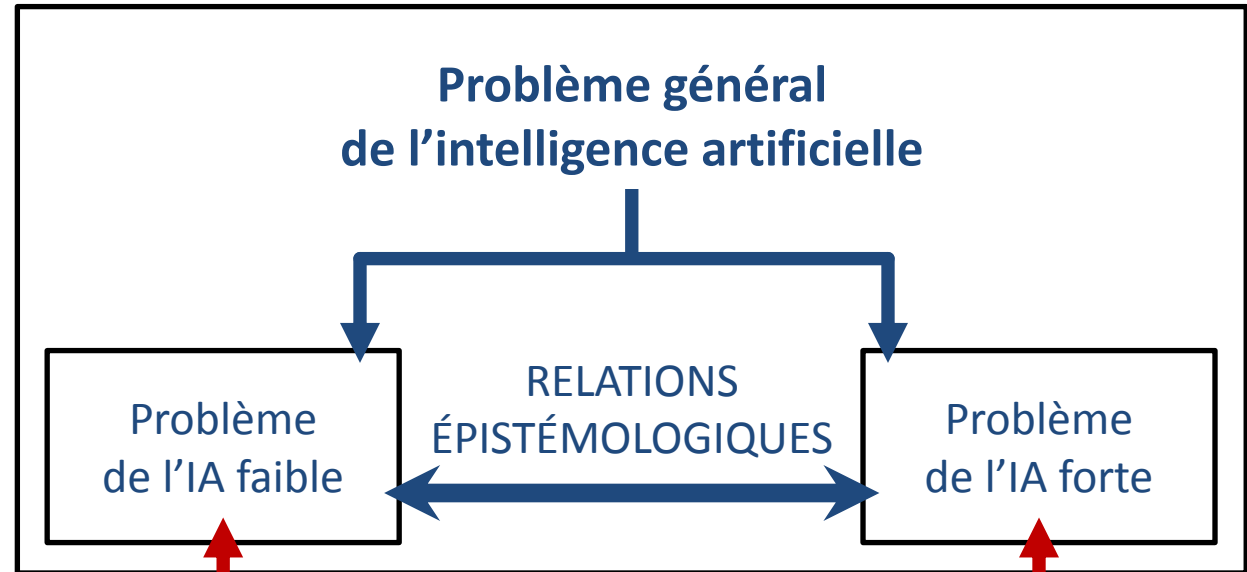


Disciplines



Stratégie générale du mémoire

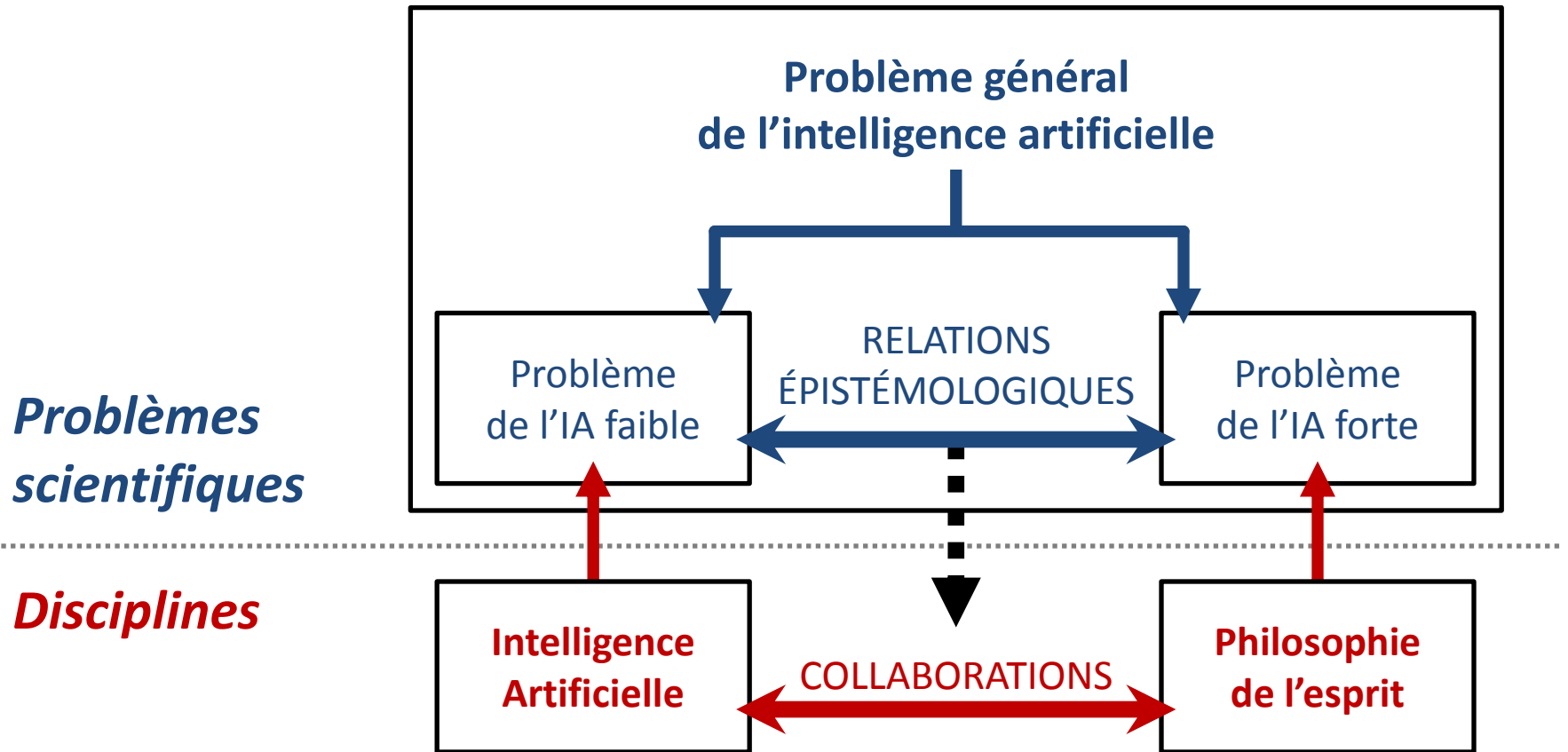
*Problèmes
scientifiques*



Disciplines



Stratégie générale du mémoire



Contenu de la soutenance

Relations



Collaborations

non IA forte → non IA faible
(Dreyfus)

IA forte → IA faible
(« Nouvelle IA »)

Application du concept
d'« émergence épistémique »
à la simulation de comportements
complexes

**La philosophie
au service de l'IA**
[Lamarche-Perrin, 2011]

IA faible → IA forte
(Levesque, van Gelder)

non IA faible → non IA forte
(Andler, Harvey)

Validation ou falsification des
modèles de la cognitions à partir de
machines concrètes

**L'IA au service de
la philosophie**
[Lamarche-Perrin, 2010]

LA PHILOSOPHIE AU SERVICE DE L'INTELLIGENCE ARTIFICIELLE

La critique du computationnalisme

- *What Computers Can't Do* [Dreyfus, 1979]
 - « non IA faible (computationnalisme) »
 - « non IA forte (computationnalisme) »
 - « non IA forte → non IA faible »

- Collaboration « négative »
 - La philosophie met en évidence les erreurs fondamentales de l'Intelligence Artificielle

La « nouvelle IA »

- Explosion paradigmatique
 - « IA forte → IA faible »
 - « IA forte (phénoménologie) »
 - « IA faible (phénoménologie) »

- Collaboration « positive »
 - La philosophie propose de nouveaux concepts fondamentaux à l'Intelligence Artificielle

Le concept d'émergence en IA

[Lamarche-Perrin, 2011]

- Relations épistémologiques
 - « non IA forte → non IA faible »
 - « IA forte → IA faible »
- Objectifs techniques
 - Simulation de systèmes complexes
 - Résolution de problèmes distribués
- Stratégie collaborative
 - Le concept d'émergence en philosophie
 - Analogie applicative
 - Le concept d'émergence en Intelligence Artificielle

Des positions métaphysiques

[O'Connor & Wong, 2006]

		Épistémologie	
		Éliminativisme Une seule science fondamentale	Non-éliminativisme Plusieurs sciences fondamentales
Ontologie	Dualisme Plusieurs substances		Vitalisme Dualisme non-éliminativiste
	Monisme Une seule substance	Mécanisme Monisme éliminativiste	Émergentisme Monisme non-éliminativiste

Des positions métaphysiques

[O'Connor & Wong, 2006]

		Épistémologie	
		Éliminativisme Une seule science fondamentale	Non-éliminativisme Plusieurs sciences fondamentales
Ontologie	Dualisme Plusieurs substances		 Vitalisme Dualisme non-éliminativiste « non IA forte (vitalisme) »
	Monisme Une seule substance	 Mécanisme Monisme éliminativiste « non IA forte (mécanisme) » 	Émergentisme Monisme non-éliminativiste « IA forte (émergentisme) »

Des méthodologies techniques

		Épistémologie	
		Éliminativisme Un seul niveau de description	Non-éliminativisme Plusieurs niveaux de description
Ontologie	Dualisme Plusieurs niveaux de modélisation		<p>SMA multi-niveau [Gil-Quijano <i>et al.</i>, 2010]</p> <p>« Blackboard systems » [Sawyer, 2001]</p> <p>Dualisme non-éliminativisme</p>
	Monisme Un seul niveau de modélisation	<p>Émergence « computationnelle » [Darley, 1994] [Bedau, 1997]</p> <p>Monisme éliminativiste</p>	<p>Détection hiérarchique [Bonabeau & Dessalles, 1997]</p> <p>Grammaires formelles [Kubík, 2003]</p> <p>Observation macroscopique [Lamarche-Perrin <i>et al.</i>, 2011]</p> <p>Monisme non-éliminativite</p>

Des méthodologies techniques

		Épistémologie	
		Éliminativisme Un seul niveau de description	Non-éliminativisme Plusieurs niveaux de description
Ontologie	Dualisme Plusieurs niveaux de modélisation	<p>« non IA forte → non IA faible »</p>	<p>SMA multi-niveau [Gil-Quiriano et al., 2010]</p> <p>« Blackboard systems » [Sageyer, 2001]</p> <p>Dualisme non-éliminativisme</p> <p>« non IA faible (x) »</p>
	Monisme Un seul niveau de modélisation	<p>Émergence « computationnelle » [Darbellay, 1994] [Benedict, 1997]</p> <p>Monisme éliminativiste</p> <p>« non IA faible (x) »</p>	<p>Détection hiérarchique [Bonabeau & Dessalles, 1997]</p> <p>Grammaires formelles [Kubík, 2003]</p> <p>Observation macroscopique [Lamarche-Perrin et al., 2011]</p> <p>Monisme non-éliminativite</p>

Des méthodologies techniques

		Épistémologie	
		Éliminativisme Un seul niveau de description	Non-éliminativisme Plusieurs niveaux de description
Ontologie	Dualisme Plusieurs niveaux de modélisation	<p>« non IA forte → non IA faible »</p> <p>« IA forte → IA faible »</p> <p>« IA forte (x) »</p>	<p>SMA multi-niveau [Gil-Quiriano et al., 2010]</p> <p>« Blackboard systems » [Sanger, 2001]</p> <p>Dualisme non-éliminativisme</p> <p>« non IA faible (x) »</p>
	Monisme Un seul niveau de modélisation	<p>Émergence « computationnelle » [Darle, 1994] [Bened, 1997]</p> <p>Monisme éliminativiste</p> <p>« non IA faible (x) »</p>	<p>Détection hiérarchique [Bonabeau & Dessalles, 1997]</p> <p>Grammaires formelles [Kubík, 2003]</p> <p>Observation macroscopique [Lamarche-Perrin et al., 2011]</p> <p>Monisme non-éliminativite</p>

Les concepts importés en IA

- **Principe de parcimonie**
À capacités égales, on doit préférer un modèle simple à un modèle compliqué.
- **Épiphénoménisme**
Les phénomènes émergents ne sont pas des entités indépendantes, mais des aspects particuliers du niveau microscopique.
- **Approche ascendante**
Les phénomènes émergents sont produits par le niveau microscopique. Méthodologiquement, il faut donc partir de ce niveau et non l'inverse.
- **Subjectivisme**
Les phénomènes émergents dépendent d'un mode de connaissance donné. Ils sont « dans l'œil de l'observateur. »
- **Pragmatisme**
Les phénomènes émergents n'existent pas en soi. Ils sont des abstractions créées par le scientifique et doivent donc être évalués en fonction du contexte général de l'analyse.

**L'INTELLIGENCE ARTIFICIELLE
AU SERVICE DE LA PHILOSOPHIE**

Évaluer les modèles cognitifs

[Lamarche-Perrin, 2010]

- « IA forte \rightarrow IA faible »
- « IA faible \rightarrow IA forte »

		Test comportemental	
		« IA faible (x) »	« non IA faible (x) »
Modèle cognitif	« IA forte (x) »	Vrai positif	Faux négatif
	« non IA forte (x) »	Faux positif	Vrai négatif

Évaluer les modèles cognitifs

[Lamarche-Perrin, 2010]

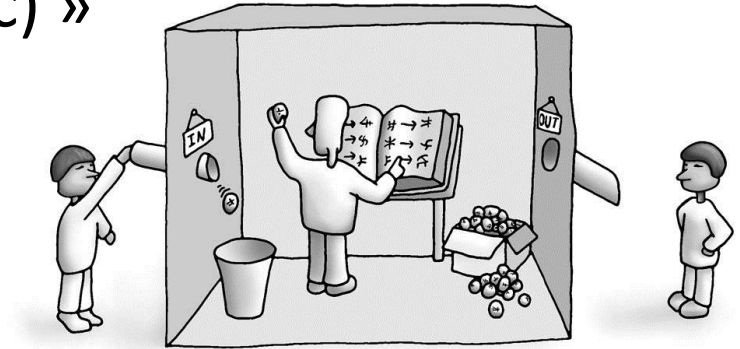
- « IA forte \rightarrow IA faible »
- « IA faible \rightarrow IA forte »

		Test comportemental	
		« IA faible (x) »	« non IA faible (x) »
Modèle cognitif	« IA forte (x) »	Vrai positif	Faux négatif
	« non IA forte (x) »	Faux positif En contradiction avec « IA faible \rightarrow IA forte »	Vrai négatif

La « chambre chinoise »

[Searle, 1980]

- « IA faible (cc) et non IA forte (cc) »
- Trois conséquences possibles
 1. « non (IA faible \rightarrow IA forte) »
 2. « non IA faible (cc) »
 3. « IA forte (cc) »
- « IA faible \rightarrow IA forte »
 - La nécessité en pratique de l'IA forte
 - De nouvelles hypothèses des travail



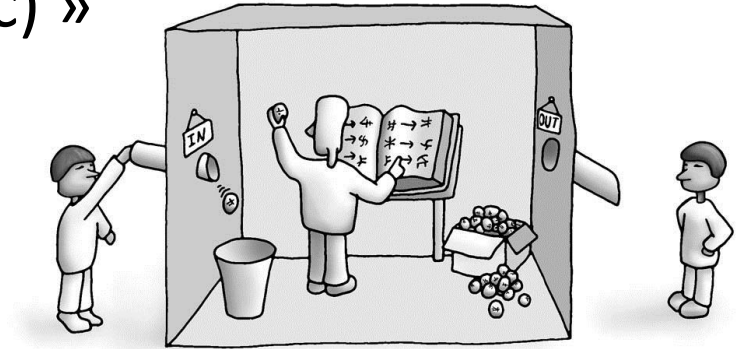
[Levesque, 2009]

[van Gelder, 1995]

La « chambre chinoise »

[Searle, 1980]

- « IA faible (cc) et non IA forte (cc) »
- Trois conséquences possibles
 1. « non (IA faible \rightarrow IA forte) »
 2. « non IA faible (cc) »
 3. « IA forte (cc) »



- « IA faible \rightarrow IA forte »
 - La nécessité en pratique de l'IA forte
 - De nouvelles hypothèses des travail

[Levesque, 2009]

[van Gelder, 1995]

\rightarrow Validation du computationnalisme

Évaluer les modèles cognitifs

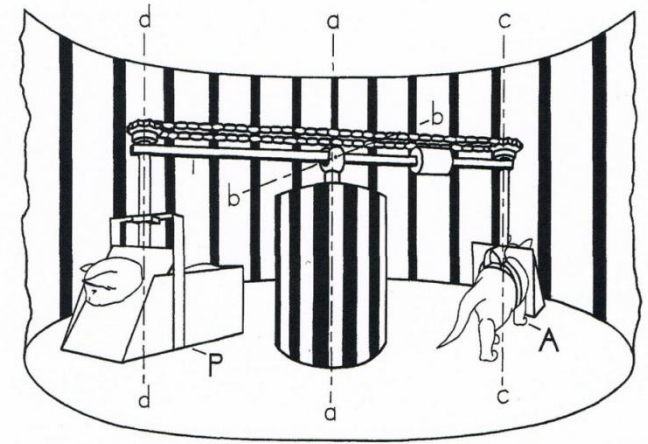
[Lamarche-Perrin, 2010]

- « IA forte \rightarrow IA faible »
- « IA faible \rightarrow IA forte »

		Test comportemental	
		« IA faible (x) »	« non IA faible (x) »
Modèle cognitif	« IA forte (x) »	Vrai positif	Faux négatif En contradiction avec « IA forte \rightarrow IA faible »
	« non IA forte (x) »	Faux positif	Vrai négatif

Les « chatons aveugles » [Held & Hein, 1958]

- « non IA faible (ca) et IA forte (ca) »
- Trois conséquences possibles
 1. « non (IA forte \rightarrow IA faible) »
 2. « IA faible (ca) »
 3. « non IA forte (ca) »
- « non IA faible \rightarrow non IA forte »
 - L'échec pratique du cartésianisme
 - La falsification des théories

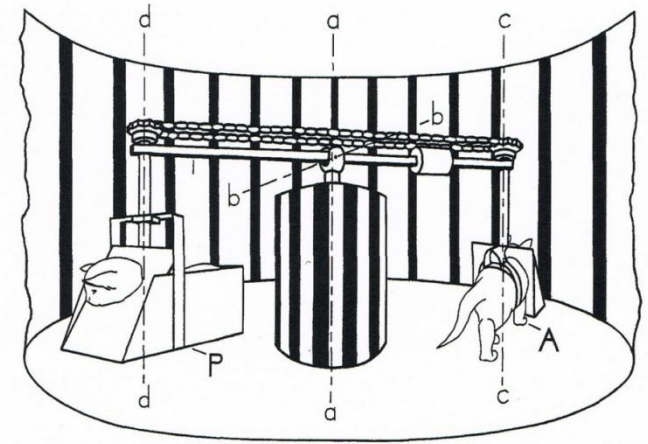


[Dreyfus, 1979]

[Andler, 1984] [Harvey, 2000]

Les « chatons aveugles » [Held & Hein, 1958]

- « non IA faible (ca) et IA forte (ca) »
- Trois conséquences possibles
 1. « non (IA forte → IA faible) »
 2. « IA faible (ca) »
 3. « **non IA forte (ca)** »
- « non IA faible → non IA forte »
 - L'échec pratique du cartésianisme
 - La falsification des théories



[Dreyfus, 1979]

[Andler, 1984] [Harvey, 2000]

→ **Falsification des modèles classiques de la cognition**

CONCLUSION

Thèse défendue

- Il existe des relations épistémologiques entre le problème de l'IA faible et celui de l'IA forte.
- Ces relations permettent de définir des stratégies collaboratives d'inspiration ou de falsification.
- Pour arguments : des exemples de collaborations, dans le passé et aujourd'hui.

Bilan des collaborations

	Démarche positive	Démarche négative
La philosophie au service de l'IA	« IA forte → IA faible »	« non IA forte → non IA faible »
L'IA au service de la philosophie	« IA faible → IA forte »	« non IA faible → non IA forte »

Contraposées

Réciproques

Réciproques

Merci pour votre attention

- Dreyfus, H.L. 1979. *Intelligence Artificielle : mythes et limites*. [What Computers Can't Do: The Limits of Artificial Intelligence, 2nd ed.] Vassallo-Villaneau, R.-M. (trad.), Andler, D. (pref.), Perriault, J. (pref.). Paris : Flammarion, 1984.
- Harvey, I. 2000. « Robotics: Philosophy of Mind Using a Screwdriver. » *Evolutionary Robotics: From Intelligent Robots to Artificial Life*, vol. III, p. 207-230.
- Held, R., Hein, A. 1958. « Adaptation of disarranged hand-eye coordination contingent upon re-afferent simulation. » *Perceptual-Motor Skills*, vol. 8, p. 87-90.
- Lamarche-Perrin, R. 2010. *Le Test de Turing pour évaluer les théories de l'esprit*. Mémoire de Master, Kistler, M. (dir.). Grenoble : Université Pierre-Mendès-France, sept. 2010.
- Lamarche-Perrin, R. 2011. « Conceptualisation de l'émergence : dynamiques microscopiques et analyse macroscopique des SMA. » *Atelier Futur des Agents et des Multi-Agents (FUTURAMA'11)*. Chambéry : Plateforme AFIA 2011, mai 2011.
- Levesque, H.J. 2009. « Is It Enough to Get the Behavior Right? » *International Joint Conference on Artificial Intelligence (IJCAI'09)*, p. 1439-1444.
- O'Connor, T., Wong, H.Y. 2006. « Emergent Properties. » *In Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/properties-emergent/>, mis en ligne le 24 sept. 2002, révisé le 23 oct. 2006, consulté le 1 mai 2011.
- Searle, J.R. 1980. « Minds, Brains, and Programs. » *In* Boden, M.A. (éd.). 1990. *The Philosophy of Artificial Intelligence*. Oxford University Press, p. 67-88.
- van Gelder, T. 1995. « What might Cognition be if not Computation? » *Journal of Philosophy*, vol. 92, n°7, p. 345-381.