



HAL
open science

Modeling cognitive processes within a creative problem-solving task: from symbolic to neuro-symbolic approaches in computational learning sciences

Chloé Mercier

► **To cite this version:**

Chloé Mercier. Modeling cognitive processes within a creative problem-solving task: from symbolic to neuro-symbolic approaches in computational learning sciences. Technology for Human Learning. Université de Bordeaux, 2024. English. NNT : 2024BORD0065 . tel-04678913

HAL Id: tel-04678913

<https://theses.hal.science/tel-04678913v1>

Submitted on 27 Aug 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

THÈSE

présentée à

L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE DE MATHÉMATIQUES ET D'INFORMATIQUE

par **Chloé MERCIER**

en vue de l'obtention du

DOCTORAT EN INFORMATIQUE

**MODÉLISER LES PROCESSUS COGNITIFS DANS UNE TÂCHE DE RÉOLUTION CRÉATIVE
DE PROBLÈME : DES APPROCHES SYMBOLIQUES À NEURO-SYMBOLIQUES EN SCIENCES
COMPUTATIONNELLES DE L'ÉDUCATION**

*MODELING COGNITIVE PROCESSES WITHIN A CREATIVE PROBLEM-SOLVING TASK: FROM
SYMBOLIC TO NEURO-SYMBOLIC APPROACHES IN COMPUTATIONAL LEARNING SCIENCES*

Soutenue le 16 AVRIL 2024.

Sous la direction de **Mme Margarida ROMERO** et **M. Thierry VIÉVILLE**

Devant la commission d'examen composée de :

Mme. Armelle BRUN <i>Présidente du Jury</i>	Professeure des Universités, LORIA, Université de Lorraine.
M. Patrick GEBHARD <i>Rapporteur</i>	Directeur de Recherche, Affective Computing Group, DFKI Saarbrücken.
M. Terrence C. STEWART <i>Examineur</i>	Associate Research Officer, National Research Council of Canada.
M. André TRICOT <i>Examineur</i>	Professeur des Universités, Université Paul-Valéry, Montpellier III.
Mme. Emmanuelle VOLLE <i>Examinatrice</i>	Chargée de Recherche, FrontLab, ICM, AP-HP, Université PSL.
Mme. Margarida ROMERO <i>Directrice</i>	Professeure des Universités, LINE, Université Côte d'Azur.
M. Thierry VIÉVILLE <i>Directeur</i>	Directeur de Recherche, Mnémosyne, Centre Inria de l'Université de Bordeaux.
M. Frédéric ALEXANDRE <i>Invité (co-encadrant)</i>	Directeur de Recherche, Mnémosyne, Centre Inria de l'Université de Bordeaux.

MODÉLISER LES PROCESSUS COGNITIFS DANS UNE TÂCHE DE RÉOLUTION
CRÉATIVE DE PROBLÈME : DES APPROCHES SYMBOLIQUES À
NEURO-SYMBOLIQUES EN SCIENCES COMPUTATIONNELLES DE L'ÉDUCATION

RÉSUMÉ L'intégration de compétences transversales telles que la créativité, la résolution de problèmes et la pensée informatique, dans les programmes d'enseignement primaire et secondaire, est un défi majeur dans le domaine de l'éducation aujourd'hui. Nous postulons que l'enseignement et l'évaluation de ces compétences transversales pourraient bénéficier d'une meilleure compréhension des comportements des apprenants dans des activités spécifiques qui requièrent ces compétences. A cette fin, les sciences computationnelles de l'apprentissage (*computational learning sciences*) sont un champ en émergence qui requiert l'étroite collaboration des neurosciences computationnelles et des sciences de l'éducation pour permettre l'évaluation des processus d'apprentissage. Nous nous concentrons sur une tâche de résolution créative de problème dans laquelle le sujet est amené à construire un "véhicule" en combinant des cubes robotiques modulaires. Dans le cadre d'une action de recherche exploratoire, nous proposons plusieurs approches s'appuyant sur des formalismes symboliques à neuro-symboliques, afin de spécifier une telle tâche et de modéliser les comportements et processus cognitifs sous-jacents d'un sujet engagé dans cette tâche. Bien qu'étant à un stade très préliminaire, une telle formalisation semble prometteuse pour mieux comprendre les mécanismes complexes impliqués dans la résolution créative de problèmes à plusieurs niveaux : (i) la spécification du problème et les observables d'intérêt à collecter pendant la tâche ; (ii) la représentation cognitive de l'espace-problème, en fonction des connaissances préalables et de la découverte des affordances, permettant de générer des trajectoires-solutions créatives ; (iii) une implémentation du raisonnement par inférence au sein d'un substrat neuronal.

MOTS-CLÉS Sciences computationnelles de l'éducation, Résolution créative de problème, Neurosciences cognitives, Raisonnement et représentation des connaissances, Apprentissage par renforcement, Architectures à vecteurs symboliques.

LABORATOIRE D'ACCUEIL

Centre Inria
de l'Université de
Bordeaux
Équipe-projet
Mnémosyne
200 Avenue de la Vieille
Tour
33405 Talence

Laboratoire Bordelais de
Recherche en
Informatique (LaBRI)
Univ. Bordeaux & CNRS
UMR 5800
351 cours de la Libération
33405 Talence

Institut des Maladies
Neurodégénératives
(IMN)
Univ. Bordeaux & CNRS
UMR 5293
Bordeaux Neurocampus
146 rue Léo Saignat
33076 Bordeaux Cedex

MODELING COGNITIVE PROCESSES WITHIN A CREATIVE PROBLEM-SOLVING TASK: FROM SYMBOLIC TO NEURO-SYMBOLIC APPROACHES IN COMPUTATIONAL LEARNING SCIENCES

ABSTRACT Integrating transversal skills such as creativity, problem solving and computational thinking, into the primary and secondary curricula is a key challenge in today's educational field. We postulate that teaching and assessing transversal competencies could benefit from a better understanding of the learners' behaviors in specific activities that require these competencies. To this end, computational learning science is an emerging field that requires the close collaboration of computational neuroscience and educational sciences to enable the assessment of learning processes. We focus on a creative problem-solving task in which the subject is engaged into building a "vehicle" by combining modular robotic cubes. As part of an exploratory research action, we propose several approaches based on symbolic to neuro-symbolic formalisms, in order to specify such a task and model the behavior and underlying cognitive processes of a subject engaged in this task. Despite being at a very preliminary stage, such a formalization seems promising to better understand complex mechanisms involved in creative problem solving at several levels: (i) the specification of the problem and the observables of interest to collect during the task; (ii) the cognitive representation of the problem space, depending on prior knowledge and affordance discovery, allowing to generate creative solution trajectories; (iii) an implementation of reasoning mechanisms within a neuronal substrate.

KEYWORDS Computational Learning Sciences, Creative Problem Solving, Cognitive Neuroscience, Knowledge Representation and Reasoning, Reinforcement Learning, Vector Symbolic Architectures

HOSTING LABORATORY

Centre Inria de l'Université de Bordeaux Équipe-projet Mnémosyne 200 Avenue de la Vieille Tour 33405 Talence	Laboratoire Bordelais de Recherche en Informatique (LaBRI) Univ. Bordeaux & CNRS UMR 5800 351 cours de la Libération 33405 Talence	Institut des Maladies Neurodégénératives (IMN) Univ. Bordeaux & CNRS UMR 5293 Bordeaux Neurocampus 146 rue Léo Saignat 33076 Bordeaux Cedex
---	--	--

ACKNOWLEDGEMENTS

I am deeply thankful for the guidance I received from my advisors, Thierry Viéville and Margarida Romero; and I want to extend this gratitude to Frédéric Alexandre who also greatly contributed to this supervision. Their expertise and dedication to combine such different fields were an invaluable source of inspiration and significantly influenced the development of my research. Beyond their outstanding academic mentorship, I am profoundly grateful for their continuous support and kindness.

I also wish to thank Lisa Roux who first introduced me to the Cubelets and her version of the CreaCube ontology developed during her post-doctoral studies, which served as a basis for my first year of PhD.

I want to acknowledge the hard work of several students that I co-supervised within internships or school projects: Lola Denet, Théophane Vallaeys, Gabriel Doriath-Döhler, Louis Kohler, Waris Radji, Corentin Léger, Lucas Bardisbanian, Paul Bernard, Benjamin Hate, Morgane Laval, Adrien Villars, Amandine Solignac, Vincent Piton, Luca Pereira, Benoît Miège, Cassandra Guillaud, Tanguy Cauquil. Regardless of whether they were closely or remotely related to this thesis, all of their contributions were a valuable and inspiring addition to my research, and a great opportunity for me to practice some mentoring skills.

A PhD journey might feel lonely at times, and I was very lucky to take it along some wonderful colleagues. First, I am profoundly thankful for the camaraderie of Axel Palaude, who started his PhD soon after I did on a very similar topic. It was such nice a nice experience having a trusted colleague by my side with whom I could exchange fruitful discussions, expeditions to Nice, and share the supervision of students. I also wish to thank Hugo Chateau Laurent for sharing with me his knowledge about the brain and the Nengo framework with a contagious enthusiasm. I am very grateful towards Fjola Hyseni and Eduarda Centeno for showing me the way to Open Science, let alone inviting me to help them organizing a workshop about it. Outside of the Mnemosyne team, I also wish to thank Arnaud Prouzeau and Juliette Le Meudec from the Potioc/Bivwac Inria team for making a reality of the virtual CreaCube experiment, opening new and exciting research perspectives. Finally, I am grateful to Cédric Brun for the insightful discussions about the social implications and limitations of so-called “neuro-education”. I am not forgetting all my other colleagues and friends from the Mnemosyne team, Inria and IMN – Maeva, Snigdha, Nathan, Xavier, Naomi, Ankur, Nikos... – for the many coffee breaks which provided interesting conversations as well as much-needed respite and moments of fun, fostering a supportive, caring and engaging academic environment.

On a more personal note, during the initial stages of my PhD, I faced a challenging health issue. I want to express my sincere appreciation to my doctor, Pr Laharie, for his medical rigor and expertise. I extend my gratitude to my nurses Maud, Laurette and Cathy who provided me with home healthcare for several months, thus enabling me to overcome this obstacle and continue my academic pursuits: as we say, *mens sana in corpore sano*.

Finally, I express my deepest feelings to my family and my partner Johann, for their unwavering patience, encouragement and understanding throughout this endeavor. A special thought to my sister Zelda, who will soon complete her MD thesis!

CONTENTS

A PROPOS	1
ABOUT	5
INTRODUCTION	8
1 CREATIVE PROBLEM SOLVING: A TRIDISCIPLINARY REVIEW	19
1.1 Creative problem solving: a complex phenomenon studied at multiple levels	19
1.1.1 A definition of creative problem solving	19
1.1.2 Multiple levels of analysis	20
1.1.3 Decomposing the CPS process	24
1.2 CPS in learning sciences: from creative pedagogy to pedagogy of creativity	27
1.2.1 From primary to secondary knowledge: goal setting for intrinsically motivated learning	27
1.2.2 Acquiring knowledge through learning activities	29
1.2.3 Assessing creativity in CPS activities	32
1.3 The neural basis of creativity	36
1.3.1 Memories for learning, recall and imagination	36
1.3.2 Neuronal networks in interaction during CPS	38
1.3.3 Cognitive control: from behavior persistence to flexibility	42
1.4 Computational approaches for CPS: from symbolic to neural formalisms	46
1.4.1 The information processing approach	46
1.4.2 Semantic grounding: representing and manipulating symbols	51
1.4.3 State generation mechanisms as a search in the problem space	54
1.4.4 State evaluation mechanisms as logical or probabilistic reasoning	56
1.4.5 Regulating mechanisms	59
1.5 Main take-aways	61
2 AN ONTOLOGY APPROACH: FORMALIZING PRIOR KNOWLEDGE AND MATERIAL CONFIGURATIONS OF THE CREA CUBE TASK	63
2.1 Introduction	63
2.2 The ontology approach	64
2.2.1 Terminology	64
2.2.2 Related work	64
2.2.3 Methodology	66

2.3	Ontology Implementation	67
2.3.1	Modeling the activity process	67
2.3.2	Modeling the material states	69
2.3.3	Modeling the learner’s behavior and mental states	69
2.4	Results and perspectives of application	71
2.4.1	Modeling as an interdisciplinary tool	71
2.4.2	Reasoning about the model	72
2.5	Discussion and conclusion	74
3	A MARKOVIAN APPROACH: MODELING CPS BEHAVIORS WITHIN A SYMBOLIC STATE SPACE	77
3.1	Introduction	77
3.2	Symbolic state space specification	78
3.2.1	General framework	78
3.2.2	Finding an optimal policy	80
3.2.3	Symbolic data structure	83
3.3	Reinforcement learning on a symbolic state space.	86
3.3.1	Q-Learning on symbolic data	86
3.3.2	Preliminary results	87
3.4	Discussion: perspectives of application to the CreaCube task	89
4	REASONING ON SYMBOLS: TOWARDS A BIOLOGICALLY PLAUSIBLE MODEL?	94
4.1	Introduction	95
4.1.1	Representing neuronal activity at a symbolic level	96
4.1.2	From sensorimotor processing to logical reasoning	97
4.2	Basic design choices	98
4.2.1	From symbols to numbers	98
4.2.2	Semantic similarity	99
4.2.3	Classes and approximate Boolean properties	100
4.2.4	Statements: facts and rules	101
4.3	A first, simple but partial, mapping	102
4.3.1	Ontology numerical mapping	102
4.3.2	Relationship and membership composition	105
4.3.3	Introducing data and data structure	109
4.3.4	Effective neuronal implementation: an illustrative example for the case of hierarchical rules	111
4.4	Discussion and conclusion	112
5	DISCUSSION AND CONCLUSION	116
5.1	Perspectives and limits for learning sciences and education	116
5.2	Perspectives for computer science: towards better explainable computational cognitive models?	119
5.3	General conclusion	121
	APPENDICES	122

Contents

A SCHOOLS OF THOUGHT IN LEARNING SCIENCES	123
B THE CREATIVE PROBLEM SOLVING LITERATURE: STAGE MODELS AND DUAL NOTIONS	125
C THE ANATOMY OF THE HUMAN BRAIN	129
D RDFS ENTAILMENT RULES	132
INDEX	135
REFERENCES	137

LIST OF FIGURES

A	The CreaCube Task	2
B	Identification of the CreaCube observables.	13
1.1	Multiple levels of analysis	23
1.2	Amabile’s componential model of creativity.	26
1.3	A typology of knowledge	30
1.4	Framework chosen for the analysis of the observables	31
1.5	The preparation phase in CPS.	33
1.6	Brain networks involved in CPS	39
1.7	Metacontrol of insight	43
1.8	Metacontrol between goal flexibility and goal persistence	44
1.9	Task switching	45
1.10	The problem space, illustrated on the Tower of Hanoi	48
1.11	Ill-defined problem-solving as trajectory generation.	50
1.12	The three levels of symbolic abstraction in the CreaCog framework.	52
1.13	The “vanilla” reinforcement learning paradigm	59
1.14	The Probe model for task sets	60
2.1	Diagram of the CreaCube ontology.	68
2.2	Example of a SPARQL query on the material states.	73
2.3	Example of SPARQL queries on the learner’s behavior.	74
3.1	The classical Reinforcement Learning paradigm	79
3.2	The Q-Learning paradigm	83
3.3	Ontology of the symbolic environment used by Radji et al. (2023).	87
3.4	Comparison of Q-Learning and Symbolic Q-Learning performances with varying ρ	89
3.5	Perspective: a reinforcement symbolic learning paradigm applied to the CreaCube task.	90
3.6	The Inverse Reinforcement Learning paradigm	91
3.7	Perspective: Hidden Markov Models applied to the CreaCube task.	92
3.8	An illustration of dynamic-time wrapping	92
4.1	Neural activity as manifold	97
4.2	Representation of partial truth $\tau \in [-1, 1]$, in relation to necessity and possibility.	100
4.3	A minimal example of ontology.	111
4.4	Results of a Nengo simulation performing inference on a minimal ontology.	113

List of Figures

4.5	A proposition of architecture to implement an ontology reasoner in a Nengo neuronal simulator.	114
A.1	Schools of thought in Learning Sciences.	124
B.1	A comparison of creative process models.	125
C.1	Human brain lobes.	129
C.2	Some vocabulary regarding the names of the brain regions.	129

LIST OF TABLES

- 3.1 Markovian frameworks and their use-cases 79
- B.1 The S1/S2 system. 126
- B.2 Other dual notions encountered in CPS. 127
- C.1 Anatomy of the human brain (1) 130
- C.2 Anatomy of the human brain (2) 131
- D.1 RDF/RDFS entailment rules. 133

A PROPOS

L'intégration de compétences transversales, telles que la créativité, la résolution de problèmes et la pensée informatique dans les programmes scolaires du primaire au secondaire est un défi majeur dans le domaine de l'éducation d'aujourd'hui. Nous postulons que l'enseignement et l'évaluation des compétences transversales pourraient bénéficier d'une meilleure compréhension des comportements des apprenants dans les activités spécifiques qui requièrent ces compétences. Dans le cadre d'une action de recherche exploratoire, nous proposons une approche tri-disciplinaire, s'appuyant sur les sciences de l'apprentissage, les neurosciences cognitives et l'informatique, pour formaliser une tâche créative de résolution de problème et le comportement d'un sujet engagé dans une telle tâche.

CONTEXTE

COMPÉTENCES TRANSVERSALES DANS L'ÉDUCATION Parmi les compétences dites "du 21e siècle" qui doivent être abordées dans l'enseignement du primaire au secondaire, la créativité et la résolution de problèmes sont deux compétences transversales qui sont étroitement liées et qui peuvent être étudiées conjointement (VOOGT et ROBLIN, 2012). Ceci implique la conception de nouvelles activités pédagogiques et l'évaluation de leur efficacité, y compris l'étude du comportement des apprenants dans ces activités. Nous ciblons ici une tâche de résolution de problème, présenté sous la forme d'un problème ouvert avec une consigne ambiguë et accompagné de matériel inconnu de l'apprenant. Cette tâche demande ainsi aux apprenants de s'engager dans une démarche créative pour générer des idées nouvelles et de les mettre en œuvre pour réussir à résoudre le problème. En outre, le matériel associé à la tâche est constitué de cubes robotiques tangibles nécessitant de les manipuler et générer, par essai-erreur, des hypothèses pour comprendre leur comportement. Ceci pourrait ainsi permettre de développer la pensée informatique et l'esprit critique, deux autres compétences du 21e siècle particulièrement intéressantes dans l'ère numérique d'aujourd'hui (ROMERO, DAVID et al., 2019).

LA RÉOLUTION CRÉATIVE DE PROBLÈMES DU POINT DE VUE DES SCIENCES COGNITIVES. Nous nous concentrons sur un problème *mal défini* dans lequel, contrairement au cadre classique de résolution de problèmes proposé par NEWELL et SIMON (1972), les états de l'espace du problème ne sont que partiellement observables par l'apprenant, y compris l'état du but (qui n'est pas nécessairement unique). De plus, les opérateurs possibles sur l'espace du problème sont découverts par l'apprenant au fur et à mesure qu'il manipule l'environnement et en perçoit les caractéristiques physiques comme des "potentialités

d'actions" (ou *affordances* (GIBSON, 1979)). Par conséquent, il est attendu de l'apprenant qu'il alterne entre deux stratégies principales : l'*exploration*, qui vise à expérimenter de nouvelles alternatives ou à agir sur l'environnement pour générer de nouveaux stimuli, et l'*exploitation*, qui se réfère à l'utilisation des connaissances existantes (sémantiques ou procédurales) dans une situation donnée. En relation avec cette distinction, nous pouvons introduire la double notion de *pensée convergente* et *divergente*, qui ont été décrites comme étant au coeur de la cognition créative (ALEXANDRE, 2020a) : la première est orientée vers la génération d'idées nouvelles, tandis que la seconde se concentre sur la sélection d'un candidat efficace comme solution du problème.

MATÉRIEL ET MÉTHODES



FIGURE A : La tâche CreaCube

LA TÂCHE CREA CUBE. ROMERO, DAVID et al. (2019) décrit une tâche de résolution de problèmes appelée CreaCube (Fig. A), qui peut être donnée dans le contexte d'une salle de classe pour initier les enfants à la résolution créative de problèmes et à la pensée informatique. Dans cette tâche, l'apprenant se voit présenter deux points de repère et quatre cubes robotiques modulaires qu'il peut assembler, pour construire une structure qui répondra à la contrainte de la consigne, à savoir "construire un véhicule autonome". L'ambiguïté de la consigne rend ce problème mal défini, comme mentionné précédemment.

Au stade actuel de nos travaux, quelques centaines de vidéos ont été annotées et encodées sous forme de données JSON représentant la chronologie des événements observés (y compris les configurations de cubes testées par l'apprenant, les problèmes rencontrés, etc.).

UNE APPROCHE PAR LES SCIENCES COMPUTATIONNELLES DE L'APPRENTISSAGE. Afin de mieux enseigner ces compétences, nous partons de l'hypothèse qu'une meilleure compréhension des mécanismes sous-jacents de la cognition humaine est nécessaire. À cette fin, la science cognitive computationnelle offre à la fois un cadre théorique et des résultats expérimentaux récents qui éclairent ces processus cognitifs. Par exemple, les notions d'exploration et d'exploitation sont précisément formalisées dans les paradigmes d'apprentissage par renforcement. Sont-ils adaptés pour modéliser les phases de pensée divergente et convergente dans une telle tâche? De même, la représentation ontologique

des connaissances en informatique fournit des outils permettant de caractériser précisément des concepts et leurs relations, afin de raisonner logiquement. Peut-on représenter la tâche à l'aide de ces outils, et en quoi cela peut-il être utile? Est-ce pertinent de modéliser le raisonnement d'un apprenant par l'inférence ontologique, et est-ce plausible d'un point de neuro-biologique?

CONTRIBUTIONS

UNE ONTOLOGIE POUR FORMALISER LES CONNAISSANCES STRUCTURÉES ET LES CONFIGURATIONS MATÉRIELLES. Nous avons tenté de formaliser la tâche de CreaCube en utilisant la modélisation ontologique (MERCIER, ALEXANDRE et al., 2022). L'ontologie résultante est alignée sur des ontologies de haut niveau pour s'assurer que le vocabulaire est bien spécifié, ce qui permet une meilleure communication entre les domaines de recherche convoqués. Cette formalisation a été réalisée étape par étape dans le but d'encapsuler les connaissances du domaine utilisées par les chercheurs en sciences de l'apprentissage lorsqu'ils analysent une telle expérience, afin de la rendre plus robuste et reproductible. Au-delà de cet objectif de spécification, nous proposons de réaliser des inférences à l'aide des raisonneurs disponibles pour mieux guider l'analyse qualitative des données collectées.

MODÉLISATION DU PROCESSUS DE RÉOLUTION CRÉATIVE DE PROBLÈME PAR APPRENTISSAGE PAR RENFORCEMENT SUR DES DONNÉES SYMBOLIQUES. Pour aller plus loin, nous proposons d'étudier comment caractériser la pensée convergente-divergente et les comportements d'exploration et exploitation dans une telle tâche. Au niveau computationnel, le paradigme est celui de l'apprentissage par renforcement avec une récompense finale (réussite de la tâche) et des récompenses intermédiaires (découverte d'affordances, résultat partiel par rapport à l'objectif), paradigme largement utilisé dans les neurosciences cognitives pour modéliser les fonctions de contrôle exécutif de haut niveau (ALEXANDRE, 2020a). Nous proposons ici de considérer une structure de données d'entrée symbolique (représentant l'état de l'espace-problème) tandis que la "récompense" pourrait être calculée dynamiquement à l'aide d'une distance d'édition entre la représentation de l'état actuel et celle de l'état-but (MERCIER, ALEXANDRE et al., 2021).

MISE EN ŒUVRE BIOLOGIQUEMENT PLAUSIBLE D'INFÉRENCES SYMBOLIQUES. Enfin, nous étudions comment les mécanismes de raisonnement logique peuvent être mis en œuvre dans les réseaux de neurones à impulsions (MERCIER, CHATEAU-LAURENT et al., 2021). Nous nous concentrons sur le raisonnement déductif tout en envisageant la possibilité d'appliquer de tels formalismes au raisonnement inductif et abductif, tous ces procédés pouvant potentiellement être utilisés dans la résolution de problèmes complexes. Nous considérons une architecture symbolique vectorielle (*Vector Symbolic Architecture (VSA)*) permettant de manipuler des symboles encapsulés dans vecteurs numériques porteurs d'informations sémantiques, ce qui nous permet de décrire les implémentations neuronales à un niveau algébrique. Cette approche neuro-symbolique illustre comment les précédents mécanismes cognitifs peuvent émerger du calcul distribué.

PERSPECTIVES POUR LES SCIENCES DE L'ÉDUCATION

Cette approche tri-disciplinaire, combinant les sciences de l'éducation, les neurosciences cognitives et l'informatique, pourrait permettre d'étudier de manière plus systématique le comportement d'un apprenant engagé dans une tâche de résolution créative de problèmes, grâce à l'utilisation d'une représentation informatique de cette tâche et d'un modèle cognitif de l'apprenant, utilisant à la fois des formalismes symboliques et numériques.

A l'heure actuelle, nos propositions consistent en des applications-jouets préliminaires; en effet, nous n'avons pas suffisamment de données correctement annotées sur la tâche CréaCube pour leur appliquer des méthodes de type apprentissage par renforcement inverse (i.e., interpréter le comportement d'un apprenant en inférant la récompense en supposant un tel paradigme). Pour pallier cela, nous avons commencé deux nouvelles collaborations afin de (i) développer des modèles permettant l'annotation automatique des données; (ii) développer une simulation de la tâche sur PC ou en réalité virtuelle, permettant de collecter automatiquement les données (bien que l'expérience utilisateur soit différente, les mêmes analyses pourraient être envisagées).

Une autre perspective consiste à utiliser des modèles de Markov cachés (*Hidden Markov Models (HMM)*) où chaque séquence d'activité peut être considérée comme une chaîne de Markov avec des états cachés (qui, selon notre hypothèse, correspondent à des représentations internes de l'apprenant, par exemple ses croyances, ses objectifs, ses émotions) à partir desquels nous observons une certaine sortie (c'est-à-dire les données JSON collectées) avec une certaine probabilité d'émission à partir de laquelle nous pourrions déterminer la stratégie de résolution de l'apprenant. Les algorithmes de regroupement (*clustering*) pourraient également être envisagés, afin de "structurer" l'espace du problème et de trouver des régularités qui aideraient à définir une distance appropriée entre les états du problème.

Bien qu'elle en soit à un stade très préliminaire, une telle formalisation ouvre la voie à la science de l'éducation computationnelle, qui semble prometteuse pour mieux appréhender la complexité des processus de résolution créative de problèmes à plusieurs niveaux: (i) la spécification du problème et des observables d'intérêts collectés pendant la tâche; (ii) la représentation cognitive de l'espace du problème, en fonction des connaissances préalables et de la découverte des affordances, permettant de générer des trajectoires de solutions créatives; (iii) l'implémentation de mécanismes de raisonnement au sein d'un substrat neuronal. Une telle compréhension pourrait être utile pour concevoir les activités pédagogiques les mieux adaptées à l'enseignement des compétences transversales, tout en tenant compte des différences individuelles dans l'apprentissage.

ABOUT

[extended abstract]

Integrating transversal competencies, such as creativity, problem solving and computational thinking, into the K-12 curricula is a key challenge in today's educational field. We postulate that teaching and assessing transversal competencies could benefit from a better understanding of the learners' behaviors in specific activities that require these competencies. As part of an exploratory research action, we propose a tri-disciplinary approach, drawing from learning science, cognitive neuroscience and computer science, to formalize a creative problem-solving task and the behavior of a subject engaged in such a task.

CONTEXT

TRANSVERSAL COMPETENCIES IN EDUCATION. Among the so-called "21st century skills" that are to be addressed in K-12 education, creativity and problem solving are two transversal competencies that are closely intertwined and can be studied in conjunction (Voogt and Roblin, 2012). In order to better teach these competencies, we make the assumption that a better understanding of the underlying mechanisms of human cognition is needed. For that purpose, computational cognitive science offers both a theoretical framework and recent experimental results enlightening such cognitive processes. We target here a problem-solving task which requires learners to engage in a creative behavior to generate novel ideas and implement them to succeed in solving an unknown problem situation. Besides, this activity involves tangible robotic cubes that appeal to visuo-spatial intelligence and are expected to help learners develop computational and critical thinking, two more 21st century skills that are especially interesting in today's digital era (Romero, David, et al., 2019).

CREATIVE PROBLEM SOLVING FROM A COGNITIVE SCIENCE PERSPECTIVE. We focus on *ill-defined* problems in which, as opposed to the classical problem-solving framework proposed by Newell and Simon (1972), the problem space states are only partially observable by the learner, including the goal state (which is not necessarily unique). Moreover, possible operators on the problem space are discovered by the learner as they manipulate the environment and perceive its physical features as "potentialities of actions" or *affordances* (Gibson, 1979)). Therefore, the learner is expected to alternate between two

main strategies: *exploration*, which aims at experimenting with new alternatives or acting on the environment to generate new stimuli, and *exploitation*, which refers to the use of existing knowledge (semantic or procedural) in a given situation. In relation to this distinction, we can introduce the dual notions of *convergent* and *divergent thinking*, which have been described as central to creative cognition (Alexandre, 2020a): the former is directed towards novel idea generation, while the latter focuses on selecting an effective candidate solution to the problem.

MATERIAL AND METHODS

THE CREACUBE TASK. Romero, David, et al. (2019) described a problem-solving task called CreaCube (Figure A), which can be given in the context of a classroom to initiate children to creative problem solving and computational thinking. In this task, the learner is presented with two waypoints and four modular robotic cubes that they can assemble together, to build a structure that will answer the guideline constraint, namely to “build an autonomous vehicle”. The ambiguity of the guideline makes this problem *ill-defined* (as defined earlier).

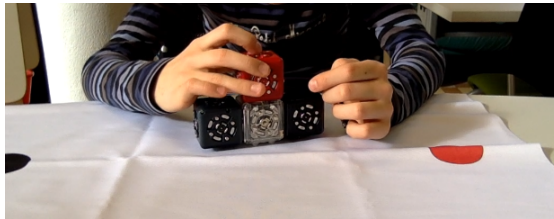


Figure A: The CreaCube task

At the present stage of our work, a few hundreds of videos have been annotated and encoded as JSON data sequences representing the observed timeline of events (including the cube configurations tested by the subject, the problems encountered, etc.).

CONTRIBUTIONS

AN ONTOLOGY TO FORMALIZE STRUCTURED KNOWLEDGE AND MATERIAL CONFIGURATIONS. We have attempted to formalize the CreaCube task using ontology modeling (Mercier, Alexandre, et al., 2022). The resulting ontology is aligned on upper ontologies to ensure that the vocabulary is well specified, allowing for a better communication between the summoned research fields. This formalization was achieved step by step with the aim of encapsulating domain knowledge used by learning science researchers when they analyze such an experiment, in order to make it more robust and reproducible. Beyond this specification purpose, we suggest performing inferences using available reasoners to better guide the qualitative analysis of the collected data.

MODELING THE CREATIVE PROBLEM-SOLVING PROCESS THROUGH REINFORCEMENT LEARNING ON SYMBOLIC DATA. A step further, we propose to study how to characterize convergent-divergent thinking and exploration-exploitation behaviors in such a task. At a computational level, the paradigm is reinforcement learning with a final reward (success of the task) and some intermediate rewards (discoveries of affordances, partial result regarding the goal), which is extensively used in cognitive neuroscience to model high-level executive control functions (Alexandre, 2020a). Here, we propose to consider a symbolic input data structure (accounting for the problem space state) while the “reward” could be dynamically computed using an edit distance between the current state and the goal state (Mercier, Alexandre, et al., 2021).

BIOLOGICALLY PLAUSIBLE IMPLEMENTATION OF SYMBOLIC INFERENCE. Finally, we investigate how reasoning mechanisms can be implemented in spiking neuronal networks (Mercier, Chateau-Laurent, et al., 2021). We focus on deductive reasoning and discuss whether this could also apply to inductive and abductive reasoning, as all these mechanisms can potentially be summoned in complex problem solving. We consider a vector symbolic architecture (VSA) providing a way to manipulate symbols embedded as numeric vectors that carry semantic information, allowing us to describe neuronal implementations at an algebraic level. This development illustrates how the former cognitive mechanisms can naturally emerge from distributed calculus, yielding neuro-symbolic computations.

PERSPECTIVES FOR LEARNING SCIENCES

This tri-disciplinary approach, combining learning science, cognitive neuroscience, and computer science, could provide a more systematic way to study the behavior of a learner engaged in a creative problem-solving task, through the use of a computational representation of this task and a cognitive model of the learner, using both symbolic and numerical formalisms. Despite being at a very preliminary stage, such a formalization paves the way for computational learning science, which seems promising to better understand complex processes involved in creative problem solving at several levels: (i) the specification of the problem and the observables of interest to collect during the task; (ii) the cognitive representation of the problem space, depending on prior knowledge and affordance discovery, allowing to generate creative solution trajectories; (iii) an implementation of reasoning mechanisms within a neuronal substrate. Such an understanding could be useful to design pedagogical activities best suited to teach transversal skills, while taking into account individual differences in learning.

INTRODUCTION

TEACHING AND ASSESSING TRANSVERSAL SKILLS IN K-12 EDUCATION

Transversal competencies, which are sometimes referred to as “twenty-first century skills”, include (but are not limited to) **creativity**, **problem solving**, **computational thinking**, **critical thinking**¹ and **collaboration**² (Romero, Lepage, et al., 2017). Integrating these skills in the K-12 curricula is a key challenge in today’s education (OECD, 2012; 2017; 2023) as studied by the field of learning sciences (Alexandre, Becker, et al., 2020; Romero, Lepage, et al., 2017). We postulate that teaching and assessing transversal competencies could benefit from a better understanding of the learners’ behaviors and underlying cognitive processes in specific activities that require these competencies.

These skills are intrinsically related to each other, which leads to interdependent definitions. In this work, we will mainly focus on the first two aforementioned skills as a combination, namely **creative problem solving** (CPS), while keeping in mind that understanding this skill at an individual level could be helpful as a starting point to better understand other competencies, e.g, collaborative problem solving.

In our context, **creativity** refers to the process of producing something new and adequate with regard to the task being performed (e.g., a problem being solved) (Alexandre, 2020a; Dietrich, 2004; Guilford, 1967). **Problem solving** in everyday life takes many different forms, depending on the task and context; in the general sense, it has been formalized at a computational level (Newell and Simon, 1972) as finding a path from an initial state (the current situation) to a final state (the goal), while taking into account some constraints (e.g., task requirements). **Computational thinking** uses tools inspired by the field of computer science (e.g., algorithms, information coding and data representation) to (for example) solve problems or design systems (Wing, 2006) (see also Lodi (2020) for a recent review). By extension, computational thinking can also be considered to include technology literacy (Jacob and Warschauer, 2018). Furthermore, techno-creative activities (Heiser et al., 2022) have the potential to develop computational thinking while also engaging learners in **critical thinking** by encouraging them to form and test hypotheses or interpret the outcome of their actions (Alexandre, Becker, et al., 2020). We have therefore chosen to focus on a problem-solving activity involving robotic artifacts, as developed in the sequel. Such an activity can also be adapted into group (2-4 people) activities,

¹*Critical thinking* is defined as “a form of directed, problem-focused thinking in which the individual tests ideas or possible solutions for errors or drawbacks. It is essential to such activities as examining the validity of a hypothesis or interpreting the meaning of research results.” (*APA Dictionary of Psychology*)

²*Collaboration* refers to “the act or process of two or more people working together to obtain an outcome desired by all” (*APA Dictionary of Psychology*)

making it interesting to study collaboration mechanisms and develop social and interpersonal skills. However, we will study this task at an individual level and not at the group level, as collaborative problem solving is another large field of research that we will not address in this work.

METACOGNITION FOR SELF-REGULATION These transversal competencies also need to be considered in relation to (meta-)cognitive processes; *metacognition* referring to “the awareness of one’s own cognitive processes, often involving a conscious attempt to control them” (*APA Dictionary of Psychology*). Pintrich (2000) has shown that students who are self-regulating, in other words, capable of setting their own goals and plans, and try to monitor and control their behavior in line with these goals are more likely to do well in school. More specifically, meta-learning or self-regulated learning is a meta-cognitive process which involves the awareness of one’s own learning processes, which is key to be able to transfer skills from a given problem-solving task to another (De Grave et al., 1996), while Callan et al. (2021) have investigated the relationship between self-regulated learning and key indicators of creativity in the context of problem solving. Here, we hypothesize that better formalizing and understanding the learner upstream to meta-learning may contribute to better pedagogical resource design and help teach this critical awareness of one’s own learning process.

LEARNING WITH TANGIBLE OBJECTS At a more concrete level, we aim to study creative problem solving through screen-less activities, also known as unplugged activities (Huang and Looi, 2021); for instance, using tangible artifacts, such as pedagogical robots or connected objects (Romero, David, et al., 2019), or even everyday objects. Using tangible material allows to observe behavioral patterns (Romero, David, et al., 2019) which, if properly characterized, may help infer cognitive processes. This paradigm has demonstrated its efficiency in education, especially for computational thinking initiation (Huang and Looi, 2021). However, in contrast to online activities on a computer, a non-trivial setup is required to track the learning process through the activity. One of the challenges is the use of efficient measurement devices of reasonable cost. It also appears that a precise model of the task allows us to define specific observables that are thus more robust to estimate (Romero, David, et al., 2019). Another obstacle is the bias that is induced by the fact that the learner is observed and monitored. One solution is to take this as a chance to involve the learner in their own learning process, as it is the case with Open Learner Models (Bull and Kay, 2010). This is yet to be properly studied in this context of tangible manipulation; see Barnabé et al. (2020) for a discussion on these points. Ethical issues of collecting human data in a way that ensures data privacy also needs to be considered, to ensure that only required data is effectively collected.

THE OBSERVATION RELEVANCE CHALLENGE Properly characterizing observable behavioral patterns can be difficult because we are working on a relatively small batch of data (around a hundred individuals, compared to the thousands of data used in classical statistical methods). For the results to be meaningful, we propose to introduce prior

information upstream to the analysis of the observables, taken from existing theoretical models in learning sciences. For this purpose, **ontology** modeling is regarded as an appropriate tool to describe structured knowledge, as we will develop in Chapter 2.

A COMPUTATIONAL LEARNING SCIENCES APPROACH

LEARNING SCIENCES, AN INTERDISCIPLINARY FIELD SUITED FOR COMPUTATIONAL MODELS Learning sciences³ are an interdisciplinary field aimed at better understanding how humans learn, with concrete applications towards the implementation of innovations in education and the improvement of instructional methodologies.

In the mid-1950s, researchers from several fields (namely psychology, philosophy, linguistics, computer science, neuroscience and anthropology) began to develop theories of mind based on complex representations and computational procedures (see also Appendix A for a timeline recapitulating the evolution of learning sciences towards computational models). In the meantime, computational cognitive architectures propose to implement operational algorithms to solve problems. Throughout the following decades, researchers in computer science (and some of its sub-fields referred to under the umbrella term “Artificial Intelligence”) often drew inspiration from human cognition and findings from cognitive neuroscience and psychology, to develop diverse algorithms capable of handling a specific set of tasks. In turn, this led to new frameworks of interest to model and study behaviors and cognitive processes in animals, including humans. A famous example is **reinforcement learning**, a computational paradigm inspired by Pavlovian conditioning, which is now often used in neuroscience: typically, **TD-learning** to model dopaminergic firing. In a nutshell, these disciplines benefit from one another when putting in perspective their results (Alexandre, Becker, et al., 2020).

Recent computational approaches combine problem solving with modular and evolutive knowledge representation in order to be able to solve **ill-defined** problems. Beyond problem solving, some models also replicate tasks assessing two modes of thinking involved in the creative process, either **divergent thinking** or **convergent thinking**, or both. We focus here on a specific learning task instead of considering human learning at a general level to offer a precise and operational formalization that may be translatable to other learning tasks.

WHAT FOR? The finality of developing a computational model of a creative problem-solving task involving computational thinking (Romero, Lepage, et al., 2017) aims to contribute, as a long-term perspective, to a better understanding of fundamental notions regarding natural versus artificial so-called intelligence, as studied in Alexandre, Becker, et al. (2020), potentially yielding a better understanding of our human intelligence. Such a model is intended to be applicable to the observables (including the learner’s behavior) and learning analytics, and aims at interpreting them. Indeed, phenomenological

³In French, the discipline is usually referred to as “sciences de l’éducation”, although the term “sciences de l’apprentissage” is sometimes preferred to include all learning contexts beyond school settings.

models already exist — for example, the activity theory (Engestrom, 2000) provides a framework that is appropriate to the context of this study: it specifies an activity as performed by a subject through the mediation of a tool to achieve a certain outcome. Based on such phenomenological frameworks, our present study is positioned at an operational level.

We postulate that understanding the underlying mechanisms of human cognition could be beneficial to better teach these competencies (not necessarily in a prescriptive way, but rather as a more integrative framework to design and evaluate related learning activities; this will be discussed in Chapter 5). For that purpose, cognitive neuroscience offers both a theoretical framework and recent experimental results enlightening such cognitive processes. Of course, these results are not directly translatable to a classroom set-up and should be put in relation to cognitive and educational psychology, considering more ecological approaches.

THE CREA CUBE TASK: AN ILL-DEFINED PROBLEM TO STUDY HUMAN CREATIVE PROBLEM SOLVING

This task has been designed considering the cognitive science literature regarding problem solving tasks (Newell and Simon, 1972; Schraw et al., 1995), but addressing here considerations regarding *ill-defined* problems, and manipulation of tangible objects.

Research on human problem solving has focused on a diversity of tasks but most of all on well-defined problems, such as the chess game or the Tower of Hanoi, which allows us to define task models and problem-solving methods (Newell and Simon, 1972). Nonetheless, problem solving in everyday life is typically *ill-defined* in the sense that the goal is only partially defined or may have to be changed throughout the resolution, the initial state is only partially observable, and the problem constraints are discovered on the way towards the solution. Consequently, solving these problems requires us to consider not only prior knowledge but also the contextual knowledge that is built throughout the task process based on the different operations that the subject performs (Romero, 2017). Moreover, the exploration of the means given to solve the problem should facilitate the generation of new *stimuli*, which can then help the learner to make progress towards the goal by fulfilling sub-goals and validating or invalidating hypotheses. This is what we would like to consider here: in particular, we aim to consider the problem-solving process instead of considering only performance scores of the task.

The CreaCube study⁴ focuses on problem-solving strategies using modular robotic cubes, targeting children between 8 and 12 years old. The guideline given to the participants indicates to “build an autonomous vehicle, composed of four items, able to move from a point A to another point B”, while a set of four modular robotic cubes that differ in their appearances (e.g., different colors) and features (e.g., wheels, switch etc.) is presented. The learner is expected to understand that the four items are the four cubes on the table and to assemble these cubes into a certain configuration that will be able

⁴See <https://creamaker.wordpress.com/> for full details

to move autonomously. It is worth noting that this problem does not have a unique solution because several configurations may satisfy the goal requirements. Furthermore, additional goals may emerge, such as understanding the cube behaviors, exploring alternative outcomes with those items, preserving the ego with respect to the task achievement, experiencing playful pleasure, and so on. The task is relatively easy in the sense that most participants come up with a solution in less than 15 minutes but complex to model because of the large problem space and the lack of specification (e.g., the possible actions, i.e. operators on the problem space states, are not clearly stated) (Romero, 2017), which makes it an **ill-defined** problem according to the previous definition.

In this activity, the participant has likely no prior experience of the modular robotic cubes. To complete the task, there is thus a need for exploration, defined here from a learning sciences perspective, as a way to gather information from the environment (Kaplan and Flum, 2010). Throughout the task, the participant moves towards the goal by mobilizing prior knowledge that appears to make sense in relation to their experience with the material being manipulated and the requirements defined by the guideline. This includes understanding the artifacts, their physical and technological characteristics, leading to formulate hypotheses on the material and define sub-goals to solve the problem. This process is better explained using the notion of **affordances** (Gibson, 1977), i.e., the practical possibilities offered by the cubes, as we will further develop in the sequel.

The problem is modeled based not only on the knowledge needed to solve it but also on the initial states of the hardware and the final state for its success, as detailed in Fig. B. We can see the different observables taken into account (i.e., the possible configurations of the cubes—disassembled or assembled into a certain shape), the identification and conceptualisation of **affordances**, the outcomes of actions performed (e.g., at the motion level), and also elements regarding the subject's emotions or attitude (perseverance, abandonment) towards the task. These observables will serve as a basis for modeling the activity in the form of structured knowledge.

The data model that is generated from this interface has been developed as a hierarchical data structure (in JSON syntax at the file format level), with both the raw and computed data, as well as the description of each type of information and its relationships to other types. This allows formal manipulation and representation of the collected information, as a first structuring step towards an ontology representation, and integrated as an interactive platform to facilitate the ergonomics of manual video analysis (Romero, Vieville, et al., 2021). Each video for the CreaCube activity is encoded as a temporal sequence of states, corresponding either to a configuration of the activity material or to a state of the learner engaged in the activity, as explained above (Fig. B). This is done manually by the experimenter using a dedicated interactive interface.

CREACUBE 2. Activity									
AS00									
AS01									
AS02									
AS03									
F000	F002	F010	F011	F012	F013	F014	F020		
F021	F022	F024	F030	F034	F040	F042	F044		
F050	F051	F060	F061	F064	FXX				
F000-SBIW-T	F000-SIWB-T	F000-SWBI-T	F000-BSIW-T	Example: F020-IBWS-T 				P01. Imbalance	
F000-BIWS-T	F000-BWSI-T	F000-IWBS-T	F000-IBWS-T	[...] (every permutation is encoded for each shape)				P02. Rotation	
F000-ISWB-T	F000-WBSI-T	F000-WSBI-T	F000-WIBS-T					P03. Wrong direction	
U00. Play instructions	U01. Stop playing instructions	U02. Question instructions	U03. No cubes in hand (no manipulation)	U04. Hands up with 1 cube	U05. Hands up with 2 cubes	U06. Hands up with 3 cubes	U07. Hands up with 4 cubes	FL01 Rotate the cube without changing its position	P04. Reverse (outward)
E01. Ecstasy/Joy/Serenity	E02. Admiration/Trust/Acceptance	E03. Terror/Fear/Apprehension	E04. Amazement/Surprise/Distracted	E05. Grief/Sadness/Pensiveness	E06. Loathing/Disgust/Boredom	E07. Rage/Anger/Annoyance	E08. Vigilance/Anticipation/Interest	FL02 Reposition the cube into the same configuration	P05. Reverse (to the person)
A01. Wheels	A02. Magnets	A03. Switch	A04. "Eyes"	A05. Sensors	S01. OFF	S02. ON	T01. No test	T02. Drop Out / Abandon	T03. Succeed
									P06. Colour association
									P07. Connection
									P08. Does not move (wheels)
									P09. Does not move (switch)
									P10. Does not move (sensor)
									P11. Does not move (inv)

Figure B: The CreaCube observables, reproduced from (Mercier, 2023) and adapted from the annotation interface CreaCube interface designed by Margarida Romero and developed by Eloïse Duhot-Prévoit (ANR CreaMaker). This figure describes a part of the system states to which is added, for example, the identification of each cube (recognizable by its color: dark blue battery, black sensor, white motor, red inverter) and the states of the cubes (e.g., “connected/disconnected” or “on/not on wheels”). This sub-ensemble of the possible states corresponds to the observables that have been chosen to analyse the activity (see <http://aide-line.inria.fr/public/doc/vid.mp4> for a video of the experiment.)

Introduction

An example of encoded data generated from the interface follows:

```
{
  "idVideo": "v0001",
  "nbParticipants": "1",
  "events": [
    {
      "time": "12",
      "tclicks": [
        "U00",
        "START1"
      ]
    },
    {
      "time": "58",
      "tclicks": [
        "AF01"
      ]
    },
    {
      "time": "73",
      "tclicks": [
        "AF04"
      ]
    },
    {
      "time": "77",
      "tclicks": [
        "AF03"
      ]
    },
    {
      "time": "95",
      "tclicks": [
        "AF02"
      ]
    },
    {
      "time": "121",
      "tclicks": [
        "U00",
        "S02"
      ]
    },
    {
      "time": "146",
      "tclicks": [
        "F030-BSWI-F",
        "P02"
      ]
    },
    [ ... ]
    {
      "time": "232",
      "tclicks": [
        "F030-BIWS-F",
        "P02"
      ]
    },
    {
      "time": "268",
      "tclicks": [
        "F030-BIWS-T"
      ]
    },
    {
      "time": "274",
      "tclicks": [
        "T03"
      ]
    }
  ],
  "idParticipant": "p0001",
  "age": 11
}
```

The methodology to generate an organized set of learning analytic log from video analysis has been developed from a learning sciences perspective through the [ANR CreaMaker](#) project to study problem solving activities ([Romero, Vieville, et al., 2021](#)). Beyond the analysis of observables for research purposes, these videos can also be used to train the teachers who would be willing to put this activity into practice within their classes ([Albero and Guérin, 2014](#)).

ORGANIZATION AND CONTRIBUTIONS OF THE THESIS

We propose a tri-disciplinary approach drawing from learning sciences, cognitive neuroscience, and computer science in order to provide a more systematic way to study the behavior of a learner engaged in a creative problem-solving task, through the use of a computational representation of this task and a cognitive model of the learner, using both symbolic and sub-symbolic formalisms.

We start with a literature review of creative problem-solving studies from these three disciplines (Chapter 1): from educational psychology and the pedagogy of creativity, to the cognitive neuroscience of creativity and related brain processes, concluding with a review of the computational formalisms that have been used to decipher and operationalize the creative problem-solving process.

Our subsequent contributions are organized by three levels of analysis related to Marr's hierarchy (Kitcher, 1988): (i) Chapter 2 describes the construction of an ontology to better specify the CreaCube task and the observables of interest to collect during the task; (ii) in Chapter 3, we investigate how a cognitive representation of the problem space, depending on prior knowledge and **affordance** discovery, allows to generate creative solution trajectories; (iii) in Chapter 3, we implement deductive reasoning mechanisms within a neuronal substrate and discuss to what extent these could be extended to inductive and abductive reasoning, or even analogy.

Finally, in Chapter 5, we reflect on the perspectives and limitations of our approach, which, despite being still quite preliminary, seems promising to clarify concepts that are used in the three aforementioned disciplines with some nuances, to describe cognitive processes: as a matter of fact, deciphering these nuances might be a first step in understanding, operationalizing and finally teaching transversal skills such as creativity and problem-solving. This could be useful to design pedagogical activities best suited to teach transversal skills, while taking into account individual differences in learning. Nevertheless, we need to keep in mind that findings from computational and cognitive modeling are usually not directly applicable to the classroom without the empirical input of the teachers: validation and adjustments "on the field" are to be expected.

HOW TO READ THIS MANUSCRIPT

This work contains many definitions due to its anchoring in various disciplines, and we aim for it to be accessible to a researcher coming from any of those disciplines.

We have therefore tried to make it easier to navigate through these numerous notions⁵. When a **word** is colored in dark red, clicking on it will make you jump to its definition in context (where the same *word* will be in italic font). An index of all notions is also available at the end of the thesis ([Index](#)).

⁵Credits to the amazing [knowledge](#) LaTeX package!

COLLABORATIONS

This whole manuscript results from a strong collaboration with my advisors Margarida Romero^{*†} and Thierry Viéville^{†*} as well as my co-advisor Frédéric Alexandre^{*}. Axel Palaude^{*}, who I like to call my “co-PhD” peer, has also contributed to some of the ideas developed in Chapters 1 and 3.

As a matter of fact, Chapter 1 results from a collaborative review between the five of us, resulting in two papers in revision at the time of writing (Alexandre, Mercier, Palaude, Romero, and Vieville, 2024; Romero, Palaude, et al., 2024) and a poster presented at the 7th annual meeting of the *Society for the Neuroscience of Creativity (SfNC 2022)* (Alexandre, Mercier, Palaude, Romero, and Viéville, 2022). With regard to this work, Lola Denet’s internship was also very valuable to decipher the role of intrinsic motivation in learning (Denet, 2021) in collaboration with Inria Flowers team.

Chapter 2 is adapted from a conference paper presented at *ICDL 2021* (Mercier, Roux, et al., 2021) which was extended as a journal article for *IEEE TCDS* (Mercier, 2023). Lisa Roux is gratefully acknowledged for her first version of the ontology drafted as part of her post-doctoral studies (Roux et al., 2020).

Chapter 3 stems from an idea we first presented at *ICANN 2021* (Mercier, Alexandre, et al., 2021). We then elaborated on this idea thanks to a collaboration with Axel Palaude (Palaude et al., 2023) as well as Louis Kohler’s internship and two student projects that Axel and I co-supervised (Bernard et al., 2023; Radji et al., 2023).

Chapter 4 originated as a collaboration with Hugo Chateau-Laurent[†], co-author of a work we presented at the KR workshop on *Knowledge Representation for Hybrid & Compositional AI (KRHCAI 2021)* (Mercier, Chateau-Laurent, et al., 2021). Terrence Stewart[§] is gratefully acknowledged for his feedback on this work. We also wish to thank Gabriel Doriath Döhler[¶] for clarifying the use of VTB algebra (Doriath-Döhler, 2021) and Théophile Vallaeys[¶] for exploring possibility and necessity representation in VSAs (Vallaeys, 2021), during their respective undergraduate internships.

Finally, we wish to acknowledge researchers from the LINE^{*} for the CreaCube experimentations and their interventions at the recurring CreaComp meetings as well as Cédric Brun[¶], for his epistemology and philosophy of science perspective, which helped nurturing Chapter 5 by providing a broader view. Another more technical perspective was brought by the Inria Potioc/Bivvac team, thanks to Arnaud Prouzeau[†] with whom I co-supervised a student project to draft a virtual desktop version of the CreaCube task (Villars et al., 2023), and Juliette Le Meudec[†] who took upon the development of this task in virtual reality as part of her internship (Le Meudec et al., 2024), which may serve as a basis for her starting PhD to study collaborative mechanisms in virtual environments.

*LINE (Laboratory of Digital Innovation for Education), Université Côte d’Azur, Nice, France. (All affiliations indicated here refer to the ones mentioned at the time of the collaboration and may have changed since.)

†Inria, Mnemosyne team, Bordeaux, France.

‡Inria, Potioc/Bivvac team, Bordeaux, France.

§Centre for Theoretical Neuroscience, University of Waterloo, Canada.

¶Ecole Normale Supérieure, PSL University, Paris, France.

¶CNRS, Institute of Neurodegenerative Diseases, UMR 5293, Bordeaux, France.

CONTRIBUTIONS

OPEN-SOURCE CODE

Clustering experiments on the CreaCube task data (ongoing):

<https://gitlab.inria.fr/line/aide-group/creadata>

Ontology specifying the CreaCube task:

<https://gitlab.inria.fr/line/aide-group/creaonto>

How to simulate ontology inferences in a Semantic Pointer Architecture (using Nengo/Spaun):

<https://gitlab.inria.fr/line/aide-group/onto2spa>

JOURNAL ARTICLES

Mercier, C. (2023). An Ontology to Formalize a Creative Problem Solving Activity. *IEEE Transactions on Cognitive and Developmental Systems* 15:4, pp. 1891–1904. ISSN: 2379-8939. DOI: [10.1109/TCDS.2022.3210234](https://doi.org/10.1109/TCDS.2022.3210234).

CONFERENCE PAPERS

Mercier, C., L. Roux, M. Romero, F. Alexandre, and T. Vieville (23, 2021). “Formalizing Problem Solving in Computational Thinking : an Ontology approach”. In: *2021 IEEE International Conference on Development and Learning (ICDL)*. IEEE, Beijing, China / Virtual, pp. 1–8. ISBN: 978-1-72816-242-3. DOI: [10.1109/ICDL49984.2021.9515660](https://doi.org/10.1109/ICDL49984.2021.9515660).

Mercier, C., F. Alexandre, and T. Viéville (2021). “Reinforcement Symbolic Learning”. In: *Artificial Neural Networks and Machine Learning – ICANN 2021*. Ed. by I. Farkas, P. Masulli, S. Otte, and S. Wermter. Lecture Notes in Computer Science. Springer International Publishing, Bratislava, Slovakia / Virtual, pp. 608–612. ISBN: 978-3-030-86380-7. DOI: [10.1007/978-3-030-86380-7_49](https://doi.org/10.1007/978-3-030-86380-7_49).

Mercier, C., H. Chateau-Laurent, F. Alexandre, and T. Viéville (11, 2021). “Ontology as neuronal-space manifold: towards symbolic and numerical artificial embedding”. In: *KRHCAI-21@KR2021*. Hanoi, Vietnam / Virtual.

POSTERS AND PRESENTATIONS

Palaude, A., C. Mercier, and M. Romero (2, 2023). “Représenter l’apprentissage humain dans des problèmes ouverts”. Presentation. Dataquittaine 2023 - IA, Recherche Opérationnelle & Data Science. Kedge Business School Bordeaux, France.

Alexandre, F., C. Mercier, A. Palaude, M. Romero, and T. Viéville (12, 2022). “The role of cognitive processes in creative problem solving: a computational approach”. Poster. 7th annual meeting of the Society for the Neuroscience of Creativity (SfNC 2022). United States / Virtual.

Mercier, C., F. Alexandre, and T. Viéville (31, 2022). “Ontology as manifold: towards symbolic and numerical artificial embedding”. Webinar. Workshop on VSA and hyperdimensional computing - Winter 2022 session, Luleå University and UC Berkeley. Lulea, Sweden / Virtual.

Introduction

Mercier, C., M. Romero, F. Alexandre, and T. Viéville (2, 2022). “Apprentissage par renforcement sur des connaissances symboliques structurées : modéliser la résolution créative de problèmes”. Poster. CJC-SCo 2022 - Colloque des jeunes chercheur·se·s en sciences cognitives. ENS, Paris.

REPORTS BY SUPERVISED STUDENTS

Bernard, P., B. Hate, and M. Laval (27, 2023). *Symboling : utiliser des structures symboliques dotées d'une métrique*. Research Report RR-9499. Inria & Labri, Univ. Bordeaux, p. 18.

Radji, W., C. Léger, and L. Bardisbanian (23, 2023). *Early Empirical Results on Reinforcement Symbolic Learning*. Research Report RR-9509. Bordeaux, France: Inria & Labri, Univ. Bordeaux, ENSEIRB Bordeaux INP, p. 9.

Villars, A., A. Solignac, V. Piton, L. Pereira, B. Miège, C. Guillaud, and T. Cauquil (2023). *Une activité de robotique éducative transposée dans un environnement virtuel*. ENSEIRB-MATMECA, Bordeaux-INP.

Denet, L. (18, 2021). “Analyse de la motivation intrinsèque au cours d’une activité de résolution de problèmes.” Master Thesis. Bordeaux, France: Inria & Labri, Université Bordeaux.

Doriath-Döhler, G. (2021). *Implémentation neuronale d'inférences logiques dans une SPA*. Research report in preparation. Inria Bordeaux Sud-Ouest.

Vallaëys, T. (9, 2021). *Généraliser les possibilités-nécessités pour l'apprentissage profond*. Research Report RR-9422. Inria, p. 1.

1 CREATIVE PROBLEM SOLVING: A TRIDISCIPLINARY REVIEW

1.1 CREATIVE PROBLEM SOLVING: A COMPLEX PHENOMENON STUDIED AT MULTIPLE LEVELS

1.1.1 A DEFINITION OF CREATIVE PROBLEM SOLVING

In psychology, *problem solving* is defined as “the process by which individuals attempt to overcome difficulties, achieve plans that move them from a starting situation to a desired goal, or reach conclusions through the use of higher mental functions, such as reasoning and creative thinking” while creative thinking denotes “the mental processes leading to a new invention, solution, or synthesis in any area” and *creativity* refers to “the ability to produce or develop original work, theories, techniques, or thoughts” (*APA Dictionary of Psychology*). More specifically, we consider here “personal” everyday creativity in the sense of (Boden, 1998) rather than “historical”¹ creativity found in culturally significant achievements. Besides, we aim to analyze creativity in the context of problem-solving tasks, as it can be experienced during learning activities. Therefore, we will rather adopt the operational definition of *creativity* as the process of producing something that is both² novel (i.e., original and unexpected) and appropriate (i.e., useful and adequate regarding the task constraints) with regard to the task being performed (e.g., a problem being solved) (Alexandre, 2020a; Dietrich, 2004; Guilford, 1967).

This clarifies the definition of *creative problem solving* (CPS) as the process of solving a problem using creativity. We focus on *ill-defined* CPS tasks, corresponding to problem-solving tasks in which the subject also needs to disambiguate different stages of the problem-solving process towards a potential solution which could not be anticipated by the initial problem state. We postulate that such disambiguation also requires creativity, from the problem posing to the problem-solving enaction within a specific task and socio-cultural context (Engestrom, 1987), mediated by the subject’s body and actions but also by all the different types of analogical or digital artifacts engaged in the development of a creative performance or product (Leroy et al., 2021).

¹also called Big-C creativity by some authors (Kaufman and Beghetto, 2009)

²Creative production may be neither valuable nor new for an entire population, in a given era of time, and so on. As this description could also include production by chance (a random natural structure from a natural phenomenon, for instance), it should be completed with the capacity of the creators to assess their work in a particular context, with a certain degree of skill, judgment, or understanding if not flair (Gaut, 2010).

1 Creative Problem Solving: A tridisciplinary review

1.1.2 MULTIPLE LEVELS OF ANALYSIS

Creativity has already been studied in learning sciences (Leroy et al., 2021) and in other related disciplines, such as psychology (Boden, 1998), cognitive neuroscience (Alexandre, 2020a), and computer science (Oltețeanu, 2020) as well as philosophy (Gaut, 2010). Nevertheless, as pointed out by (Kupers et al., 2019), current studies on CPS need to overcome the macro analysis of creativity, perceived as a static personal trait, to better characterize creativity as a dynamic process to be assessed at the micro level, e.g. within a task. We aim to achieve this objective by contributing to a better understanding of creativity within complex ill-defined CPS tasks, through a multi-disciplinary approach using computational and neuro-cognitive formalisms. This approach allows us to propose a multi-factor theoretical description of such cognitive human activities³. Based on this, we aim to operationalize a computational framework of creative processes during CPS learning tasks, focusing on an ill-defined CPS task performed by a unique learner.

MULTIPLE LEVELS IN SOCIAL AND LEARNING SCIENCES

A shift in education emerged in the 1960s when creativity research started to enrich teaching objectives and methods (Torrance and Goff, 1989). Researchers such as Guilford and Torrance proposed to use creativity, perceived as a personal trait and a key dimension of intelligence (Guilford, 1967), to measure “potential” and detect “gifted” students who might have not performed so well at IQ tests (Alabbasi et al., 2022; Cramond et al., 2005). As a result, new frameworks were developed to study creativity in psychology, encompassing – but not limited to – an educational context. Rhodes (1961) pointed out four aspects of creativity which are interrelated and need to be studied in conjunction, namely “the 4 Ps”, as reviewed, for example, by Tang and Hanneghan (2011):

- (i) The Person: one’s personality, intellect, temperament, traits, habits, attitudes, values, skills, and behavior⁴;
- (ii) The Problem: the mental activities that occur during thinking;
- (iii) The Product: the creative outcome of such mental activities;
- (iv) The Press: the socio-cultural environment.

Buiding on this pioneering work, Kupers et al. (2019) defined creativity, considering a child student, as an iterative process of interrelations between the child, the task, and the social environment, organized in a hierarchical way, yielding top-down “constraints”

³Of course, creative problem-solving is not only exhibited by humans: other species (most notoriously apes or corvids) develop creative uses of tools to solve problems (Lefebvre, 2013). Non-biological organisms (so-called artificial intelligence) may also display creative behaviors, as reviewed in the sequel.

⁴In this work, we use a restrictive definition of **behavior** as an observable sequence of **actions** displayed by the subject, following classical **reinforcement learning** paradigms. We thus do not include mental activity, despite these notions being deeply intertwined.

1.1 Creative problem solving: a complex phenomenon studied at multiple levels

and bottom-up “emergences”. Further clarification in the framework proposed by [Kupers et al. \(2019\)](#) concerns the different levels of creativity study: from the micro-level within a task (behavior) to the macro-level considering “creative personalities” across tasks; and how they can be evaluated (e.g., by product assessment or personality questionnaires), whether the study is static (measurement at a given time) or dynamic (measurement of the evolution of creativity over time), and whether the proposed intervention is unidirectional (simple evaluation of its effect on creativity) or takes place within a causality process (examining the chain of causes and effects over time).

Taking a step further, [Mansfield et al. \(1978\)](#) highlighted the fact that studies of the effectiveness of creativity training methods generally suffer from methodological weaknesses, for example the Hawthorne effect (individuals modify aspects of their behavior in response to their awareness of being observed). Regarding creativity training, they observed that considering only **divergent thinking** (defined in the next section) as a creative process yields biased results, whereas it is a complete process influenced by cognitive, motivational, personality, and situational factors.

While cognitive neuroscience advances the understanding of how the brain functions in creative tasks ([Alexandre, 2020a](#); [Daikoku et al., 2021](#)), education and psychology researchers undertaking creative studies have developed a diversity of approaches in which we can find studies at the cognitive level ([Hao et al., 2016](#); [Lubart and Sternberg, 1995](#); [Radel et al., 2015](#)) or the behavioral level ([Nemiro et al., 2017](#)). Learning scientists have focused more often on situated learning tasks in ecological contexts of education in some cases to analyze the creative process ([Savic, 2016](#)), to consider the emergence of the creative process at different social scales as represented in Fig. 1.1, namely the individual, small-group, or classroom level ([Mathisen et al., 2004](#); [Stinkeste et al., 2021](#)), and to observe differences with respect to culture, personality, and recognized values ([Basantia, 2017](#); [Lubart, 1998](#)). These approaches also lead to applications in pedagogy, as reported by [Ni et al. \(2014\)](#), in which the theoretical approach successfully helped students to improve their creativity.

These different levels of analysis of creativity create diversity not only of conceptual but also of methodological models in the study of this phenomenon in education. To advance in the consideration of a multilevel approach to creativity, we consider a pluralistic epistemological approach ([Turkle and Papert, 1992](#)) in the study of creativity in domain-specific tasks, thus being restrained to an activity-oriented approach ([Albero and Guérin, 2014](#); [Romero, Vieville, et al., 2021](#)), and we focus on ill-defined problem-solving tasks engaging one learner. This allows us, as stated in the introduction, to consider a behavioral time sequence in which divergent and convergent thinking processes appear at different moments. Despite this specificity, we acknowledge that studying human creativity also requires the consideration of a socio-cultural perspective to analyze how knowledge is shaped by prior experiences, the cultural context, and the activity system in which the subjects are developing their creative processes, as briefly reviewed below.

While learning scientists focus on diverse ecological and social scales, studying the brain in cognitive neuroscience is performed at different temporal, spatial and topological scales (Betzel and Bassett, 2016). We reproduce part of this hierarchy in Fig. 1.1 (leaving out the topological aspect).

This multi-level modeling can be related to symbolic versus sub-symbolic modeling; that is to say, from high-level “human-readable” representation, to low-level numerical implementation in neuronal structures. As reviewed by Glennerster (2002), Marr and Gibson pointed out, in separate works, that cognitive processing (e.g., vision) passes through a series of stages of information processing, each corresponding to a different representation, from neural input activity (e.g., a retinal image), to the representation of a scene as real-world objects. A key idea in both theories is that contingent regularities in the physical environment, known by the system, explain how the visual system determines the shape and location of objects in the world on the basis of the retinal input. This generalizes to other sensorimotor skills. Such knowledge is to be represented as symbolic information (what we mean by **symbols** is defined in 1.4). In line with this idea, Betzel and Bassett (2016) show that multi-scale topological structure modeling requires a qualitative symbolic description of the “communities” hierarchy (e.g., clusters or attractors), in addition to their qualitative properties in the sense of dynamic system theory. This means that the description requires a non-trivial symbolic data structure, and not only a numerical representation of the behavior.

Marr’s original formulation referred to these 3 levels as (i) the “computational” level, i.e. the specification or description of the problem the system is solving; (2) the “algorithmic” level, i.e. the representations and manipulations used by the system to solve that problem; and (3) the “implementation” level, i.e. how that algorithm is computed in the physical hardware of the system. However, as pointed out by V.G. Hardcastle and K. Hardcastle (2015), the notion of “computational” has highly evolved with the progress in computational neuroscience. More than a mere description of the problem to solve, the computational paradigm is at the core of the interaction between the biophysical implementation and higher-level behavioral stages (e.g., Kitcher (1988)). That is why we prefer to use the terms “specification”, “representation/manipulation” and “implementation”, less ambiguous with regard to today’s use of computational and algorithmic paradigms, while still in line with Marr’s idea. We put this three stages in relation to modeling from neuronal to cognitive and behavioral aspects, and numerical to symbolic approaches, as sketched out in Fig.1.1. This also lets us clarify that the numerical, i.e., quantitative, aspects mainly stand on the biophysical side, while symbolic representation mainly stands on the behavioral stage. The key point is that computational modeling should intrinsically be able to manage both representations conjointly. This corresponds to the development proposed in Chapter 4.

1.1 Creative problem solving: a complex phenomenon studied at multiple levels

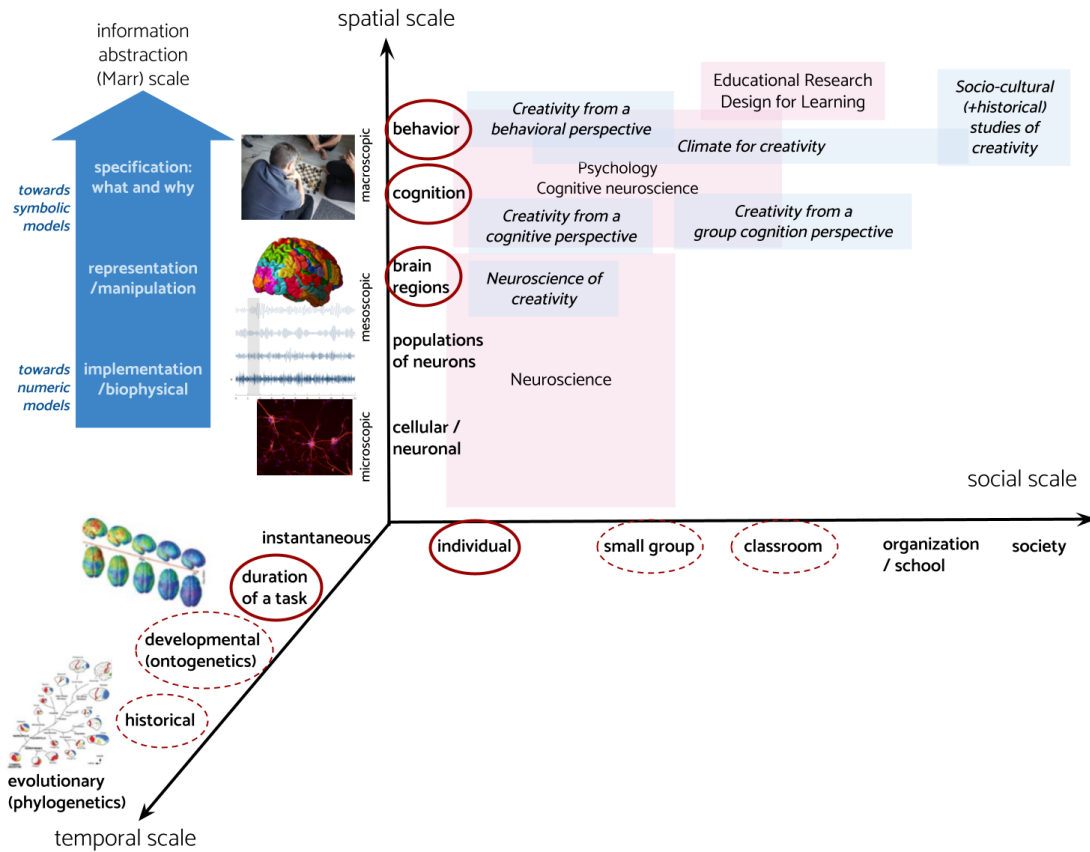


Figure 1.1: Multiple levels of analysis (adapted and synthesized from Romero (2022), Lodge et al. (2017), Betzel and Bassett (2016) and Kitcher (1988)). We consider here 3 dimensions to distinguish between different scientific angles to study the CPS process: (a) a spatial scale, (b) a temporal scale and (c) a social scale⁵. Beside the spatial scale, we put in perspective the levels of Marr as a scale of abstraction of information, in relation to symbolic and numerical modeling. The areas circled in red are the ones we are going to focus on in this thesis, while the dotted circled areas could be the closest potential perspectives of this work.

1 Creative Problem Solving: A tridisciplinary review

WHERE ARE WE?

A natural follow-up question is: at which level(s) is this thesis situated? Here, our geometric formulation of CPS is rather “micro” (within the scope of a task, performed by one individual), and we consider the creative process as a dynamic causality process, therefore being complementary to what is generally studied, according to [Kupers et al. \(2019\)](#) for children or [Thurlings et al. \(2015\)](#) at the level of teachers’ innovative behaviors and pedagogical innovation. From a computational neuro-cognitive point of view, we aim to give a symbolic specification of the task (Chapter 2) and investigate sub-symbolic to numerical approaches to better understand knowledge representation and reasoning within the CPS process, thus target the behavioral and cognitive level (Chapter 3) and to some extent, implementation within neuronal populations (Chapter 4).

1.1.3 DECOMPOSING THE CPS PROCESS

FIRST APPROACHES TO CPS MODELING Prior to the cognitive revolution and its advances in computational models ([G. Mandler, 2002](#)), the two main approaches to study were associationism and – starting from the 1930s – Gestaltism ([J. M. Mandler and G. Mandler, 1964](#)) (see also Appendix A). The former holds that cognitive processing in the mind consists of following a chain of “ideas”, that is, interconnected cognitive representations. This does not explain, however, how one can find a solution to a problem that has never been seen before. Conversely, the Gestalt approach considers cognitive representations as coherent patterns than separate ideas merely connected by associations. Gestaltists are interested in the nature of perception and how we can make sense of patterns, e.g., reconstitute the contour of an incomplete shape. Similarly, they hypothesize that we can also “see” the structure of a solution to a problem through a sudden illumination called **insight**, and point out this empirical phenomenon, despite not being able to explain why it occurs. These two theories are still underlying current approaches to CPS modeling, including more recent computational models based on information processing (see 1.4).

HOW MANY STAGES? Many attempts to decompose the CPS process have been proposed, as reviewed e.g. in [Howard et al. \(2008\)](#) and [Childs et al. \(2022\)](#).

One of the first models is the one of [Wallas \(1926\)](#), who identifies four stages:

- (i) **Preparation:** the problem is “investigated in all directions” ([Wallas, 1926](#)) in order to collect as much information as possible about the problem.
- (ii) **Incubation:** a period during which unconscious and involuntary mental processes occur while the learner is either relaxing or working on other problems.
- (iii) **Illumination:** mental processes that occurred during incubation spontaneously come together in the form of a recognizable and worthwhile idea, also referred to as **insight** by other authors (e.g., [Z. Zhang et al. \(2016\)](#)) and in this thesis.
- (iv) **Verification:** deliberate mental processes involved to convert the idea into its intended outcome.

1.1 Creative problem solving: a complex phenomenon studied at multiple levels

This model leans on the Gestalt approach by introducing the “illumination” stage, corresponding to a moment of insight. However, other authors such as Max Wertheimer (as reported in [Wertheimer \(2000\)](#)) argue that creativity can also be achieved through more deliberate ways: he suggests a distinction between “productive thinking” which is spontaneous (such as insight) and reproductive thinking whereby one proceeds algorithmically or by trial and error.

[Guilford \(1967\)](#)’s formulation of the creative process relies on two main subprocesses of divergent and convergent thinking, consistent with the definition of creative thinking as the act of producing new and appropriate ideas ([Guilford, 1967](#)). *Divergent thinking* (DT) is directed towards idea generation (fostering originality and novelty), while *convergent thinking* (CT) focuses on selecting an effective candidate solution to the problem (ensuring adequation). These two processes have been characterized by measurable indicators ([Jung et al., 2013](#)) despite some contradictions among studies which might be due to a lack of refinement in the process breakdown ([Dietrich and Kanso, 2010](#)).

Interestingly, these models of creative problem solving mirror the ones developed in the educational context, the most prominent one being [Polya \(1945\)](#)’s four-phase model consisting in:

- (i) Understanding the problem, related to estimating the present state and choosing a goal or subgoal;
- (ii) Devising a plan;
- (iii) Carrying out the plan;
- (iv) Looking back on the observed outcome compared to the predicted outcome.

These phases still resonate in the PISA framework for CPS ([OECD, 2017](#)) and similar phase-oriented approaches to studying problem solving as a competency in education. For example, [Zelazo et al. \(1997\)](#) considered the definition and representation of the problem mentally, the development of a solution strategy, the monitoring or regulation of the problem-solving progress, and the evaluation of the solution. The looking-back step leads to an iterative process of improving the solution, although this was not explicit in [Polya \(1945\)](#)’s presentation. The planning aspect of problem solving in this model, and the “incubation/illumination” creative stages of [Wallas \(1926\)](#), are both considered in creative problem solving.

In a nutshell, creative problem solving is thus decomposed into multiple cycles ([Alexandre, 2020a](#); [Dietrich and Haider, 2017](#); [Hennessey and Amabile, 2010](#)) of a DT phase in which new ideas or responses are generated, and a CT phase in which the appropriateness of these solutions is evaluated. As developed in more detail by [Amabile \(1996\)](#), these two phases are preceded by a preparation phase and ends with the creative outcome, as illustrated in Fig. 1.2. Other models have been proposed but most of them are built on the same core stages (see the recapitulative table of [Howard et al. \(2008\)](#) reproduced in Appendix B).

According to [Amabile \(1996\)](#), the problem presentation corresponds to the task identification, driven by internal or external **stimuli**; the preparation step involves building up

1 Creative Problem Solving: A tridisciplinary review

or reactivating relevant knowledge and **policies** or **heuristics**; the idea generation step corresponds to the divergent phase, that is, search in the memories and the immediate environment of possibilities; while the idea evaluation step refers to the convergent phase, that is, the response test against available knowledge and criteria; and the outcome corresponds to observing the expected achievement, considering that there is no reasonable response, or deciding that some progress has been made and looping back to one of the previous steps ⁶.

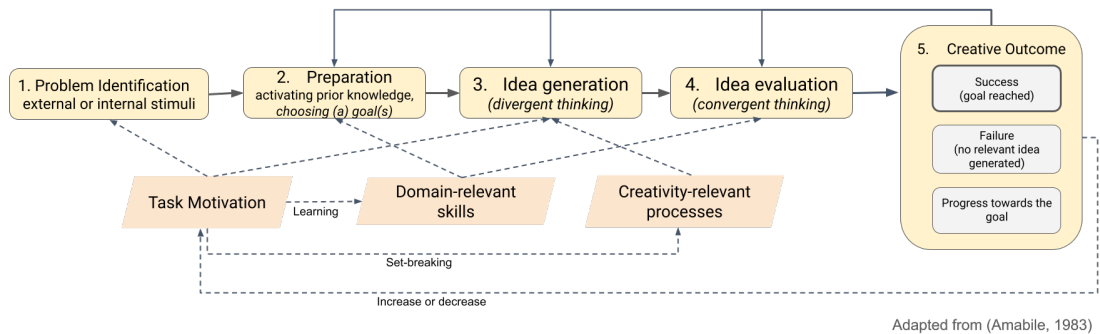


Figure 1.2: The componential model of creativity proposed by Amabile (1983), updated in Amabile (2012).

Creative elements are generated during the divergent phase: new ideas are generated in a way that may require the most obvious responses to be inhibited (which is referred to as “outside-the-box thinking”). Existing assumptions are questioned, and unexplored possible solutions are considered. This approach includes estimating the partially observable current situation, that is, which state we are in, and identifying the available resources and the possible actions to reach the goal, that is, what the possible path toward the chosen goal or subgoal is.

Ideas elaborated in the divergent phase are evaluated during the convergent thinking phase, and then the internal representation, and potentially the chosen goal, is reviewed or refined, as during the initial preparation phase, yielding an iterative process.

⁶This recurrent process was evoked as follows by Amabile (1983): “As tasks become more complex, the application of this outline to the production of creative responses on those tasks also becomes increasingly more complex. Work on any given task or problem may involve a long series of loops through the process, until success in a final product is achieved. Indeed, work on what seems to be a single task may actually involve a series of rather different subtasks, each with its own separate ‘solution.’ These subtasks may be hierarchically arranged, and the completion of any single subtask may in itself involve several runs through the process until success is finally achieved. For example, the superordinate goal of ‘writing a poem’ involves several subtasks, including finding a theme, deciding on a meter to use, choosing major and minor guiding images, inventing metaphors and similes, writing particular words, phrases, and lines. Each of these can be seen as a task or a subtask whose achievement is necessary for successful poetry-writing. Thus, success on a task depends in part upon the outcomes of subtasks and the difficulty of achieving success on those subtasks.”

1.2 CPS in learning sciences: from creative pedagogy to pedagogy of creativity

Amabile (1996) also considered creativity in context regarding task-intrinsic and task-extrinsic **motivation**, involved when engaging in the task during the problem presentation step and during the **DT** phase. Liu et al. (2016) considers that motivation is strongly involved during the preparation step, including goal setting, but also during the evaluation step. The componential approach also identifies domain-relevant skills versus transversal creative abilities, the former being mainly involved during preparation and evaluation and the latter during generation.

DUAL PROCESSES IN INTERACTION In the previously reviewed approaches, we have already identified some dual processes involved in creativity: spontaneous insight versus deliberate thinking, explicit versus implicit knowledge, convergent versus divergent thinking. In the following sections, we will review how these dual notions are understood in light of the learning sciences and cognitive neuroscience. A recapitulative table can be found in Appendix B.

1.2 CPS IN LEARNING SCIENCES: FROM CREATIVE PEDAGOGY TO PEDAGOGY OF CREATIVITY

Creativity in education is summoned at two levels. On the one hand, a creative approach to disciplinary learning at both the teacher and the student level is of great benefit, for instance, to optimize student engagement or to widen the ways to learn better (Leroy et al., 2021). On the other hand, creativity on its own is one of the important 21st-century skills and is thus a competence in itself (Engeness, 2020).

1.2.1 FROM PRIMARY TO SECONDARY KNOWLEDGE: GOAL SETTING FOR INTRINSICALLY MOTIVATED LEARNING

PRIMARY VS SECONDARY KNOWLEDGE How to teach creative problem solving is particularly challenging as it could be considered to fall under what is referred to as **primary knowledge**, as opposed to **secondary knowledge** (Lespiau and Tricot, 2022): according to this distinction made in evolutionary psychology, *primary knowledge* is “naturally” acquired because humans need it to survive and adapt to a changing world, whereas *secondary knowledge* includes the academic knowledge that requires more time and effort to learn (Sweller et al., 2019). However, while primary knowledge is easier to learn, it is also, by essence, mostly tacit in the sense of Nonaka (1994), that is to say “highly personal and hard to formalise, making it difficult to communicate to others or to share with others”. It may thus seem paradoxical to teach such a skill.

As a matter of fact, while being mostly tacit, creative problem solving also includes *methods*, that is to say, i.e. high-level procedures that can be applied to a large number of tasks, and can be explicitly taught to some extent ⁷. However, we postulate that such

⁷Such methods include, for instance, **analogical reasoning**, **deductive reasoning**, **inductive reasoning** or **abductive reasoning**; we detail these further in the sequel)

1 Creative Problem Solving: A tridisciplinary review

explicit methods are best acquired when integrated to an activity that does not explicitly appeals to them. Building secondary knowledge upon primary knowledge by presenting primary knowledge activities first, has been studied, for instance, in the context of teaching logical reasoning (Lespiau and Tricot, 2022). This study also seemed to illustrate the existence of a continuum between the knowledge types with, at one extreme, primary knowledge met in our direct environment and, at the other extreme, secondary knowledge; the intermediate step would be the primary knowledge that is not part of our direct environment (Roussel et al., 2017). This could justify the use of tangible objects that are unknown to the learner but include familiar elements (such as wheels or magnets) (Romero and Barma, 2022).

Here, we need to clarify the difference between a *task* and an *activity*: a task necessarily holds a goal, whereas an activity merely refers to being active and doing something, while the goal may be undefined or ill-defined. An activity may therefore include several tasks. In the activity theory, an activity is still associated with an outcome, but this outcome may have not been defined before the start of the activity (Sannino and Engeström, 2018). In the following, we may sometimes refer to learning activities as learning tasks when they hold a specific learning goal.

MOTIVES AND GOALS From a socio-cultural perspective, in problem-solving activities, the initial tensions of a problematic situation lead to cognitive dissonance between conflicting motives and related instruments to reach the goal (Sannino and Laitinen, 2015). These conflicting motives could be overcome through an auxiliary *motive* (e.g., an object of *motivation*) that helps people to make a decision (Sannino and Engeström, 2018). Goal setting thus involves choosing this auxiliary motive. A famous example is the experiment of making someone wait inside a room with a clock (Sannino and Engeström, 2018): most people will be conflicted between keeping on waiting or leaving the room, and choose a specific position on the clock as an auxiliary motive (e.g. “At 3:15, if nothing happens, I will leave”), thus “delegating” their decision to an external element. Such interactions between the agent and the environment, including socio-cultural dimensions, correspond to additional constraints to be taken into account when developing a computational model of problem solving, even if the possibility for taking such aspects into consideration will be rather limited for the model to be tractable.

Complex problem solving typically involves a hierarchy of goals, each goal yielding some sub-goals to achieve in order to complete the task (Eppe et al., 2022). We also wish to consider goal orientation (Pintrich, 2000), distinguishing between mastery goals and performance goals. Participants concerned with *mastery* are oriented to develop task-related self-improvement (Poortvliet, 2016), while those focusing on *performance* aim to achieve the task objective using “known ways to quickly implement knowledge and skills that have already been mastered” (Seijts and Latham, 2005).

We consider here a cognitive framework in which these goals can be organized around four fundamental questions (Alexandre, 2021) that relate to the learner’s cognitive functions: (i) *What* is the object of the goal? (ii) *Why* does it answer some current motivation or specification? (iii) *Where* is it located and how can it be accessed? And, (iv) *how* can

1.2 CPS in learning sciences: from creative pedagogy to pedagogy of creativity

it be manipulated and more generally addressed by a skill? This description of the goals represent an operational way to deal with all semantic, motivational and sensori-motor aspects of a goal to be considered in problem solving, consistent with what is known about associative neural circuits, as detailed in 1.3.

MOTIVATION AND ENGAGEMENT *Motivation* is the drive by which we pursue motives and goals. Traditionally, a distinction has been made between *extrinsic motivation*, oriented towards external motives such as the expectation of punishment or reward, and *intrinsic motivation* which derives from interest or curiosity in the activity itself. In this sense, the acquisition of **primary knowledge** is thought to be more intrinsically motivated, making it easier to learn, whereas the motivation to learn **secondary knowledge** is often more extrinsic (Lespiau and Tricot, 2022). In Moerland et al. (2018), extrinsic motivation is assimilated to homeostasis (that is, keeping some internal variables at a steady state, such as temperature or sugar level) and intrinsic motivation to appraisal (that is, the result of evaluations of incoming **stimuli** according to personal relevance), following the literature in affective computing.

This dichotomy has since been replaced by a more complex continuum of motivational styles from extrinsic to intrinsic (Molinaro and A. G. E. Collins, 2023). This is the case in self-determination theory (SDT), a theoretical framework for addressing human motivation that has shown empirical support over the the last decades (Ryan et al., 2022). As a matter of fact, contrarily to a common intuition, recent studies show that the difference between intrinsic and extrinsic motivation is not much correlated with academic performance (as reviewed e.g. in Sander et al. (2018)) ; however, overall motivation is correlated with academic performance but an even greater correlation exists between academic failure and amotivation, i.e., the absence of motivation (Denet, 2021; Sander et al., 2018). Therefore, fostering motivation remains crucial for the success of students.

Another point of interest is the level of attention and engagement into the activity, essentially theorized by the ICAP framework, which categorizes learning activities from passive to active to constructive to interactive (Chi, 2009). The ICAP hypothesis predicts that students learn more as they become more engaged with the learning materials, and has been empirically validated in the context of three specific engagement activities: note taking, concept mapping and self-explaining (Chi and Wylie, 2014). Following these footsteps, Romero, Laferrière, et al. (2016) have proposed an extension of this framework adapted to digital and technological learning activities.

1.2.2 ACQUIRING KNOWLEDGE THROUGH LEARNING ACTIVITIES

A TYPOLOGY OF KNOWLEDGE The acquisition of knowledge is tightly related to the tasks that involve it. In the context of instructional design, Musial and Tricot (2020) propose to consider not only the knowledge to acquire when designing a learning activity, but a couple (K, T_i) where where K is the knowledge needed to solve tasks T_i . Similarly, the couple (T, K_i) for a given task T and all K_i needed to achieve this task is to be considered when one focuses on the task itself. This formulation allows to define, on the

one hand, **declarative knowledge** (K, T_i) and **procedural knowledge** (T, K_i). This first dimension (that we could characterize as explicit-implicit⁸ is completed by considering a second dimension of knowledge, from the most task-specific to the most general. A typology of knowledge along these two dimensions is represented in Fig. 1.3. Such a typology is related to different types of memories in interaction, as reviewed in 1.3, and is useful to distinguish between different stages of learning and the processes involved: for instance, a learner may first *memorize* a literal trace, that is to say, a piece of knowledge that has been presented to them without understanding how to apply it in a given context (*comprehend*) or how to generalize it to other contexts (*conceptualize*). We can note that what we previously referred to as **methods** (e.g. analogical reasoning) are general procedural knowledge that can be explicitized (as opposed to automatism that are completely implicit).

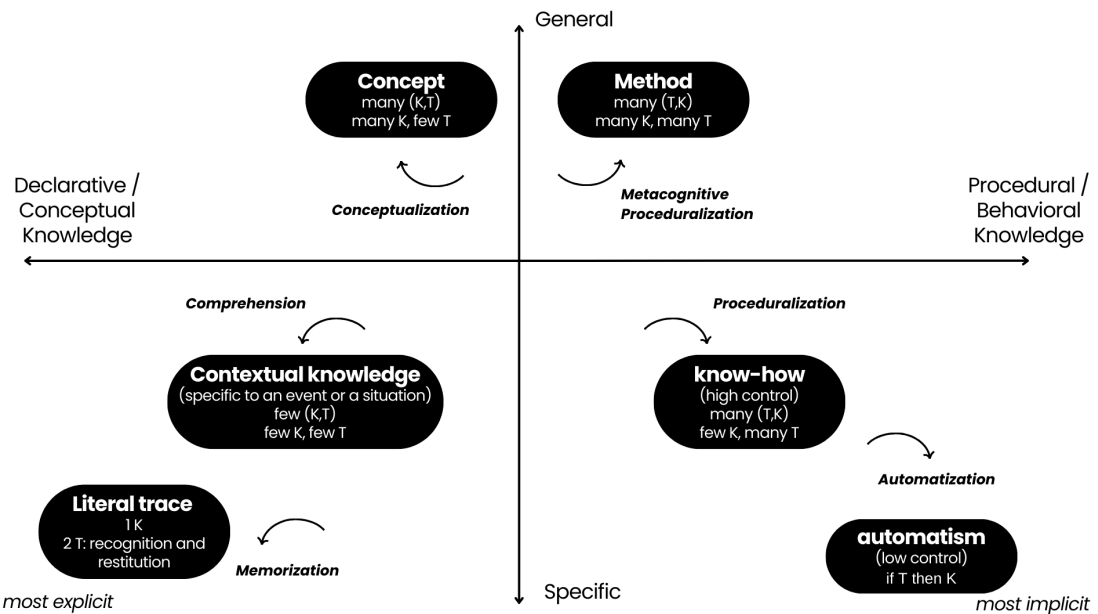


Figure 1.3: A typology of knowledge, adapted and translated from Musial and Tricot (2020). This typology follows two main dimensions: specific-general, and explicit-implicit. We here assimilate Musial and Tricot (2020)’s declarative and procedural knowledge to, respectively, conceptual and behavioral knowledge in the sense of Dillenbourg and Self (1992). We distinguish between Musial and Tricot (2020)’s two types of proceduralization by introducing metacognitive proceduralization, following Conway-Smith et al. (2023), to denote the process of forming **methods**.

⁸“Implicit cognition’ refers to unconscious influences reflecting perception, memory, and learning, without subjective phenomenal awareness” (Reingold and Ray, 2006); see also Appendix B, especially Table B.1)

1.2 CPS in learning sciences: from creative pedagogy to pedagogy of creativity

LANGUAGE AS A METACOGNITIVE TOOL The distinction between explicit and tacit knowledge raises the question of how language mediates this transition. Constructivism and social constructivism (see also Appendix A) highlighted the role of language in development, learning, and self-regulation: Piaget (1997) observed that young children tend to vocally describe their ongoing activities, and he suggested that such “egocentric language” has the purpose to coordinate sensorimotor processes. Vygotsky and Luria (1980) later added that egocentric language was not only a descriptive tool but also a way for children to generate novel goals. For Vygotsky, as reported by Engestrom (1987), language is progressively integrated as inner speech through *internalization*. Engestrom (1987) extends this definition of *internalization* as a process of embedding explicit knowledge into tacit knowledge. Language is thus a metacognitive tool, allowing to explicitize tacit processes that have been or are about to be internalized, and to set oneself some goals in order to plan one’s behavior. Language implies the manipulation of *symbols*, as reviewed in 1.4.

MONITORING LEARNING ACTIVITIES

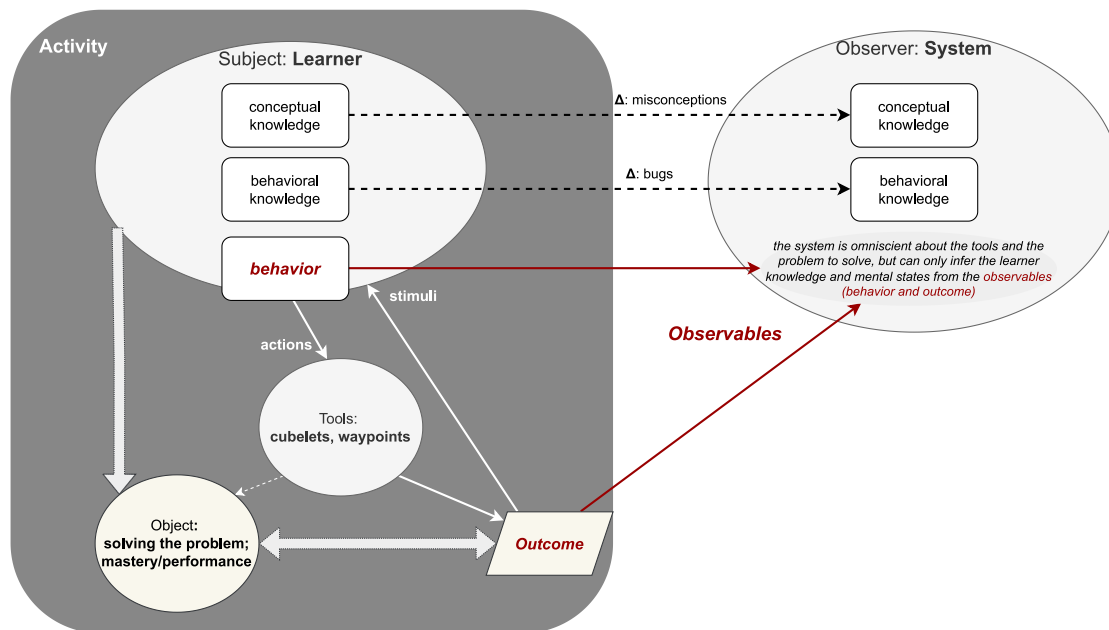


Figure 1.4: Framework chosen for the analysis of the observables, adapted from Dillenbourg and Self (1992) completed by the activity theory (Engestrom, 2000). This framework corresponds to the concepts from learning sciences taken into account for the construction of the ontology. The observables are the ones described in Fig. B in the Introduction.

Thanks to the development of tutoring systems, new frameworks have been developed for monitoring learning activities and modeling the learners within these activities.

1 Creative Problem Solving: A tridisciplinary review

Dillenbourg and Self (1992) proposed a general framework for learner modeling in problem-solving activities, considering two dimensions: on the one hand, they distinguished the system from the learner and defined concepts regarding the system's representation of the learner; and on the other hand, they established a distinction between behavior, behavioral knowledge, and conceptual knowledge. *Behavioral knowledge* refers to a set of rules and primitives that can be used to infer some behavior for a given problem, whereas *conceptual knowledge* contains the definition of the concepts underlying the behavioral knowledge. For the sake of simplification, we will assimilate these to the formerly defined *procedural knowledge* and *declarative knowledge* from Musial and Tricot (2020)⁹. This representation allows us to precisely define, for example, *misconceptions* (i.e., the difference between the learner's and the system's conceptual knowledge) or *bugs* (i.e., the difference between the learner's and the system's behavioral knowledge). In our case, even if we are not implementing a tutoring system, we are in the situation of a system observing the learner's behavior and inferring their knowledge. Moreover, the problem is ill-defined from the point of view of the learner, but it can be well-specified from the point of view of the system to guide the analysis of the observables, which is what we are trying to achieve here.

This framework does not really address the use of tangible artifacts in problem-solving activities. The activity theory (see e.g. Engestrom (2000) for a review) fills in the blank by introducing the notion of tools that play the role of mediator for the activity, which enables the learner (subject) to pursue a goal (object). Fig. 1.4 shows how we incorporate these notions into the learner modeling framework proposed by Dillenbourg and Self (1992). In CreaCube, the tools are the cubes and the waypoints, and the main object is to solve the problem (e.g., build a vehicle), which can be broken down into (possibly concurrent) several sub-goals. In what follows, we will also explicitize how conceptual and behavioral knowledge can be understood in light of the distinction that is made in cognitive neuroscience between different types of memory.

1.2.3 ASSESSING CREATIVITY IN CPS ACTIVITIES

CREATIVITY AS A REQUIREMENT FOR ILL-DEFINED PROBLEM SOLVING

Focusing on ill-defined CPS tasks corresponds to considering situations in which prior knowledge does not allow people to engage in conservative¹⁰ behavior (only exploiting existing knowledge and how-to) but requires the generation of novel ideas and their implementation to succeed in solving an unknown problem (Romero, DeBlois, et al., 2018). This is especially interesting in the context of digital interactive and manipulative technologies, which requires subjects of all ages to deal with unknown artifacts. In well-defined problem-solving tasks, more prior knowledge can support the identification of the problem and the use of existing relevant knowledge to solve the problem (Brand-Gruwel et al., 2005). Nevertheless, in ill-defined problems which require the ability to

⁹Here, conceptual knowledge includes high-levels concepts but also contextual knowledge, i.e. concepts with a limited scope that apply only to specific situations.

¹⁰As opposed to exploratory (Romero, DeBlois, et al., 2018).

1.2 CPS in learning sciences: from creative pedagogy to pedagogy of creativity

think outside the box, prior knowledge may inhibit early ideas of a solution (Leroy et al., 2021). In other words, we hypothesize that ill-defined problem solving requires creativity¹¹, as experimentally observed, except in the limited case of completely random behavior yielding (or not) a solution by chance.

CREATIVITY WITH TANGIBLE OBJECTS

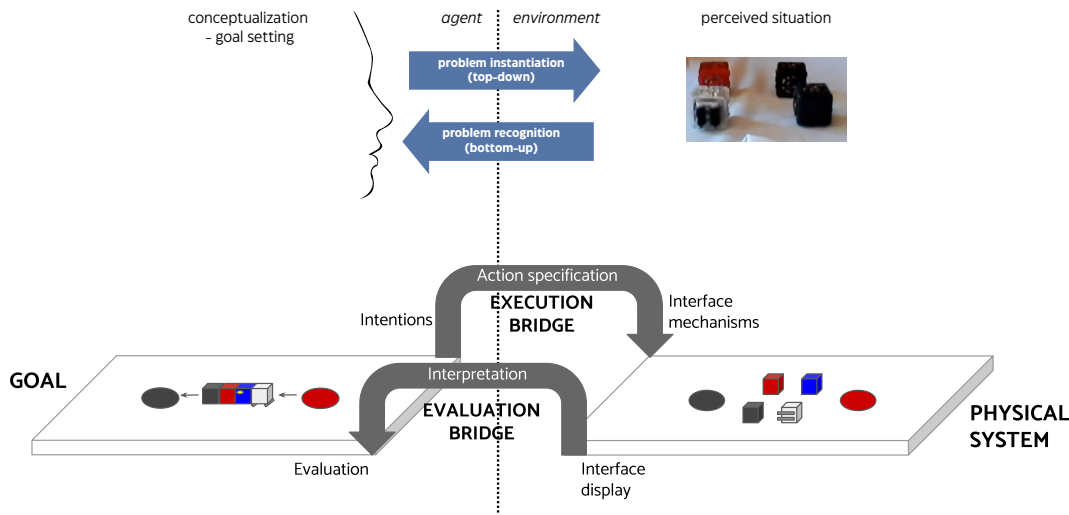


Figure 1.5: Top-down and bottom-up mechanisms at stake during the preparation phase in CPS, in relation the gulfs of execution and evaluation defined by Norman (2013). The bottom figure is adapted from Romero (2023).

Building with modular bricks is based on visuo-spatial constructive play objects (Ness and Farenga, 2007) used for engaging learners into tangible programming brick games (Kalmpourtzis and Romero, 2020; McNerney, 2004) with modular educational robots (Leroy et al., 2021; Romero, 2019). In the context of an ill-defined problem solving task (Schraw et al., 1995), the affordances of the robotic elements will influence the actions of the participant (Jamone et al., 2016) towards the objects (observing, grasping, assembling). Indeed, as opposed to a well-defined problem, the legal operators that can be considered to act on the problem states are not specified here. Instead, they are suggested by the physical features of the artifacts: for example, the blue cube has a switch on one of its faces, appealing to be activated. The exploration of one's environment (as developed in the next paragraph) aims to discover such *affordances* defined by Gibson (1977) as opportunities offered by the environment. Norman (2013) adapted this concept to the context of

¹¹Of course, this is not exclusive of the fact that well-defined tasks may also (when not “must” also) be solved with some creativity, this issue being beyond our present problem position.

human–computer interaction to refer to “action possibilities” that are readily perceivable by a subject. In Norman (2013)’s definition, affordances depend not only on the physical capabilities of the subject, but also on their goals, beliefs, and past experiences. Through their interactions with the task material, the learner relates the perceived physical properties of the artifacts to prior knowledge stored in long-term memories (e.g., a switch can be flipped to activate a device) and generates new hypotheses. **Affordances** are a way to make sense of **stimuli**, and a key point in the creative process is to be able to think of different appropriate ways to use the material to solve the problem, using mechanisms such as **analogy** to transfer knowledge from a task to another (Alexandre, 2020a; De Grave et al., 1996; Guilford, 1967). This is to be put in relation to the preparation phase in the sense of Amabile (2012), whereby one recognizes a problem from both the external stimuli and the internal goals and prior knowledge, thus combining both top-down and bottom-up mechanisms (as detailed in 1.3). Norman (2013) refers to the gulf of execution (gap between a user’s goal for action and the means to execute that goal) and the gulf of evaluation (gap between an external stimulus and the time a person understands what it means). These mechanisms are illustrated in Figure 1.5.

COMMON TASKS AND ASSESSMENT DIMENSIONS OF CREATIVITY

Several tests have been developed to assess human creativity, that have then been used in computational creativity evaluation. Some tasks, such as the **AUT** or the **RAT**, focus respectively on divergent and convergent thinking as described in the sequel, while batteries of tests such as the popular **TTCT** aim to consider multiple aspects of the creative process.

DT ASSESSMENT The Free Association Task, also called Verbal Fluency Test is a popular verbal **DT** task: starting from a given concept (category in VFT), subjects are asked to say whatever comes to their minds within a limited amount of time (see e.g. Denervaud et al. (2021) or Lopez-Persem et al. (2023)).

The *Alternative Uses Test* (**AUT**) (Guilford, 1967) is another popular **DT** task that consists in finding **affordances** as defined previously: participants are given an everyday object, and they have a few minutes to think of as many possible uses for that object as they can come up with.

In both of these tests, the participants are evaluated according to several criteria, the most common ones being the four dimensions originally introduced by Guilford (1967):

- (i) **Fluency** or **Fluidity**: The number of ideas the participant can come up with;
- (ii) **Flexibility**: The number of conceptual domains the answers relate to, or the number of distinctly conceptually different ideas the answers relate to;
- (iii) **Originality** or **Novelty**: Originality is generally a frequency measure that shows how uncommon the solutions are, as compared to the uses other participants came up with. Novelty is assessed by human judges;

1.2 CPS in learning sciences: from creative pedagogy to pedagogy of creativity

- (iv) *Elaboration*: How detailed these answers are (there is often a trade-off between *Fluency* and *Elaboration*).

The *Torrance Test of Creative Thinking* (TTCT) (Torrance, 1974; 1990) was built upon this work to conceive a battery of verbal and figural tasks, with or without stimuli. Tasks include, for instance, finding unusual uses of objects (similar to the AUT), inventing stories, improving a given product, sketch pictures starting from a given shape, etc. The TTCT is assessed using 5 norm-referenced scores, including Guilford's dimensions (discarding flexibility because of its high correlation with fluency), with the addition of *abstractness* (regarding titles in the figural tasks) and *tolerance to ambiguity* (referred to as *resistance to premature closure* by the author). The TTCT is one of the most well-known and widely used creativity assessment tools (Besançon et al., 2011; Cramond et al., 2005), although its validity has been questioned by some authors (Almeida et al., 2008; Kim, 2006).

In the CreaCube task, DT dimensions have been assessed as follows (Kohler and Romero, 2023): *fluidity* is computed as the number of configurations made, regardless of differences between configurations; *flexibility* refers to configurations that are different, for a given participant during the activity, from other configurations; and *originality* is computed considering configurations appearing in less than 5% of the totality of them.

CONVERGENT THINKING ASSESSMENT The TTCT includes tasks that appeal more to convergent thinking, such as figure completion, and the last versions introduce a checklist of 13 creative strengths (Torrance, 1974; 1990), some of them being not only related to DT (Alabbasi et al., 2022), e.g. *expressiveness* and *articulateness* in the context of story-telling, or *synthesis* in the context of incomplete figures (the ability to combine several incomplete figures and complete them within a same drawing).

Amabile (1982) proposed a method to evaluate both dimensions of creativity called the Consensual Assessment Technique (CAT), which relies on the domain-specific expertise of several judges to assess the novelty and the appropriateness of a products. This comes from the postulate that the average should counteract the subjectivity of the judges, however this test has not been used a lot in creativity research, making it hard to assess its validity (Jeffries, 2012).

Another test, purely verbal, is expected to focus only on the convergent thinking part, making it a kind of convergent counterpart of the Free Association Test: in the *Remote Associates Test* (RAT) (Mednick, 1968), the participants are given 3 words and are asked to come up with a fourth word that is associated to all three of them (e.g. manners, round and tennis → table).

CREATIVITY: A STABLE TRAIT OR A NEURAL PROCESS TO BE CONSIDERED IN CONTEXT?

Kupers et al. (2019) mentioned that, whereas children's creativity is often conceptualized as a static and stable trait, it is probably more realistic and fruitful to study it at the micro-level as a dynamic process with moment-to-moment interactions between the child and the environment. In addition, authors have claimed that interventions could be more beneficial if they were related to chains of causes and effects in the underlying processes.

1 Creative Problem Solving: A tridisciplinary review

Elucidating the neural bases of creativity is an interesting way to propose such a description at the micro and dynamic levels. In addition, such a study can help substantially in answering related questions, like understanding the influences between creativity and motivation or considering the role of behavioral disorders (e.g., hyperactivity or autism spectrum disorders) in children's creativity (see [Khalil and Moustafa \(2022\)](#) for a recent review).

1.3 THE NEURAL BASIS OF CREATIVITY

1.3.1 MEMORIES FOR LEARNING, RECALL AND IMAGINATION

CPS involves recalling knowledge and experience in order to recognize a problematic situation, generate ideas and apply adequate skills to find a solution. The role of memory in creativity has been extensively studied ([Gerver et al., 2023](#)), more precisely the interactions between memories that are of several types ([Alexandre, 2020b](#)). A distinction is traditionally made between short-term (i.e., working) memory and long-term explicit (or declarative) and implicit memories, with regard to how the information is stored in the brain. This supports the typology of knowledge presented in [Figure 1.3](#) and explains in more details the multiple sub-processes involved in human learning.

WORKING MEMORY *Working memory* is considered not simply as a system for the transitory storage of information but also as a processing system ([Cowan, 2008](#)). Incoming stimuli perceived by sensory modalities are encoded in a sensory "buffer" (sometimes called sensory memory, e.g. by [Cowan \(2010\)](#)) where they persist very briefly. *Stimuli* consist of external or internal information which is received and activates sensory areas, including *exteroception* (the perception of the situation by, for example, the visual or auditory cortical areas), *proprioception* (the perception of one's own body and movement) and *interoception* (the sensation of pain, pleasure, and emotions in the *insular cortex*¹²). Beyond immediate perception, regularities and rules are extracted from past experience, and different kinds of learning shape cerebral circuits' ability to be exploited in creative problem-solving tasks. After being filtered through attentional mechanisms, some of them can be temporarily retained in working memory for future use and manipulation. This mainly controlled by the *prefrontal cortex* (PFC). For example, when solving a problem, working memory makes it possible to store traces useful to the processes involved in carrying out the task, as well as updating them as the task progresses. Because these processes often require prior information, working memory also enables retrieval from longer-term memories.

SEMANTIC MEMORY Associative learning in the *sensory cortex* takes care of the multi-modal sensory dimensions of situations and learns to associate the most frequently

¹²In this work, we mention some major brain regions of interest regarding CPS. A popularized cartography of the human brain is reproduced in [C](#) to facilitate reading for non specialists.

linked. Building on this, the sensory cortex is associated with the elaboration of *semantic memory*, a declarative long-term memory in which the organization of knowledge extends beyond association: a hierarchy of concepts is built from experience and can be found in a semantic network, with both the idea of hierarchy (one specific class belonging to a more general class, like a dog being an animal, the relation being “is-a-kind-of”), and the idea of associated relations and properties (visual or auditory characteristics but also more abstract as a name or a link to the owner). It is also worth noting that, on the surface of the sensory cortex, the features and properties related to the identification of the corresponding “objects” are mapped onto the ventral regions (called the *what* pathway), whereas those related to their localization or use are mapped onto the dorsal regions (the *how* pathway). The associative nature of semantic memory is believed to be of primary importance in creative thinking (Dietrich, 2004; Jung et al., 2013) (see also Gervert et al. (2023) for a meta-review) as a function of both educational experience (Denervaud et al., 2021) and lifespan (Cosgrove et al., 2021).

EPISODIC MEMORY *Episodic memory* involves a cerebral region called the *hippocampus* (or more generally the hippocampal formation) and corresponds to memorizing *episodes*, i.e. multimodal neural activity patterns related to personally experienced situations within their spatio-temporal contexts and associated responses and outcomes. Episodes including a strong emotional dimension (for example associated with an error, novelty, or other kinds of reinforcement) correspond to stronger traces, whereas episodes in which nothing special happened have the tendency to be forgotten quickly. To store such information, the hippocampus receives as an input a compressed summary of the activities of most regions of the cortex at the moment of the episode and has a powerful mechanism associated with its recurrent architecture that allows the *binding* of the references to these multimodal elements to form a unique trace: frequently repeated segments of temporal inputs are concatenated into single conceptual units (Gobet and Sala, 2019). Episodic memory retrieval is a dynamic process that draws upon the sequential ability to reconstruct past experiences from corresponding cues, by a phenomenon called *replay* (Rabinovich et al., 2023). Later, when a similar episode (or a part of it, referred to as a cue) is experienced, the hippocampus is able to recall the full initial episode and to reactivate the corresponding cortical regions, yielding the same various sensations associated with that episode. Furthermore, when similar episodes are stored, an internal mechanism within the hippocampal formation allows the brain to detect that these episodes bind common features. In a process called *consolidation* (Eichenbaum, 2017), if this *binding* is not yet represented in the *semantic memory*, these episodes will be sent back by replay to the cortex during specific off-line moments (particularly during sleep), to train the cortex actively and help in the formation of new concepts binding these features, improving the quality of the information representation in the semantic network.

PROCEDURAL MEMORY *Procedural memory* is an example of implicit memory: it stores know-how and integrated procedures that have been automatized, such as motor skills.

1 Creative Problem Solving: A tridisciplinary review

These procedures are gradually learned through a transfer from declarative and working memories.

In problem-solving tasks, the brain figures out the problem situations by associating its sensory representations with a physical or mental behavior encoded in the motor and *premotor cortex* (respectively for elementary and integrated action plans) and in the *orbito-frontal cortex* (for giving an emotional value to a situation and making a decision accordingly). More precisely, as previously mentioned, this goal-directed organization can be broken down into *what-why-where-how* loops (Alexandre, 2021) in which the value of an action is evaluated in terms of general preference (*what*) and motivation (*why*) to decide both a general goal and a next-step sub-goal, allowing the generation of rules for action (*where* and *how*).

The elaborated multiscale distributed representation of sensory information is associated through training with a variety of behaviors, corresponding to responses that can be given in the physical and mental worlds. What is specifically considered here is the capacity to anticipate the resulting situation when the response is triggered. This is the case for procedural learning, mainly associating the motor and *premotor cortex* with the dorsal (*how*) cortical regions. Throughout this learning, the consequences of actions in the real world can be anticipated. This gives rise to the concept of *affordances*, whereby the perception of a situation can pre-activate possible actions and anticipate their outcomes (Gibson, 1977). Each object is always associated with some “what to do with” properties. This is also the case for respondent conditioning, associating the *orbito-frontal cortex* with the *what* cortical regions. Here, the decision to give a certain emotional value to a situation allows us to anticipate the corresponding reinforcement (reward or punishment).

In both cases, such learning involves loops associating the cortex with the *basal ganglia*¹³. After a certain time, this process results in what is called the dominant behavior, the behavior generally triggered in the corresponding situation. After extensive training, it can even become a habitual behavior in the sense that the response is automatically triggered when the situation is perceived, with no anticipation of the forthcoming outcome. This response is called *stimulus-based* behavior, as traditionally opposed to *goal-directed* behavior. It can be remarked that such automatic behaviors, which are very frequent in a stable and predictable world, are exactly the opposite to what is sought in a creative intention oriented towards novelty and appropriateness. Therefore, in most cases, cognitive control is needed to inhibit the dominant behavior and promote less usual – goal-directed – behaviors, as detailed further.

1.3.2 NEURONAL NETWORKS IN INTERACTION DURING CPS

Observing brain activity with imaging devices (e.g., fMRI and MEG) during elementary cognitive tasks has led to the definition of large-scale brain networks, associating widespread brain regions. Among them, three are particularly active in some steps of

¹³This learning involves dopaminergic firing and is often modeled by *TD Reinforcement learning*, taking the prediction error as reward.

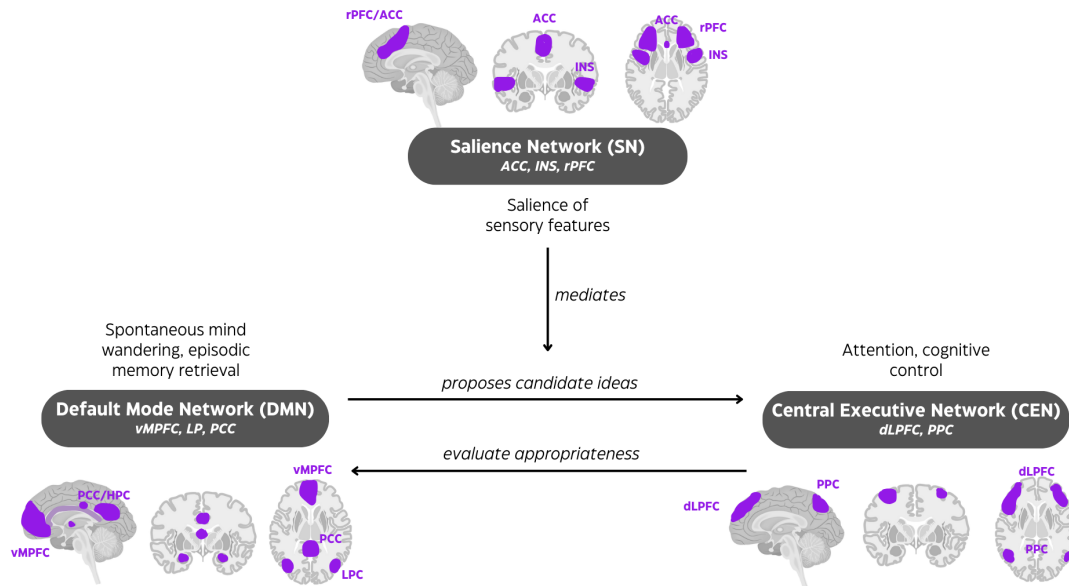


Figure 1.6: Large-scale brain networks involved in CPS, according to [Beatty et al. \(2016\)](#) and supported by [Dietrich \(2004\)](#), [Jung et al. \(2013\)](#) and [Kounios and Beeman \(2014\)](#).

The **DMN**, involved in mind wandering and idea generation, is mainly composed of the ventro-medial prefrontal cortex (vMPFC), the lateral parietal cortex (LPC) and the precuneus cortex (PCC). Thanks to the connection of the PCC to the hippocampal regions (precisely the caudal part of the parahippocampus), it enables episodic memory retrieval and replay.

The **CEN**, involved in cognitive control, is mainly composed of the dorso-lateral prefrontal cortex (dLPFC) and the posterior parietal cortex (PPC). There is a slight asymmetry between the right and left hemisphere, the dLPFC and PPC of the dominant hemisphere (left here) being more active in convergent mode.

Finally, the **SN**, involved in attention and control, is mainly composed of the rostral prefrontal cortex (rPFC), the anterior cingulate cortex (ACC) and the **insular cortex** (INS - primarily the anterior insula AI).

1 Creative Problem Solving: A tridisciplinary review

the creative process (Beaty et al., 2016). The *default mode network* (DMN) corresponds to brain regions that are active during spontaneous thought and mind wandering. It includes regions surrounding the hippocampus (precisely the precuneus cortex), together with the ventro-medial prefrontal cortex and the lateral parietal cortex; a primary function of this network is *episodic memory* retrieval. The *central executive network* (CEN) includes the dorso-lateral prefrontal cortex and is activated for the control of sustained attention and planning, thus playing a key role in *working memory*. The *salience network* (SN) mainly consists of the *insular cortex* and the *anterior cingulate cortex* (ACC). It monitors the salience of *stimuli*, integrating a variety of internal and external information. It is hypothesized to play the role of mediator activating a switch between the DMN and CEN (Uddin, 2015). Beaty et al. (2016) reported a major association between the DMN and the CEN during creativity, in which the DMN proposes candidate ideas (divergent phase) and the CEN stands for the evaluation of their appropriateness (convergent phase).

SPONTANEOUS VERSUS DELIBERATE PROCESSES A distinction is made between the spontaneous emergence of an idea, as a “bounce” effect of the creative approach, and a deliberate search for new ideas, including questioning existing beliefs or knowledge about the problem to solve. In a more spontaneous mode, new ideas are generated from the default network and episodic memory (Kounios and Beeman, 2014), subsequently evaluated by the CEN. In a more deliberate mode, the SN and CEN can take into account the characteristics of the task and orient the DMN toward a more systematic exploration of possible ideas. Dietrich (2004) related the first mode to *intuition* and discussed the correlation of age with this dual process of imposing constraints to be appropriate and relaxing them to be creative, with a more spontaneous approach observed in childhood. Kounios and Beeman (2014) also reported a neural inhibition of visual inputs by the CEN to reduce distracting inputs, facilitate the retrieval of weak solutions, and evoke actors influencing *insight* (see below), like mood (positive affects and reduced anxiety favoring a long-term view and broadened semantic processing; see also Diamond and Ling (2016) on these topics).

IDEA GENERATION Jung et al. (2013) insisted on the important role of episodic memory in the replay of retrospective and *prospective memories* and of interoception in the insular cortex to set the selection of replays toward more original ideas. Amnesic patients suffering from episodic memory loss due to hippocampal damage have been reported to perform poorly on the TTCT (Jung-Beeman et al., 2004). As a matter of fact, the *hippocampus* is not only able to recall a full initial episode or an episode with common features but episodes also able to replay a partial episode or composite episodes made up of several pieces, in other words to predict unrealized episodes through simulation (Stachenfeld et al., 2017). The *prefrontal cortex* plays a fundamental biasing role towards hippocampal replay, these two structures being deeply interconnected (Eichenbaum, 2017). In one direction, the hippocampus can provide arbitrary cue binding to the PFC, allowing to refine the contexts in which the dominant behavior should be inhibited and replaced

by specific rules. In the other direction, the prefrontal cortex can control the retrieval of memories in the hippocampus in certain contexts by suppressing the recall of inappropriate memories. Altogether, this forms the basis of *prospective memory* (Buckner, 2010), a phenomenon whereby the prefrontal cortex can control the hippocampus step by step to make it produce a virtual (i.e., not really experienced) trajectory within the mapping of previous episodes. This generative capability of the hippocampus thus allows for imagination and generation of candidate ideas, for which metacognitive judgement will help selecting those that can lead to an efficient and creative process or outcome. Altogether, as pointed out by Schlichting and Preston (2015), it is interesting to observe that imagination is achieved through the recombination of prior memories and is consequently not so “novel.”

REGULATION Such recombinations are performed under the control of both spontaneous and deliberate regulatory processes, allowing to behave in a creative way instead of reusing existing habitual behaviors, especially when these become inappropriate. At this stage, a rather subtle difference is to be made between regulation and evaluation: during the **DT** phase, the role of the deliberate system is not to evaluate the result of the spontaneous generation process but rather to regulate the way in which the generation process is performed (Tubb and Dixon, 2014).

This involves judgment along both “hot” (affective) and “cold” (analytic) dimensions (Dietrich (2004), emotion allowing the comparison of different signals using emotional value as a common currency (D.J. Levy and Glimcher, 2012). Tubb and Dixon (2014) maps this distinction onto an explicit-implicit dimension: implicit judgement is fast and affective, resulting in what is sometimes called a “gut feeling”, whereas explicit judgement is more demanding but provides additional information such as the value function gradient.

IDEA EVALUATION Regarding the control of idea appropriateness, the same distinction between the ventral (*what*, semantic representation) and the dorsal (*how*, perception for action) regions of the **sensory cortex** can be made for the **PFC** (Bunge, 2004). In the PFC, the former is in charge of biasing the activity of the ventral part of the sensory cortex, setting the focus on “objects” and dimensions that will be of specific interest for the contextual rule, and keeping the representation of the desired goal active. The latter takes care of the organization of the generated behavior, including over time, so that the global behavior obeys the constraints of the ongoing task and displays appropriate temporal characteristics. More specifically, Dietrich (2004) reported that the dorsal lateral **PFC** is involved in its syntactic aspect, whereas the dorsal medial **PFC** is for the control and inhibition of common ideas (Maysless et al., 2015) (also useful in the regulation of **DT** phases, as reviewed earlier).

In addition, the posterior to anterior axis of the lateral **PFC** seems to correspond to a concrete to abstract axis. At the *how* level, Badre (2008) observed that more concrete and proximal in time rules are represented on the posterior side, whereas more anterior regions display more abstract and temporally extended rules, leading to a hierarchical

1 Creative Problem Solving: A tridisciplinary review

view of the representation of rules, in which more abstract rules can control more concrete ones. At the *what* level, this concrete to abstract axis corresponds to the level of abstraction of the representation (O'Reilly, 2010).

INTUITION AND INSIGHT As far as the spontaneous aspect is concerned, a distinction is made between *insight* (the *aha* or *eurêka* effect), which is a sudden and unexpected apprehension of the solution, and *intuition*, which is mainly characterized by reaching, more gradually, a solution with neither an explicit representation nor a clear explanation of how it emerged (Zander et al., 2016). Z. Zhang et al. (2016) observed that *insight* is related to a sudden appearance of *what* could be the solution, after an implicit restructuring of mental processes and tacit knowledge associations; whereas—in experimental paradigms—*intuition* is more related to the spontaneous selection of a solution among alternatives, with an emerging perception of coherence (Zander et al., 2016). Using the RAT task, Jung-Beeman et al. (2004) found a gamma-band oscillation associated with conscious retrieval in insight problem solving. In addition, they observed an alpha burst preceding the gamma burst, which may reflect unconscious solution-related processing (Jung-Beeman et al., 2004). In line with this observation, using fMRI, W. Zhang et al. (2020) reported that a diffuse and weak activation of the superior temporal gyrus (STG) precedes the moment of insight, which is followed by a strong activation of STG in coherence with the burst of positive surprise (as illustrated on Figure 1.7).

Taking a step further with regard to *metacognition*, Metcalfe and Wiebe (1987) examined the difference in metacognitive judgment processes between insight and non-insight problem solving by using metacognitive judgments such as “warmth” ratings of perceived closeness to the solutions. Their results of higher warmth ratings during non-insight problem solving than during insight problem solving support the view that the unexpected realization of the solution is a distinguishing characteristic of insight problem solving. More recently, Gilhooly and Murphy (2005) have observed that deliberate processes are required in both insight and non-insight problem solving, but their roles differ: specifically, deliberate processes are more involved in strategy switching and inhibition in insight problem solving, whereas they support systematic search processes in non-insight tasks.

1.3.3 COGNITIVE CONTROL: FROM BEHAVIOR PERSISTENCE TO FLEXIBILITY

Here, let us first underline the difference between persistence / flexibility with regard to a goal, and persistence / flexibility with regard to a behavior. In creative behavior, we are aiming for flexible behavior (possibly inhibiting the dominant behavior) while staying persistent towards a goal, ensuring a novel and appropriate response. The common distinction between *stimulus-based* and *goal-directed* behaviors thus needs a little more nuance, since creative behavior should always be more or less goal-directed. Instead, let us review in more details the top-down and bottom-up mechanisms involved in attention and allowing to bias the behavior with regard to goal persistence.

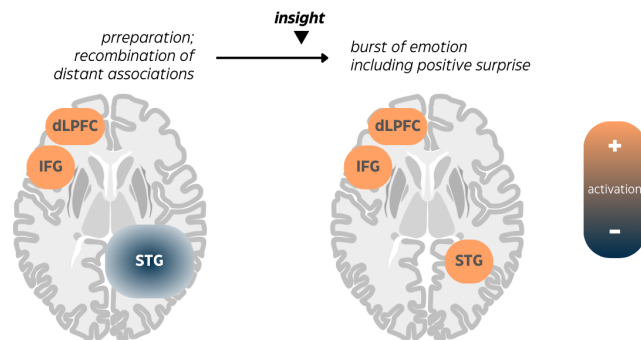


Figure 1.7: Metacontrol of insight, adapted from [W. Zhang et al. \(2020\)](#). During an insight trial of the RAT, [W. Zhang et al. \(2020\)](#) reported a diffuse and weak activation of the superior temporal gyrus (STG) preceding the moment of insight, which is followed by a burst of positive surprise and a strong activation of STG.

TOP-DOWN VERSUS BOTTOM-UP BIAS

Following [Ede et al. \(2020\)](#), the distinction proposed between **goal-directed** and **stimulus-based** behaviors is highly related to sources of attentional selection, that is, to focus voluntarily on things that are relevant to our goals rather than involuntarily capturing salient events in the external world. These two sources jointly influence the selection of internal memory representations.

It is thus not possible to restrain stimulus-based mechanisms to a simple bottom-up process because attentional mechanisms the brain are, from the beginning, modulated by top-down processing: in the posterior (temporal/parietal) cortex, this top-down process is modulatory and maps the injection of prior information onto the involved perceptual stimulus processing ([Friston, 2003](#)). These attentional mechanisms are of primary importance during the preparation phase as defined in [Figure 1.2](#).

Following this track, in line with [W. Zhang et al. \(2020\)](#)'s metacontrol approach, both persistent (i.e., characterized by a strong top-down bias and competition between goals) and flexible (i.e., characterized by some weak top-down bias and weak competition between goals) behaviors are goal-directed behaviors. Furthermore, goal-directed behavior is hierarchical as soon as the problem-solving task becomes complex, as pointed out by, for example, ([Eppe et al., 2022](#)).

According to [W. Zhang et al. \(2020\)](#), creative cognition in divergent and convergent thinking is modulated by metacontrol states, as described in [Fig. 1.8](#). **DT** and insight solutions are enhanced by goal flexibility whereas **CT** seems to benefit from metacontrol biases toward goal persistence.

To this end, the **prefrontal cortex** is a major cerebral region to behave beyond stimulus-driven dominant behavior and to promote more flexible behavior driven by internal analysis. This is achieved through two internal processes ([O'Reilly et al., 2014](#)), encountered previously: the medial part monitors errors, suggesting that the dominant be-

1 Creative Problem Solving: A tridisciplinary review

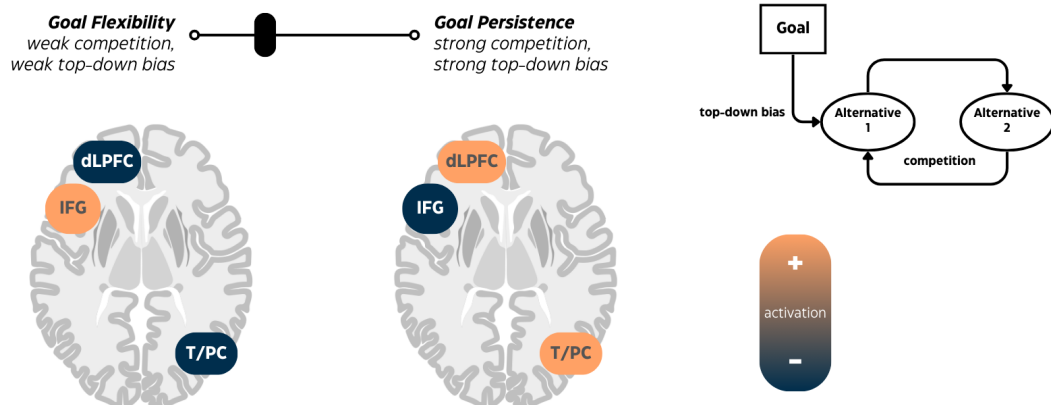


Figure 1.8: Metacontrol between goal flexibility and goal persistence, adapted from [Hommel and Wiers \(2017\)](#) and [W. Zhang et al. \(2020\)](#). Flexibility towards the goal is characterized by a weak top-down bias and a weak mutual competition between alternative decisions. [W. Zhang et al. \(2020\)](#) report that flexibility is promoted by a weak activation of the dorsolateral prefrontal cortex (dLPFC) together with the temporal/parietal cortex (T/PC) while the inferior frontal gyrus (IFG) is strongly activated in the dominant hemisphere (represented on the left here) in comparison with the non-dominant hemisphere. The opposite pattern is observed for goal persistence.

havior might not be adopted in certain circumstances, and aims to predict when these errors might occur to inhibit the dominant behavior accordingly; while the lateral part learns new contextual rules, which are better adapted to new specific cases, based on interactions with episodic memory.

In both processes, this is implemented with the specific mechanism of **working memory**, in which prefrontal neurons display sustained activity to evaluate histories of activity, maintain constant inhibition of the dominant behavior, and bias the activity of the sensory cortex to promote other more adapted rules instead. In particular, the biasing might correspond to increasing the saliency of some (classically supposed) minor features to orient the behavior toward responses dealing with them. Behavior switching is much more efficient than learning and unlearning a new behavior and is crucial in CPS.

The prefrontal cortex implements biasing activity to orient other cortical regions toward non-dominant behavior, rather than implementing a behavioral rule *per se*. The consequence is genericity, and, in the framework of reasoning by analogy, for example, it should be rather easy to adapt the contextual rule to another similar biasing of activity in another cortical region. The **task set** model reviewed below proposes a model for such behavior switching.

TASK SWITCHING: BETWEEN EXPLORATION AND EXPLOITATION

Task sets (see, e.g., Donoso (2013) for an introduction) are basic executive units subserving cognitive control, formalised as instantiating (hidden) states of the environments corresponding to stable external contingencies over time. Task-sets are temporal abstraction including *action sets* which in turn comprise a *selective model* mapping stimuli onto actions ($S \rightarrow A$) and a *predictive model* predicting action outcomes from stimuli ($S, A \rightarrow O$). The selective model can be organised hierarchically according to additional cues (C), the likelihood (LH) of which is encoded into a *contextual model* given external states associated with task sets. The anatomical mapping of these functions onto the frontal lobes is illustrated on Fig.1.9. *Task sets* thus constitute an operational model of cognitive control to promote more flexible behaviors, which can be implemented and simulated within computational models using *reinforcement learning*, as developed in the next section 1.4.

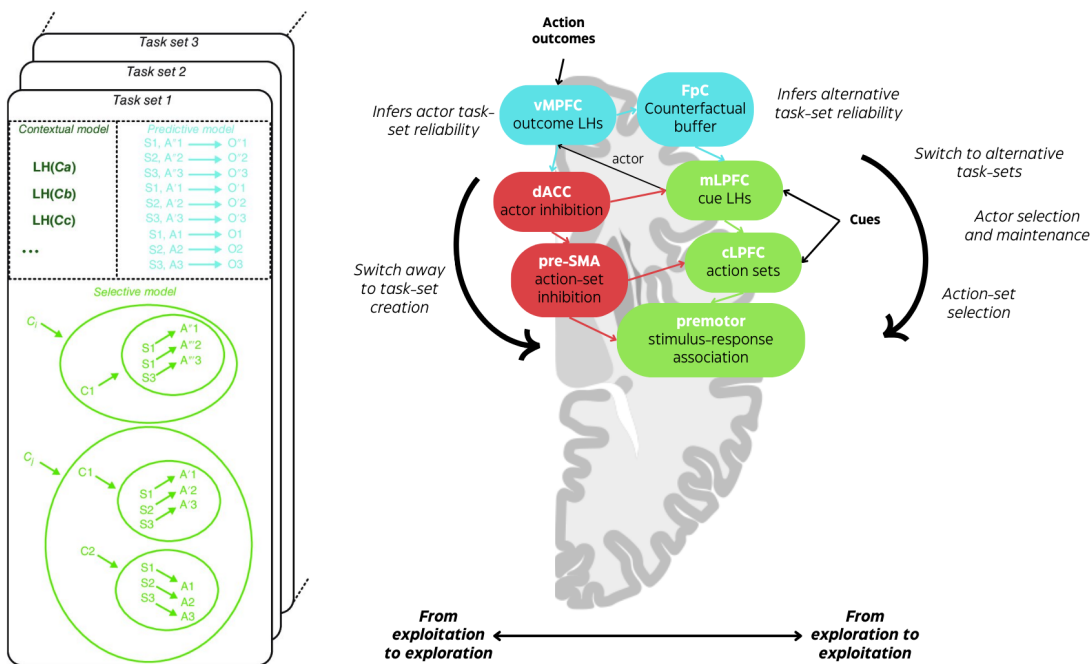


Figure 1.9: Task switching, adapted from Domenech and Koehlin (2015) and Duverne and Koehlin (2017). The lateral regions comprise the fronto-polar cortex (FpC), the middle and caudal lateral prefrontal (mLPFC and cLPFC) and the premotor (PM). The medial regions comprise the *orbito-frontal* (not represented here but involved in giving an emotional value to a situation and making a decision accordingly), ventromedial prefrontal (vMPFC), the dorsomedial prefrontal (dMPFC) including the dorsal *anterior cingulate cortex* (dACC) and the pre-supplementary motor area (pre-SMA).

Having highlighted the main psychological and neuro-cognitive notions inherent to creative problem solving, let us now review computational approaches aiming to operationalize such mechanisms.

1.4 COMPUTATIONAL APPROACHES FOR CPS: FROM SYMBOLIC TO NEURAL FORMALISMS

Following the cognitive revolution in the 1950s (G. Mandler, 2002), novel cognitive and computational approaches emerged to study creative problem solving.

This implicit/explicit duality reviewed earlier can be paralleled to an analogous duality between symbolic and numerical AI (Garcez and Lamb, 2020), the former relying on symbolic knowledge representation and reasoning, while the latter corresponding to sub-symbolic or numerical computation (as it is the case in most machine learning algorithms, including deep learning).

While the first proposed computational models for problem-solving relied on symbolic manipulation (e.g., Newell and Simon (1972) as introduced below), novel approaches in deep learning display impressive capabilities for playing non-trivial board games such as Go¹⁴ in which symbolic rules are not enough to account for intuitive recognition of potentially winning configuration (in a manner similar to *insight*). Furthermore, recent advances in generative models, such as generative artificial neural networks, diffusion models and transformers (Epstein et al., 2023), enabled the release of DALL-E¹⁵ or Midjourney¹⁶, which are able to provide “creative” (as in novel and appropriate) responses to a textual prompt in a pictorial form (photographs or paintings), or in the case of ChatGPT¹⁷ in a verbal form (e.g., poems).

For this computational review, we chose to focus on models that do not need extensive data training (hence avoid generative and deep-learning approaches) for reasons we develop in Chapter 5. Nonetheless, as reviewed in Garcez and Lamb (2020), recent advances in AI propose to integrate both symbolic and numerical approaches allowing for fruitful interactions between the two paradigms. We thus focus on symbolic paradigms for which a sub-symbolic representation could be considered, allowing to use numerical computations in a frugal way while still considering biological plausibility (Kajić et al., 2017; T. Stewart and Eliasmith, 2011)

1.4.1 THE INFORMATION PROCESSING APPROACH

THE PROBLEM SPACE HYPOTHESIS

In *Human Problem Solving*, Newell and Simon (1972) proposed an information processing approach to problem solving; along with a formalization of well-defined problems. This

¹⁴As performed by AlphaGo (<https://deepmind.google/technologies/alphago/>)

¹⁵<https://openai.com/dall-e-2>

¹⁶<https://www.midjourney.com>

¹⁷<https://openai.com/chatgpt>

approach, based on the computer metaphor, consists of applying a search algorithm to a mental representation of what they call the **problem space** (1.4.1). They illustrated this formalization with the well-known problem of the Tower of Hanoi (Newell, 1981) (Figure 1.10).

With this formulation, solving the problem thus consists of searching the problem space using *heuristics* in its computational sense¹⁸, that is to say, functions that rank alternatives in search algorithms at each branching step based on available information to decide which branch to follow. A heuristic may not always give the best solution but allows to approach it when exploring the whole problem space would be too greedy.

Problem space characterization in well-defined problems

For Newell (1981), the *problem space* consists of:

- a set of symbolic structures called the *states* of the space
- a set of *operators* over the space, i.e. functions (possibly partially) defined over the state space and producing states as outputs. Sequences of operators thus define *paths* that unfold into sequences of states.

A *problem* is specified by:

- a set of initial states,
- a set of goal states,
- a set of path constraints,

and consists of finding a path in the problem space, starting at any initial state and ending at any goal state while satisfying the path constraints (Newell, 1981).

The interest of such an information processing approach is that it allows to test problem solving algorithms using computer simulations: Newell and Simon (1972) have implemented such problems and heuristics allowing to solve them within the General Problem Solver (GPS), which would later on inspire other computational cognitive ar-

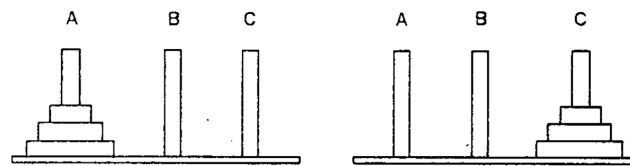
¹⁸In psychology, heuristics are more considered as mental shortcuts that ease the cognitive load of making a decision. Such shortcuts include using trial and error, a rule of thumb or an educated guess, but to our best knowledge, there is no consensus on a typology or precise definition. Polya (1945) propose several examples of heuristics, quoting: “

- If you are having difficulty understanding a problem, try drawing a picture.
- If you can't find a solution, try assuming that you have a solution and seeing what you can derive from that (“working backward”).
- If the problem is abstract, try examining a concrete example.
- Try solving a more general problem first (the “inventor’s paradox”: the more ambitious plan may have more chances of success).

”

chitectures such as SOAR (J. Laird et al., 1986; J. E. Laird et al., 2017) or ACT-R (J. R. Anderson et al., 1997).

The key principle, which the GPS relies on, is the *physical symbol system hypothesis* (PSSH) (Newell and Simon, 1972), that is, the problem space is represented as units of information called **symbols** that the GPS can manipulate and combine into structures (expressions) using logical rules. This is the basis for what we now call symbolic artificial intelligence.



Task and Problem Statement

A board has three pegs, A, B and C (as shown at left). On Peg A are N disks of different sizes (3 in the diagram), in order with the largest at the bottom. The task is to get all the disks on Peg C in the same order (as shown at right). A disk may be moved from any peg to another, providing (1) that it is the top disk on its peg and (2) that it cannot be put on top of a disk smaller than itself.

Figure 1.10: The problem space, illustrated on the Tower of Hanoi, reproduced from Newell (1981). The definition from 1.4.1 applies as follows:

- *States:* Arbitrary configurations of the N disks on the three pegs.
- *Operators:* (i) Move a disk by removing it from a peg and putting it on another peg; (ii) Recognize a configuration as an instance of a pattern.
- *Initial state:* The configuration shown in the diagram on the left.
- *Goal state:* The configuration shown in the diagram on the right.
- *Path constraints:* No disk may be placed on a smaller disk.

Despite being a significant advance in the computational study of problem solving, Newell and Simon (1972)'s approach remains limited to well-defined problems such as the Tower of Hanoi. In the next section, we propose to extend this formalization to ill-defined problems. Another obstacle is the combinatorial explosion which may arise with a very large problem space (*a fortiori* in ill-defined problems). This is where divergent and convergent processes may come into play to find heuristics that will narrow down the space to explore. Finally, the symbolic representation raises the non-trivial question of *semantic grounding*, that is to say, how to make sense of symbols and relate them to sensorimotor processes, as reviewed in the sequel.

1.4 Computational approaches for CPS: from symbolic to neural formalisms

A GEOMETRICAL DEFINITION EXTENDED TO ILL-DEFINED PROBLEM SOLVING

CONSIDERING CREATIVITY: THE CONCEPT SPACE When attempting to define creativity, Boden (2004) introduced another notion called the *concept space*, that Wiggins (2006a) and Ritchie (2012) have formalized into computational terms. The idea of the concept space is that generating new creative concepts can be considered as a search process similar to the ones used in problem solving within the problem space. Does that mean that the concept space and the problem space refer to the same notion? According to Wiggins (2006b), the creative system framework is more a generalisation of the problem space search into a geometrical space: in particular, it admits expansion, which requires an **open-world** assumption to allow for the creation of new ideas, and uses the notion of relative comparison for evaluation rather than absolute Boolean values. In the following, we will review what kinds of creative search algorithms can be performed in such a space. From now on, when referring to the problem space, we use a generalized geometrical definition allowing for such extensions.

CONSIDERING ILL-DEFINED TASKS The spatial model of problem solving extends to complex problem-solving tasks in the sense of Eppe et al. (2022). In such a case, a complex problem is solved in several steps by generating subgoals on the trajectory. Eppe et al. (2022) considered both a temporal modular approach (i.e., a sequential decomposition of the path) and a spatial modular approach (i.e., a multi-scale decomposition of the state space), in addition to concurrent policies, to solve the problem (i.e., alternative routes in the problem space), yielding hierarchical decompositions.

Beyond complex problem-solving tasks, we are interested here in ill-defined problems for which we propose a computational formulation following Newell and Simon (1972)'s criteria, given below. In such tasks, it is not possible to apply a predefined **heuristic** or **policy**, that is, a predefined set of rules allowing learners to choose the next action given the observed state, from the initial to the final state. Indeed, understanding the initial state, choosing some final goal parameters, and adapting the trajectory generation as a function of the evolution of the problem state are also part of the task. The solution is thus adaptive and emerges from an important process of disambiguation throughout the problem-solving task.

PROBLEM SOLVING AS TRAJECTORY GENERATION We summarize here the geometrical approach of problem solving as the selection of a trajectory within a (possibly continuous) problem space, as illustrated in Figure 1.11. Interestingly, the states can be abstract and can include contextual information about the state of the environment but also about the agent (learner), including their internal (e.g., emotional) state. Selecting a trajectory corresponds to deciding, as a step in the problem-solving process, to operate on the problem space both at the external level (e.g., moving an object) and at the internal level (e.g. modifying some internal value of the learner) (Alexandre, 2020a). Our representation of problem solving in Figure 1.11 is also consistent with the work of Hay et al. (2017)

considering the multiplicity of paths that can be generated from the problem state to a goal state as part of the “solution search” process.

Problem space characterization in ill-defined problems

Adapting (Newell and Simon, 1972)’s formulation of problem solving, we consider ill-defined problems as problems for which one or several of the following non-exclusive possibilities exist:

- the open-ended goal is not well defined but simply constrained by certain requirements, making the final state not unique and/or not known from the beginning;
- the initial state is only partially known (and observable) to the learner;
- there is no explicit operation to move from one state to another, but potential actions are discovered by the learner when interacting with the environment, for instance discovering new objects and perceiving their features as *affordances*, thus inferring how to use them (Gibson, 1977) with regard to their goals (Norman, 2013).

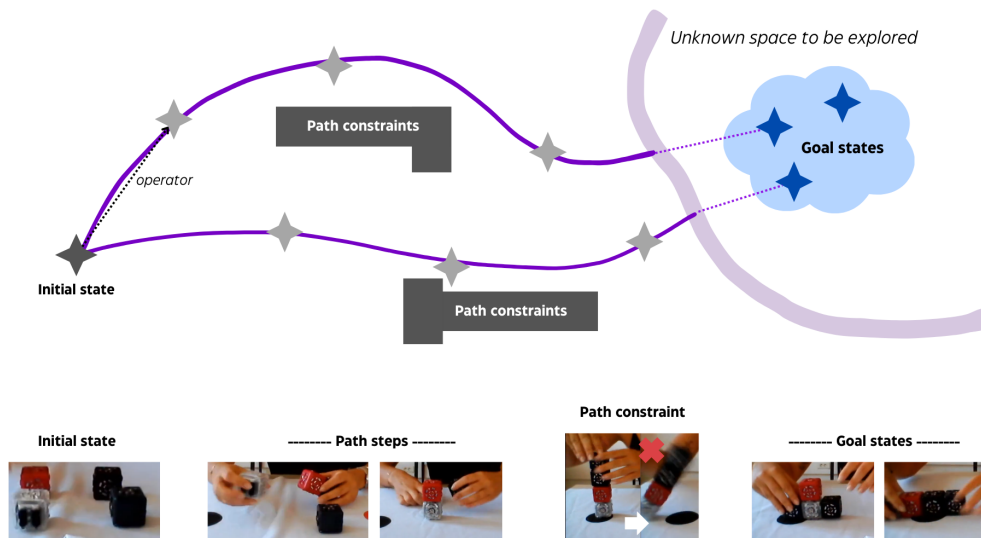


Figure 1.11: Ill-defined problem-solving as trajectory generation. We hereby extend Newell and Simon (1972)’s definition of problem-solving and the problem space to ill-defined problems. In this context, operators are limited by the learner’s procedural knowledge and discovery of affordances. The goal could be loosely defined, thus potentially reached in multiple states. Examples from the CreaCube task illustrate this definition. (Adapted from Alexandre, Mercier, Palaude, Romero, and Viéville (2022)).

1.4 Computational approaches for CPS: from symbolic to neural formalisms

The initial present state is a point in this state space resulting from the problem-posing activity of the learner commencing the task. The goal is a region of this state space, constrained by requirements to fulfill (this region can be reduced to a point if the goal is to reach a given state). The problem-solving task process corresponds to a (not necessarily unique) path, that is, transforming the initial present state into a final one; each feasible path being constrained by certain requirements.

This spatial representation corresponds to a formal computer science definition of problem solving with generic algorithms as soon as the problem has been embedded in a continuous space. Such optimization algorithms formulate the problem as the generic minimization of “potential,” yielding a biologically plausible and computationally efficient mechanism (see, e.g., [Viéville \(2002\)](#)), which is also deeply related to computational neuroscience models; the hippocampal brain structures, for instance, are involved in both navigation tasks and planning, with similar functional characteristics ([Stachenfeld et al., 2017](#)). Of course, such mechanisms require to equip the space with a definition of **distance**, as developed in Chapter 3.

1.4.2 SEMANTIC GROUNDING: REPRESENTING AND MANIPULATING SYMBOLS

In the brain, it is usually admitted that data is represented a distributed rather than localist way, through activation of multiple (rather than single) neurons ([McClelland and Rogers, 2003](#)). However, at the computational level, a symbolic specification is useful at the level of abstraction chosen for this study, although we will also investigate how such a representation can be implemented and manipulated into more biologically plausible architectures.

FROM SENSORIMOTOR FEATURES TO CONCEPTS

Before going any further, let us first review what is a *symbol*. This non-trivial notion is the ground of our human knowledge, and when daring to consider “symbolic representation”, we need to clarify in detail what is defined here.

The semiotic approach as reviewed in [De Villiers \(2007\)](#), introduces a hierarchy of “signs”, through which an object or concept (signified) is represented by a signifier, based on three different levels of abstraction: the “icon” built from the sensible likeness with the object; the “index” built from concrete relationships between given objects (e.g., a weathercock indexing the wind direction and strength), and finally the “symbol” in the semiotic sense, which is an arbitrary designation (such as a word) for the signified object, and thus must be culturally learned.

At the syntactic level, a symbol is therefore no more than the label (or identifier) of an object or concept. It has a “meaning” in the sense of [Harnad \(1990\)](#), as reviewed and discussed in [Taddeo and Floridi \(2005\)](#), when it is semantically *grounded*, understood as the process of embedding symbolic computations onto real-valued features ([Badreddine et al., 2021](#)), thus providing a semantic interpretation or model (in the sense of a model of a set of logical assertions) of the symbolic system, which involves the capacity to pick referents of concepts, including their **affordances**. As discussed in, e.g., [Raczaszek-Leonardi](#)

and Deacon (2018), the key problem is how a symbolic representation can emerge from sensorimotor features and interaction with the environment. Enactive and developmental processes contribute to the formation of stable visuo-spatial symbols (as reviewed, e.g., in Ness and Farenga (2007)).

Such a multi-level representation, corresponding to various levels of abstraction, is in line with the “quality dimensions” proposed by Gärdenfors (2004) i.e., features with some typed value. A step further, Freksa (2015) and Oltețeanu (2020) argue that such features are anchored in feature spaces built around different sensorimotor dimensions, following a mild abstraction approach, as represented in Figure 1.12.

It is interesting to note that, in the brain, sensory and motor information seem to be encoded using a common representation, accounting for mirror neurons (Hommel, Müseler, et al., 2001). Oltețeanu (2020)’s framework incorporates this idea by defining **affordances** as a feature space at the same level of representation than other sensory information (e.g. visual, such as color).

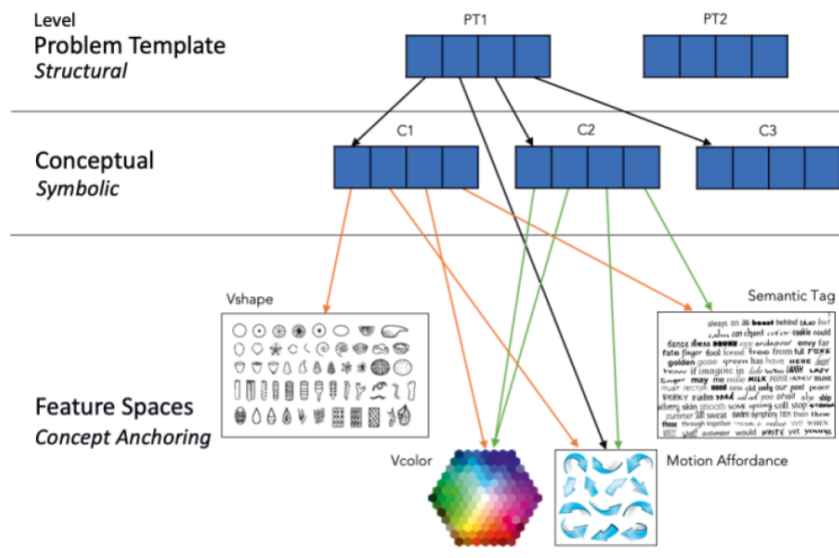


Figure 1.12: Three levels of symbolic abstraction, as proposed in Oltețeanu (2020)’s CreaCog framework.

Gärdenfors (2004) also hypothesizes that a concept is a convex region around prototypical examples. This formulation, compatible with our geometric formulation of the problem space, implies the existence of “default” attributes; for example, not all birds fly (a penguin, for instance, does not fly), but when we are asked to think of any bird, the example of a flying bird is more likely to come into mind than a penguin; which supposes that a prototypical bird has, by default, the capacity of flying as an attribute.

A concept is thus a bipartite notion corresponding to: (i) a prototype for this concept (e.g. in our example, the node “bird” defines a typical bird) with features accounting for “default” values; and (ii) a region of the concept space, corresponding to all individ-

1.4 Computational approaches for CPS: from symbolic to neural formalisms

uals having features that are compatible with a certain specification called *schema*. Such a notion of schema¹⁹ is relevant in our context because **DT** might be implemented as relaxing some of these constraints while **CT** would consist of verifying that the resulting new data are still compliant with what could be expected to attain the desired goal. This could be generalized to other kinds of semantic constraints (e.g., bounds on the weight or constraints on the feature type), still defining convex regions or structural constraints (e.g., conditional constraints given some feature values).

The **open-world** assumption required by Wiggins (2006b) also implies the definition of a generic concept, namely a thing that is an empty concept and thus for which nothing is known. This makes it quite interesting to implement the “thinking-out-of-the-box” process because it could be related to relaxing some features to consider a less specified concept and then exploring other feature values, as also discussed by Wiggins (2006b). We will develop this idea in the next section.

In the end, this geometrical representation of concepts anchored on features allow us to define an ill-defined CPS task, the initial state or the goal not being a precise value but a region; the wider the region, the fewer constraints on it. We can also define a deterministic notion of partial knowledge, with features lying in a given region of values rather than having precise values.

MEMORY AS SEMANTIC NETWORKS

Having defined concepts and features (e.g., the two lower layers of Olteteanu (2020)’s framework), we need to develop how relationships between concepts can be expressed and manipulated (e.g., the higher-level layer).

Associative processes are typically modeled by random walks in semantic networks (accounting for the structure of **semantic memory**), as developed by Zemla and Austerweil (2018) or Lopez-Persem et al. (2023); that is, modeling semantic knowledge as a network of concepts. From this semantic network approach, high connectivity and low modularity allow for more **flexibility** (Denervaud et al., 2021; Kenett and Faust, 2019). However, these networks do not account for other kinds of memory organizations mechanisms beyond associativity (J. Mandler, 2010), such as hierarchical and sequential relationships, or different levels of abstraction.

McClelland and Rogers (2003) proposed to use four different predicates in order to link the concepts to their features: hierarchical taxonomy (**is-a**), capability (**can**), extrinsic features (**has**), and intrinsic features (**is**). Such a representation corresponds to a semantic graph where the nodes are concepts or features, and the edges are predicates, which can be hierarchical or relational. It is an augmentation of semantic networks proposed earlier, in that edges are annotated and directed. In terms of concepts as defined by Gärdenfors (2004), hierarchical taxonomic relationships mean that, if a concept is subsumed by another concept, it is a region enclosed in the parent concept region and thus “inherits” its features.

¹⁹In the sense of XML-schema.

1 Creative Problem Solving: A tridisciplinary review

This is actually very similar to a standard knowledge graph representation as it is formalized, e.g., in the Semantic Web (RDFS or OWL) with the RDF model using triples of the form <subject> <predicate> <object> (e.g., <bird> <is-a> <vertebrate>).

Such a representation is, for instance, used by Han et al. (2018) to implement computer-assisted analogy reasoning using an OWL ontology. OWL builds on both RDF triples on the one hand, and on the other hand description logics, a fragment of first-order logic which also has the benefit of directly implementing the open-world assumption mentioned previously; that is, all that is not specified is unknown. For this reason, this representation is also effective in the sense that it can be manipulated by symbolic logical reasoners (see also Chapter 2 for some examples).

At a higher level, such a specification allows meta-reasoning about rule mechanisms since such rules are also structured objects using concepts. Similarly, we could extend this formalism to sequences: in that case, a given instance of behavior would correspond to a sequence of rule applications. This proposal corresponds to the third level of the theoretical framework of CreaCogs (Oltețeanu, 2020) with problem templates structured around the concepts that are anchored in sensorimotor features.

1.4.3 STATE GENERATION MECHANISMS AS A SEARCH IN THE PROBLEM SPACE

Dietrich and Haider (2017) discussed several computational mechanisms of DT, such as evolutionary algorithms and Bayesian-based prediction that could implement some of these modalities, while Wiggins (2006b) and Ritchie (2012) established that generating new creative concepts can be considered as a search process.

RANDOM WALKS As mentioned earlier, random walks in semantic networks (Zemla and Austerweil, 2018) are search algorithms that often use to model, mainly spontaneous, associative processes; but there are also more deliberate associative processes that we will now consider.

BISOCIATION AND CONCEPTUAL BLENDING Bisociation, defined by Dubitzky et al. (2012) as the association of two concepts from different domains perceived simultaneously, allows the generation of novel ideas. Conceptual blending (Fauconnier and M. Turner, 2003) is a particular case of bisociation: starting from a partial match between two concepts (e.g., sharing a subset of common features), conceptual blending consists of selecting a few features from the first concept that do not belong to the second concept and adding them to that second concept, resulting in the emergence of a new structure and new meaning.

Oltețeanu (2020)'s CreaCog framework is primarily based on bisociation. It has been used to replicate results of the AUT, producing plausible answers that were comparable to findings in humans (Oltețeanu, 2020). As pointed out by Mekern et al. (2019), this model could account for inter-individual differences in the way its feature space is organized: the search for affordances is dependent on the size and number of feature spaces in the knowledge base (i.e., semantic memory): the more feature spaces are considered, the more divergent the search can be.

1.4 Computational approaches for CPS: from symbolic to neural formalisms

ANALOGY Another example is *analogy*, or analogical reasoning, defined as translating a known relationship between two concepts in a source context or domain from a target domain given an “analogy” between concepts (e.g., “Robin is to Batman what Sancho is to Don Quixote”). Analogy is key to, e.g., transfer a solution to a known problem, to another problem of the same structure (P.N. Johnson-Laird, 1989; Keefer and Landau, 2016).

The Retriever model (Han et al., 2018) has been developed as a tool to assist creative design by generating new ideas of uses for everyday objects. This model relies on analogical reasoning performed on an ontology based on ConceptNet²⁰. The results have indicated that the tool could significantly improve the **fluency** and **flexibility** of idea generation for groups who were using it.

CONVERGENT ASSOCIATIONS Although this rather pertains to **CT**, we wish to mention here how associative processes similar to the one described before can be used in a more convergent fashion, in order to perform a task such as the **RAT**. As reviewed by Mekern et al. (2019), this has been recently tested in different computational paradigms. In the comRAT model also based on the CreaCog framework (Oltețeanu, 2020), when a triple of words is provided, all related word pairs in the comRAT knowledge base are activated and an associative search is performed to find an answer present in the three word pairs at once. If an answer can not be found, an alternative answer matching only two word associations could be returned. In the general cognitive architecture Soar (Schatz and J. Laird, 2018), an agent can retrieve information from a semantic memory, either by providing a specific cue that is matched against elements in long-term memory (Cued Retrieval model), or an agent can use a general cue and leverage spreading activation to retrieve words based on context as defined by the contents of working memory (Free Recall model). Such spreading activation has been used in more biologically plausible way within a spiking neural network (Kajić et al., 2017) using a sub-symbolic encoding based on a VSA approach similar to the one already mentioned in T. Stewart and Eliasmith (2011). All three models – comRAT, SOAR and the VSA approach – showed comparable results to human data on the **RAT**. **CT**, however, relies not only on associativity and spreading activation of nodes in a semantic network, or by finding common features between concepts. We may also need more sophisticated ways of evaluating the relevance of an idea using reasoning, as reviewed below.

GEOMETRIC FORMULATION In a geometric formalism, projection may be an adequate model of bisociation: indeed, in a conceptual space, it is akin to projecting features of a concept onto another concept, replacing some features of the latter with features from the former, resulting in a new “blended” concept.

Conversely, analogy would correspond to a translation. This is well formalized in terms of vectors in the case of embeddings such as *word2vec* (Ethayarajh et al., 2019): for instance, given the vectors **batman**, **robin** and **donquixote**, the vector **donquixote** – **batman** + **robin** should yield a representation of Don Quixote’s companion Sancho.

²⁰<https://conceptnet.io/>, a collaborative commonsense knowledge graph.

1 Creative Problem Solving: A tridisciplinary review

In a nutshell, at the geometric level, **DT** consists of extrapolating new concepts from existing ones, either by projection (bisociation) or translation (analogy). On the other hand, associative processes described to perform the **RAT** are more similar to interpolation: indeed, we only consider the subset of common features between two (or three, in the **RAT**) concepts.

1.4.4 STATE EVALUATION MECHANISMS AS LOGICAL OR PROBABILISTIC REASONING

Reasoning refers to determining whether a conclusion follows from premises (P.N. Johnson-Laird et al., 1999), and is a key mechanism in problem solving. We make the distinction between:

- *deductive reasoning*, which is the process of determining the formal logical consequence of some assumptions considered as true or approximately true;
- *inductive reasoning*, which is the process of inferring some general principle from a set of knowledge and plausible induction rules;
- *abductive reasoning*, which is the process of inferring an explanation of some assertions, i.e., hypothesizing the precondition of a consequence.

Deductive logic has a long history of formalization in Western civilizations, starting from Aristotle (Oaksford and Chater, 2009). As pointed out by Mangan (1978), such a mechanism includes a cultural bias, since not all cultures feel the need to develop formal logical operation competencies, and obviously most people do not use such formal operations in all aspects of their lives. Furthermore, as thoroughly studied by, e.g., Purves et al. (2001), the experience of conscious or subconscious emotion has a powerful influence on rational decisions, including the choice of alternatives in deductive reasoning.

Nonetheless, as discussed in, e.g., Keefer and Landau (2016), especially for goal-driven behavior, deductive reasoning is deeply interleaved with heuristic deduction, which we need in order to solve problems in everyday life; while *inductive reasoning* and *abductive reasoning* as defined by Pierce are key to the hypothesis generation and evaluation epistemic process (Boutilier and Beche, 1995; Rodrigues, 2011). Therefore, our understanding is that human cognition must implement some kind of such (potentially loosely) reasoning mechanism. Piaget (1997) viewed formal logical reasoning as defining the end-point of cognitive development, and contemporary psychology of reasoning has focused on comparing human reasoning against logical standards. This has been studied, for instance, by Wason (1968), using a card-turning problem formulated as: “If a card has a vowel on one side, it has an even number on the other side. Select those cards that you definitely need to turn over to find out whether or not they violate the rule: A D 4 7”. The card-turning task has been replicated multiple times, showing that people do not typically use classical logic (in this case, the *modus ponens* and *modus tollens* rule) to draw their conclusion (Oaksford and Chater, 2009).

Moreover, ill-defined situations call for a certain tolerance for ambiguity, thus a need to consider the notion of partial knowledge and its related degree of belief, that is, the estimation of the available knowledge by a subject.

1.4 Computational approaches for CPS: from symbolic to neural formalisms

PROBABILITIES AND BAYESIAN INFERENCE For the last century (see, e.g., [Smithson \(1970\)](#) for a presentation), human evaluation of knowledge has been deeply driven by a weak notion of probability, with two key ideas:

1. that the degree of belief can be expressed by a number between 0 and 1 (often given as a percentage between 0% and 100% of “plausibility,” “chance,” or any other qualification of a belief evaluation);
2. that, for a given event, some knowledge about the possibility and necessity of its occurrence can be acquired through the observation of a count of its past occurrences.

Since brain functions often involve integrating different sources of information, and since processing inherently involves uncertainty, Bayesian inference can be a good descriptive framework for a number of neuroscience study topics ([O’Reilly et al., 2012](#)). For example, it can be applied to perceptual decisions, during which the subject attempts to categorize an object or estimate a measure of it on the basis of uncertain sensory information. The authors point out, however, that the human brain is unlikely to operate entirely according to Bayesian inference, given the high cost in terms of computing power, incompatible with **working memory** limitations. [Oaksford and Chater \(2009\)](#) argue that, although people are typically poor at numerical reasoning about probability, human thought is sensitive to subtle patterns of qualitative Bayesian, probabilistic reasoning. As a result, it has been suggested that non-Bayesian learning might be preferred to Bayesian learning in situations where the latter offers little added value. According to this perspective, the brain could therefore use non-Bayesian heuristics to approximate Bayesian inference, in all situations where these heuristics offer a reasonable compromise between computational demand and performance. This may also be put in relation to spontaneous and deliberate evaluation processes, as evoked in 1.3.

MODAL LOGIC AND MENTAL MODELS Another hypothesis accounting for human reasoning is modal logic, privileged by [Piaget \(1997\)](#). Modal logic is based on possibility and necessity, rather than probability (see [Dencœux et al. \(2020\)](#) for an introduction of the general possibility theory).

Subsequent research has shown that people perform better when the card-turning problem is presented in concrete form, such as detecting underage drinkers²¹ ([Cox and Griggs, 1982](#)). According to [P. Johnson-Laird and Ragni \(2019\)](#), this illustrates the multiplicity of semantic interpretations used in modal reasoning, the main ones being:

- *Alethic* knowledge, regarding relations between two sets of possibilities, including necessary and possible inferences, and causal relations.

²¹For instance, in this case, the problem would be formulated as: “If a person is drinking beer, then the person must be over 19 years of age. Select the card or cards you would definitely need to turn over to determine whether or not the people are violating the rule: “Drinking a beer”, “Drinking a coke”, “16 years old”, “22 years old”

1 Creative Problem Solving: A tridisciplinary review

- *Deontic* knowledge concerns social relations governing obligations and permissions, which speech acts themselves can create.
- *Epistemic* knowledge regarding scientific interpretations of observations, varying on a scale of degrees of belief from impossible to certain (akin to naive probabilities as in [Khemlani, Lotstein, et al. \(2015\)](#)).

Based on a series of empirical investigations, [P. Johnson-Laird and Ragni \(2019\)](#)'s mental model theory provides an account of deductive reasoning in which people build concrete mental models of the situation described in the premises. Mental models represent possibilities as iconic representations of what is common to their infinite realizations, which also takes inspiration from the Gestalt approach, with a representation adapted for "fast thinking" and one for "slow thinking" (in the sense of [Kahneman \(2011\)](#) which is related to the implicit-explicit or spontaneous-deliberate distinction made earlier, with nuances reviewed in Appendix B); the "fast" representation being related to *insight*.

DIFFERENT EVALUATION RULES FOR DIFFERENT CONTEXTS Overall, these findings point to the domain specificity of reasoning, that is, the idea that people use the concrete context of the problem to guide their reasoning rather than apply the same general logical rules to all problems. Furthermore, experimental results in cognitive psychology suggested that there are situations in which people reason about uncertainty using the rules of possibility theory rather than those of probability theory but still in line with it ([Raufaste et al., 2003](#)). Another key aspect ([Dubois and Prade, 2015b](#)) is the fact that we can interpret possibility (and necessity) measures in terms of probabilities of rare events, especially crucial events or events of substantial interest.

[Tettamanzi et al. \(2017\)](#) have related possibility theory and ontology-based representations and proposed an example of effective combining rules, while [Dubois and Prade \(2015a\)](#) discussed how such combining rules are constrained by the theory. At the data representation level, this corresponds to attributing belief value to the ontology statements (see also Chapter 4) while the possibility theory provides rules to combine those belief values. At the cognitive level, this means that the estimated feature values are now weighted by a level of belief combining the belief with prior information and a level of confidence in the stimulus.

In a geometric model, both the current state approximative estimation and the targeted goal correspond to regions of the state space, as schematized in Figure 1.11. Each of them is estimated with regard to both bottom-up, external stimuli, and top-down, internal information (that is, prior knowledge as well as emotional and cognitive considerations), as illustrated in Figure 1.5. Furthermore, the estimated value is both imprecise, that is, partially defined, and approximate, that is, bound to some uncertainty (in the probabilistic case) or precision (in the deterministic case).

Given these two ingredients, in a probabilistic framework, the notions of prior versus posterior probability, for example in Bayesian approaches, implement such ideas with a numeric data representation. Such a computing mechanism, as made explicit by [Viéville, Lingrand, et al. \(2001\)](#), for example, corresponds to a geometric projection of the incoming

stimulus onto the region specified by prior information. In other words, given a stimulus and prior information, the estimated value is the state value that is as close as possible to the stimulus and compatible with the prior information. Our geometrical representation is thus relevant at this level.

1.4.5 REGULATING MECHANISMS

As observed by Mekern et al. (2019) and W. Zhang et al. (2020), convergent and divergent mechanisms are often studied in isolation, offering little chance for direct comparison. As a matter of fact, creative intention and exploration are not enough; regulatory cognitive processes are also required in CPS in order to switch from phase to phase as proposed in the componential approach (Figure 1.2, using behavior switching as reviewed in 1.3. For this purpose, **reinforcement learning** (RL) constitutes an interesting starting point.

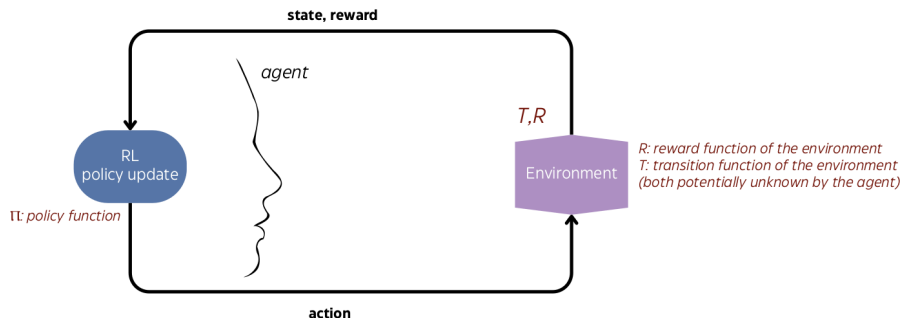


Figure 1.13: Illustration of the “vanilla” reinforcement learning paradigm. See Chapter 3 for details.

Without getting into the details of the framework (described in Chapter 3), let us introduce the principle on Figure 1.13. In a classical RL set-up, an agent interacts with the environment until it reaches some terminal state. At a given discrete time, it perceives a part of the environment, i.e., a stimulus, including a reward, which leads it to compute the next action²². One such traversal is called an **episode** and consists of a sequence of states, actions, and rewards that the agent got on his path. The sequence of actions defines the **behavior** of the agent. This behavior follows a **policy** function π (potentially stochastic) that predicts the next action based on the current state. Choosing between making the best decision given the current information or starting exploring and finding more information, is known as the **exploration/exploitation** dilemma.

²²In psychology, an *action* is a self-initiated sequence of movements, usually with respect to some goal (*APA Dictionary of Psychology*). This obviously leads to a difficulty of determining the level of abstraction used to specify such an action, as pointed out in the activity theory (Engestrom, 2000). In RL, the set of possible actions is typically predefined and fixed.

1 Creative Problem Solving: A tridisciplinary review

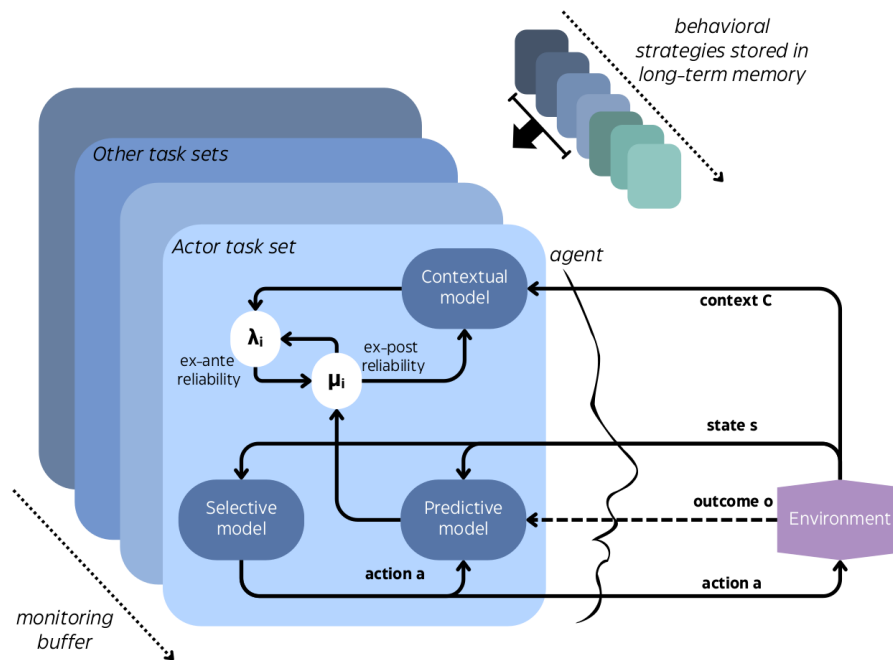


Figure 1.14: The Probe model for task sets, adapted from [A. Collins and Koechlin \(2012\)](#). The monitoring buffer comprises a limited number of task sets, each indexing a behavioral strategy stored in long-term memory and comprising: a selective, predictive and contextual mapping (M.). A single task-set (the actor) is used to guide the action, and updated by reinforcement learning. However, different task-sets are continuously monitored, and their ex-ante reliability signals (λ_i) are used to decide on state and actor changes. If the model is in exploitation and the actor's λ falls below a certain threshold, it enters exploration and creates a *probe*. If it is in exploration and one of the alternative task-sets goes above that threshold, this task-set becomes the new actor and the probe is destroyed. If it is the probe that goes over the threshold, the probe becomes the new actor (translated from [Donoso \(2013\)](#)). See [A. Collins and Koechlin \(2012\)](#) or more details on how reliability signals (λ_i and μ_i) are calculated from contextual and predictive models.

While the notion of exploration and exploitation is well known in cognitive science and in machine learning at the level of reinforcement learning, they are two distinct conceptualizations that we will now make explicit. In problem-solving activities, subjects alternate between two main strategies: exploration, which aims at experimenting with new alternatives or acting on the environment to generate new **stimuli** ([Kaplan and Flum, 2010](#)), and exploitation, which is the use of existing knowledge (declarative or procedural) in a given situation. At this level of description, we can draw parallels with these notions in reinforcement-based machine learning, but in this second domain exploration is mostly random, whereas in problem solving as studied in learning sciences, exploration is seen as a **divergent thinking** process for generating new internal representations (ideas or stimuli) and new behaviors by recombining prior knowledge

or behaviors and/or by manipulating the environment. This being said, although exploration/exploitation is not superposable with divergent/convergent thinking, it could be an account for behavior switching as an operationalization of the **task set** model introduced in Chapter 1.3, such as the Probe model (see e.g., [A. Collins and Koechlin \(2012\)](#) or [Donoso \(2013\)](#)). In this approach, represented in Figure 1.14, each task set corresponds to a **Q-learning** mechanism (i.e., a quite simple RL algorithm). In Chapter 3, we also propose to consider extend Q-learning by coupling it to a symbolic model of the environment.

1.5 MAIN TAKE-AWAYS

Sharing a common acceptable and operational framework of creative problem solving between three communities constitutes an important challenge. Through this overview of CPS as studied in the learning sciences, cognitive neuroscience and computational modeling, we have identified some major notions and their similarities or nuances throughout the disciplines. From implicit to explicit knowledge representation and manipulation, convergent vs divergent thinking, exploration vs exploitation; these dual notions correspond to rather sophisticated complementary aspects of creativity but are sometimes ambiguous. To this end, we elaborated two recapitulative tables [B.2](#) and [B.1](#) which can be found in Appendix B.

Our motive is to demystify human CPS using computational models, with a further perspective to provide new tools of assessment. This has been done in other works, for example, [Denervaud et al. \(2021\)](#) has used the semantic network approach developed in [Kenett and Faust \(2019\)](#) to measure the impact of Montessori pedagogy on **DT**. This could have implications on education beyond the simple understand of CPS mechanisms and be helpful e.g. to address learning inequalities. The present review enlightens what design choices could be of interest for modelling the kind of task we target.

First, we generalized the usual geometric formalism of problem solving to an abstract trajectory generation in a state space. We explicitized what ill-defined means with such a definition of problem solving and how creative processes unfold dynamically within the problem-solving task.

Regarding data representation, we agree with the mild abstraction approach adopted by [Freksa \(2015\)](#) and [Olteteanu \(2020\)](#) whereby concepts are anchored onto physical features. This is especially important in the case of modelling a task such as CreaCube, which uses tangible objects and is not merely a verbal creativity task (such as the **RAT** or the **AUT**). We also think that this symbolic representation should reflect the organization of memory as evoked by [Eichenbaum \(2017\)](#), thus integrate not only an associative structure but also sequential and hierarchical organizations. For this purpose, the Semantic Web technologies (e.g. **RDFS**, **OWL**) seem interesting as an existing standard to formalize semantic graphs with hierarchical taxonomy and other kinds of relationships. Logical reasoners already exist to perform semantic evaluation using deductive inference for **CT**, while other processes relevant for **DT** such as analogy can also be implemented (e.g. [Han et al. \(2018\)](#)). The representation of different degrees of belief re-

quired to tolerate the ambiguity inherent to problem solving could be considered in a way similar to what has been done in [Tettamanzi et al. \(2017\)](#) regarding modal notions of possibility and necessity (that we also proposed to relate to a probabilistic framework in [Viéville and Mercier \(2022\)](#)).

We target here a symbolic representation that has more explanatory potential for the purpose we are aiming at, but we are also interested in the biological plausibility of such a representation into sub-symbolic architectures such as using VSAs, as developed in Chapter 4. The important point is that this representation could be processed explicitly (through logical inference) but also implicitly thanks to hybrid mechanisms (such as Bayesian inference or RL algorithms computing numerical rewards on symbolic states) and we also need to equip the space with a metric (i.e., a **distance**) in order to perform search and evaluation mechanisms, as we will develop in Chapter 3. Finally, a metacontrol mechanism should be implemented to switch behaviors according to the context, simulating an exploration-exploitation paradigm allowing to combine divergent and convergent thinking.

Therefore, considering the description of creative process in the literature already mentioned, it seems that a relatively simple paradigm such as reinforcement learning (combined with an ontology), the biological plausibility of which has been studied in depth, seems more reasonable than more sophisticated frameworks aiming for generalized problem solving, such as the IDyOT architecture ([Wiggins, 2020](#)) implementing the creative framework of [Wiggins \(2006a\)](#) following [Boden \(2004\)](#), or the SOAR architecture ([J.E. Laird et al., 2017](#)), in which a hierarchy of goals and problem spaces is generated, with operator selection and evaluation mechanisms. Such general architectures can indeed account for the CPS process discussed here but at a level of complexity that might be oversized.

2 AN ONTOLOGY APPROACH: FORMALIZING PRIOR KNOWLEDGE AND MATERIAL CONFIGURATIONS OF THE CREACUBE TASK

2.1	Introduction	63
2.2	The ontology approach	64
2.2.1	Terminology	64
2.2.2	Related work	64
2.2.3	Methodology	66
2.3	Ontology Implementation	67
2.3.1	Modeling the activity process	67
2.3.2	Modeling the material states	69
2.3.3	Modeling the learner’s behavior and mental states	69
2.4	Results and perspectives of application	71
2.4.1	Modeling as an interdisciplinary tool	71
2.4.2	Reasoning about the model	72
	Relating the configuration encodings to the material states	72
	Inferring the knowledge of the learner from the observables	73
	Predicting the learner’s behavior	74
2.5	Discussion and conclusion	74

2.1 INTRODUCTION

Modeling across several disciplines often constitutes a terminological challenge, hence the need to agree on a shared vocabulary. To this end, ontologies provide a computational tool for structured knowledge representation, allowing to define concepts and relationships between them at a symbolic level.

Besides, the CreaCube task being rather innovative and unprecedented, it is useful to give a precise specification of it. This includes the previously introduced observables, describing how the material works, but also synthesizing the first observations made by experimenters in terms of knowledge necessary to understand the task and solve the problem. Such an ontology allows to share the task efficiently with other researchers¹, as it includes relevant information about the task in a standard format that can be processed by other computer programs.

¹This ontology is publicly available here: <https://line.gitlabpages.inria.fr/aide-group/creaonto/>

2 An Ontology approach

The following chapter is adapted from [Mercier, Roux, et al. \(2021\)](#) and [Mercier \(2023\)](#). We describe here our methodology to develop this ontology, explicitize the entities that have been selected and present some possible perspectives of applications.

2.2 THE ONTOLOGY APPROACH

2.2.1 TERMINOLOGY

Ontologies allow us to represent structured knowledge by defining concepts, as well as relationships and hierarchies between these concepts. In the following, we use the *Web Ontology Language* (OWL) terminology, as used in the Semantic Web.

OWL terminology

- (i) **Individuals** represent atomic, real-world objects.
- (ii) **Classes** represent concepts; a class may therefore be a collection of individuals (which are called instances of the class).
- (iii) Individuals may be linked by relationships, which are labeled by **properties**. Properties may have a **domain** (i.e., the class of individuals that they can be applied to) and a **range** (i.e., the class that they can take values from).

OWL also defines both class hierarchy and property hierarchy. This formalism allows us to structure and specify data that can be provided in a standard such as the Resource Description Framework (RDF).

OWL specifications are based on description logics, providing in some cases² computational completeness and decidability (a didactic introduction can be found in [Allemang et al. \(2020\)](#)). This makes it possible to perform logical reasoning to validate the model, as well as logical inference to find all of the assertions that can be deduced from the user-defined statements.

2.2.2 RELATED WORK

ONTOLOGIES TO MODEL LEARNING ACTIVITIES A recent systematic review ([Stancin et al., 2020](#)) of ontology use in education has observed a growing trend in the contribution of ontologies to educational systems in the last five years, and has identified the following common use cases:

- Describing learning domains,
- Describing learner data,

²Several OWL “profiles” are available, each of them providing a different trade-off between expressivity and reasoning power. The OWL-DL profile fully adheres to description logic principles and is thus decidable.

- Modeling and managing curricula,
- Describing learning services.

In our case, we are particularly interested in works that attempt to model both the learning domain and the learner data. This has been done especially in intelligent tutoring systems and adaptive learning systems, as reviewed in [George and Lal \(2019\)](#), where semantic inference is performed on learning analytics and the learning content to tune the resources that are presented to the learner (e.g., in [Nouira et al. \(2019\)](#)). These studies are inspiring but they differ from our approach in some important aspects.

Most reviewed approaches attempt to give a general model the learner (i.e., in a diversity of learning tasks). From that perspective, learner modeling is often directed towards classification of learners into several learner profiles with regard to a set of characteristics, and by doing so to select the more adequate resources for a given learner. This is also the case for knowledge representation beyond ontology approaches, such as those related to the xAPI and game-based learning strategies ([Torrente et al., 2009](#)). The cognitive trajectories introduced by [Fuchs and Henning \(2018\)](#) are an interesting approach to better understand the learner behavior, at a general level or in a specific task.

Another point is that a large majority of these tasks are well defined (e.g., well defined knowledge or how-to acquisition), whereas we consider here a more open ill-defined problem solving task. We can, however, cite the work of the [Ludo Game Model Ontology \(LUDO\)](#) ([Tang and Hanneghan, 2011](#)) to describe serious games. These activities can be presented as open problems, and (despite being not really reusable here, because it does not involve tangible artifacts) the ontology itself includes both a model of the game (task) and a model of the player (learner) behavior.

Eventually, the goal of the existing ontologies is to improve the learner's performance ([George and Lal, 2019](#)), which is less easy to define in our problem solving context (solving the problem faster does not mean that we learn a lot from it) and is not our main purpose, which is to better *understand* how we solve such an open-ended problem.

With respect to these works, we introduce here computational and cognitive neuroscience knowledge in the modeling process, which does not seem to be a current practice, and we also consider open-ended problem solving, which is seldom addressed.

ONTOLOGIES IN NEUROSCIENCE AND COGNITIVE SCIENCE The Cognitive Atlas³ ([Poldrack et al., 2011](#)) is a collaborative initiative that aims to unify the vocabulary used in cognitive science and indexes cognitive tasks described in the literature. It includes over 800 cognitive concepts but does not encompass much logic, aiming for a more flexible approach. Its task model is partially based on CogPO ([J.A. Turner and A.R. Laird, 2012](#)), which is another ontology that focuses on the description of experimental paradigms used in neuroscience (mainly fMRI) tasks. While CogPO might be slightly too specific and medical-oriented for our problem, the Cognitive Atlas encloses many of the cognitive concepts that we may need to refer to in subsequent work. It also seems that the

³See <http://www.cognitiveatlas.org/> for a browsable version, and on BioPortal for a reusable ontology version.

2 An Ontology approach

concepts we target here are not (to the best of our knowledge) addressed at the level of ontology modeling.

Beyond ontology-based methods, a wide spectrum of approaches, out of the scope of the present work, have been recruited to model ill-defined creative problem solving; for instance, using description logic, extended with probabilistic approaches (Lieto, Perrone, et al., 2019). This example is interesting because the authors have made the distinction between different types of knowledge and memories, and considered its integration in a cognitive architecture (Lieto and Pozzato, 2020).

CONSIDERING FOUNDATIONAL ONTOLOGIES A good practice in ontology modeling is to base our model on a foundational ontology to better specify the design choices by relating our concepts to well-established formalized semantics and facilitate the integration with other ontologies. Neuroscience ontologies often rely on recommendations set by the [Open Biological and Biomedical Ontology \(OBO\) Foundry](#), which is itself built on the principles of [Basic Formal Ontology \(BFO\)](#), as is the case for CogPO. However, the ontological choices made in BFO are meant to represent real-world elements and not, for example, mental states. Integration to biomedical ontologies is not necessarily a requirement for us, and the Cognitive Atlas, in that matter, is rather agnostic; therefore, a more appropriate choice for us would likely be the [Descriptive Ontology for Linguistic and Cognitive Engineering \(DOLCE\)](#) (Masolo et al., 2003) which is general enough to be linked with both the cognitive and learning science aspects of our formalization work. In brief, it is a descriptive ontology of particulars that “aims at capturing the ontological stands that shape natural language and human cognition, based on the assumption that the surface structure of natural language and the so-called commonsense have ontological relevance” (Masolo et al., 2003), which may better fit which our paradigm.

2.2.3 METHODOLOGY

[Noy and McGuinness \(2001\)](#) describes a simple methodology to build an ontology in seven steps:

- (a) Determine the domain and scope of the ontology.
- (b) Consider reusing existing ontologies.
- (c) Enumerate important terms in the ontology.
- (d) Define the classes and the class hierarchy.
- (e) Define the slots (i.e., the properties shared by a class).
- (f) Define the facets of each slot (i.e., the possible values that a property can take).
- (g) Create instances.

Considering (a) the scope defined in the previous paragraph, we have chosen to (b) rely on a lightweight version of the DOLCE foundational principles (DOLCE+DnS UltraLite) and reuse concepts defined in the Cognitive Atlas. In the following, we describe (c,d,e,f) the terms used to define the concepts and properties in our ontology (the slots refer here to the properties shared by a class, and the facets to the possible values that a property can take). Finally, we explicitize (g) how we instantiated those classes.

Based on these elements, the ontology has been designed at two levels. At the task material level, all cubes configurations and other physical elements of the setup have been identified and encoded, factorizing equivalent configurations (e.g., up to a rotation), and reducing the state space to a few qualitative relevant values. At the learner modeling level, considering the previously published studies (Romero, 2017; 2019; Romero, David, et al., 2019), our work has been to formalize what has been experimentally determined by learning scientists and education experts from the first experiments, including the chosen pertinent observables. In order to do so, we have attempted to rely on theoretical learning science and cognitive science frameworks, and to identify the concepts that were the most relevant for our context. Observing experiment courses and discussing with the learning experimenters allowed us to capture those different aspects, both at the task level and at the learner level, and specify them as ontology entities.

2.3 ONTOLOGY IMPLEMENTATION

We implemented the ontology using the software Protégé⁴. First, we formalized the classes corresponding to the material environment of the task, and the classes referring to the learner's cognition and behavior. Then, we have defined specific classes and properties to include the observables of the task, which are situated in time, such as events and situations (in the sense of DOLCE+DnS Ultralite). Finally, we have instantiated these classes, listing their individuals based on the task design itself and the observations collected during the experiments, according to the methodology described in 2.2.3.

2.3.1 MODELING THE ACTIVITY PROCESS

The CreaCube activity itself is a *dul:Process*; that is to say, considered in its evolution. Each scene of interest annotated in the videos is a *dul:Situation*, which we call an *ActivityState*.

Activity states are transitional and situated in time. To account for this transitionality, we use the Description & Situation ontology pattern⁵ of DOLCE+DnS Ultralite (we now use the prefix *dul:* for pre-defined entities).

A *dul:Situation* is, as per the entity specification, a relational context that is created by an observer on the basis of a *dul:Description*. Consequently, a *dul:Description* is itself a conceptualization that provides an interpretation for a set of entities in a given situation.

⁴<https://github.com/protegeproject>

⁵<http://ontologydesignpatterns.org/wiki/Submissions:DescriptionAndSituation>

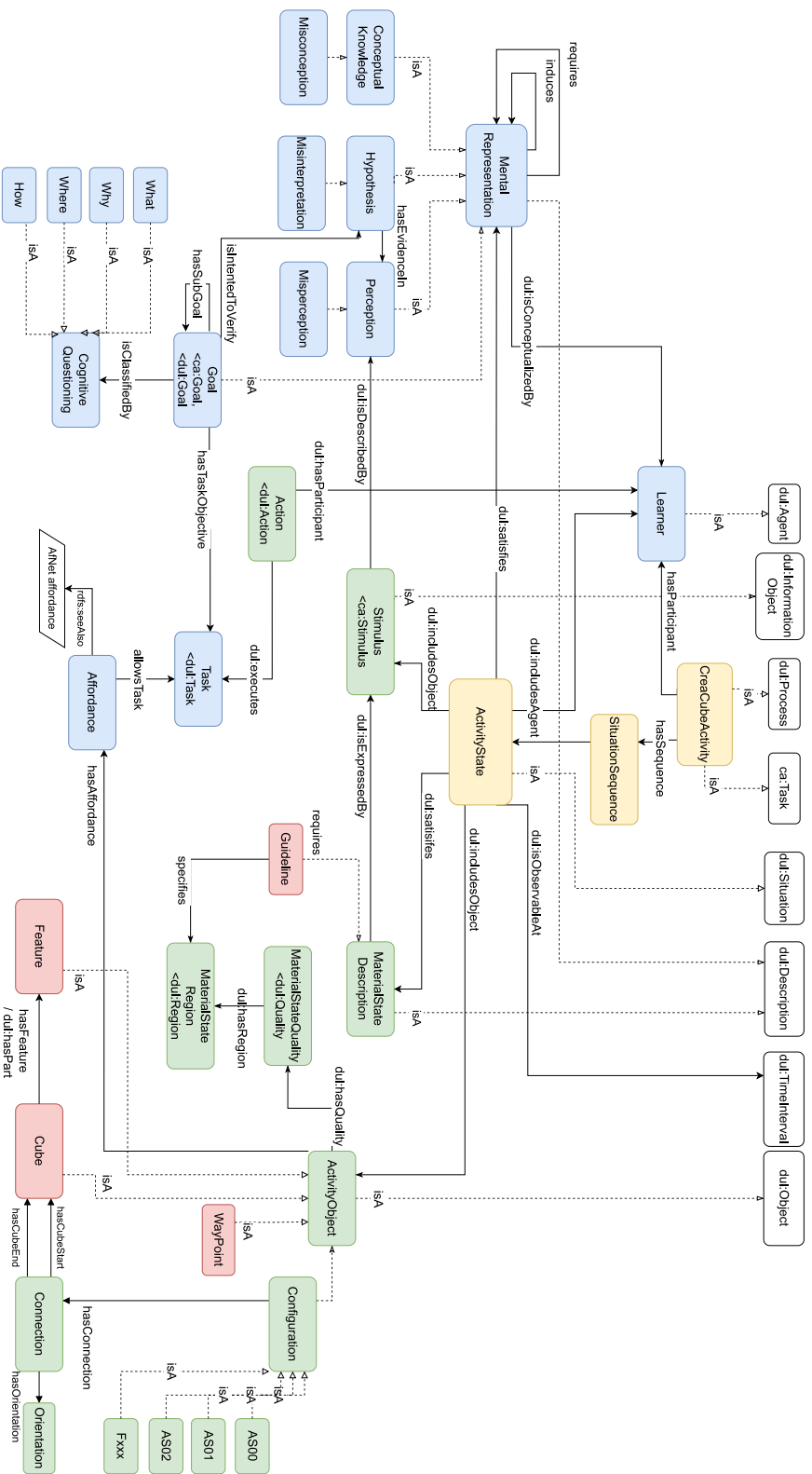


Figure 2.1: A partial diagram exposing the main classes defined in our ontology. The tools mediating the activity are highlighted in red boxes. In green are the other classes related to the observables. Those in blue characterize the learner's behavior and mental representations. Some classes extend the ones defined in the Cognitive Atlas (prefixed by ca:) and/or in the DOLCE+Dns Ultralite ontology (prefixed by duli:). The symbol "<" in labels stands for "isA" when the node is not expanded for concision purposes.

We consider here two aspects of the activity situation: the learner’s mental representation (their representation of the scene) and the material states, from the point of view of an expert of the activity (cf Fig1.4). We therefore assign, for each *ActivityState*, two kinds of *dul:Description*: a *MaterialStateDescription* and a *LearnerMentalRepresentation*.

2.3.2 MODELING THE MATERIAL STATES

The task involves several tools (in the sense of the activity theory), which can consist in *dul:PhysicalObjects* (cubes, waypoints) or *dul:SocialObjects* (guideline recording). Physical objects have *features* that can themselves be considered as physical objects constitutive of the cubes (e.g., switch, wheels, magnets etc.). The cubes can be assembled into multiple configurations (encoded by a series of letters and numbers as per Fig. B) characterizing the material states. As opposed to transitional material states, physical objects have a proper space region and mass. Their state is represented by the values taken by certain properties called *dul:Quality* (e.g., color) in a given qualitative *dul:Region* (e.g., red): for example, a switch has the quality of being either ON or OFF.

Material states are characterized by a specific configurations of those physical objects, their quality values, and correspond to the observables shown on Fig 2.1. We have also listed the possible permutations of the cubes within each configuration shape, so that each configuration encoding (e.g., “F000-BSIW-T”) is characterized by its connections. These connections are themselves specified by the cubes they connect along what spatial axis. This allows us to give a better description of the material states and relates them to their JSON encodings.

These material states represent the problem space of the task in the sense of [Newell and Simon \(1972\)](#). Indeed, the goal states of the task are those where the cubes are in a configuration that satisfies the guideline (which is itself a *dul:Description* describing the requirements for a state to be a solution state). The participants may interact with the material and change its state through their actions: this is formalized as follows in the sequel.

2.3.3 MODELING THE LEARNER’S BEHAVIOR AND MENTAL STATES

ACTIONS AND AFFORDANCES Some actions are only available to the learner in some particular states of the material environment. These states define the preconditions of the action. Similarly, post-conditions define changes of state that an action results in. These pre- and post-conditions also constitute knowledge to acquire. They are represented in the ontology by properties relating a *dul:Task* (referring to the execution of an action) to an *ActivityState* (i.e. the *dul:Situation* before and after the action has been performed). Actions are suggested by the physical objects—or their features—that the learner relates to their knowledge, which we introduced earlier as *Affordances*. Affordances allow a *dul:Task* to be executed. When appropriate, we related those affordances to the ones defined in AfNet ([Varadarajan and Vincze, 2013](#)), a semantic network aiming at categorizing structural affordances based on physical properties of objects.

2 An Ontology approach

STIMULI AND PERCEPTIONS Throughout the activity, the learners receive various stimuli from the material environment. Stimuli are sometimes modeled as events in DOLCE-based ontologies, as is the case in the Stimulus-Sensor-Observation ontology (Janowicz and Compton, 2010). Indeed, in the CreaCube activity, stimuli are most often the result of events occurring when participants proactively interact with the environment. For instance, when they flip the switch on the blue cube, a diode lights up as a response to their action. However, the learner may not always perceive stimuli as soon as a triggering event occurs, and stimuli might remain in the environment for some time before the learner focuses their attention on it (e.g., the diode that lights up is very small and some participants only notice it later). Therefore, we prefer to view them as sensory information input provided to the learner, rather than events themselves; hence, extending the *InformationObject* class. When they are filtered by the attentional focus of the learner, traces of these stimuli are kept active during the task before being eventually consolidated as knowledge. These traces constitute *Perceptions*: they are a conceptualization of the perceived stimuli, thus a kind of *dul:Description*.

KNOWLEDGE RETRIEVAL AND HYPOTHESES To complete the task, the learner needs to retrieve knowledge that has been acquired in the course of past experiences and learning. Most of the time, this knowledge is activated through the discovery of affordances. For instance, after finding of the wheels, the learner might retrieve the fact that wheels on a vehicle must be in contact with the ground to allow it to move. It can also be activated by the concepts evoked in the guideline.

We previously distinguished between conceptual knowledge and procedural knowledge, before deciding to represent only *ConceptualKnowledge* in the ontology because behavioral knowledge consists of procedures and inference rules that are, in our case, not declarative. However, we accounted for the result of inferences with the class *Hypothesis*, representing hypotheses drawn by the learner by contextualizing pieces of conceptual knowledge with regard to their perceptions, using procedural knowledge.

ConceptualKnowledge, *Perceptions* and *Hypotheses* are all mental representations that the learner forms to make sense of the environment. These items extend the class *dul:Description* and are interdependent (i.e. some of them can be inferred or induced from others). These dependencies are represented by ontological properties, such as *requires* (logical necessity), *induces* (inductive reasoning) and *isEvidenceFor/hasEvidenceIn* (e.g., when a perception tends to confirm a hypothesis).

Instead of directly identifying discrepancies at the behavioral knowledge level (which were called “bugs” in Dillenbourg and Self (1992) and on Fig 1.4), we identified *misperceptions* and *misinterpretations*. Misperceptions happen when internal representations of stimuli formed by the learner are erroneous; for example, if the learner misheard the guideline, did not see a visible feature on the cubes, or mistook the dark blue cube for another black cube. Misinterpretations refer to erroneous hypotheses that can be inferred, either from misperceptions or from some conceptual knowledge not transferable to the situation at hand. For example, the fact that wheels automatically start to turn after the user has pulled the vehicle backwards is only applicable in certain situations—

most often, this mechanism is found in children car toys. However, in this task, trying to press down on the white cube by manually rolling it backwards will not cause it to move forward autonomously. Such misinterpretations can be corrected after verification showing a lack of evidence in favor of the hypothesis. Nonetheless, if it comes from a misconception that the underlying conceptual principle is a general truth that is valid in any situation, then it might hinder the completion of the task: as long as the learner keeps on trying to make the wheels work this way, they might not look for other clues. This makes it clear that the learner needs to understand which pieces of knowledge are effectively applicable in the context of the activity and relevant to complete it.

GOALS The learner's *Goals* are a subclass of *dul:Goals*, which are in turn "the Description of a Situation that is desired by an Agent" as per the DUL specification. The combination of a *Goal* and an *Affordance* thus allow to define a *Task*, e.g., a practical way to achieve the *Goal*. Goals are broken down into the four queries described in 1.3 (namely, *What*, *Why*, *Where*, *How*) and towards goal orientation (mastery, performance). As described in section 1.3, goals are a way for the participant to organize their understanding of the task and plan their actions hierarchically. For instance, testing hypotheses defines sub-goals that will help to solve the problem. Global goals (essentially responding to *What* and *Why*, in relation to high-level executive functions (Alexandre, 2021)) may be instructed by an external direction—in this case, the guideline—or a metacognitive reflection, inducing (*Where*, *How*) sensori-motor mechanisms (Alexandre, 2021) thus yielding sub-goals at a more local level. This goal hierarchy is represented in our ontology through the property *hasSubGoal*.

These are clearly preliminary prior assumptions to bootstrap a first version of the ontology, with the perspective to adapt them from further observation by learning sciences experts and logical validation by ontological reasoners.

2.4 RESULTS AND PERSPECTIVES OF APPLICATION

2.4.1 MODELING AS AN INTERDISCIPLINARY TOOL

This work allowed us to confirm the multiple interest of working with ontological modeling when we bring together three disciplines to tackle a problem, as follows:

- Terminological interest to begin with: beyond periphrases and phenomenological descriptions in each discipline, constraining us to define things through lexical choices and fully specified properties and relations between these key words, forcing us to clearly posit what we are talking about.
- Interest in formalization: this approach allows, without even using algorithmic reasoning skills at this stage, to take stock of what can be defined rigorously to formalize completely.

Defining a well-formulated ontology is therefore already by itself a structuring exercise, an epistemological tool in a way, before even using it. As with any modeling, it

2 An Ontology approach

gives an exhaustive and explicit view of what is to retain compared to what is neglected. Once the specification has been set, the use of available reasoners, such as Pellet or Hermit, allows to check its consistency and query the data on the consequences that may or may not be drawn from it.

This being done, we also have an immediate check of the model properties, in particular its logical coherence and soundness. This not only allows us to verify the absence of contradictory statements but we also experimented that if our definitions are not sufficiently specified, then we can easily deduce spurious consequences (e.g., mixing categories of concepts). We thus have an operational tool to verify the completeness of our formalization.

An illustrative example of how a JSON video annotation can be translated into this formalism is given [here](#).

2.4.2 REASONING ABOUT THE MODEL

A query language, such as SPARQL⁶ in the case of RDF data, combined with a reasoner (in our case, Pellet⁷), allows us to retrieve specific information—asserted or inferred—from our model. This raises some perspectives of application both at the material level and at the learner level. Let us illustrate each aspect in the following examples, the former being already operational and the latter being still at an exploratory stage.

RELATING THE CONFIGURATION ENCODINGS TO THE MATERIAL STATES

At the material level, this ontology can be used to retrieve physical information from configuration encodings. In the JSON annotations of the videos, configurations tested by the subject are encoded as strings of characters and letters (e.g., “F000-BSIW-T”) that are not intelligible at a first glance, and hard to relate to real-world connections between the cubes. The ontology does this strenuous work for us: by querying a configuration by its encoding, we can retrieve the associated connections between the cubes. Corresponding ontology assertions are based on simple geometric considerations and the design of the cubes provided by the manufacturer. Conversely, we can also query the knowledge base on configurations which have, for instance, a connection between the blue cube and the red cube. Integrating an ontology-based module to the annotation interface might help to avoid encoding mistakes. A step further, we could also add some transformation rules between those configurations to validate the pre- and post-conditions of the learner’s actions to verify the consistency of the annotations on a sequential scale. For example, if the learner has disconnected two cubes from each other, then we should see that there is one less connection in the new configuration.

⁶We used the Protégé Snap-SPARQL plugin which allows us to perform logical inference and answer queries not only based on asserted data but also inferred data: <https://github.com/protegeproject/snap-sparql-query>

⁷<https://www.w3.org/2001/sw/wiki/Pellet>

The screenshot shows a window titled "Snap SPARQL Query:" with a text area containing the following SPARQL query:

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX : <http://www.semanticweb.org/mnemosyne-line/aide/creacube/material#>

SELECT ?config ?connection ?cubeStart ?cubeEnd ?orientation WHERE {
  ?config rdfs:label "F002-IBWS" .
  ?config :hasConnection ?connection .
  ?connection :hasStartCube ?cubeStart .
  ?connection :hasEndCube ?cubeEnd .
  ?connection :hasOrientation ?orientation .
}
```

Below the query is an "Execute" button. The results are displayed in a table with the following columns: ?config, ?connection, ?cubeStart, ?cubeEnd, and ?orientation.

?config	?connection	?cubeStart	?cubeEnd	?orientation
F002-IBWS	ConnectionWzB	WhiteCube	BlueCube	Top
F002-IBWS	ConnectionSzW	BlackCube	WhiteCube	Top
F002-IBWS	ConnectionBzl	BlueCube	RedCube	Top

Figure 2.2: Example of a SPARQL query on the material states: Retrieving connections and cubes for a given configuration.

INFERRING THE KNOWLEDGE OF THE LEARNER FROM THE OBSERVABLES

At the learner behavior level, the main perspective of this work would be to interpret observables and infer, for example, the participants' goals and associated plans or motivations. Although more data and further investigation is needed to validate the relevance of the ontology inferences on this aspect, we propose here the general idea of such an application.

Storing domain knowledge about behavioral processes in an ontology allows to consistently interpret new data. For example, we may observe a new action which changes the state of the switch, and wonder what motivated the learner to execute this action. As illustrated by Fig. 2.3, a SPARQL query on the model helps us retrieve that the action may have been either a simple reaction to the stimulus of seeing the switch, in the case of a **stimulus-based** behavior, or, in a **goal-directed** mode, the result of a more elaborate reasoning, such as an attempt to understand the function of the blue cube (thus testing the hypothesis that it is indeed a switch) or, this hypothesis being verified, to activate the movement of the vehicle. However, deciding which of the possible scenarios returned by the reasoner (i.e., valid regarding logical consistency) is the right interpretation for a particular learner in a particular situation, remains unclear. We would probably need more complex models such as Bayesian networks (see, e.g., Hidden Markov Models in Chapter 3, Table 3.1) allowing to rank scenarios by probability. Another difficulty is the high computational complexity of assertions regarding the learner behavior, which

2 An Ontology approach

does not scale well with a large knowledge base and may lead to poor reasoning performances.

```

Snap SPARQL Query:
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX creamtrl: <https://gitlab.inria.fr/lineaide-group/creaonto/-/blob/master/creacube_material#>
PREFIX : <https://gitlab.inria.fr/lineaide-group/creaonto/-/blob/master/creacube_behavior#>

SELECT ?action ?stimulus ?goal ?supergoal ?knowledge ?hypothesis
WHERE {
  ?action :changesTheStateOf creamtrl:Switch .
  OPTIONAL {
    ?action :mayBeAResponseTo ?stimulus .
    ?stimulus rdfs:comment ?stimulus .
  }
  OPTIONAL {
    ?action :mayBeDoneBecauseOf ?goal .
    OPTIONAL {
      ?goal :mayBeASubtaskOf ?supergoal .
    }
    OPTIONAL {
      ?goal :mayBeInstructedBy ?k .
      ?k rdfs:comment ?knowledge .
    }
    OPTIONAL {
      ?goal :mayBeIntendedToVerify ?h .
      ?h rdfs:comment ?hypothesis .
    }
  }
}

```

action	stimulus	goal	supergoal
:FlipSwitch	The blue cube has a switch	:TFlipSwitch	:TFindFunctionCubeBe
:FlipSwitch	The blue cube has a switch	:TSwitchOnF	:TMakeConfigMoveAutonomously
:FlipSwitch	The blue cube has a switch	:TSwitchOnF	:TMakeConfigMoveAutonomously
:FlipSwitch	The blue cube has a switch	:TSwitchOnF	:TMakeConfigMoveAutonomously

knowledge	hypothesis
Switches must be 'on' to activate an electronic device	The blue cube is a switch and allows to activate the vehicle
The blue cube has a switch (assuming the color is perceived) /	
Switches are used to activate an electronic device	
Switches must be 'on' to activate an electronic device	

(a) An example SPARQL query

(b) SPARQL results returned for the query (a): We enquired about which actions change the state of the blue switch, what stimuli may have provoked them, or what goal they might be aiming at, and if that goal might be the sub-goal of another one, as well as what knowledge they are using and what hypothesis might be tested.

Figure 2.3: Example of SPARQL queries on the learner's behavior.

PREDICTING THE LEARNER'S BEHAVIOR

Instead of querying the model using backward chaining as in the previous example, we could also use forward chaining to simulate the learner's deduction, hence predicting their behavior. For example, we might wonder how the learner would react to the visual stimulus of seeing the switch. In a **stimulus-based** mode, they would likely try to flip it immediately (assuming that the knowledge relative to switches being flipped has become a procedural knowledge, that is to say automated into an almost reflex behavior). In a **goal-directed** mode, we may infer that the learner will elaborate the hypothesis that the switch can be used to activate the vehicle, which will result in planning goals (flip the switch, connect the cube to the structure) and sub-goals (they have to grasp the cube before connecting it to the others).

This ontology thus provides a subset of plausible behaviors. However, we still need a way to decide which of these behaviors to choose to simulate the learner. Mechanisms to explain what the learner's attention is more focused on in a given situation are still to be developed. Moreover, we need to take into account inter-individual differences: some participants might be more interested in achieving a performance goal, thus appealing for a predominant exploitation strategy, whereas others might pursue a mastery goal and try to understand as much as possible about the environment of the task, displaying more exploratory behaviors.

2.5 DISCUSSION AND CONCLUSION

This study introduces the formalization of a creative problem-solving task. Within this type of task, we need to consider both exploration and exploitation strategies directed

toward different concurrent goals (e.g., performance versus mastery). These goals refer to an objective (*What*) but we also need to consider the *Why*; that is, extrinsic motivation (e.g., task performance) and intrinsic motivation (e.g., increase knowledge). They result in concrete plans to achieve them (*Where* and *How*). The execution of such plans is the observable behavior of the learner that we attempt to capture and formalize in our ontology. Through this process, we take into account the stimuli received during the interactions of the learner with the modular robotic cubes, and the perception and interpretation of the different technological affordances that will support the problem-solving activity through concrete actions to be executed. The learner's prior knowledge is mobilised to understand current perceptions and transform knowledge into a way that makes sense within the context of activity (Musial and Tricot, 2020), formulating and testing in-task hypotheses.

This formalization was achieved step by step with the aim of encapsulating domain knowledge used by learning science researchers when they analyze such an experiment, in order to make it more robust and reproducible. The purpose of this description is not to remain phenomenological; that is, driven by a theoretical framework expressed in human language to guide the analysis of the data. We hereby reformulate these notions at a computational level to give us the means to process them in a more systematic way using an algorithmic approach, as discussed here, specifying the concepts invoked in a formal language.

As stated in the introduction, we model the learners' behavior and the artefact configurations during a creative problem-solving activity that is ill-defined at the participant level, while the activity state space is entirely formalized at the observer level. As such, our approach is voluntarily specific, while the proposed methodology is easy to generalize, as pointed out in (Barnabé et al., 2020) where three other computational thinking learning tasks have been also considered. From the paradigm choice to rely on an ontology modeling to the general knowledge about cognition when engaged in a creative problem solving task, an important part of this formalization translates to another learning task. Relying on a robust foundational ontology such as DOLCE also ensures compatibility and facilitates such knowledge transfer. Besides, our paradigm is based on frameworks that are already widely used in learning sciences such as the activity theory and the ITS learner/observer representation, combined with a cognitive basis.

With respect to the state of the art, briefly reviewed previously, our approach applies to tangible activities, modeling both the task and the learner in the task. This approach is rather disruptive with respect to learning science because, to the best of our knowledge, there has not been such modeling of an ill-defined task using an ontology. This is only a first step but it already shows that formalizing such a problem-solving activity is a complex but attainable work. From a practical point of view, this ontology has proven useful to collaborate with other teams; for instance, we are working towards developing a version of the task in virtual reality (Le Meudec et al., 2024) and automatizing the annotation process of the video. For these two applications, having a formal description of the task at hand has been a valuable asset.

2 *An Ontology approach*

To complete this description, one of the next steps would be to better formalize the relationship between affordances and knowledge (e.g., we could draw inspiration from [Asprino et al. \(2017\)](#) who proposed a design pattern involving the use of frames⁸). We also consider to address the emotional aspect, such as how to identify observable emotions (e.g., the ones listed in Fig. B) and link inferred cognitive states to emotional states. Finally, we need to make use of reasoning mechanisms to better infer the subject internal motivation and behaviors from the experiment observables (e.g., learner state observation, post-activity evaluation) and conduct statistical analyses on the results to assess the validity of such an application.

⁸in the sense of <https://framenet.icsi.berkeley.edu/fndrupal/frameIndex>

3 A MARKOVIAN APPROACH: MODELING CPS BEHAVIORS WITHIN A SYMBOLIC STATE SPACE

3.1	Introduction	77
3.2	Symbolic state space specification	78
3.2.1	General framework	78
	Markov Decision Processes	78
	Reinforcement Learning (with a pseudo-episodic memory)	79
3.2.2	Finding an optimal policy	80
	Temporal Difference Learning (TD-Learning)	80
	Model-free vs model-based RL	82
	Q-Learning	82
3.2.3	Symbolic data structure	83
	Input structure	83
	A symbolic metric space	84
	Inferring other elements from input	85
3.3	Reinforcement learning on a symbolic state space.	86
3.3.1	Q-Learning on symbolic data	86
3.3.2	Preliminary results	87
3.4	Discussion: perspectives of application to the CreaCube task	89
	Towards inverse reinforcement learning	89
	State clustering, trajectory clustering	91

3.1 INTRODUCTION

Complex problem solving involves representing structured knowledge, reasoning and learning, all at once. In [Mercier, Alexandre, et al. \(2021\)](#), we introduce the idea of applying a reinforcement learning paradigm to a symbolic representation of a concrete problem-solving task, modeled here by an ontology. This chapter develops this idea and proposes a small proof-of-concept as well as perspectives of further applications and connections to other frameworks.

Reinforcement Learning (RL) is a computational paradigm whereby an agent interacts with its environment and learns to perform sequences of actions by maximizing some notion of cumulative reward ([Sutton and Barto, 2018](#)). This family of models have been extensively used in cognitive neuroscience to model high-level executive control functions (e.g., [Domenech and Koehlin \(2015\)](#) and [Rmus et al. \(2021\)](#)).

3 A Markovian approach

In Chapter 2, we have introduced the operationalization of a creative problem-solving task, via the construction of an ontology. The state is defined by the configuration of the objects manipulated during the task and some other observables regarding the learner. The key point is that, in our case, the internal state of the subject and the external state of the task material constitute a complex structure, modeled here by a set of “statements”; that is to say, in the ontological vocabulary, entities typed by classes and linked together through relationships labeled by properties. Unlike usual mechanisms based on Markov chains, the state space to consider in our RL setup is thus not reduced to an unordered finite enumeration. We would like to study here to what extent reinforcement learning could be designed on such state spaces.

In “symbolic reinforcement learning” (e.g. [Garnelo et al. \(2016\)](#)), deep neural networks transform raw perceptual data into a symbolic representation which is then fed to a symbolic module that might perform, for instance, action selection. Other approaches such as [Ma et al. \(2020\)](#) propose architectures where a numerical reinforcement algorithm communicates with a reasoner. Here we would like to explore another track and make the reinforcement algorithm work directly on the symbolic data space itself in a more integrative way, which we propose to call “reinforcement symbolic learning”.

3.2 SYMBOLIC STATE SPACE SPECIFICATION

3.2.1 GENERAL FRAMEWORK

MARKOV DECISION PROCESSES

This typical setting for RL is the *Markov Decision Process* (MDP) framework.

Markov Decision Process (MDP)

A Markov decision process is a 4-tuple $(\mathcal{S}, \mathcal{A}, P_a, R_a)$, where:

- \mathcal{S} is a set of states called the state space,
- \mathcal{A} is a set of actions called the action space (alternatively, \mathcal{A}_s is the set of actions available from state s),
- $T_a(s, s') = \mathbf{Prob}(s_{t+1} = s' \mid s_t = s, a_t = a)$ is the probability that action a in state s at time t will lead to state s' at time $t + 1$ (i.e., the transition function),
- $R_a(s, s')$ is the immediate reward (or expected immediate reward) received after transitioning from state s to state s' due to action a .

Some extensions of this framework allow to involve multiple agents or to infer an optimal behavior from partially observable states (see Table 3.1). All of these frameworks have in common the *Markov property*, which states that a future state s' must only depend on the current state s and not on its predecessor: in other words,

$$\mathbf{Prob}(s_{t+1} | s_t, a_t) = \mathbf{Prob}(s_{t+1} | s_0, \dots, s_t, a_t)$$

i.e. the transition function is memoryless. A workaround for this is to introduce some sort of **episodic memory**, as detailed below.

	Agency (control over the state transitions)		
	No agent	1 agent	Multiple agents
States totally observable	Markov chain	Markov Decision Process (MDP)	Markov Stochastic Game
States partially observable	Hidden Markov Model	Partially Observable Markov Decision Process (POMDP)	Partially Observable Stochastic Game (POSG)

Table 3.1: Markovian frameworks and their use-cases

The state and action spaces may be finite or infinite, for example the set of real numbers. Some processes with countably infinite state and action spaces can be reduced to ones with finite state and action spaces.

REINFORCEMENT LEARNING (WITH A PSEUDO-EPISODIC MEMORY)

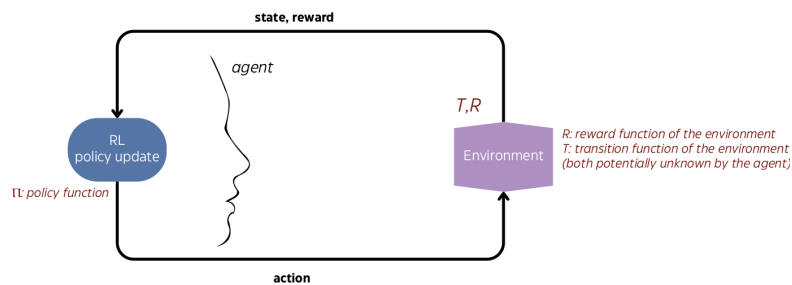


Figure 3.1: The classical Reinforcement Learning paradigm

In such a **MDP** set-up, our agent interacts with the environment until it reaches some terminal state at time T (potentially infinite). At a given discrete time, it perceives a part of the environment, i.e., a stimulus, including a reward. It infers elements (e.g., causes) from this input cue, including the computation of the next action. One such traversal is called an **episode** and consists of a sequence of states, actions, and rewards that the agent got on his path. The sequence of actions defines the **behavior** of the agent.

3 A Markovian approach

In our case, we consider a potentially hypermnesic agent for which any previous input, including rewards, might be part of the internal state. This choice is directly related to the notion of **episodic memory**, an episode here being represented by an event sequence (a list of times of occurrence of atomic events). This frees us from the Markovian constraint taking into account not just one step in the past, at the cost of a multiplicative increase of the state space. This is going to be manageable thanks to the hierarchical structure of our state space, which encompasses a lot of information without the need for an exhaustive enumeration.

3.2.2 FINDING AN OPTIMAL POLICY

The agent's behavior follows a **policy** function π (potentially stochastic) that predicts the next action based on the current state:

$$\pi : \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$$

$$\pi(a, s) = \mathbf{Prob}(a_t = a \mid s_t = s)$$

This **policy** can be either deterministic or stochastic. The goal of the agent is to optimize the **policy**, that is to say, to decide which action a to take in a certain state s that will result in the best possible reward at a more or less long-term.

In practice, rewards are usually not immediately observable. There are even settings where the only reward happens at a terminal state and no intermediate reward is observed. Instead, we use the discounted return R , which is the combination of discounted rewards in the future (future rewards are valued less).

$$R = \sum_{t=0}^T \gamma^t r_{t+1}$$

γ is a discount factor accounting for the importance of future rewards: a factor of 0 means that the agent only considers current rewards, whereas a factor of 1 makes it strive for long-term rewards (with the risk of a cumulative reward becoming infinite, resulting in diverging computed values).

Hence, what is actually maximized is the **state value** function V^π of a state s defined by:

$$V^\pi(s) = \mathbb{E}_{a \sim \pi} [R \mid s_0 = s]$$

that is to say, the expected return of a trajectory starting from state s and following the **policy** π .

TEMPORAL DIFFERENCE LEARNING (TD-LEARNING)

The Temporal Difference Error is the difference between the ultimate correct reward and our current prediction:

$$\begin{aligned} \text{TDE} &= V_t^* - V_t \\ \text{TDE} &= V_t^* - V_t \\ &= r_{t+1} + \gamma V_{t+1} - V_t \end{aligned}$$

In *TD-learning*, the value is updated following the Bellman equation ¹:

$$\begin{aligned} V(s) &\leftarrow V(s) + \alpha \text{TDE} \\ &= V(s) + \alpha \left(\underbrace{r + \gamma V(s')}_{\text{The TD target } V_t^*} - V(s) \right) \end{aligned}$$

The learning rate α (or step size) determines to what extent newly acquired information overrides old information. A factor of 0 makes the agent learn nothing (exclusively exploiting prior knowledge), while a factor of 1 makes the agent consider only the most recent information (ignoring prior knowledge to explore possibilities). In fully deterministic environments, a learning rate of $\alpha_t = 1$ is optimal. When the problem is stochastic, the algorithm converges under some technical conditions on the learning rate that require it to decrease to zero. In practice, often a constant learning rate is used, such as $\alpha_t = 0.1$ for all t .

Values and policies are updated iteratively:

$$\begin{aligned} V(s) &:= \sum_{s'} P_{\pi(s)}(s, s') (R_{\pi(s)}(s, s') + \gamma V(s')) \\ \pi^*(s) &:= \operatorname{argmax}_a \left\{ \sum_{s'} P_a(s, s') (R_a(s, s') + \gamma V(s')) \right\} \end{aligned}$$

EXPLORATION VS. EXPLOITATION When interacting with its environment, if the agent only takes actions leading to immediate rewards, actions and states not on the greedy path may not be sampled sufficiently and potentially better rewards would stay hidden from the learning process. Choosing between making the best decision given the current information or starting exploring and finding more information, is known as the *exploration/exploitation* dilemma.

A greedy *policy* (choosing the best value each time) would mean full-on *exploitation*, which can make us stuck in the local minima. Conversely, full-on *exploration* would mean that we would need a lot of time to collect the needed information. In order to balance this dilemma, we can resort an ϵ -greedy *policy* as explained below.

ON-POLICY VS OFF-POLICY On-policy methods solve the exploration/exploitation dilemma by including randomness, meaning that non-greedy actions are selected with some

¹The Bellman equation, foundational to dynamic programming, is commonly used for optimization problems and can be used iteratively: the current value to optimize will be updated by itself + learning rate * error.

3 A Markovian approach

probability. These policies are called ϵ -greedy policies as they select random actions with an ϵ probability and follow the optimal action with $1-\epsilon$ probability.

While on-policy algorithms try to improve the same ϵ -greedy policy that is used for exploration, off-policy approaches have two policies: a behavior policy b and a target policy π . The behavioral policy is used for exploration and episode generation, while the target policy is used for function estimation and improvement. Therefore, the target policy π gets a “balanced” view of the environment and can learn from potential mistakes of b while still keeping track of the good actions and trying to find better ones. The distribution “mismatch” between the two policies can be facilitated by a technique called importance sampling.

MODEL-FREE VS MODEL-BASED RL

RL algorithms are also categorized with regard to how they represent the environment. *Model-based* algorithms maintain a representation of the environment beyond the state and action space, typically the transition and reward function. The agent is thus able to plan the consequences of its action (post-conditions). In contrast, *model-free* algorithms do not maintain such an explicit model, but rather a chunked representation of values, computed from the previous events and rewards. The *model-free / model-based* dichotomy has traditionally often been mapped onto models of *habitual* and *goal-directed* behavior, but recent works suggest that such a mapping is less relevant than e.g. value-free vs value-based algorithms (indeed, habitual behaviors can result from a simple association between stimuli and actions, without any kind of reinforcement through reward, due to Hebbian plasticity) (Miller et al., 2018); besides, *model-free* and *model-based* learning are probably not that segregated in the brain, and may rely on shared components or representations (A. G. E. Collins and Cockburn, 2020).

Q-LEARNING

Value evaluations in *TD-learning* require a full episode, hence it can be useful to define *state-action values* to immediately assess the quality of a situation. The *state-action value* of the pair (s,a) under π is defined by:

$$Q^\pi(s, a) = \mathbb{E}_{a \sim \pi} [R \mid s_0 = s, a]$$

where R now stands for the expected return associated with first taking action a in state s before following π from the next state s' .

Q-learning is a popular off-policy learning algorithm, illustrated on Figure 3.2. It updates its state-action values using the state-action value of the next state s' and the greedy action a' . In other words, it estimates the return (total discounted future reward) for state-action pairs assuming a greedy policy is followed (by always choosing the action with the maximum value), despite actually following an ϵ -greedy policy:

3.2 Symbolic state space specification

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \underbrace{\alpha}_{\text{learning rate}} \left(R_t - \gamma \max_a Q(s_t, a) - Q(s_t, a_t) \right) \quad (3.1)$$

$$= Q(s_t, a_t) + \underbrace{\alpha}_{\text{learning rate}} \left(r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \right) \quad (3.2)$$

Here, we update the function using both a_t , which is chosen by the ϵ -greedy behavior **policy** π , and $\max_a Q(s_t, a)$ which is selected by the greedy target **policy**.

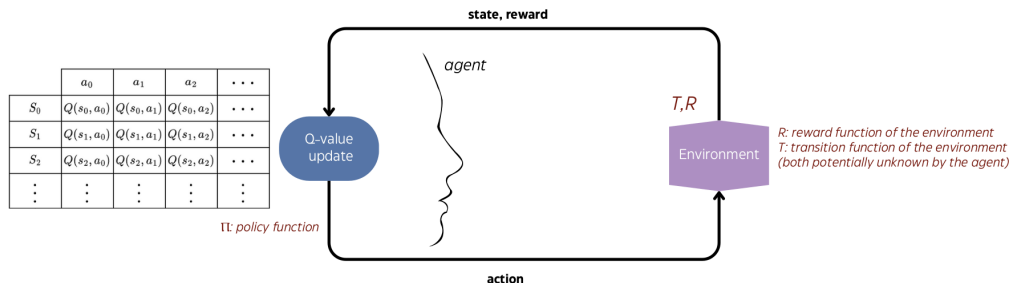


Figure 3.2: The Q-Learning paradigm

Q-learning has proven to be a powerful technique for a variety of problems, including game playing, robotics, and control systems. However, one limitation of Q-learning is that it can struggle to generalize to new and unseen states. This is because Q-learning relies on an explicit representation of the state space, which can be difficult to achieve for complex and high-dimensional environments. To avoid this issue, techniques combining Q-learning with deep neural networks, such as the deep Q-networks (DQN) approach (Mnih et al., 2015), have been proposed.

The approach proposed here, following Mercier, Alexandre, et al. (2021), is less ambitious and more economical: To what extent could classical reinforcement learning algorithms not interact but intrinsically manipulates symbolic information? Considering Q-learning with state and action being symbolic data structures, how can we adapt the former algorithmic scheme?

3.2.3 SYMBOLIC DATA STRUCTURE

INPUT STRUCTURE

At the computing level, we aim at manipulating the symbolic representation of knowledge as hierarchical data structure, as introduced in Mercier and Vieville (2023). Concepts are anchored in an input/output, i.e., stimulus/response, framework, which might consist of sensorimotor feature spaces (colored regions) corresponding, for example, to different sensor modalities. Literal values are taken among a finite enumeration of qualitative values (e.g., a color set) or quantitative values (i.e., finite precision bounded values). They are completed by meta-information that is not explicitly manipulated by

3 A Markovian approach

the agent but is used for process specification or interpretation (e.g., the weight unit and bounds).

At the modeling level, we follow [Gärdenfors \(2004\)](#), with the idea that an individual resource can be defined by “feature dimensions,” i.e., attributes with some typed value. For instance, a bird could be the following. The used syntax is a **weak form of the JSON syntax**, which could be equivalently represented as **RDF triples**².

```
bird: {
  is_a: vertebrate
  can: { sing fly eat: { worm fish } }
  has: { feather beak }
  is: { weight: { min: 0.010 max: 50 unit: kilogram } }
}
```

This illustrative example is sufficient to allow us to detail the main characteristics of our representation. Concepts can be hierarchical, e.g. `bird` is a sub-concept of `vertebrate`. Some features are intrinsic properties, and others are relations. A property can be qualitative (e.g. `beak`, or quantitative (e.g., the `weight`). The features can be hierarchical, either because the value is an enumeration (e.g., `can` takes a value in `{sing, fly}`) or because the value has some metadata attributes (e.g., `weight`). Such a data structure defines a “concept” in the sense of [Gärdenfors \(2004\)](#) and corresponds to a geometric conceptual space as reviewed in Chapter 1.

Using Vector Symbolic Architecture implemented at the neural spiking assembly level thanks to the Neural Engineering Framework ([Eliasmith, 2013](#)), such a cognitive symbolic data structure can be implemented as biologically plausible memory, allowing to manipulate it conjointly at both a symbolic and numeric level ([Mercier and Vieville, 2023](#)), as developed in Chapter 4.

A SYMBOLIC METRIC SPACE

Distance

A metric or *distance* function is a function $d(\cdot, \cdot)$ which satisfies the following rules:

- The distance between an object and itself is always zero: $\forall x, d(x, x) = 0$
- The distance between distinct objects is always positive: $\forall x \neq y, d(x, y) \geq 0$
- Distance is symmetric: $\forall(x, y), d(y, x) = d(x, y)$.
- “Intermediate stops can’t speed you up”, also known as the triangle inequality: $d(x, z) \leq d(x, y) + d(y, z)$.

A metric or distance allows to defined *geodesics*, i.e. the shortest path between two points. A *geodesic* can be defined for a Riemannian *manifold* (e.g. a surface in Euclid-

²See also <https://www.w3.org/TR/rdf-json/>

ian geometry) but also has a sense in graph theory, where the distance between two vertices in a graph is the number of edges in a shortest path (also called a **graph geodesic**) connecting them.

Edit distance is originally defined as a string metric, i.e. a way of quantifying how dissimilar two strings are to one another. They can be thus extended to sequences or even (under certain constraints) tree structures.

Edit Distance (strings)

Given two strings a and b on an alphabet Σ , the **edit distance** $d(a, b)$ is the minimum-weight series of edit operations that transforms a into b .

One of the simplest sets of edit operations is that defined by Levenshtein (1966):

- Insertion of a single symbol. If $a = uv$, then inserting the symbol x produces uxv .
- Deletion of a single symbol x changes uxv to uv .
- Substitution of a single symbol x for a symbol $y \neq x$ changes uxv to uyv .

In this definition, each of these operations has a unit cost. The Levenshtein distance is thus equal to the minimum number of operations required to transform a to b .

Ouangraoua and Ferraro (2009), following K. Zhang (1996), proposed an extension of edit distances to semi-ordered labeled trees.

An input s is thus a tree data structure, equipped with a *partial semi-order* compatible with an *extended semi-distance*³. This means that two values may be equal, indistinguishable (i.e., too close to be ordered, thus equal or not), comparable or incomparable (i.e., too different to be compared). The distance between two inputs is defined as the minimal cost of editing sequences transforming one input into another, i.e., an **edit distance**. We consider *edit operations* given an input (l+) adding, (l-) deleting or (l#) changing a value in a list, (t+) defining, (t-) undefining or (t#) changing a value in a tuple, each of these operations having a user-defined positive cost. This mechanism also allows, when comparing two inputs, to make explicit which node has been added, deleted, or changed. Thus, it offers a “geodesic”, i.e., a path of transformation from one structure to another, allowing to interpolate intermediate input structure between both of them.

INFERRING OTHER ELEMENTS FROM INPUT

Each data structure is translated in terms of RDF statements as follows. Each tuple is a “subject” and each named value corresponds to a relationship labeled by a “property”,

³The key point is that we consider restricted edit distances preserving the tree filiation, computable in polynomial time (Ouangraoua and Ferraro, 2009), which would not have been the case otherwise, or if considering the tree as a general graph or ontology portion.

the value being the “object” targeted by the relationship. This transformation allows us to generate an **ontology**, offering the possibility to perform inferences, thus implementing the learner behavior at a pure symbolic level. Conversely, each RDF ontology graph may be mapped back onto the data structure, with tuples having the value undefined if the corresponding statement is absent from the ontology after reasoning, and defined as the property object-s otherwise — hence the need for a pre-defined schema.

3.3 REINFORCEMENT LEARNING ON A SYMBOLIC STATE SPACE.

3.3.1 Q-LEARNING ON SYMBOLIC DATA

Let us consider a concrete example on a Q-learning set-up with ϵ -greedy exploration. At each step this function is updated, using the weighted average of the old value and the new information, while the action is chosen to either maximize the reward, given the state, or with a small probability randomly explore new actions. This algorithm is known as **model-free**, however, we are using it in a non-conventional way with inferences we generate from prior information on symbolic states, bringing it somewhat closer to model-based algorithms.

The key point is that, given the largeness of the state space, each state is very likely different from another, so that one state is very likely visited once, making it impossible to use the usual update rule on tabulated values $Q[s_t, a_t]$. However, given a new state s_t and reward r_{t+1} , we can easily update all preceding state-action values in a neighborhood.

Considering an exponential weighting of radius ρ , for a learning rate α , a discount factor γ , this writes, for all state-action values:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha e^{-\frac{\overbrace{d(s_{t+1}, s_t)}^{\text{edit distance}}}{\underbrace{\rho}_{\text{neighborhood radius}}}} \left(\underbrace{r_{t+1}}_{\text{immediate reward}} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \right)$$

where $d(s_{t+1}, s_t)$ stands for the predefined edit distance between both states. During an epoch (or episode) of T steps, it means that we have to compute $O(T^2)$ edit distances, and state-action value updates, but this complexity depends only on the trajectory length, rather than on the state space itself.

The computation of the maximum reward prediction $\max_a Q(s_{t+1}, a)$, requires to both (i) interpolate the state-action value given available tabulated values and (ii) enumerate action candidates, including unprecedented actions. For (ii) consider a set of potential actions a_k including previous actions, predefined prototypical actions, and putative actions generated by an external process. For (i) we can use an exponential interpolation, consistent with the previous design choice:

3.3 Reinforcement learning on a symbolic state space.

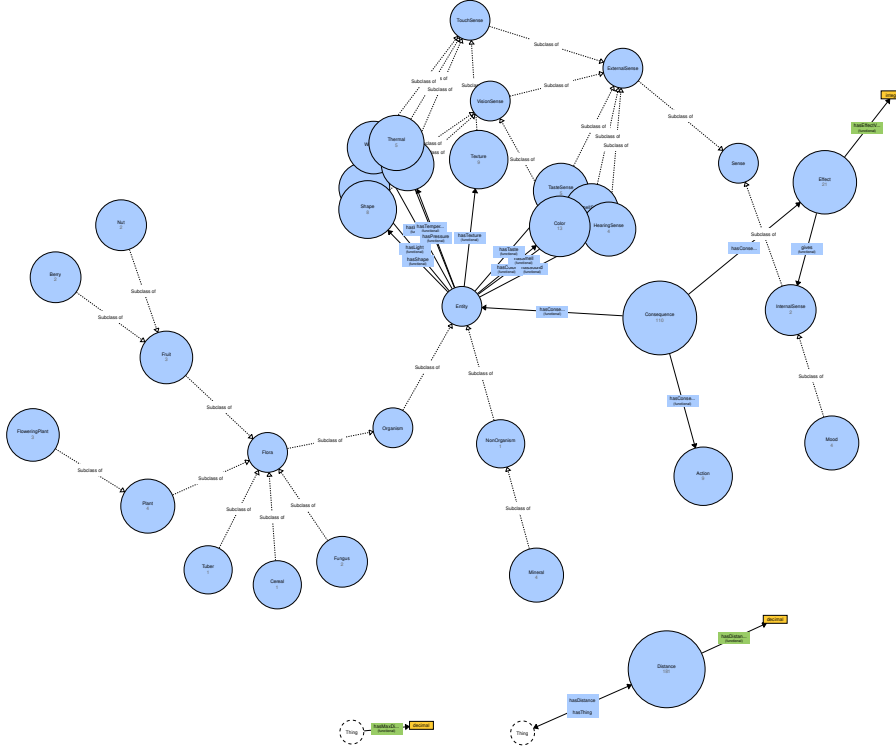


Figure 3.3: Ontology of the symbolic environment used by Radji et al. (2023).

$$Q(s, a_k) = \sum_{s_t, a_t} e^{-\frac{d(s, s_t) + d(a_k, a_t)}{\rho}} Q[s_t, a_t] / \sum_{s_t, a_t} e^{-\frac{d(s, s_t) + d(a_k, a_t)}{\rho}}$$

3.3.2 PRELIMINARY RESULTS

As part of a student project we supervised, Radji et al. (2023) have applied some of the ideas introduced earlier. They have implemented an environment, described as an OWL ontology, illustrated in 3.3, in which an agent needs to explore the environment and choose the best actions for its survival. Their code is open-source and available at <https://github.com/riiswa/symbolic-rl>.

SET-UP At each step, the agent receives an observation (among 6400 possible combinations) which is composed of an integer that represents the internal state of the agent, and another integer that represents the entity that the agent encounters in the environment. Each entity has some attributes (defined as ontology properties) that belong to specific sensory modalities (similar to feature spaces in Olteteanu (2020)’s representation). The internal state of the agent is a vector calculated from homeostatic and emotional variables which can have either a positive effect (energy, health, joy) or a negative one

3 A Markovian approach

(anger, sadness, fear). The agent must then choose an action which will have a postcondition (defined in the environment ontology) updating the agent’s internal state. The **edit distance** used in this study quantifies the similarity between two observations, defined as the average of the distances between their internal states and their associated external entities. More specifically, Radji et al. (2023) propose to calculate the distance between two entities using two sub-metrics: a “tree distance”, defined as the sum of the length of the ancestors of each individual in the ontology; and a “property distance” which is an empirical distance within a given feature space (for example, in the color space, the distance between yellow and orange might be defined as 1, while the distance between yellow and red might be defined as 2). The distance between two internal states $d(i_1, i_2)$, is calculated as the usual Euclidean normalized distance of their two representation vectors. The agent takes as argument the distance function to compute the distance between two observations, and for faster computation, distances between each observation are pre-computed and stored in a matrix of size 256^2 . At each step, the agent updates the Q-table (of size $256 \times 25 \times 9$ in this set-up) using equation 3.3.1.

RESULTS Radji et al. (2023) trained both classical Q-Learning agents and symbolic Q-Learning agents for 10000 steps, using a linear epsilon greedy policy with the same hyperparameters (except for the radius ρ which exists only for symbolic Q-Learning), and evaluated the agents performances every 10 training steps. The evaluation involved performing 100 rollouts in 100 environments with fixed seeds, and the score of an agent was calculated as the sum of the cumulative rewards obtained during each rollout.

Figures 3.4 show the evolution of the scores of the different agents throughout the training. Unsurprisingly, the Q-Learning agent struggles to learn, which is simply due to the fact that the Q-Table is too large. The Symbolic Q-Learning agent seems to have much better generalization capabilities and needs far fewer steps in the environment to perform. Radji et al. (2023) propose several hypotheses of interpretation:

- With a very small ρ for example (0.01) Symbolic Q-Learning will not update the Q-Table significantly, so the performance will be close to that of classical Q-Learning.
- With a slightly larger radius (0.04) it is possible to obtain performances that are significantly better than the classical Q-Learning, so we see that the ρ hyperparameter is very sensitive, and a small change can drastically change the results.
- When ρ is a bit too large (0.06) the symbolic Q-learning agent seems to try to generalise too much, and probably has to update values that it shouldn’t, which makes its learning go quite badly, and even performs worse than Q-Learning. The results are quite unstable compared to smaller values of ρ . A potential improvement could be to lower the radius over time, in order to “stabilize” the agent’s generalization process.

Radji et al. (2023)’s work thus constitutes an interesting proof of concept of combining Q-learning with a symbolic ontological description of the environment, ensuring that

3.4 Discussion: perspectives of application to the CreaCube task

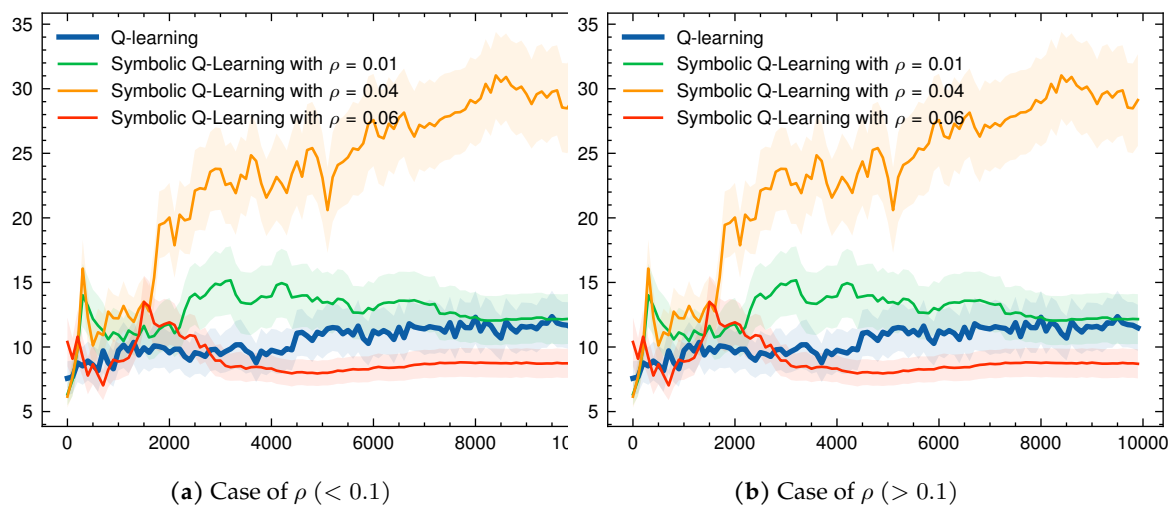


Figure 3.4: Comparison of Q-Learning and Symbolic Q-Learning performances with varying ρ

the state space remains tractable. The results we obtained with Symbolic Q-Learning suggest that this approach could be a promising direction for our research perspectives.

3.4 DISCUSSION: PERSPECTIVES OF APPLICATION TO THE CREA CUBE TASK

How could we use the previous framework to interpret the observed behavior of a learner engaged in a creative problem-solving task such as CreaCube, from a Markovian perspective?

TOWARDS INVERSE REINFORCEMENT LEARNING

A perspective of such a framework would be to characterize convergent-divergent phases and goal orientation in the CreaCube task (or some other creative problem solving task). At the cognitive level, we need to consider both exploration and exploitation strategies related to convergent and divergent thinking as described earlier. At a computational level, the paradigm is reinforcement learning (RL) with a final reward (success of the task) and some intermediate rewards (discoveries of affordances, partial result regarding the goal), more precisely intrinsically motivated reinforcement learning (Singh et al., 2010).

The key point is that the reward is not provided by the environment but computed by the agent itself, using both a representation of the state of the environment, and the beliefs and goals integrated by the agent. For example, we could imagine that an agent oriented towards performance would want to solve the problem as fast as possible, and would find it more rewarding when the distance between (i) its representation of the state of the environment and (ii) its representation of a possible goal implied by the guideline, decreases as much as possible. Conversely, an agent oriented towards mas-

3 A Markovian approach

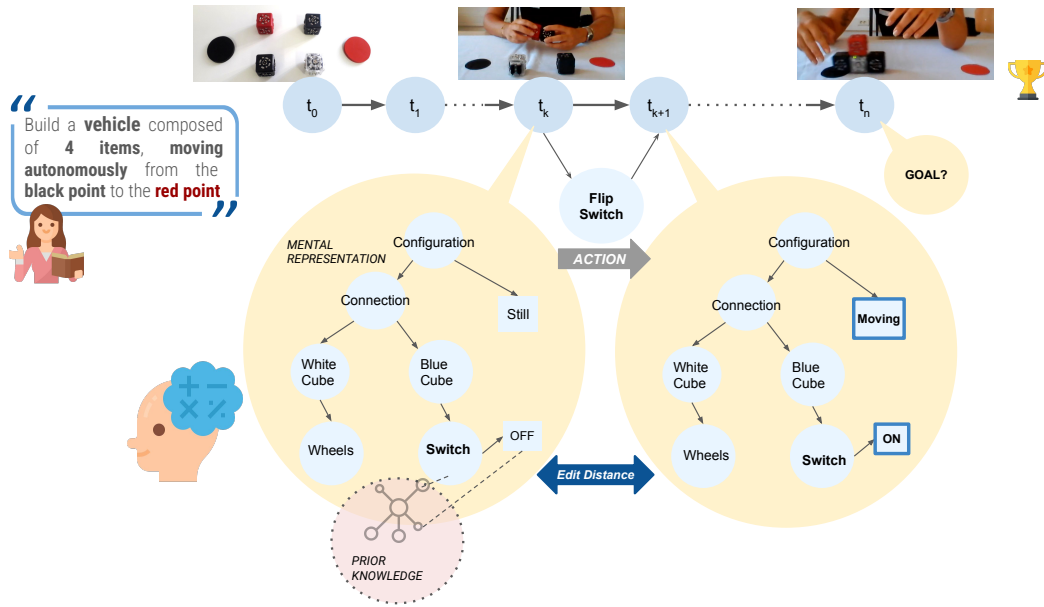


Figure 3.5: Perspective: a reinforcement symbolic learning paradigm applied to the CreaCube task.

tery would seek to find new states that are very semantically distant from the present state, in order to gather more information (using some entropy measure). Kohler and Romero (2023) have formalized an **edit distance** between CreaCube configurations at the static material level, that we could partially use in the calculus of the edit distance proposed in our context (we would need to add other information regarding the state of the configuration, e.g. is the switch on, is the configuration moving, etc).

More precisely, we aim to evaluate the cost toward a given goal from the previously defined distance between a current state and the estimated subgoal state. Regarding the performance-related goal, we thus make the strong assumption that the shorter this distance, the lower the potential cost, that is, that edit distance operations can be related to actions generating a trajectory in the state space. We also consider that the reward provided by a subgoal is related to the estimated distance to the desired goal; the shorter this distance, the higher the reward. Combining the cost and reward provides a rule for choosing the goal: the one with the maximal reward at the minimal cost. Such a common currency is directly related to emotion, as studied by Dietrich (2004), reported by Alexandre (2020a) and reviewed in Chapter 1 (1.3). Regarding mastery-related goals, we must introduce the notion of missing information, which corresponds to the fact that we cannot apply a behavioral rule because we cannot decide whether a precondition is to be applied, that is, because the current state of the knowledge is not in a region where the information is available⁴. This generates an internal subgoal, which is to minimize

⁴Given a sensorimotor feature or meta information, the region where the feature is defined corresponds simply to a region where the feature value is any defined value but the default “undefined” value.

the distance to such a region, and is thus still a distance minimization problem. This design choice is coherent with the fact that, with performance-related goals, we may apply a rule, even if the precondition is not false but not totally verified, to attempt to reduce the distance to the final goal, while, with mastery-related goals, intermediate subgoals are added.

The next step would be to infer the reward function of an observed agent, given its observed behavior, using Inverse Reinforcement Learning (IRL) as illustrated in Figure 3.6 (see, e.g., [Arora and Doshi \(2020\)](#) for a survey).

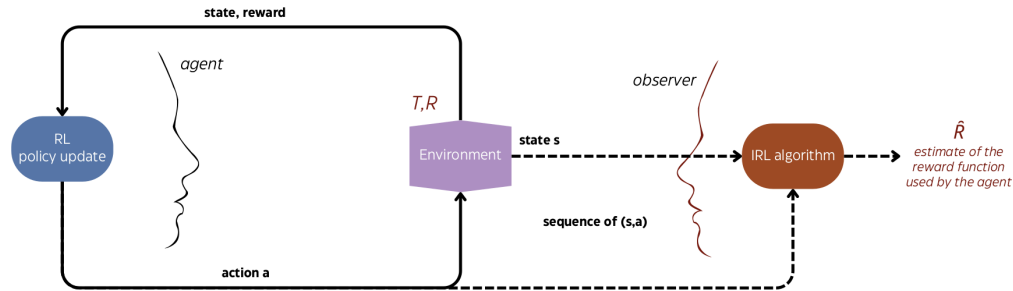


Figure 3.6: The Inverse Reinforcement Learning paradigm

Regarding the characterization of convergent-divergent behaviors, we could in first approximation map it on exploration-exploitation behavior switching, using the Probe model of task sets.

A step further, we could implement a more sophisticated RL algorithm using task sets, to allow for different behaviors depending on the context.

In first approximation, we could consider that the contexts correspond to the stages of componential model of [Amabile \(2012\)](#): preparation, generation, or generation regulation, enforced by selecting a **task set** of a corresponding category. Given a current stage, for example generation, there is a concurrent process between carrying on and stopping to switch to the evaluation stage, the result of which could be to consider success, failure, or starting another task.

However, what if the **task sets** contexts do not correspond to the componential model's stages? Instead of testing every possibility, we could instead derive a certain number of contexts by using other methods such as clustering, reviewed below.

STATE CLUSTERING, TRAJECTORY CLUSTERING

As a quantitative step towards the analysis of the collected data, we propose to fit the observed problem solving trajectory on Bayesian models using statistical methods, while we can also benefit from logical inferences that are allowed by our ontology representation. The goal is to infer the internal state of the learner (e.g. their beliefs, goals, emotions) from the observed state. A first idea is to use Hidden Markov Models (HMM) (see Table 3.1) where each activity sequence could be seen as a Markov chain with hidden states (that, we hypothesize, correspond to internal representations of the learner)

3 A Markovian approach

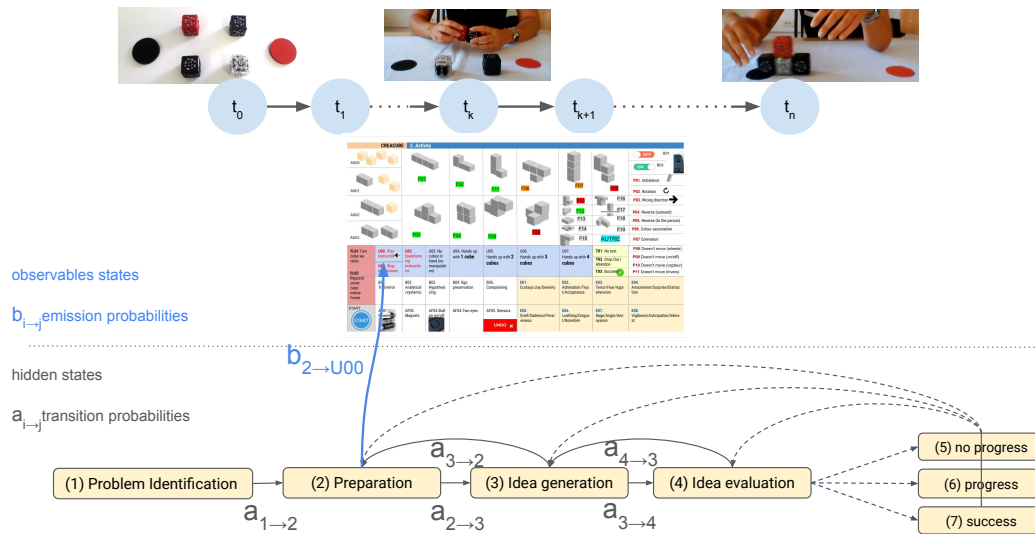


Figure 3.7: Perspective: Hidden Markov Models applied to the CreaCube task. (We have not represented all possible probabilities of transition and emission for clarity purpose.)

from which we observe a certain output (i.e. the collected JSON data) with a certain probability of emission. This would allow us to determine the probabilities of transitions between (hidden) states, and guide us on the metric we should use to generate our trajectory (e.g. edit distance as suggested in the previous paragraph). An illustrative example is given on Figure 3.7. Clustering algorithms on hidden states could then be considered, as a way to “structure” the problem space and find regularities that would help define an appropriate distance between states. These hidden states might correspond to CPS componential stages, or other kinds of stages that are specific to the task, and might constitute different contexts as the ones used by *task sets*.

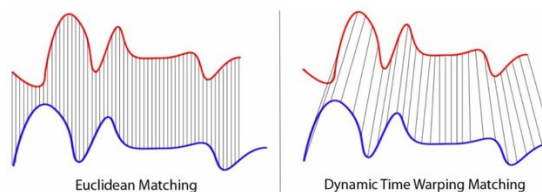


Figure 3.8: An illustration of dynamic-time wrapping, reproduced from [Martin et al. \(2022\)](#)

We could also compare trajectories themselves using sequence clustering, in order to distinguish several “profiles” of CPS behaviors among learners. To achieve that, we suggest using Dynamic-Time Wrapping (DTW), which has the benefit of comparing the shapes of trajectory regardless of translation or minor dilatation in time (see Figure

3.4 Discussion: perspectives of application to the CreaCube task

3.8). In [Palaude et al. \(2023\)](#), we have proposed to use such trajectory clustering in relation to common assessment dimensions of **divergent thinking** (e.g., **fluidity**, **flexibility**, **originality**).

These two proposals do not take into account timing, which could be of interest in the study of the CPS process. Regarding this, [Köhler et al. \(2021\)](#) have considered using Hawkes processes to model the timed changes of behavior. Drift-diffusion models, which are commonly used to model reaction times in neuroscience and psychology experiments ([Bogacz, 2007](#)), could also be taken into account.

4 REASONING ON SYMBOLS: TOWARDS A BIOLOGICALLY PLAUSIBLE MODEL?

4.1	Introduction	95
4.1.1	Representing neuronal activity at a symbolic level	96
4.1.2	From sensorimotor processing to logical reasoning	97
4.2	Basic design choices	98
4.2.1	From symbols to numbers	98
4.2.2	Semantic similarity	99
4.2.3	Classes and approximate Boolean properties	100
4.2.4	Statements: facts and rules	101
4.3	A first, simple but partial, mapping	102
4.3.1	Ontology numerical mapping	102
	A Semantic Pointer Triplestore	102
	Using the VTB algebra	103
4.3.2	Relationship and membership composition	105
	Hierarchical representations	105
	Relational representations	107
	Other intrinsic OWL properties	108
4.3.3	Introducing data and data structure	109
	Representing literals	109
	Data container	110
4.3.4	Effective neuronal implementation: an illustrative example for the case of hierarchical rules	111
	Nengo architecture	112
	Results	112
	Limits of the current approach	112
4.4	Discussion and conclusion	112

The following chapter is based on a conference article presented at the Workshop on Knowledge Representation for Hybrid and Compositional AI at KR2021 ([Mercier, 2023](#)), augmented with updates submitted to the Cognitive Computation journal ([Mercier and Vieville, 2023](#)).

4.1 INTRODUCTION

It is generally admitted that human problem solving involves tightly interleaved numerical but also symbolic (including logical) computations. In order to model some specific tasks such as CreaCube, which involves cognitive functions as diverse as sensorimotor skills and exploitation of (sometimes partial) explicit and tacit knowledge, we postulate there is a need for an integration of both numerical and symbolic paradigms which cannot be achieved with separate modules.

In hybrid systems, both symbolic and numerical components are involved (Lallement et al., 1995), but their contributions are usually kept separate in distinct aspects of the system. This is the case for example with ontology-based deep learning (Hohenecker and Lukasiewicz, 2020; Petrucci et al., 2016; Phan et al., 2017) or “black-box” cooperation between deep-networks and ontology reasoners (Ayadi et al., 2019; Jiménez et al., 2018). However, geometric mapping of ontologies onto Euclidean spaces or manifolds (Eidooon et al., 2008; Tous and Delgado, 2006; Xiao et al., 2015) allows to perform reasoning at both a symbolic and to a limited extent numerical level, while the relationship between neural network representation and manifolds is well established in both machine learning (Chui and Mhaskar, 2018; Zhu et al., 2017) and computational neuroscience (Duncker and Sahani, 2021; Langdon et al., 2023).

Since McCulloch and Pitts (1943), the idea that neural activity is computational and that cognition emerges from neural computations has led some authors to argue that a purely numerical approach could implement the needed unified system (see, e.g., Pulvermüller (2013) for a general review), as the brain seems to manipulate symbolic information with neuronal units generally considered as numerical systems. Yet it is a wide and still open issue to establish the numerical primitives required for modeling symbolic representation and manipulation (see Alexandre (2019) for a general discussion). As a result, in the last few years, models using numerical mechanisms to process symbolic information such as logical computation (Shi et al., 2020) or reasoning (Riegel et al., 2020) have flourished.

Given a symbolic graph representation, there are numerous vectorization approaches, such as translational distance models or random walk based methods (see Wang et al. (2021) for a review regarding approximate statistical reasoning, or Sajjad et al. (2019) regarding representation learning from data) including taking into account dynamic knowledge such as Sauerwald and Zanetti (2019) or combining vector representations and inference rules such as Guo et al. (2016). Other approaches propose to integrate or interface real-valued first-order logic (e.g., probabilistic) computations with machine learning algorithms (see Garcez and Lamb (2020) for a general review and, e.g., Cohen et al. (2017) for an example). Our approach is complementary in the sense that we focus on both a biologically plausible neuronal framework, and a symbolic formalism based on triples (as defined in 4.2.4) that seems appropriate to model human reasoning in psychology (McClelland and Rogers, 2003).

Specific questions may arise when trying to integrate for example relations between ontology specifications and partial knowledge (Tettamanzi et al., 2017), allowing one to

4 Reasoning on symbols: towards a biologically plausible model?

consider not only true or false knowledge, but relative degree of truth. In most machine learning algorithms, this is usually represented by probabilities. However, the human “level of truth” seems to be different and related to other notions such as possibility and necessity, related to a given modality, that is a context, a given time, and so on, which is also considered as representative to what is modeled in educational science and philosophy (see [Smith \(1994\)](#) while [Rusawuk \(2018\)](#) proposes a discussion).

4.1.1 REPRESENTING NEURONAL ACTIVITY AT A SYMBOLIC LEVEL

As a possible entry to a unified approach, *Vector Symbolic Architectures* (VSA) were introduced as a way to manipulate symbolic information represented as numeric vectors (see e.g. [S. D. Levy and Gayler \(2008\)](#) for an introduction). VSAs have been proven helpful to model high-level cognition and account for multiple biological features ([Eliasmith, 2013](#); [Gayler, 2003](#)), providing vectors that can be manipulated within networks of spiking neurons, as demonstrated by [Eliasmith \(2013\)](#). This approach makes a significant step towards the unification of symbolic and sub-symbolic processing in that it provides a way to translate the former into the latter. Consequently, complex knowledge representation in the form of compositional structures that are traditionally restricted to symbolic approaches can now be distilled in numerical and even neural systems ([Crawford et al., 2016](#)).

How can we represent a symbol in a neuronal assembly? A localist representation (one neuron or neuron group represented by a symbol) does not correspond to what is observed in the brain, and the basic idea is that a symbol corresponds to a pattern of activity distributed over the whole assembly. Let us consider a spiking neuron network and quantify its activity using, e.g., the neuron rates or higher-order statistics (see, e.g., [Cessac and Viéville \(2008\)](#) for a discussion). As developed by [Eliasmith and C. H. Anderson \(2002\)](#), this includes timing codes and population codes (i.e., relative timing codes between neurons). This high-dimensional set of bounded quantitative values can be collected and normalized, as a unitary stochastic vector in a high-dimensional space (with a few thousand dimensions for a biological neuronal map and often a few hundred dimensions at the simulation level), thus defining a *Semantic Pointer Architecture* (SPA). This builds upon a particular case of VSA and the Neural Engineering Framework (NEF) ([Eliasmith and C. H. Anderson, 2002](#)). The NEF provides a set of principles for implementing such an architecture through synaptic connections, including a time representation in spiking neuron systems (rather than, e.g., other representations based on synchrony within the neural assembly; (see [Eliasmith and C. H. Anderson \(2002\)](#) for technical details). This framework is a rather scalable alternative for a biologically plausible implementation of VSA, and it has already been implemented into a simulator called Nengo ([Bekolay et al., 2014](#)).

In the present study, we consider these developments as prerequisites and will simply consider that neural assembly activity is represented by a high-dimensional unary stochastic vector. We also need to specify transformations and define them at this abstract algebraic level. We will mainly consider the hetero-association mechanism, as developed in [T. C. Stewart et al. \(2011\)](#), and functional transformations, as detailed in [Eli-](#)

smith and C. H. Anderson (2002); their development is based, in a nutshell, on parameterized kernel-based approaches.

Manifolds and neural networks

In mathematics, a *manifold* is a topological space that locally resembles Euclidean space near each point.

The activity of a large number of neurons can often be captured by a much smaller number of key dimensions. Therefore, in the neuroscience literature, manifold usually refers to a subspace of neuronal activity, identified using a dimensionality reduction method (see Fig. 4.1).

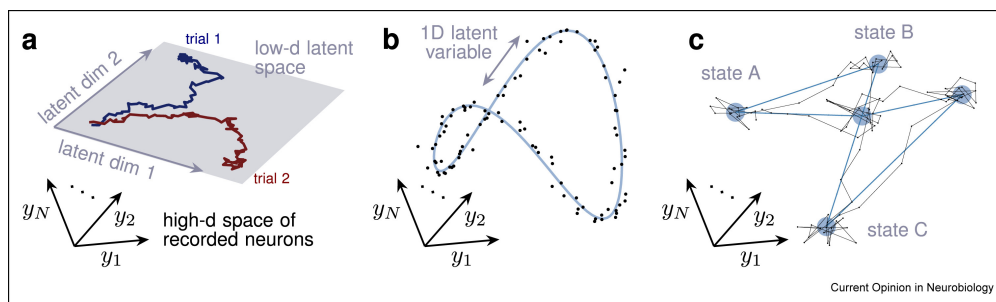


Figure 4.1: From neural activity to manifold: figure reproduced from [Duncker and Sahani \(2021\)](#). (a): The activity of the neuronal population evolves in a high dimensional coordinate space, where each axis corresponds to the firing rate of a recorded unit. (b) Dimensionality reduction techniques can be applied to obtain a low-dimensional latent description of the high-dimensional population activity. This is illustrated here by a one-dimensional manifold, embedded non-linearly in the high-dimensional space of recorded neurons. Similar topologies have been identified in the head-direction cell system. (c): Illustration of a manifold representing different state values, where each state engages different groups of neurons.

4.1.2 FROM SENSORIMOTOR PROCESSING TO LOGICAL REASONING

As review in Chapter 1, logical reasoning with symbols emerges through a complex interplay of cognitive processes, including enactive and developmental processes, involving sensori-motor associations which contribute to the formation of stable symbols or concepts (as reviewed, e.g., in [Ness and Farenga \(2007\)](#)).

In particular, creative problem solving requires sophisticated reasoning mechanisms, such as **deductive**, **inductive** and **abductive reasoning**. We need to make deductions to solve problems in everyday life, while the elements we briefly reviewed here demonstrate that such deductions are not boolean (either true or false) but related to a given context, weighted by a certain level of belief, and biased by motivational elements in a wide sense. Furthermore, as discussed in, e.g., [Keefer and Landau \(2016\)](#), especially for

4 Reasoning on symbols: towards a biologically plausible model?

goal-driven behavior, deductive reasoning is deeply interleaved with heuristic deduction, which involves analogy and metaphor. In this study, we show how such **deductive reasoning** could emerge from biologically plausible neuronal implementation, and discuss to what extent this could be extended to approximate deductive reasoning, in addition to **inductive** and **abductive** reasoning mechanisms.

Considering knowledge representation and reasoning, the capabilities of Semantic Web modeling languages, such as the **Resource Description Framework Schema (RDFS)** and the **Web Ontology Language (OWL)** (see, e.g. [Allemang et al. \(2020\)](#) for a recent didactic reference) is a rather accessible and powerful way to manipulate symbolic data. To what extent could such a mechanism be implemented in a biologically plausible architecture such as the **SPA**? In order to contribute to this issue, we propose here some design choices allowing us to implement a **RDFS** specification using **VSA**s. We illustrate this on a simple example and also discuss to which extent **OWL** specification could benefit from the same method.

4.2 BASIC DESIGN CHOICES

4.2.1 FROM SYMBOLS TO NUMBERS

In order to represent symbolic information, we use **RDFS** to structure knowledge representation. It is based on the **RDF** data model, which represent knowledge as **triples**, as made explicit in section [4.2.4](#).

Resource Description Framework Schema

In the *Resource Description Framework Schema (RDFS)*, the *universe of discourse* is made of *resources*, referenced by some universal resource identifier (IRI), i.e. a fixed lexical token. To structure this universe of discourse, we consider:

- (i) *individuals* that refer to real-world concrete or abstract objects, or
- (ii) *literals* to characterize individuals using data attributes, i.e., numerical values, character strings, or any structured information such as dates
- (iii) *concepts* and *roles* (namely *classes* and *properties*) that allow to structure the knowledge about individuals.

Before going further, we can point out that the present definition follows the **RDF/RDFS** framework, with the following variants:

- we conflate *name* with both IRI and blank node, since on the one hand blank node can be eliminated,¹ and on the other hand because we only process the information locally at this stage, thus avoiding considering all issues regarding distributed information between different sources;

¹Using a standard process related to skolemisation.

- we do not consider (i) semantic web specific literal (e.g., `rdf:XMLLiteral`), or (ii) utility and annotation or other human-targeted properties (e.g., `rdfs:seeAlso`) at this stage;
- we will introduce both containers, i.e., ordered or unordered sequences, and collections, i.e., chained lists, later in these specifications, but in a somehow different form, adapted to the numerical representation and obvious to map on RDF representations;
- we do not consider all XSD data-types, but will introduce a precise notion of numerical values and will detail how to represent structured data in our framework.

The RDFS framework allows to formalize entailment rules, used for deductive reasoning, that are detailed in Appendix D.

At the numerical level, each resource is implemented as a randomly drawn fixed unit d -dimensional vector, $\mathbf{x} \in \mathcal{R}^d$, and the key idea is to study to what extent symbolic reasoning on resources may correspond to algebraic operations implemented via numerical computations (that we explicitize in 4.3.1, based on the SPA framework introduced in Eliasmith (2013)). Typically, $d \simeq 100 \cdots 1000$ and we expect to manipulate $k \simeq 100 \cdots 10000$ resources. A similarity measure is now introduced in order to semantically compare two vectors.

4.2.2 SEMANTIC SIMILARITY

Classically, the cosine similarity (i.e., normalized dot product, denoted \cdot) is used to compute the semantic similarity between two unit vectors:

$$\mathbf{x} \cdot \mathbf{y} \stackrel{\text{def}}{=} \mathbf{x}^\top \mathbf{y}$$

where \mathbf{x}^\top denotes the transpose of \mathbf{x} .

The key property is that, provided that the space dimension d is large enough, two randomly chosen different vectors will be approximately orthogonal. More precisely,

$$\mathbf{x} \cdot \mathbf{y} \sim \mathcal{N}(0, O(1/d)),$$

i.e., follows a centered normal distribution (Schlegel et al., 2020), while by construction $\mathbf{x} \cdot \mathbf{x} = 1$.

Based on this, most VSA approaches consider that two vectors \mathbf{x} and \mathbf{y} are semantically equivalent when this similarity τ equals to 1, but with different ways to interpret the result:

- **Closed-world assumption:** Anything, that cannot be stated as true is false, thus $\tau \in \{0, 1\}$, often obtained by projection and rectification (to avoid negative values). This is the most common interpretation in VSAs.

4 Reasoning on symbols: towards a biologically plausible model?

- **Open-world assumption:** Anything might be true unless it can be proven false, prompting a need to characterize unknown statements; furthermore the notion of negation is either not defined (as in the **RDFS** model, yielding only monotonic reasoning²), or defined at a higher level (as in **OWL** and more generally in description logics, that we also consider as a perspective of this work).

Here we consider open-world reasoning and we enrich the notion of being either *true* or *false*, by a numeric representation of partial knowledge, as illustrated in Fig 4.2. The *true* value corresponds to 1 (fully possible and fully necessary), the *false* value to -1 (neither possible nor necessary, i.e., impossible) and the *unknown* value to 0, which corresponds to a fully possible but absolutely not necessary value. Inbetween these extremes, negative values correspond to partially possible but unnecessary values and positive values to fully possible and more or less necessary values.

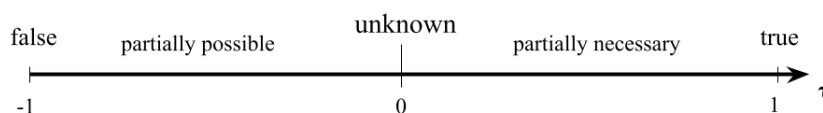


Figure 4.2: Representation of partial truth $\tau \in [-1, 1]$, in relation to necessity and possibility.

This representation has been designed to be compatible with the ternary Kleene logic, beside being also coherent with respect to the possibility theory³ (not developed here, please refer to (Denceux et al., 2020) for a general introduction). This deterministic representation of partial knowledge can be generalized in order to also include a probabilistic representation (using a 2D representation), as developed in Viéville and Mercier (2022).

4.2.3 CLASSES AND APPROXIMATE BOOLEAN PROPERTIES

The first kind of concept to structure the knowledge is the hierarchical notion of *class*, which is equivalent to the notion of Boolean property, defining the class of all individuals enjoying (or not) this property, and defining the property that an individual belongs (or not) to a given class.

In RDFS this translates into, given a scoring τ as introduced previously:

$$x \text{ rdf:type } c \triangleright \tau$$

²*Monotonic* reasoning means that former conclusions cannot be withdrawn considering new knowledge.

Deductive reasoning is typically *monotonic*, while *abductive reasoning* is non *monotonic*.

³To make the link explicit, given necessity ν and possibility π , with $\nu \leq \pi$ by construction, while $\nu > 0 \Rightarrow \pi = 1$ and $\pi < 1 \Rightarrow \nu = 0$, we have the one to one correspondence with our representation using $\tau \in [-1, 1]$:

$$\tau \stackrel{\text{def}}{=} \nu + \pi - 1 \text{ with } \begin{cases} \pi &= 1 + H(-\tau) \tau \\ \nu &= H(\tau) \tau, \end{cases}$$

We assume in the following that classes and individuals are in disjoint semantic sets. Although this is not necessarily the case in RDFS, this assumption is compatible with description logics (i.e. the OWL-DL level of specification) we target to use in future developments, and ensures to avoid self-reference paradoxes.

At the numeric level, they both correspond to unit vectors, with the following interpretation:

- For a given vector \mathbf{x} encoding an individual, the vector $\mathbf{c} = \mathbf{x}$ corresponds to the singleton class $\mathcal{C} = \{\mathbf{x}\}$.
- For a given class vector \mathbf{c} the individual vector $\mathbf{x} = \mathbf{c}$ can be interpreted as a “prototype” for this class.

At a geometric level, this can be interpreted as covariant/contravariant duality, and the similarity between a class covariant vector \mathbf{c} and an individual contravariant vector \mathbf{x} may be interpreted as the fact this individual approximately belongs, or not, to the class.

4.2.4 STATEMENTS: FACTS AND RULES

RDF Triples

In the *Resource Description Framework* (RDF), knowledge about the universe of discourse is structured into statements of the form subject predicate object, called *triples*, where subject and object are two resources linked by a relationship (property) explicitized by the predicate.

We introduce *weighted triples* of the form:

subject predicate object $\triangleright \tau$

with a value τ as discussed previously, which generalizes usual RDF statements, introducing a scoring value (refer to (Tettamanzi et al., 2017) for an introduction and a recent literature review). This is easily stated in the RDF language itself, using reification, while in our case, it is an intrinsic feature of our design. More precisely, when considering RDFS where only true (but not false) assertions can be stated, we have $\tau \in [0, 1]$, while when generalizing to the OWL language where negation can be stated we will use $\tau \in [-1, 1]$, this setting being compatible with both levels of specification.

It is worth noting that such statements are of three kinds, although this is not explicit in the model:

- entity data-property literal $\triangleright \tau$
defines *attributes* allowing to specify some entity attributes as literal values (e.g., numbers, dates, strings...);

4 Reasoning on symbols: towards a biologically plausible model?

- entity object-property entity $\triangleright \tau$
defines *relations* between entities;
- concept predicate resource $\triangleright \tau$
defines *meta-properties* about classes or properties.

4.3 A FIRST, SIMPLE BUT PARTIAL, MAPPING

4.3.1 ONTOLOGY NUMERICAL MAPPING

A SEMANTIC POINTER TRIPLESTORE

Let us suppose we need to represent the following weighted statements:

$$\begin{aligned} \text{subject}_1 \text{ predicate}_1 \text{ object}_1 &\triangleright \tau_1 \\ \text{subject}_2 \text{ predicate}_2 \text{ object}_2 &\triangleright \tau_2 \\ \text{subject}_2 \text{ predicate}_3 \text{ object}_3 &\triangleright \tau_3 \end{aligned}$$

(note that the same subject subject_2 is used in the two last statements).

The first step towards a Semantic Pointer representation is to encode each resource by a randomly sampled vector on the unit hypersphere of our d -dimensional vector space: let us denote $\mathbf{s}_i, \mathbf{p}_i, \mathbf{o}_i$ the respective vector representations of the resources $\text{subject}_i, \text{predicate}_i, \text{object}_i$.

Next, we may store each of these statements into an associative memory (T. C. Stewart et al., 2011), similar to a hash table, where each “key” would be a resource and the corresponding value would express the statements for which this resource is a subject:⁴

$$\mathbf{s}_1 \rightarrow \tau_1 \mathcal{B}(\mathbf{o}_1, \mathbf{p}_1)$$

$$\mathbf{s}_2 \rightarrow \tau_2 \mathcal{B}(\mathbf{o}_2, \mathbf{p}_2) + \tau_3 \mathcal{B}(\mathbf{o}_3, \mathbf{p}_3)$$

for which we need to introduce the following operations:

1. A scalar multiplication $\mathbf{a} = \tau \mathbf{b}$ that scales a vector \mathbf{b} by a factor of τ and preserves its direction.
2. A vector superposition (e.g. element-wise addition) $\mathbf{a} = \mathbf{b} + \mathbf{c}$ that results in a vector \mathbf{a} with $\mathbf{a} \cdot \mathbf{b} = 1$ and $\mathbf{a} \cdot \mathbf{c} = 1$, assuming \mathbf{b} and \mathbf{c} are orthonormal.
3. A binding operation $\mathbf{a} = \mathcal{B}(\mathbf{b}, \mathbf{c})$ that outputs a vector \mathbf{a} that is not collinear with either \mathbf{b} or \mathbf{c} .

VSA^s and, in particular, the SPA implement such operations (Eliasmith, 2013). Several choices are available for the binding operation. The most classical one is the circular convolution $\mathbf{a} = \mathbf{b} \otimes \mathbf{c}$, which is used in Holographic Reduced Representations (HRR) (Plate, 1995), one of the first examples of VSA. This operation is commutative, associative

⁴In RDF/RDFS, this kind of database is known as a triplestore, as it stores statements in the form of triples.

and distributive. In order to retrieve an element vector \mathbf{b} from the resulting vector \mathbf{a} , the circular convolution can be used with the approximate convolutive inverse of the other element (\mathbf{c}^{-1}): $\mathbf{b} \approx \mathbf{a} \otimes \mathbf{c}^{-1}$.

However, commutativity and associativity could lead to serious misunderstandings. Let us illustrate this point with an example: considering the statements “Luigi eats this Pizza” and “this Pizza has a topping of Mozzarella”, encoded as:

$$\text{Luigi} \rightarrow \text{eats} \otimes \text{Pizza} \quad (4.1)$$

$$\text{Pizza} \rightarrow \text{hasTopping} \otimes \text{Mozzarella} \quad (4.2)$$

By injecting the Pizza subject from the statement (2) into the Pizza object in the statement (1), we can infer the new statement:

$$\text{Luigi} \rightarrow \text{eats} \otimes (\text{Pizza} \otimes (\text{hasTopping} \otimes \text{Mozzarella})) \quad (4.3)$$

Under this form, (4.3) may be interpreted as “Luigi eats this Pizza which has a topping of Mozzarella”, but the right member can be rewritten as:

$$\text{Luigi} \rightarrow \text{eats} \otimes \text{Pizza} \otimes \text{hasTopping} \otimes \text{Mozzarella} \quad (\text{associativity}) \quad (4.4)$$

$$= \text{hasTopping} \otimes \text{Mozzarella} \otimes \text{eats} \otimes \text{Pizza} \quad (\text{commutativity}) \quad (4.5)$$

which would be interpreted as “Luigi has a topping of Mozzarella which eats a Pizza”. Note that circular references are easily managed with this setup, thanks to the binding mechanism generating always an almost orthogonal combined vector.

To overcome this issue, a possible “work-around” consists in adding new vectors **predicate** and **object** accounting for the functions in the triple, as well as an index t_i (represented by a vector \mathbf{t}_i) to each statement and store the triples as:

$$\mathbf{s}_1 \rightarrow \tau_1 \mathbf{t}_1 \otimes (\mathbf{p}_1 \otimes \text{predicate} + \mathbf{o}_1 \otimes \text{object})$$

$$\mathbf{s}_1 \rightarrow \tau_2 \mathbf{t}_2 \otimes (\mathbf{p}_2 \otimes \text{predicate} + \mathbf{o}_2 \otimes \text{object})$$

$$+ \tau_3 \mathbf{t}_3 \otimes (\mathbf{p}_3 \otimes \text{predicate} + \mathbf{o}_3 \otimes \text{object})$$

Such a representation similar to what has been done by [Crawford et al. \(2016\)](#) to represent the Wordnet database. However, this formalism is quite heavy and we would miss some interesting properties induced by binding directly the predicate to the object.

Instead, we will use another binding operation offered by the **Vector-derived Transformation Binding (VTB)** algebra, described by [Gosmann and Eliasmith \(2019\)](#) and also implemented in Nengo.

USING THE VTB ALGEBRA

The **Vector-derived Transformation Binding (VTB)** algebra ([Gosmann and Eliasmith, 2019](#)) is neither commutative nor associative, but distributive and bilinear: this is of interest

4 Reasoning on symbols: towards a biologically plausible model?

for our specific use case, as explained previously. We need to define the binding and unbinding operations in such an algebra.

THE BINDING OPERATION With d the dimensionality of the vector space, which the VTB requires to be square, and $d'^2 = d$, the binding operation is defined, following (Gosmann and Eliasmith, 2019):

$$\mathcal{B}(\mathbf{x}, \mathbf{y}) \stackrel{\text{def}}{=} \mathbf{B}_{\mathbf{y}} \mathbf{x}$$

where $\mathbf{B}_{\mathbf{y}}$ is block-diagonal matrix defined as

$$\mathbf{B}_{\mathbf{y}} \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{B}'_{\mathbf{y}} & 0 & \dots & 0 \\ 0 & \mathbf{B}'_{\mathbf{y}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{B}'_{\mathbf{y}} \end{bmatrix}$$

where

$$\mathbf{B}'_{\mathbf{y}} = d^{\frac{1}{4}} \begin{bmatrix} y_1 & y_2 & \dots & y_{d'} \\ y_{d'+1} & y_{d'+2} & \dots & y_{2d'} \\ \vdots & \vdots & \ddots & \vdots \\ y_{d-d'+1} & y_{d-d'+2} & \dots & y_d \end{bmatrix}$$

With these notations, the first relation above becomes

$$\mathbf{s}_1 \rightarrow \tau_1 \mathbf{B}_{\mathbf{p}_1} \mathbf{o}_1$$

This design choice enjoys the following interesting properties:

- The relation is not commutative in \mathbf{x} and \mathbf{y} , in the general case, as required.
- The relation is bilinear in \mathbf{x} and \mathbf{y} (this is obvious for the former, and easily verified on the matrix form for the latter).
- The relation is approximately left and right invertible as discussed now.

THE RIGHT UNBINDING OPERATION In order to retrieve an element vector from the bound vector, we need to bind it to the inverse of the other element vector. Unlike HRR and the circular convolution, the VTB algebra does not provide two-side inverses: there is no left inverse for VTB.

The right approximate inverse \mathbf{y}^{\sim} must enjoy the following property:

$$\forall \mathbf{x}, \mathcal{B}(\mathcal{B}(\mathbf{x}, \mathbf{y}), \mathbf{y}^{\sim}) = \mathbf{B}_{\mathbf{y}^{\sim}} \mathbf{B}_{\mathbf{y}} \mathbf{x} = \mathbf{x}.$$

The key point here is that since the coefficients of \mathbf{y} are chosen randomly and independently identically distributed, the matrix $\mathbf{B}_{\mathbf{y}}$ is almost orthogonal, thus:

$$\mathbf{B}_{\mathbf{y}}^{\top} \simeq \mathbf{B}_{\mathbf{y}^{\sim}},$$

4.3 A first, simple but partial, mapping

which corresponds to simply permuting the elements of \mathbf{y} .

Consequently, the right identity vector \mathbf{i}_B such that $\mathbf{B}_{\mathbf{i}_B} = \mathbf{I}$, writes explicitly:

$$[\mathbf{i}_B]_i = \begin{cases} d^{-\frac{1}{4}} & \text{if } i = (k-1)d' + k, 0 < k \leq d' \\ 0 & \text{otherwise.} \end{cases}$$

where $[\mathbf{i}_B]_i$ stands for the i -th coordinate of the vector. We get \mathbf{i}_B by “unfolding” the identity matrix $I_{d'}$ line by line and concatenating it d' times (and adjusting the resulting vector by a $d^{-\frac{1}{4}}$ factor).

This allows us to check if a property is a predicate for a given resource and, if so, retrieve the corresponding object (thus performing *right unbinding*):

$$\mathbf{x} \rightarrow \mathbf{z} = \mathcal{B}(\mathbf{y}, \mathbf{p}) = \mathbf{B}_p \mathbf{y}$$

if and only if $\mathbf{B}_p^\top \mathbf{z} = \mathbf{B}_p^\top \mathbf{B}_p \mathbf{y} = \mathbf{y}$ and \mathbf{y} belongs to the vocabulary. Specifically, “ \mathbf{y} belongs to the vocabulary” means that there exists a vector \mathbf{v} in the vocabulary such that $\mathbf{v} \cdot \mathbf{y} \simeq 1$. Otherwise the result is simply undefined, leading to the conclusion that the predicate is undefined for the given resource.

THE LEFT UNBINDING OPERATION A step beyond what is proposed by [Gosmann and Elia-smith \(2019\)](#), we may also want to check if two resources are linked to each other by a predicate and, if so, retrieve the corresponding property (referred to as *left unbinding*). In order to do that, we first need to “flip” the order of the terms involved in the binding. This is achievable by performing the following operation:

$$\mathcal{B}(\mathbf{a}, \mathbf{b}) = \mathbf{B}_{\leftrightarrow} \mathcal{B}(\mathbf{b}, \mathbf{a})$$

considering the matrix $\mathbf{B}_{\leftrightarrow}$ defined as:

$$[\mathbf{B}_{\leftrightarrow}]_{ij} \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } j = 1 + \left\lfloor \frac{i-1}{d'} \right\rfloor + d'[(i-1) \bmod d'] \\ 0 & \text{otherwise.} \end{cases}$$

Thanks to this, we obtain:

$$\mathbf{x} \rightarrow \mathbf{z} = \mathcal{B}(\mathbf{y}, \mathbf{p}) = \mathbf{B}_p \mathbf{y}$$

if and only if $\mathbf{B}_y^\top \mathbf{B}_{\leftrightarrow} \mathbf{z} = \mathbf{B}_y^\top \mathbf{B}_y \mathbf{p} = \mathbf{p}$ and \mathbf{p} belongs to the vocabulary.

4.3.2 RELATIONSHIP AND MEMBERSHIP COMPOSITION

HIERARCHICAL REPRESENTATIONS

The most common deductive rule is likely the inheritance rule, a particular syllogism that is well understood as soon as formal reasoning emerges in children ([Smith, 1994](#)).

RDF/RDFS provides two preset properties to account for hierarchical representations between individuals and classes:

4 Reasoning on symbols: towards a biologically plausible model?

- $x \text{ rdf:type } c$, expresses that the individual x belongs to the class c .
- $c \text{ rdfs:subClassOf } c'$ expresses that the class c is a subclass of the class c' , or in other words, that the concept c is included in the class c' . This means that all individuals that belong to c automatically belong to c' .

Consequently, some entailment rules⁵ can be used to disclose implicit memberships:

- inheritance (referred to as *rdfs9* by the W3C):
 $x \text{ rdf:type } c \wedge c \text{ rdfs:subClassOf } c' \implies x \text{ rdf:type } c'$
- transitivity of *rdfs:subClassOf* (*rdfs11*):
 $c \text{ rdfs:subClassOf } c' \wedge c' \text{ rdfs:subClassOf } c'' \implies c \text{ rdfs:subClassOf } c''$

The Semantic Pointer representation that we proposed earlier, using the **VTB** algebra, allows us to implement such rules. For example, the inheritance rule can be expressed as follows: $x \rightarrow \mathbf{B}_{\text{type}} c$ and $c \rightarrow \mathbf{B}_{\text{subClassOf}} c'$ implies

$$x \rightarrow \mathbf{B}_{\text{type}} c + \mathbf{B}_{\text{type}} \mathbf{B}_{\text{subClassOf}} c' + \mathbf{B}_{\text{type}} c'$$

A sufficient condition for this implication to be true would be $\mathbf{B}_{\text{type}} \mathbf{B}_{\text{subClassOf}} = \mathbf{B}_{\text{type}}$, which would mean **subClassOf** = **i_B**; that may not be desirable as it would lead to many wrongly inherited memberships.

Instead, we prefer using a second associative memory to encode entailment rules such as $\mathbf{B}_{\text{type}} \mathbf{B}_{\text{subClassOf}} \rightarrow \mathbf{B}_{\text{type}}$, as schematized in Fig.4.5. However, we cannot store matrices in the associative memory, only vectors. But we can observe that both of these matrices are block-diagonal, and of the form

$$\mathbf{B}_z = \begin{bmatrix} \mathbf{B}_z' & 0 & 0 \\ 0 & \mathbf{B}_z' & 0 \\ 0 & 0 & \mathbf{B}_z' \end{bmatrix},$$

thus correspond to a given vector \mathbf{z} which can be stored instead into the associative memory.

In this particular case, we would therefore store the association:

$$\text{type} \odot \text{subClassOf} \rightarrow \text{type}$$

where we introduce a *vector composition* operator \odot defined as: $z = a \odot b$ such that:

$$\mathbf{B}_z' = d^{\frac{1}{4}} \mathbf{B}_a' \mathbf{B}_b',$$

yielding $\mathbf{B}_z = \mathbf{B}_a \mathbf{B}_b$.

Similarly, we can express the transitivity of $c \text{ rdfs:subClassOf } c'$:

$$c \rightarrow \mathbf{B}_{\text{subClassOf}} c'$$

⁵A recapitulative table of RDFS entailment rules as defined by the W3C can be found in Appendix D.

and

$$c' \rightarrow \mathbf{B}_{\text{subClassOf}} c''$$

implies

$$c \rightarrow \mathbf{B}_{\text{subClassOf}} c' + \mathbf{B}_{\text{subClassOf}} \mathbf{B}_{\text{subClassOf}} c'' + \mathbf{B}_{\text{subClassOf}} c''$$

As a consequence, the inference rule will be stored as:

$$\text{subClassOf} \circ \text{subClassOf} \rightarrow \text{subClassOf}$$

Then, by iterating on such inference rule, we easily obtain the transitive closure of hierarchical properties inferred from `rdf:type` and `rdfs:subClassOf` statements. At the distributed implementation level, this means using a loop between the second associative memory encoding inference rules and the first associative memory encoding the input statements, as schematized in Fig.4.5, on the last page. This general principle allows the system to compute the closure of the reasoning rules. i.e., fixed point iteration, adding any new statement derived from the application of the rules, until a fixed point is reached.

Let us now discuss, to which extents we can generalize the present formalism, to other kinds of relationships.

RELATIONAL REPRESENTATIONS

The previous mechanism allows us to express any entailment of the form:

$$(x \text{ p } y) \wedge (y \text{ p}' \text{ z}) \implies (x \text{ p}'' \text{ z})$$

by storing the rule

$$\mathbf{p} \circ \mathbf{p}' \rightarrow \mathbf{p}''$$

thus implemented by an operator product, for some predicate \mathbf{p} , \mathbf{p}' , \mathbf{p}'' . In other words, we may express any entailment based on **predicate chaining**.

A step further, we need to implement property restrictions, i.e. at the RDFS level, property range `rdfs:range` and domain `rdfs:domain`, e.g., inference of the form:

$$(x \text{ p } y) \wedge (p \text{ rdfs:domain } c) \implies (x \text{ rdf:type } c)$$

and

$$(x \text{ p } y) \wedge (p \text{ rdfs:range } c) \implies (y \text{ rdf:type } c)$$

and subproperty hierarchy of the form:

$$(x \text{ p } y) \wedge (p \text{ rdfs:subPropertyOf } q) \implies (x \text{ q } y).$$

for the `rdfs2`, `rdfs3` and `rdfs7` entailment rules, respectively.

These rules belong to Description Logic Programming (DLP), roughly speaking at the intersection of OWL and Logic Programming with Horn clauses (Grosz et al., 2003), allowing to implement the complete RDFS language targeted here and beyond an interesting part of OWL axioms (Levesque, 1986).

4 Reasoning on symbols: towards a biologically plausible model?

In order to derive such generalization, we need non trivial algebra. Let us rewrite, for a given triple:

$$(x \text{ p } y) \text{ i.e., } x = \mathbf{B}_p y$$

using tensorial notations:

$$\begin{aligned} x &= {}^p_o \mathbf{B} p y, & p &\simeq {}^s_o \mathbf{B} x y, & y &\simeq {}^s_p \mathbf{B} x p, \\ x &= {}^o_p \mathbf{B} y p, & p &\simeq {}^o_s \mathbf{B} y x, & y &\simeq {}^p_o \mathbf{B} p x, \end{aligned}$$

these notations being well defined because, as stated in the previous section, since (i) the operator is bi-linear, (ii) the $\mathbf{B}_{\leftrightarrow}$ operator allows to swap property and subject, (iii) the operator is approximately invertible, while we also can define:

$$\begin{aligned} x &= {}^p_s \mathbf{B} p, & p &\simeq {}^s_p \mathbf{B} x, & y &\simeq {}^s_o \mathbf{B} x, \\ x &= {}^o_s \mathbf{B} y, & p &\simeq {}^o_p \mathbf{B} y, & y &\simeq {}^p_o \mathbf{B} p, \end{aligned}$$

introducing projection (e.g., $x = {}^p_s \mathbf{B} p$ simply states that the subject x has a property p for some unknown object value, and so on). Explicitly deriving these expressions is a perspective of the present preliminary work.

If we are able to explicitly compute the previous forms of the \mathbf{B} which exist, thanks to the stated algebraic properties, it would allow us to translate, at least the three previous inference rules at the numerical level. However, in this case, at the implement level, we do not set a single value to accumulate all possible derived results. Given this precision, since:

$$x = {}^p_s \mathbf{B} p, \quad p = {}^p_o \mathbf{B} \text{ domain } c$$

we may implement

$$x = {}^p_s \mathbf{B} {}^p_o \mathbf{B} \text{ domain } c$$

while $x = {}^p_o \mathbf{B} \text{ type } c$, for any x and any c , we derive:

$${}^p_s \mathbf{B} {}^p_o \mathbf{B} \text{ domain} = {}^p_o \mathbf{B} \text{ type}$$

and similarly:

$${}^p_o \mathbf{B} {}^p_s \mathbf{B} \text{ range} = {}^p_o \mathbf{B} \text{ type}$$

while for the last inference rule, we obtain:

$${}^p_o \mathbf{B} {}^p_s \mathbf{B} \text{ subPropertyOf} = {}^p_o \mathbf{B},$$

using the same kind of algebra.

Since our design choice leads to a multi-linear relation between the three statement elements, by construction we can also easily define reification, i.e., define the statement s itself, via a relation of the form:

$$s = {}^s_p \mathbf{B} x p y$$

which is to be explicitized, and then obtain the `rdf:subject`, `rdf:predicate`, or `rdf:object` from the calculated statement, by further algebraic combinations.

OTHER INTRINSIC OWL PROPERTIES

Beyond our objective of integrating RDFS axioms, it appears that the properties of the chosen algebraic VTB structure allows us to directly implement some of the OWL fea-

tures. We already mentioned in 4.2 the possibility of inferring any predicate chaining, allowing to implement OWL2 property chains (`owl:ObjectPropertyChain`, i.e., property defined by appending two other properties) which generalizes this mechanism. Further OWL features include:

- $\neg c$ the complement of a class c (`owl:complementOf`)
- $c_1 + c_2$ the union of two classes c_1 and c_2 , this definition may be extended to any number of classes (`owl:unionOf`)
- the inverse p^\sim of a property p (`owl:inverseOf`)
- if a property p is encoded as a vector \mathbf{p} such that the matrix \mathbf{B}_p is symmetric, we can directly conclude that this property is symmetric. Indeed, the transpose of a symmetric matrix is equal to itself; since the matrices are considered orthogonal, the transpose is also the approximate inverse, therefore the approximate inverse of a symmetric matrix is equal to itself (`owl:SymmetricProperty`)

Further investigating how other intrinsic OWL properties can be integrated is a perspective of the present work.

4.3.3 INTRODUCING DATA AND DATA STRUCTURE

REPRESENTING LITERALS

A step further, we not only define relationships between entities (i.e., “ObjectProperty”) but also relate an entity to a literal (i.e., “DataProperty”), i.e., a quantitative or qualitative value.

Regarding numerical values, it appears that we can easily define multidimensional numerical values of the form:

$$\mathbf{v} = \mathbf{e}_1^{k_1} + \mathbf{e}_2^{k_2} + \dots$$

where \mathbf{e}_i is a identifier representing the k_i -th component and $\mathbf{e}_i^{k_i}$ stands for the i -th iterate of the binding operator (Komer et al., 2019). This allows to define integer values, and even more.

If the binding is performed using a convolution operator, it is easily shown that this generalizes to complex numbers.⁶

In our case, we propose an alternative and consider only “physical” numerical values r , i.e., bounded in $[\min, \max]$ and with a finite precision ε , which is practice, the case for any physical value. We must thus write:

$$r \stackrel{\text{def}}{=} k \frac{\varepsilon}{\max - \min} + \min, k \in \{0 \dots \lfloor \frac{\max - \min}{\varepsilon} \rfloor\}$$

considering that two values v_1, v_2 with $|v_1 - v_2| < \varepsilon$ are indistinguishable. Such strong

⁶In a nutshell, because it can be related to a numerical exponentiation via a Fourier transform (Komer et al., 2019).

4 Reasoning on symbols: towards a biologically plausible model?

specification is particularly useful for numerical calculus (normalized estimations, spurious value detection, estimation precision threshold, ...) and in our case, it allows to consider only the related positive integer value k .

Regarding string literals or qualitative values (e.g., Boolean “true” or “false” value), at this stage we simply propose to define one identifier (i.e. a string literal and a random unit vector) for each value.

DATA CONTAINER

In RFD and in any knowledge specification, we need to define data containers (i.e., `rdfs:Container`), mainly either unordered (i.e., `rdf:Bag` or “set”) or ordered (i.e., `rdf:Seq` or “list”).

More precisely, RDF/RDFS provides several classes to represent collections:

- (i) `rdf:Container`, which includes the following subclasses:
 - (i.i) `rdf:Seq`, which is an ordered container,
 - (i.ii) `rdf:Bag`, which is an unordered container,
 - (i.iii) `rdf:Alt`, which contains a set of “alternatives”, the first element of which being the default, and the other elements constituting an unordered collection
- (ii) `rdf:List` completed by:
 - the properties `rdf:first` and `rdf:rest`
 - the instance `rdf:nil` corresponding to the empty list

In our representation, ordered sets (`rdf:List` or `rdf:Seq`) could all be represented as lists in the sense of RDF, in a recursive manner, by storing the following associations:

$$\begin{aligned} \mathbf{l} &\rightarrow \mathcal{B}(\text{rdf:List}, \text{rdf:type}) \\ &+ \mathcal{B}(\mathbf{l}_1, \text{rdf:first}) \\ &+ \mathcal{B}(\mathbf{l}', \text{rdf:rest}) \end{aligned}$$

$$\begin{aligned} \mathbf{l}' &\rightarrow \mathcal{B}(\text{rdf:List}, \text{rdf:type}) \\ &+ \mathcal{B}(\mathbf{l}_2, \text{rdf:first}) \\ &+ \mathcal{B}(\mathbf{l}'', \text{rdf:rest}) \end{aligned}$$

...

$$\begin{aligned} \mathbf{l}^{(n-1)} &\rightarrow \mathcal{B}(\text{rdf:List}, \text{rdf:type}) \\ &+ \mathcal{B}(\mathbf{l}_n, \text{rdf:first}) \\ &+ \mathcal{B}(\text{rdf:nil}, \text{rdf:rest}) \end{aligned}$$

Furthermore, unordered sets of value are obviously represented by simple addition, i.e.:

$$\text{bag} = \mathbf{x}_1 + \mathbf{x}_2 + \dots$$

allowing to easily implement membership property, element addition and deletion.

A step ahead, “Alternative” containers (i.e., `rdf:Alt` in the RDFS sense) are simply implemented by a combination of the two previous representations, where `rdf:first` points to an addition. Similarly, we could define vectors like indexed containers like in (Eliasmith, 2013).

4.3.4 EFFECTIVE NEURONAL IMPLEMENTATION: AN ILLUSTRATIVE EXAMPLE FOR THE CASE OF HIERARCHICAL RULES

In order to illustrate these developments and validate the fact the proposed formalism is effectively compatible with a biologically inspired approach, we have considered a very small example of ontology (schematized in Fig. 4.3) inspired by the so-called “pizza” tutorial ontology⁷ (Horridge, 2011). To that end, we used the Nengo simulator (Bekolay et al., 2014), which provides an effective implementation of the Neural Engineering Framework (NEF) and the Semantic Pointer Architecture (SPA). It enables building large-scale bio-inspired models of spiking neurons by connecting together reusable modular components (see Bekolay et al. (2014) for more details). Therefore, instead of directly manipulating the neurons, we can implement our architecture using only those modular neural networks and the available connecting operations. Among other features, it encompasses binding and unbinding operations defined within the VTB algebra, as well as associative memories which allow us to store and recall patterns organized into SPA vocabularies. Theoretical details underlying the implementation of such associative memories are described in T. C. Stewart et al. (2011).

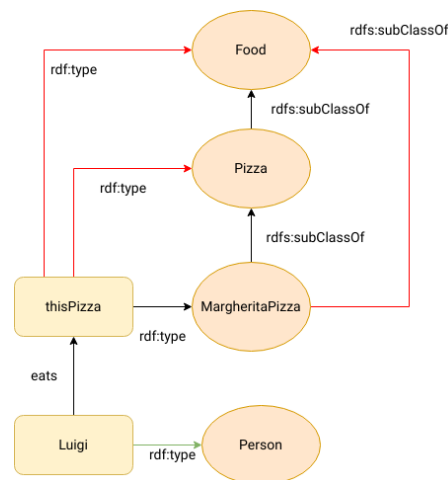


Figure 4.3: An example of a very simple ontology with two individuals, black arrows correspond to factual statements input in the data base and red arrows to inferred statements. Rectangular boxes stand for individuals, round boxes for classes and arrows are labeled by properties.

⁷<https://github.com/owlcs/pizza-ontology>

4 Reasoning on symbols: towards a biologically plausible model?

NENGO ARCHITECTURE

To illustrate the use of these features with an example, we considered two associative memories as schematized in Fig. 4.5.⁸ We created a Nengo vocabulary containing all the resources of our ontology encoded as vectors, and we stored asserted memberships and relationships between these resources into a first associative memory (AM1). In a second associative memory (AM2), we stored the class inheritance entailment rule corresponding to *rdfs9* in RDFS. Such an architecture could store more rules based on predicate chaining but, for the sake of simplicity, we restricted our example to this one only.

The idea is to query the knowledge base (stored into AM1) against the entailment rules (stored into AM2) according to some specific cues: our network can therefore be seen as a question answering system, where the question to answer is raised by these cues. For instance, in our example, the system receives a subject cue (*thisPizza*) and a predicate cue (*type*), which induce the question “what is the type of *thisPizza*”. A third cue indicates which rule to use: in this case, the class inheritance entailment (in practice, instead of a third cue, the network could keep a buffer of several rules to test successively).

RESULTS

In order to visualize the different steps of the data processing, we plotted the similarities of the module outputs at several points of the architecture, compared against the symbols in our vocabulary (Fig 4.4). From the initial set of statements, we are able to infer that *thisPizza* belongs to the class *Pizza*.

LIMITS OF THE CURRENT APPROACH

This preliminary study is to be completed at different levels. At a technical level, the tensorial generalization in section 4.3.2 has been stated at an abstract level, and has yet to be effectively implemented in order to target all RDFS mechanisms. The fact that we rely on Description Logic Programming (DLP) restrains the possibility to take into account the whole decidable part OWL2 specification and we aim to surpass this limit by further studying the intrinsic algebraic properties of our representation. We will also, as in human cognitive processes, consider “approximate” reasoning about undecidable facts, by better considering the semantic offered by the possibility/necessity score used here. Finally, we made explicit how literal representation is easily possible, but a lot of work is still to be done to have this part fully operational.

4.4 DISCUSSION AND CONCLUSION

What has been presented here is a proof of concept of mapping between RDFS (with some OWL extensions) ontology specification and a biologically plausible numerical

⁸The code is openly shared at <https://gitlab.inria.fr/line/aide-group/onto2spa>

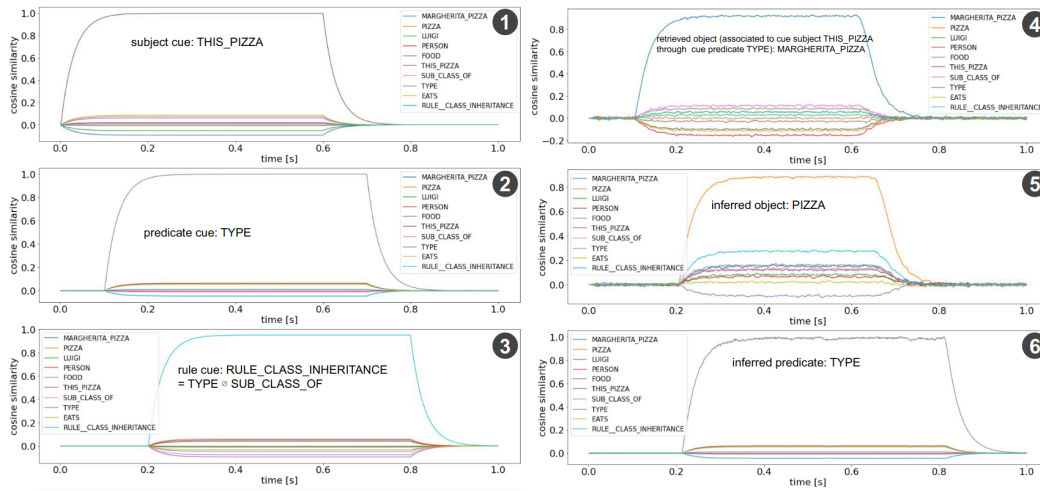


Figure 4.4: The numerical simulation results. For each signal, we plot its cosine similarity against the vocabulary. The cues guiding the reasoning (under the form of square inputs) are plotted on the left, and processing steps are on the right (refer to Fig. 4.5 for the location of each signal, marked by the circled numbers). We can observe the correctness of the inferred object and predicate (respectively PIZZA and TYPE, plotted on 5 and 6) bound together to characterize the subject THIS_PIZZA.

implementation based on VSAs, with a preliminary partial numerical experimentation. The key point is to demonstrate that symbolic knowledge, as possibly expressed in languages such as RDFS and OWL, can be also represented and manipulated with a numerical neuronal formalism of associative memory. It is worthwhile to note that both induction, as, for instance, in Domingos, 1996, and abduction could be formalized by inference rules (see, e.g., Lakkaraju and Y. Zhang, 2000 for a computational example and Shanahan, 2000 for a contribution regarding temporal reasoning). However, RDFS alone only allows *monotonic* reasoning, which is insufficient for abductive reasoning; hence the need for more generic OWL mechanisms for which non *monotonic* semantics have been proposed (Knorr et al., 2012). Our position is that reasoning in the brain is mainly related to the construction of mental models, as discussed in, e.g., Khemlani, Barbey, et al., 2014, considering common-sense reasoning. Such a framework better corresponds to human reasoning than pure logic reasoning, including modal logic reasoning Ragni and P.N. Johnson-Laird, 2018. In any case, a mechanism to construct, enrich, and modify such a mental model must be described. Regarding the genericity of this demonstration, it is worth mentioning that G. Mandler (2011) propose that human memory can represent three classes of structures (in short, associations, sequences and relational structures), the three of them being discussed here. By plausibility, we indeed do not claim that this is directly coded “as is” in the brain, but that what corresponds to symbolic processing in the brain may be represented by such processing, as discussed by Eliasmith (2013).

We are well-aware that, in our representation, the ontology and inference rules are “hard-coded” within the associative memories; the long-term perspective would be to

4 Reasoning on symbols: towards a biologically plausible model?

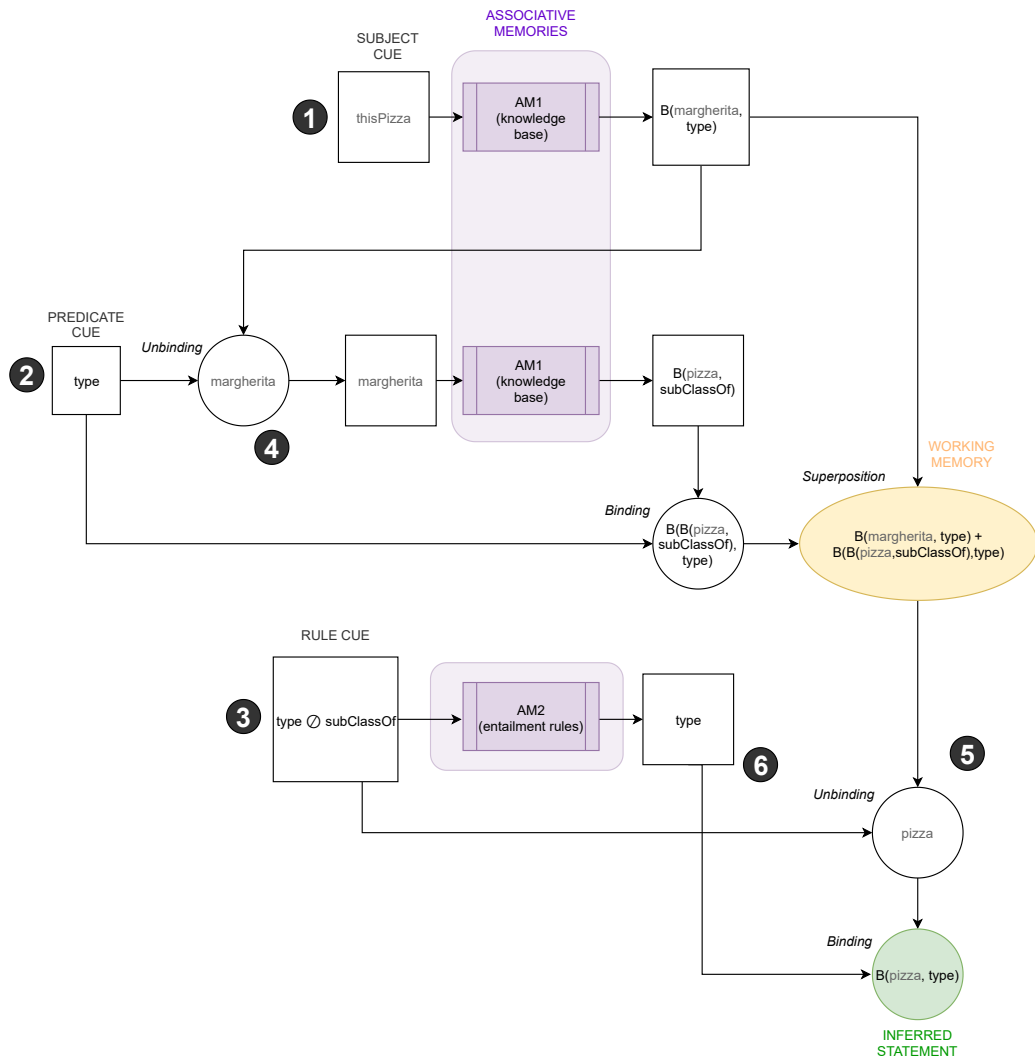


Figure 4.5: A proposition of architecture to implement our ontology and inference rules in a neuronal system, using the modules provided by the Nengo simulator. A subject cue (thisPizza) and a predicate cue (type) prompt the question “what is the type of thisPizza”; a third cue indicates the rule to use (class inheritance entailment). The network retrieves information in each associative memory and combines them to infer that thisPizza is not only a MargheritaPizza but also, through class inheritance, a Pizza. (Data is in square boxes and their processing in round boxes. Rectangular purple boxes account for associative retrievals. The circled numbers are markers locating which signals are plotted on Fig. 4.4.)

learn those associations. As new observations come along, we wish to store the newly observed relationships between individuals. A step further, logical inference as we proposed is a way to learn new associations, both memberships and relationships, within the ontology itself: for now, the newly inferred statements are only kept into the working memory, but we could store them for future use. Last but not least, we also wish to learn the inference rules themselves, by spotting some patterns that are repeated within the ontology. We presume that this vector formalism will facilitate learning, as it is the representation used by most learning algorithms, although a concrete implementation for this feature is still an open question. In order to remember inferred relationships, the naive approach would be to change the connections in the associative memory or even add new ensembles by hand. However, the neuroscience and machine learning fields provide more sophisticated and biologically plausible solutions to this problem. For example, [Voelker et al. \(2014\)](#) used a combination of unsupervised and supervised learning to learn new key-value associations in spiking neurons online.

Regarding concrete applications to model problem solving as a process, [T. Stewart and Eliasmith \(2011\)](#) have proposed an implementation for solving the Tower of Hanoi. The SPA is also interesting regarding creativity, as reviewed earlier, with [Kajić et al. \(2017\)](#)'s proposal which implements a version of the RAT test. This leads us to believe that such an architecture could be a good starting point to implement the multiple processes involved in creative problem solving at a biologically plausible level, although other neural dynamic approaches could also be considered (e.g., the cognon architecture proposed by [Rabinovich et al. \(2023\)](#), or recurrent neural networks such as conceptors ([Jaeger, 2017](#))).

5 DISCUSSION AND CONCLUSION

5.1 PERSPECTIVES AND LIMITS FOR LEARNING SCIENCES AND EDUCATION

The starting point for this thesis work was to discuss the potential contributions of neuro-cognitive and computational modeling in education, and more specifically in understanding and teaching transversal skills. In hindsight, what did we bring to the table?

USING LEARNING ANALYTICS TO FIT THE TRAJECTORY OF A SIMULATED LEARNER In the short term, we should first point out that a direct computational model is not a finality in itself; however, by simulating a parametrizable artificial agent, we could generate a CPS trajectory and fit it to an actual learner's trajectory. The question of *what parameters to adjust* is a key point for learning science, because it also gives a measurable account of individual differences in learning, which is useful for e.g. evaluating a skill with more nuance than merely "being able to solve a problem". Parameters could include both symbolic and numerical variables such as prior knowledge (behavioral, conceptual) and goal orientation (mastery, performance) associated with some weights. The parameters fitting best to the learner's trajectory may be validated or invalidated by pre- and post-test questionnaires. However, in order to do that, we would need much more data than we currently have (e.g. to perform Inverse Reinforcement Learning (Arora and Doshi, 2020)). To this end, we have also started two collaborations during my thesis.

First, with the help of Xlim in Poitiers, we are looking at algorithms to automatize the annotation of the CreaCube videos¹. The ontology of observables developed in Chapter 2 may be useful in this context to constraint deep-learning methods, as it has been done e.g. in Ruiz et al. (2023).

Second, thanks to Le Meudec et al. (2024)'s work, a virtual reality implementation of the CreaCube task is now available, with at least two potential applications for our team: (i) We can easily collect data on this task, for which annotations can be extracted from the program logs. Of course, the user experience is quite different, but the protocol of analysis should be very similar. (ii) We now have a physical model of the environment that a simulated agent can interact with, potentially allowing to develop artificial agents that behave in a more "embodied" way (well, still with virtual bodies).

TOWARDS A CPS FRAMEWORK APPLICABLE TO META-LEARNING In the longer term, we wish to extend such a framework of analysis to other creative problem-solving tasks, including collaborative problem solving.

¹A first prototype has been developed as part of Théo Carmes' internship, with the code available here: <https://github.com/TheoCarme/CreaCube>

Indeed, group processes are being studied in the CreaCube task from a learning sciences perspective (Cassone et al., 2019). Operationalizing this to a computational level would require a lot more work, but some elements regarding the computational and social aspects are already underlying our study: from learning and developmental aspects as developed by Vygotski et al. (1934), to the collective perspective in the activity theory as studied by Engestrom (2000), with influences on the componential model of creative problem solving developed by Engestrom (2000). We could imagine an extension of the present framework to Markovian multi-agents (see Table 3.1). Besides, the work of Le Meudec et al. (2024) is actually a preliminary step towards the study of collaborative processes in virtual reality, which is the subject of her PhD.

For obvious ethical reasons, the learner is aware of being engaged in a task with learning analytic evaluation. To what extends this could perturb the outcomes has been addressed in (Romero, 2019), in fact only marginally. Besides, this could be turned into a significant advantage, since it is a way to engage the participants in understanding their own learning process, opening opportunities to meta-learning. As discussed previously, teaching such transversal skills may increase the students' confidence by helping them gain awareness about their own learning process, and getting them motivated to "learning to learn" (UNESCO, 1996).

This could help preventing amotivation, which as shown a high correlation with school failure (Sander et al., 2018). Furthermore, a better understanding of creative problem solving may help bridging the gap between disciplines that are sometimes seen as antonymous (e.g. mathematics vs humanities) and strongly gendered, despite appealing to a common set of skills. Highlighting the the common ground for these disciplines may help balancing the gender repartition in higher education fields of study. Xiao et al. (2015) has demonstrated it is equally essential for both men and women to provide them with more creation-related cognitive activities, which is conducive to internalizing scientific motivation for both groups and the improvement of scientific innovation level. In addition, it is more effective for girls to cultivate intrinsic motivation in science through practical engagement in creative activities. This ambition is associated with the STEAM approach: this acronym stands for "Sciences, Technology, Engineering, Art and Mathematics", thus relating artistic aspects to technical disciplines formerly referred to as STEM. Such skills should also be studied considering not only gender differences but also other factors such as socio-cultural background, age (Romero, 2019) or neurodiversity (Khalil and Moustafa, 2022).

CONSIDERING OTHER TRANSVERSAL SKILLS Beyond academic purposes, transversal skills are crucial to develop the well-being and qualities of a citizen able to navigate in today's society. We have already mentioned the importance of developing a critical mind and computational literacy at the era of Artificial Intelligence, in order to avoid passive consumption of technology. A step further, we also include in transversal skills what is commonly referred to as *soft skills*, that is to say, socio-emotional skills such as empathy, which could be enhanced through collaborative activities. Some countries such as Den-

mark already promote teaching empathy at school (Mygind and Bølling, 2022), hereby preventing school harassment and bullying.

Transversal skills are all the more important in a fast-paced changing world. A few decades ago, it was common to study for one specific job. With the development of cutting-edge generative technologies, some professions may soon become outdated (although possibly replaced with new ones), hence the need to be able to adapt quickly and be able to learn new skills within a short time frame. Besides, navigating a world where the frontier between deepfakes and reality is becoming blurry and the consequences of climate change will be more and more impactful, present and future students are likely to require stronger scientific and technical knowledge as well as critical thinking (OECD, 2019).

LIMITS OF NEURO-COMPUTATIONAL MODELS FOR EDUCATION Despite the aforementioned benefits of incorporating computational and neuro-cognitive frameworks to analyse educational practices, we need to be cautious of drawing prescriptive conclusions.

First, although computational models are useful to better understand a phenomenon, it can be counter-productive to use such models to, e.g. automatize learning paths at the expense of self-led learning. In particular, adaptive learning (AI-led or not) consists in adapting the difficulty and coverage of exercises based on the students' skills to make them progress best (e.g. as in Clement et al. (2014) by taking inspiration from the notion the Zone of Proximal Development defined by Vygotsky and Luria (1980) in the context of children development). Of course, this is not a bad practice in itself, and it is basically what teachers do. However, in order to develop meta-learning, it is a good idea to make this adaptive process flexible and transparent for the learner, e.g. by using Open Learner Models (Guerra et al., 2016).

Another point of caution is minding the gap between neuro-cognitive theories, empirical practices on the field and the students' and teachers' needs. On the one hand, the foundations of educational practices sometimes suffer from a lack of evidence-based studies (Carnine, 2000; L. Zhang et al., 2022). On the other hand, it seems unlikely that neuroscience alone can improve teaching and learning, at least not without the support of psychology to bridge the gap between highly different levels of analysis as described in our first chapter 1.1.2 and as discussed by e.g. Bowers (2016) and Horvath and Donoghue (2016).

Moreover, the way such theories are diffused are often prone to misinterpretation, and numerous examples of "neuromyths" have been integrated into the popular thought or even to the educational field, either due to an erroneous diffusion of a valid study, or a lack of methodology in the study itself. Regarding creativity, a famous example is the idea that the right brain hemisphere would be intuitive and emotional while the left brain hemisphere would be analytical and reasoning oriented, with creative people using their right brain more. As reviewed in Chapter 1, creativity implies both analytical and intuitive aspects, and despite the brain being indeed asymmetrical with a lateralization of cognitive and motor functions, this asymmetry does not follow such a simple schema. Besides, an fMRI study has shown that most people do not use a hemisphere

more than the other (Nielsen et al., 2013). The confusion may come from fact that, for right-handed people, the left brain hemisphere dealing with the right-sided motricity is called the “dominant hemisphere” because it specializes in some routine tasks. Some other erroneous beliefs more specifically related to education have been debunked but remain popular, such as the numerous theories of learning styles: these theories all propose different typologies to account for individual differences in learning (e.g. the VAK model stating that we are either visual, auditory or kinesiesthetic learners), but none of them has proven efficient to actually increase academic performance (Sander et al., 2018). Regardless, there are also some successful examples of how neuro-cognitive theories were used in instructional design to develop better teaching methods. For instance, the cognitive load theory when designing exercises and problems with an appropriate presentation of the information (Sweller et al., 2019) Another example is how spaced repetition learning, based on knowledge on memory retention, has been proven to foster memorization (Kelley and Watson, 2013); even better, Kang (2016) have shown that spaced practice combined with tests could also enhance other forms of learning, including problem solving, and generalization to new situations. We can thus hope for a similar success in the development of creative (sic) pedagogical activities allowing students to better understand CPS processes so that they can approach new complex problems without getting overwhelmed, through a critical and operational lens; and beyond CPS, enhancing other transversal skills for their own confidence, well-being, and adaptation to a fast-paced changing world.

5.2 PERSPECTIVES FOR COMPUTER SCIENCE: TOWARDS BETTER EXPLAINABLE COMPUTATIONAL COGNITIVE MODELS?

Paradoxically, is the development of computational cognitive models indirectly contributing to the fast-paced evolution of technology, the risks of which we previously recalled? It is hard to predict how the field of artificial intelligence is going to evolve in the next decade. Nevertheless, studying creative problem-solving from a computational point of view also interrogates the limits of natural intelligence: this is one of the skills targeted by Artificial General Intelligence (AGI) (J. E. Laird et al., 2017). Although most researchers agree that we are still far from it, recent advances especially in the field of generative machine learning has led models to display impressive capabilities, including creative abilities, despite having no access to sensorimotor interactions and thus to any kind of semantic grounding (Harnad, 2024).

Nevertheless, when manipulating knowledge, explicit symbolic machine learning mechanism (i.e. reasoning), is much more efficient to infer knowledge from known facts, and hybrid mechanisms seem mandatory for complex tasks where non-trivial *a priori* knowledge is to be introduced (see, e.g, De Penning et al. (2011) for a realistic application). It has also great advantages when both interpretability (by an expert) and explainability (for end users) is a key feature, as discussed in Chraïbi Kaadoud et al. (2022).

These features are essential if such learning mechanisms are used to model cognitive functions. In addition, formal reasoning is by no means less costly than billions of oper-

ations of deep networks computation, as studied by [Desislavov et al. \(2023\)](#), which limit their academic use for usual research teams, beyond raised ecologic and ethical issues, discussed by the authors. The use of such “black-box” tools at the modeling level is also questionable as analyzed in [Schaeffer et al. \(2022\)](#).

Coupling both approaches should bring the required computational power to solve complex problems or model complex brain functions, in an explainable way, and with parsimonious resources consumption, and neural-symbolic computing brings together robust learning in neural networks with reasoning and explainability via symbolic representations. This has been an active area of research for many years, as recently reviewed in details by [Garcez and Lamb \(2023\)](#), who distinguish between different coupling levels summarized as follow:

- Type 1: *standard numeric computation with symbolic input/output* (e.g., word2vec nicely introduced in [Rong \(2016\)](#)). In this class of method, symbolic input is embedded in a numerical space that is able to encounter the pertinent relation between the input symbols. The symbolic processing is devoted to, e.g., a deep neural network in which output units feed a softmax (sometimes hierarchical) mechanism attributing a kind of probability to each expected symbol.
- Type 2,3: *hybrid systems with symbolic and numeric modules interaction* (e.g., AlphaGo) corresponds to loosely-coupled neural network with a symbolic problem solver up to more sophisticated interactions between both systems.
- Type 4,5: *symbolic knowledge and rules are compiled in distributed calculation* (e.g. Vector Symbolic Architectures (VSA) ([Eliasmith, 2013](#))) either using localist mapping of symbols or more complex embedding of symbolic rules, in both cases symbolic knowledge is translated into the network architecture and parameters. Relevant examples stand from localist logical neural networks ([Riegel et al., 2020](#)) to logic tensor network ([Serafini and Garcez, 2016](#)).

An additional type 6 is evoked as a perspective on ongoing research². The present work precisely aims at considering type 4 or 5 coupling levels.

Up to our best knowledge, computational neuroscience models are mainly based on either Type 1 mechanisms such as encoders, which yields an epistemological issues at the modeling level, as evoked before. Models based on Types 2 and 3 are also used, although quite exceptionally (e.g., [Spens and Burgess \(2023\)](#) who studied to what extent the GPT model can account for hippocampus replay, by training neocortical generative models; although it is not clear what is learned by such a general-purpose model regarding the hippocampus specificity, while such an approach is far from being refutable³),

²This layer is described as able to solve high cognitive functions, like attention schema, with the hope of being able to perform efficient combinatory reasoning (i.e., find approximate solutions to NP-hard problems by hybrid, de facto polynomial systems). This last point corresponds to existing algorithms as for instance reviewed in [Paschos \(2007\)](#).

³We make here a distinction between the historic notion of falsifiability and refutability, see [Olszewski and Sandroni \(2011\)](#) and [Cartwright \(1983\)](#) for a wider epistemological discussion on modeling.

while the Type 5 SPA architecture (after [Eliasmith \(2013\)](#)) introduces an intermediate algebraic abstract view of spiking neural network architectures, in order to develop biologically plausible cognitive functionalities, and human knowledge representations ([Crawford et al., 2016](#)), including high-level symbolic representation allowing reasoning as covered in Chapter 4. Beyond epistemological issues, generative models of computational art also raise plagiarism-related ethical issues, as they are often trained on human creative content and are likely to reuse such material in their outputs, even when it is protected by copyright ([Murray, 2023](#)).

5.3 GENERAL CONCLUSION

The present work, bound to the exploratory research action *AIDE* (*Artificial Intelligence for Education*) is indeed still at an exploratory stage, with numerous perspectives and only preliminary results. Following this divergent research phase, a convergent phase will be needed to study behavioral data in light of our proposed framework. We believe, however, that this approach constitutes a promising step towards computational learning sciences for a better understanding of complex creative problem-solving processes.

APPENDICES

A SCHOOLS OF THOUGHT IN LEARNING SCIENCES

Learning Sciences are a multi-disciplinary field that has been influenced by various schools of thoughts in psychology, cognitive neuroscience and artificial intelligence. We propose in Figure A.1 a recapitulative timeline organized into three major themes – (i) learning by association, (ii) learning by manipulating internal representations, (iii) learning by interacting with the environment. We chose to focus mainly on computational paradigms that are addressed in this thesis work, which may result in partial coverage, but allows for a better understanding of our perspective.

A Schools of thought in Learning Sciences

COMPUTATIONAL LEARNING SCIENCES IN THE WESTERN WORLD: A PARTIAL TIMELINE OF SCHOOLS OF THOUGHT

4th
CENTURY BC

ARISTOTELIAN LAWS OF ASSOCIATION

In *De memoria et reminiscencia*, Aristotle posits that thoughts are systematically connected within a chain, through relationships he formulates as similarity, contrast, and contiguity: association is thus the natural ability to recall past experiences that are stored in the mind from various cues.

ARISTOTELIAN FORMAL LOGIC

Ancient Greek philosophers such as Plato and Aristotle tried to understand how humans learn and reason. Plato used dialectic as both a method of reasoning and a means of philosophical training, while Aristotle is credited with the first formal study of logic, understood as the study of arguments. An argument is a series of true or false statements which lead to a true or false conclusion, and invalid arguments are called syllogisms. Aristotelian principles dominated Western logic up to the 19th century.

17th-19th
CENTURY

THE ASSOCIATIONIST SCHOOL

Members of the Associationist School included John Locke (1704), David Hume (1788), and two centuries later, Ivan Pavlov (1906). The phrase 'association of ideas' was first used by John Locke in 1689, following Thomas Hobbes, to describe how ideas can be connected. Some of the Associationist's principles anticipated the notion of conditioning and its applications in behavioral psychology.

BAYESIAN PROBABILITY THEORY

In 1763, Thomas Bayes proved a special case of what is now called the Bayes' theorem, paving the way for Bayesian inference and the Bayesian Probability theory. From this theory derived the Markov Decision Processes framework which is based Reinforcement Learning. Bayesian inference is also related to subjective logic, and hereby to adductive reasoning.

ABDUCTIVE REASONING

In addition to rigorously articulating inductive and deductive reasoning, Peirce proposed and defined the notion of abductive reasoning. His type of reasoning is frequently employed in empirical research as of today to draw conclusions from empirical observations. Peirce was also a pioneer in statistics, and realized — long before computers were invented — that electrical switching circuits could perform logical operations.

1900s

PAVLOVIAN CONDITIONING

In the early 1900s, following his experiments on dogs, Ivan Pavlov develops the concept of 'conditioned reflex' (or in his own words the conditional reflex) and the idea of 'classical conditioning' as an automatic form of learning by association. Classical conditioning can be used to alter behavioral reactions by introducing preceding conditions. Later on, classical conditioning has been examined across many different organisms, including humans, and different contexts including learning settings. Its underlying principles have influenced preventative antecedent control strategies used in the classroom.

CONSTRUCTIVISM AND MODAL LOGIC

Constructivism has its roots in educational psychology and is associated with Jean Piaget's theory of cognitive development. Piaget focused on how humans form knowledge by relating their experiences and their ideas. His views tended to focus on development of humans as individuals, rather than influenced by other persons. Piaget's theory of constructivist learning has had a significant influence on educational theories and pedagogues. Despite being sometimes confused with constructivism, an educational theory created by Seymour Papert and influenced by Piaget's concepts, constructivism does not refer to a particular pedagogy. Piaget was also interested in representing how humans think about uncertainty by probability and necessity rather than probability, following the work of his contemporaries on modal logic led by Kripke.

THE COGNITIVE REVOLUTION: COGNITIVISM AND THE BIRTH OF ARTIFICIAL INTELLIGENCE

Cognitivism emerged in the 1950s as a response to behaviorism, which cognitivists said neglected to explain cognition. In particular, Neum Chomsky argued that language acquisition could not be explained merely by habit and had to be at least partially explained by the existence of internal mental states. He instead suggested that language comprehension be explained in terms of mental rules called grammars. Meanwhile, in 1956, George Miller summarized numerous studies which showed the limited capacity of human thinking (and in particular, working memory). He proposed that such constraints can be overcome by forming synthetic mental representations (chunks) that require mental procedures for encoding and decoding the information. Cognitive psychology is thus an information-processing perspective on psychology, presupposing a mental representation and manipulation of symbols, and a specific form of mental activity of the kind advanced by computationalism. At the same time, marked by the 1956 Dartmouth conference, pioneers such as John McCarthy, Marvin Minsky, Allen Newell, and Herbert Simon founded the field of symbolic Artificial Intelligence (AI).

COMPUTATIONALISM

Computationalism is a family of views that hold that the human mind is not simply analogous to a computer program, but it is literally an information processing (i.e., computational) system that is realized (i.e. physically implemented) by neural activity in the brain. McCulloch and Pitts (1943) were the first to suggest that neural activity is computational and that neural cognitions explain cognition. Computation here is commonly understood in terms of Turing machines which manipulate symbols according to a rule, in combination with the internal state of the machine. Computationalism was vigorously disputed in the 1990s due to work by Putnam, Searle, and others.

COGNITIVE LOAD THEORY

Cognitive load theory (Sweller, 1988) is a learning theory based on the amount of working memory resources available in the human brain. It has often been used in instructional design to improve the presentation of the information to learners, as a way to enhance understanding and retention.

COGNITIVE ARCHITECTURES

Following the work of symbolic AI pioneers, several research projects of computational cognitive architectures started in the late 1980s to be pursued until the 2010s, the most famous ones being:

- SOAR (Laird 83), oriented towards (although still far from) General AI, that is, agents that can manipulate and learn knowledge to realize the full range of cognitive capabilities found in humans.
- ACT-R (Anderson 90), intended to specify how the human brain itself is organized in a way that enables individual processing modules to produce cognition.
- CLARION (Sun and Peterson 98), aimed at an integrative framework to explain and simulate cognitive and psychological phenomena.

 More recent approaches include the Nengo/Spain architecture, a neuro-symbolic architecture allowing to manipulate symbols through artificial spiking neural networks.

NEURO-SYMBOLISM

Neuro-symbolic AI is a type of artificial intelligence that integrates neural and symbolic AI architectures to address the weaknesses of each, providing a robust AI capable of reasoning, learning, and cognitive modeling. As argued by Leslie Vallan, Gary Marcus and others, the effective construction of robust computational models capable of reasoning, learning, and cognitive modeling, demands the combination of symbolic reasoning and efficient machine learning.

1930s

BEHAVIORISM

As Pavlov's research gained popularity in the West, behaviorism emerged as a response to psychoanalysis and other forms of psychology directed towards the unconscious, which frequently struggled to produce hypotheses that could be verified through experimentation. Behaviorists like J. B. Watson argued that psychology should only investigate the relationship between perceptible stimuli and perceptible behavioral reactions. Discussions about consciousness and mental images were frowned upon. Through the 1950s, behaviorism dominated the psychological scene, particularly in North America.

1950s

FIRE TOGETHER, WIRE TOGETHER

Hebbian theory is a neuropsychological theory introduced by Donald Hebb in 1949, an attempt to explain synaptic plasticity (i.e., the adaptation of brain neurons during the learning process). The theory postulates that an increase in synaptic efficacy arises from a presynaptic cell's repeated and persistent stimulation of a postsynaptic cell.

CONNECTIONISM - FIRST WAVE

The first wave appeared in the 1950s with Warren McCulloch and Walter Pitts (1943) focusing on comprehending neural circuitry through a formal and mathematical approach, and Frank Rosenblatt who introduced the Perceptron in 1958. However, in 1969, a book written by Minsky and Papert about the limitations of the original perceptron idea greatly contributed to discouraging major American funding agencies from investing in connectionist research, resulting in a period of inactivity for the field.

1960s

ASSOCIATIONISM - THE COME BACK

Associationism regain popularity in the studying of problem solving and creativity, with the idea that cognitive processing in the mind involves following a chain of associations from one idea to the next (Mednick, 1964) and the development of the Remote Associate Test (Mednick, 1962). These approaches were precursors for the study of creativity and memory organization using random walks in semantic networks.

1980s

CONNECTIONISM - SECOND WAVE

The term 'connectionist model' was reintroduced in the early 1980s by Jerome Feldman and Dana Ballard. The second wave blossomed in the late 1980s, following the 1987 book about Parallel Distributed Processing by James L. McClelland, David C. Rumelhart et al. which introduced several improvements to the original perceptron, such as intermediate processes (hidden layers) between input and output units. They also proposed the use of a sigmoid activation function, building upon the work of John Hopfield, who investigated the mathematical characteristics of such functions.

1990s

CONNECTIONISM - THIRD WAVE

The current wave has been marked by advances in Deep Learning allowing for great progress and research in tasks such as natural language processing, image classification and speech recognition, resulting in a connectionist revolution in the AI field. Up until the 2010s, neural networks were typically trained as discriminative models. Starting from 2014, innovations such as variational autoencoders and generative adversarial networks allowed to develop the first deep neural networks capable of learning generative models, followed by the first transformer models in 2017. In 2021, the release of DALL-E, followed by Midjourney and Stable Diffusion marked the emergence of artificial intelligence art from natural language prompts, questioning the nature of creativity. In March 2023, GPT-4 was released, displaying impressive capacities of mimicking human intelligence.

2000s

2010s

2020s

WE LEARN BY ASSOCIATION

WE LEARN BY MANIPULATING INTERNAL REPRESENTATIONS

WE LEARN BY INTERACTING WITH THE ENVIRONMENT

REFERENCES

1. Anderson, J. R., Leonesio, M., Anderson, C. (2017). ACT-R: A theory of adaptive learning and cognition and its relation to educational research. *European Commission, Horizon 2020, Grant Agreement No. 101019763*.

2. Anderson, J. R., Leonesio, M., Anderson, C. (2017). ACT-R: A theory of adaptive learning and cognition and its relation to educational research. *European Commission, Horizon 2020, Grant Agreement No. 101019763*.

3. Anderson, J. R., Leonesio, M., Anderson, C. (2017). ACT-R: A theory of adaptive learning and cognition and its relation to educational research. *European Commission, Horizon 2020, Grant Agreement No. 101019763*.

4. Anderson, J. R., Leonesio, M., Anderson, C. (2017). ACT-R: A theory of adaptive learning and cognition and its relation to educational research. *European Commission, Horizon 2020, Grant Agreement No. 101019763*.

5. Anderson, J. R., Leonesio, M., Anderson, C. (2017). ACT-R: A theory of adaptive learning and cognition and its relation to educational research. *European Commission, Horizon 2020, Grant Agreement No. 101019763*.

6. Anderson, J. R., Leonesio, M., Anderson, C. (2017). ACT-R: A theory of adaptive learning and cognition and its relation to educational research. *European Commission, Horizon 2020, Grant Agreement No. 101019763*.

7. Anderson, J. R., Leonesio, M., Anderson, C. (2017). ACT-R: A theory of adaptive learning and cognition and its relation to educational research. *European Commission, Horizon 2020, Grant Agreement No. 101019763*.

8. Anderson, J. R., Leonesio, M., Anderson, C. (2017). ACT-R: A theory of adaptive learning and cognition and its relation to educational research. *European Commission, Horizon 2020, Grant Agreement No. 101019763*.

9. Anderson, J. R., Leonesio, M., Anderson, C. (2017). ACT-R: A theory of adaptive learning and cognition and its relation to educational research. *European Commission, Horizon 2020, Grant Agreement No. 101019763*.

10. Anderson, J. R., Leonesio, M., Anderson, C. (2017). ACT-R: A theory of adaptive learning and cognition and its relation to educational research. *European Commission, Horizon 2020, Grant Agreement No. 101019763*.

11. Anderson, J. R., Leonesio, M., Anderson, C. (2017). ACT-R: A theory of adaptive learning and cognition and its relation to educational research. *European Commission, Horizon 2020, Grant Agreement No. 101019763*.

12. Anderson, J. R., Leonesio, M., Anderson, C. (2017). ACT-R: A theory of adaptive learning and cognition and its relation to educational research. *European Commission, Horizon 2020, Grant Agreement No. 101019763*.

13. Anderson, J. R., Leonesio, M., Anderson, C. (2017). ACT-R: A theory of adaptive learning and cognition and its relation to educational research. *European Commission, Horizon 2020, Grant Agreement No. 101019763*.

14. Anderson, J. R., Leonesio, M., Anderson, C. (2017). ACT-R: A theory of adaptive learning and cognition and its relation to educational research. *European Commission, Horizon 2020, Grant Agreement No. 101019763*.

15. Anderson, J. R., Leonesio, M., Anderson, C. (2017). ACT-R: A theory of adaptive learning and cognition and its relation to educational research. *European Commission, Horizon 2020, Grant Agreement No. 101019763*.

16. Anderson, J. R., Leonesio, M., Anderson, C. (2017). ACT-R: A theory of adaptive learning and cognition and its relation to educational research. *European Commission, Horizon 2020, Grant Agreement No. 101019763*.

17. Anderson, J. R., Leonesio, M., Anderson, C. (2017). ACT-R: A theory of adaptive learning and cognition and its relation to educational research. *European Commission, Horizon 2020, Grant Agreement No. 101019763*.

18. Anderson, J. R., Leonesio, M., Anderson, C. (2017). ACT-R: A theory of adaptive learning and cognition and its relation to educational research. *European Commission, Horizon 2020, Grant Agreement No. 101019763*.

19. Anderson, J. R., Leonesio, M., Anderson, C. (2017). ACT-R: A theory of adaptive learning and cognition and its relation to educational research. *European Commission, Horizon 2020, Grant Agreement No. 101019763*.

20. Anderson, J. R., Leonesio, M., Anderson, C. (2017). ACT-R: A theory of adaptive learning and cognition and its relation to educational research. *European Commission, Horizon 2020, Grant Agreement No. 101019763*.

SOCIAL CONSTRUCTIVISM

Social constructivism stems from the work of Lev Vygotsky, who stressed the importance of socio-cultural learning and how learners internalize interactions with more knowledgeable peers to form mental constructs through the zone of proximal development (ZPD). Expanding upon Vygotsky's theory, social constructivist psychologists developed the concept of instructional scaffolding, whereby the environment offers supports for learning that are gradually withdrawn as they become internalized.

ACTIVITY THEORY - FIRST WAVE (RUSSIAN)

In the 1930s, Alexei Leontiev developed and popularized a line of phenomenological and social science research now referred to as the Activity Theory, building upon Vygotsky's cultural-historical framework. Leontiev proposes to view human processes under the prism of activity, that is, a system in which a subject acts on an object, towards a desired outcome, and through the mediation of instruments or tools. This approach became predominant in both theoretical and applied psychology in the former USSR, including education.

GESTALT THEORY

Gestaltism is a theory of perception that focuses on the processing of whole patterns rather than individual components. The Gestaltists empirically investigated the perception of movement and contour, often using perceptual illusions. German Gestalt psychologists such as Wertheimer were also particularly involved in the study of problem solving; they proposed to distinguish between productive and reproductive thinking, corresponding respectively to spontaneous insight versus deliberation based on previous experience and knowledge.

AFFORDANCE THEORY

The term affordance was coined by American psychologist James J. Gibson in 1966, and defined as follows in 1977: "The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill. [...] It implies the complementarity of the animal and the environment." In 1988, Donald Norman adapted the concept to the context of Human-Computer Interaction to refer to "action possibilities" that are readily perceivable by a subject. In Norman's definition, affordances depend not only on the physical capabilities of the subject, but also on their goals, beliefs, and past experiences.

ACTIVITY THEORY - SECOND WAVE (SCANDINAVIAN)

The Scandinavian activity theory aims to integrate and develop concepts from Vygotsky's and Leontiev's works with Western intellectual developments such as Cognitive Science and Human-Computer Interaction. Engeström proposed to reformulate and extend the Activity Theory by introducing 'expansive learning' in the late 1990s. He added a new layer to the original activity scheme in order to better specify the case of collective activity, hereby introducing the notions of community, rules, and division of labour.

REINFORCEMENT LEARNING

Reinforcement learning (RL) emerged in the late 1980s and early 1990s, with TD-Learning proposed by Richard Sutton in 1988 and Q-Learning developed by Chris Watkins in 1992. It became an interdisciplinary area of machine learning with applications in cognitive neuroscience, concerned with how an intelligent agent takes actions in a dynamic environment to maximize a cumulative reward while considering the exploration vs. exploitation trade-off.

POST-COGNITIVISM

In the 1990s and throughout the 2000s, various new theories, often influenced by phenomenological and postmodern philosophy, challenged cognitivism and the idea that thought was best described as computation. These include (among others):

- **enactivism**: cognition arises through a dynamic interaction between an acting organism and its environment.
- **situated cognition**: knowledge is bound to activities within specific social, cultural, and physical contexts.
- **embodied cognition**: sensorimotor capacities shape cognitive functions.

Figure A.1: Schools of thought in Learning Sciences.

B THE CREATIVE PROBLEM SOLVING LITERATURE: STAGE MODELS AND DUAL NOTIONS

The CPS literature produced a variety of models describing the CPS process. However, despite their differences in terminology, they are rather similar in how they decompose the process in several stages, as shown by [Howard et al. \(2008\)](#) in a recapitulative table reproduced below.

Models	Analysis phase				Generation phase			Evaluation phase	Communication / implementation phase		
Helmholtz (1826)	Saturation				Incubation	Illumination		X	X		
Dewey (1910)	A felt difficulty	Definition and location of difficulty			Develop some possible solutions			Implications of solutions through reasoning	Experience collaboration of conjectural solution		
Wallas (1926)	Preparation				Incubation	Illumination		Verification	X		
Kris (1952)	X				Inspiration			Elaboration	Communication		
Polya (1957)	Understanding the problem	Devising a plan			Carrying out the plan			Looking Back	X		
Guilford (1957)	X				Divergence			Convergence	X		
Buhl (1960)	Recognition	Definition	Preparation	Analysis	Synthesis			Evaluation	Presentation		
Osborn (1963)	Fact-finding				Idea-finding			Solution-finding	X		
Parnes (1967)	Problem, challenge, opportunity	Fact-finding	Problem-finding		Idea-finding			Solution-finding	Acceptance-finding	Action	
Jones (1970)	Divergent Search for data Understand the problem				Transformation Pattern finding Flashes of insight			Convergent Judgement	X		
Stein (1974)	X Fact-finding				Hypothesis formulation			Hypothesis testing	Communication of results		
Parnes (1981)	Mess finding			Problem-finding	Idea-finding			Solution-finding	Acceptance-finding		
Amabile (1983)	Problem or task presentation	Preparation			Response generation			Response validation	Outcome		
Barron and Harrington (1981)	X				Conception	Gestation	Parturition	X	Bring up the baby		
Isaksen et al. (1994)	Constructing opportunities	Exploring data	Framing problem		Generating ideas			Developing solutions	Building acceptance	Appraising tasks	Designing process
Couger et al. (1993)	Opportunity, delineation, problem definition		Compiling information		Generating ideas			Evaluating, prioritising ideas	Developing an implementation plan		
Shneiderman (2000)	Collect				Relate			Create			Donate (communicate)
Basadur et al. (2000)	Problem finding	Fact finding	Problem defn.		Idea finding Diverge – converge at each stage			Evaluate and select	Plan	Acceptance	Action
Kryssanov et al. (2001)	Functional requirements		Structural requirements		Functional solutions	Analogies, metaphors		Reinterpretation	X		

Figure B.1: A comparison of creative process models proposed by, and reproduced from, [Howard et al. \(2008\)](#).

These models have in common several dual notions, beyond the well-accepted dichotomy between divergent and convergent phases. Authors have made a distinction between implicit versus explicit (or spontaneous versus deliberate) modes in interaction, to be put in relation to stimulus-based versus goal-directed creative behavior, which is likely to be influenced by the dual processing model of the fast-thinking versus slow-

thinking view of [Kahneman \(2011\)](#). The general idea is to differentiate explicit CPS, which can be reported and consciously organized, experimented with, and analyzed, from implicit CPS, with possibly the same level of efficiency but carried out automatically from the expression of the problem with no access to its constituents and its motivations. Cognitive science researchers have agreed on such distinctions, but, when entering into details, their descriptions have differed, with an established effect on the scientific community as discussed in [Zander et al. \(2016\)](#).

Table B.1: The S1/S2 System as summarized by [Tubb and Dixon \(2014\)](#) following [Stanovich and West \(2000\)](#), with supplements from [Garcez and Lamb \(2020\)](#); Reproduced from [Alexandre, Mercier, Palaude, Romero, and Vieville \(2024\)](#)

System 1/implicit	System 2/explicit
spontaneous mode	deliberate mode
associative reasoning	rule-based reasoning
parallel processing	sequential processing
holistic dialectical reasoning, involving understanding a system as a whole, considering its large-scale patterns, and reacting to them	analytic reasoning, thinking about system parts and how they work together to produce larger-scale effects
automatic processing	controlled processing
machine learning numeric processing implementation	machine learning symbolic processing implementation
short-term reactions	long-term planning
relatively undemanding	cognitively demanding
with large associative memory	with limited working memory capacity
acquisition through biology and experience (see also primary knowledge)	acquisition through cultural and formal tuition (see also secondary knowledge)
slow learning: requires a certain amount of examples to extract pattern regularities and adapt internal parameters accordingly	faster learning (a few explicit examples can be enough)
fast retrieval thanks to the acquisition of automatisms	slower retrieval; requires more cognitive load
phylogenetically evolved first	evolved recently

With the research objective of developing a computational model accounting for human learning, we need to dissect these notions precisely up to their computational meaning at the implementation level. On the one hand, as summarized by [Tubb and Dixon](#)

(2014) and detailed in Table B.1, there is the idea that these complementary dual processes can all be related to each other. Such a view is not far from the Taoist concepts of yin and yang, as discussed in Deborah Frisch’s commentary in the study by Stanovich and West (2000). We consider this to be an interesting integrated view but believe that it should be taken as an inspired outline; while (i) the two systems can act conjointly and interact, (ii) some qualities may also be related to the other system; for instance, we may consciously (thus explicitly) partially control our implicit system, as discussed by Kühberger in Stanovich and West (2000) study. On the other hand, as summarized in Table B.2, other dual aspects are to be taken into account that cannot be projected onto the S1/S2 dual axis, thus showing that the cognitive processes involved in creative complex ill-defined problem solving are multi-dimensional.

Table B.2: Other dual notions encountered in CPS. Reproduced from Alexandre, Mercier, Palaude, Romero, and Vieville (2024)

Divergent / convergent thinking	As discussed earlier, these are two temporally separated phases in creativity, in fact forming a triad with the preparation phase, as analyzed by Amabile (1996).
Fast thinking (S1) / slow thinking (S2)	As shown in Table 1 and discussed by Tubb and Dixon, 2014 and Augello et al., 2015. They considered that the divergent process decomposes into exploratory versus reflective mechanisms, respectively, with regard to the S1 versus the S2 system, and the convergent process decomposes into tacit versus analytic mechanisms, respectively.
Emotional / cognitive	According to Dietrich, 2004, the emotional versus cognitive knowledge domain is crossed with the spontaneous versus deliberate (thus S1 versus S2) domain to produce four basic types of creativity at the cognitive neuroscience level.
Semantic / syntactic	In the study by Alexandre, 2020a, the standard S1/S2 distinction was discussed at the neuro-cognitive level, with different perspectives making the difference between semantic and syntactic aspects. Furthermore, semantic value was related to emotional valuation.

Exploration / exploitation	This polysemic notion has a precise definition in reinforcement learning and has been related to divergent versus convergent processes by several authors, as reviewed by Quillien, 2019 ; to the best of our understanding, it has a different meaning. The divergent process in implicit (S1) mode was also named exploration by Tubb and Dixon, 2014 and subsequent authors, and this seems to be yet another concept.
Model-free / model-based	In problem solving involving reinforcement learning, as discussed by, for example, A. G. E. Collins and Cockburn, 2020 , learning and decision making may involve a predictive mechanism of the environment either explicitly or implicitly (with or without a model). When considering a creative process, these aspects interact with other dimensions mentioned here, as discussed by the authors.
Stimulus-driven / goal-directed	Alexandre, 2021 considered a systemic approach, distinguishing two different neuro-cognitive modalities, as discussed in the conclusion of section 3, relating to the attention to external versus internal goals (Ede et al., 2020) Ede et al., 2020 . This is also related to top-down versus bottom-up processes, which are a relevant but polysemic notion, as clarified in the text.
Flexibility / persistence	W. Zhang et al., 2020 , following Nijstad et al., 2010 and Hommel and Wiers, 2017 , drew a link between divergent thinking and cognitive flexibility, promoting “loose thinking” and creative thought as a process introducing more positive mood states, as opposed to convergent thinking, which is associated with cognitive persistence, for instance narrowing the focus of attention and increasing the top-down control. As discussed by A. Collins and Koechlin, 2012 , flexibility is related to a form of exploration (as detailed above) when referring to the choice of task sets and related actions when attempting to complete a task, while the persistence of using a given task set is related to exploitation, the link with the corresponding reinforcement learning mechanisms being made by the authors at the implementation level.
Default mode network (DMN) / central executive network (CEN)	As reviewed by Alexandre, 2020a and Dietrich and Haider, 2017 , the DMN network is involved in the spontaneous bottom-up processing mode of creativity, while the CEN network is used in creativity that emanates from the explicit system, a deliberate top-down mode.

C THE ANATOMY OF THE HUMAN BRAIN

The following recapulative figures and tables are aimed towards non neuroscientists who, like I did, get a little lost within the map of brain regions. Please note that this popularization material synthesized from [Wikipedia](#) is only meant to spare you the hassle of clicking multiple links and is by no means a scientific review.

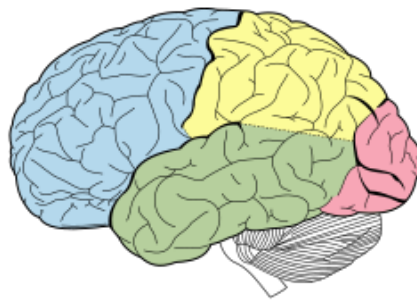


Figure C.1: Localization of the different lobes in the human brain: the **frontal lobe** (blue), the **parietal lobe** (yellow), the **occipital lobe** (red) and the **temporal lobe** (green). The **limbic lobe** is not represented because it is situated underneath the other lobes, on the medial surface of each hemisphere. The regions of the human brain within these lobes are recapitulated in the two tables on the next page. Picture reprinted from [Wikimedia](#) (public domain image).



- (a) The surface of the cortex consists of ridges called *gyri* (sg. *gyrus*) and depressions called *sulci* (sg. *sulcus*). The largest *sulci* are also called fissures. Reprinted from [Wikimedia](#) (public domain image)
- (b) We usually refer to specific regions by their location with regard to one or several of these axes: rostral-caudal (front-back), dorsal-ventral, and medial-lateral.

Figure C.2: Some vocabulary regarding the names of the brain regions.

Table C.1: Regions of the human brain within the lobes presented in Fig C.1 with matching colors. Some of them, in bold text, are mentioned in the manuscript.

Cortex	Frontal lobe	Superolateral	Prefrontal	Superior frontal gyrus Middle frontal gyrus, including Dorsolateral prefrontal cortex (dLPFC) Inferior frontal gyrus (IFG) (including Broca's area in the dominant hemisphere) Superior frontal sulcus Inferior frontal sulcus Precentral gyrus Precentral sulcus
			Precentral	Superior frontal gyrus Medial frontal gyrus Paraterminal gyrus/Paraolfactory area Straight gyrus
		Medial/inferior	Prefrontal	Orbital gyri/Orbitofrontal cortex Ventromedial prefrontal cortex (vmPFC) Subcallosal area Olfactory sulcus Orbital sulcus
			Precentral	Paracentral lobule Paracentral sulcus Primary motor cortex Premotor cortex Supplementary motor area (SMA) Supplementary eye field Frontal eye fields
		Both		
Parietal lobe	Superolateral		Superior parietal lobule Inferior parietal lobule Supramarginal gyrus Angular gyrus Parietal operculum Intraparietal sulcus Paracentral lobule	
		Medial/inferior	Precuneus cortex (PCC) Marginal sulcus	
		Both	Postcentral gyrus/Primary somatosensory cortex Secondary somatosensory cortex Posterior parietal cortex (PPC)	
Occipital lobe	Superolateral		Occipital pole of cerebrum Occipital gyri Lateral occipital gyrus Lunate sulcus Transverse occipital sulcus	
		Medial/inferior	Visual cortex Cuneus Lingual gyrus Calcarine sulcus	
Temporal lobe	Superolateral		Transverse temporal gyrus/ Auditory cortex Superior temporal gyrus (STG) (including Wernicke's area in the dominant hemisphere) Inferior temporal gyrus Superior temporal sulcus Middle temporal gyrus Icus	
			Medial/inferior	Fusiform gyrus Medial temporal lobe Inferior temporal sulcus Inferior temporal gyrus

Table C.2: Regions of the human brain within the limbic and insular lobes, as well as the sub-cortical areas.

Cortex	Limbic lobe	Parahippocampal gyrus	Entorhinal cortex (EC) Perirhinal cortex Postrhinal cortex Posterior parahippocampal gyrus Prepyriform area Subgenual area
		Cingulate cortex/gyrus	Anterior cingulate cortex (ACC) Posterior cingulate cortex Isthmus of cingulate gyrus: Retrosplenial cortex Subiculum Hippocampal sulcus Fimbria of hippocampus Dentate gyrus Rhinal sulcus
		Hippocampal formation	Hippocampus proper (HPC): Cornu Ammoni CA1, CA2, CA3, CA4 (sometimes considered part of the medial temporal lobes)
Subcortical areas	Insular cortex (INS) or insular lobe, folded deep within the lateral sulcus (the fissure separating the temporal lobe from the parietal and frontal lobes)	Other	Indusium griseum Uncus Amygdala
			Anterior insula Posterior insula
		Thalamus	
	Basal Ganglia (BG)		Striatum: dorsal striatum (caudate nucleus and putamen) ventral striatum (nucleus accumbens and olfactory tubercle)
			globus pallidus ventral pallidum substantia nigra subthalamic nucleus

D RDFS ENTAILMENT RULES

This appendix reproduces the entailment rules permitted by **RDFS**, allowing for deductive reasoning, defined in <https://www.w3.org/TR/rdf-mt/>. These rules are implemented by reasoners such as Pellet.

Rule set	Rule Name	If E contains:	then add:
simple entailment rules	<i>se1</i>	uuu aaa xxx .	uuu aaa -:nnn . where -:nnn identifies a blank node allocated to xxx by rule <i>se1</i> or <i>se2</i> . -:nnn aaa xxx .
	<i>se2</i>	uuu aaa xxx .	where -:nnn identifies a blank node allocated to uuu by rule <i>se1</i> or <i>se2</i> . uuu aaa -:nnn . where -:nnn identifies a blank node allocated to the literal lll by this rule.
special case of rule <i>se1</i> for literals	lg (literal generalization)	uuu aaa lll .	uuu aaa lll .
special case of rule <i>se1</i> for literals (RDFS)	gl (literal instantiation)	uuu aaa -:nnn . where -:nnn identifies a blank node allocated to the literal lll by rule lg. uuu aaa yyy .	uuu aaa lll .
RDF entailment rules	<i>rd/1</i>	uuu aaa lll . where lll is a well-typed XML literal .	aaa rdf:type rdf:Property . -:nnn rdf:type rdf:XMLLiteral . where -:nnn identifies a blank node allocated to lll by rule lg. -:nnn rdf:type rdfs:Literal . where -:nnn identifies a blank node allocated to lll by rule lg.
	<i>rd/2</i>	uuu aaa lll . where lll is a plain literal (with or without a language tag). aaa rdfs:domain xxx .	uuu rdf:type xxx .
	<i>rd/s1</i>	uuu aaa xxx . uuu aaa vvv .	uuu rdf:type rdfs:Resource . vvv rdf:type rdfs:Resource .
	<i>rd/s2</i>	uuu aaa yyy .	uuu rdfs:subPropertyOf xxx .
	<i>rd/s3</i>	uuu aaa vvv .	uuu rdfs:subPropertyOf uuu .
	<i>rd/s4a</i>	uuu rdfs:subPropertyOf vvv .	uuu bbb yyy .
	<i>rd/s4b</i>	vvv rdfs:subPropertyOf xxx .	uuu rdfs:subClassOf rdfs:Resource .
	<i>rd/s5</i>	uuu rdf:type rdf:Property .	vvv rdf:type xxx .
	<i>rd/s6</i>	aaa rdfs:subPropertyOf bbb .	uuu rdfs:subClassOf rdfs:Resource .
	<i>rd/s7</i>	uuu aaa yyy .	vvv rdf:type xxx .
	<i>rd/s8</i>	uuu rdf:type rdfs:Class .	uuu rdfs:type rdfs:Resource .
	<i>rd/s9</i>	uuu rdfs:subClassOf xxx .	vvv rdf:type xxx .
	<i>rd/s10</i>	vvv rdf:type uuu .	uuu rdfs:subClassOf uuu .
<i>rd/s11</i>	uuu rdf:type rdfs:Class .	uuu rdfs:subClassOf xxx .	
<i>rd/s12</i>	uuu rdfs:subClassOf vvv .	uuu rdfs:subPropertyOf rdfs:member .	
<i>rd/s13</i>	vvv rdfs:subClassOf xxx . uuu rdf:type rdfs:ContainerMembershipProperty . uuu rdf:type rdfs:Datatype .	uuu rdfs:subClassOf rdfs:Literal .	

Table D.1: The RDF/RDFS entailment rules, reproduced from <https://www.w3.org/TR/rdf11-mt>.

INDEX

- abductive reasoning, 27, 56, 97, 98, 100
- action, 20, 59
- activity, 28
- affordance, 12, 15, 33, 34, 38, 50–52
- Alternative Uses Test (AUT), 34, 35, 54, 61
- analogy, 27, 34, 55
- anterior cingulate cortex (ACC), 40, 45

- basal ganglia (BG), 38
- behavior, 20, 59, 79
- binding, 37

- central executive network (CEN), 39, 40
- closed-world, 99
- collaboration, 8
- computational thinking, 8
- convergent thinking (CT), 10, 25, 43, 53, 55, 61
- creative problem solving (CPS), 8, 19
- creativity, 8, 19
- critical thinking, 8

- declarative/conceptual knowledge, 30, 32
- deductive reasoning, 27, 56, 97, 98, 100
- default mode network (DMN), 39, 40
- distance, 51, 62, 84
- divergent thinking (DT), 10, 21, 25, 27, 34, 35, 41, 43, 53, 54, 56, 60, 61, 93

- edit distance, 85, 88, 90
- elaboration (DT dimension), 35
- episode, 37, 59, 79
- episodic memory, 37, 40, 79, 80
- exploitation, 59, 81
- exploration, 59, 81

- exteroception, 36

- flexibility (DT dimension), 34, 35, 53, 55, 93
- fluency (DT dimension), 34, 35, 55, 93

- geodesic, 84, 85
- goal-directed, 38, 42, 43, 73, 74, 82

- habitual; stimulus-based, 38, 42, 43, 73, 74, 82
- heuristic, 26, 47, 49
- hippocampus (HPC), 37, 40

- ill-defined, 10–12, 19
- inductive reasoning, 27, 56, 97, 98
- insight, 24, 40, 42, 46, 58
- insular cortex (INS), 36, 39, 40
- internalization, 31
- interoception, 36
- intuition, 40, 42

- manifold, 84, 95, 97
- Markov decision process (MDP), 78, 79
- Markov property, 78
- metacognition, 9, 42
- method, 27, 30
- model-based, 82
- model-free, 82, 86
- monotonic, 100, 113
- motivation, 27–29
- motive, 28

- ontology, 10, 64, 86
- open-world, 49, 53, 54, 100
- orbito-frontal cortex (OFC), 38, 45
- originality (DT dimension), 34, 35, 93

Index

- policy, 26, 49, 59, 80–83
- prefrontal cortex (PFC), 36, 40, 41, 43
- premotor cortex (PMC), 38
- primary knowledge, 27, 29, 126
- problem solving, 8, 19
- problem space, 47
- procedural memory, 37
- procedural/behavioral knowledge, 30, 32
- proprioception, 36
- prospective memory, 40, 41

- Q-learning, 10, 38, 61, 81–83

- reinforcement learning, 10, 20, 45, 59, 77
- Remote Associates Test (RAT), 34, 35, 42, 43, 55, 56, 61, 115
- Resource Description Framework (RDF), 54, 72, 84, 101
- Resource Description Framework Schema (RDFS), 54, 61, 98, 100, 107, 112, 113, 132

- salience network (SN), 39, 40
- secondary knowledge, 27, 29, 126
- semantic memory, 37, 53
- Semantic Pointer Architecture (SPA), 96, 98, 99, 102, 115
- sensory cortex, 36, 41
- state value, 80
- state-action value, 82
- stimulus, 11, 25, 29, 34, 36, 40, 60
- symbol, 22, 31, 48, 51

- task, 28
- task set, 44, 45, 61, 91, 92
- Torrance Test of Creative Thinking (TTCT), 34, 35, 40
- triples, 54, 95, 98, 101

- Vector Symbolic Architecture (VSA), 96, 98, 102, 113
- Vector-derived Transformation Binding (VTB), 103, 106

- Web Ontology Language (OWL), 54, 61, 64, 87, 98, 100, 107, 112, 113

- working memory, 36, 40, 44, 57

REFERENCES

- Alabbasi, A. M. A., S. H. Paek, D. Kim, and B. Cramond (2022). What do educators need to know about the Torrance Tests of Creative Thinking: A comprehensive review. *Frontiers in Psychology* 13. ISSN: 1664-1078.
- Albero, B. and J. Guérin (2014). L'intérêt pour l'activité en sciences de l'éducation. Vers une épistémologie fédératrice ? *TransFormations : Recherches en éducation et formation des adultes* 11, pp. 11–45.
- Alexandre, F. (10, 2019). De quelles façons l'intelligence artificielle se sert-elle des neurosciences ? *The Conversation*.
- (7, 2020a). "Creativity explained by Computational Cognitive Neuroscience". In: *ICCC'20 - International Conference on Computational Creativity*, p. 4.
 - (26, 2020b). Les relations difficiles entre l'Intelligence Artificielle et les Neurosciences. *Interstices*.
- Alexandre, F. (2021). A global framework for a systemic view of brain modeling. *Brain Informatics* 8:1, p. 3. ISSN: 2198-4018, 2198-4026. DOI: [10.1186/s40708-021-00126-4](https://doi.org/10.1186/s40708-021-00126-4).
- Alexandre, F., J. Becker, M.-H. Comte, A. Lagarrigue, R. Liblau, M. Romero, and T. Viéville (24, 2020). *Open Educational Resources and MOOC for Citizen Understanding of Artificial Intelligence*.
- Alexandre, F., C. Mercier, A. Palaude, M. Romero, and T. Viéville (2024). *Modeling Creative Problem-Solving tasks from a computational and neuroeducational approach*.
- Alexandre, F., C. Mercier, A. Palaude, M. Romero, and T. Viéville (12, 2022). "The role of cognitive processes in creative problem solving: a computational approach". Poster. 7th annual meeting of the Society for the Neuroscience of Creativity (SfNC 2022). United States / Virtual.
- Allemang, D., J. Hendler, and F. Gandon (2020). *Semantic Web for the Working Ontologist: Effective Modeling for Linked Data, RDFS, and OWL*. 3rd ed. Association for Computing Machinery, New York, NY, USA. 510 pp. ISBN: 978-1-4503-7617-4. DOI: [10.1145/3382097](https://doi.org/10.1145/3382097).
- Almeida, L. S., L. P. Prieto, M. Ferrando, E. Oliveira, and C. Ferrándiz (1, 2008). Torrance Test of Creative Thinking: The question of its construct validity. *Thinking Skills and Creativity* 3:1, pp. 53–58. ISSN: 1871-1871. DOI: [10.1016/j.tsc.2008.03.003](https://doi.org/10.1016/j.tsc.2008.03.003).
- Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology* 43:5, pp. 997–1013. ISSN: 1939-1315, 0022-3514. DOI: [10.1037/0022-3514.43.5.997](https://doi.org/10.1037/0022-3514.43.5.997).
- (1983). *The Social Psychology of Creativity*. Springer Verlag, New York, NY. ISBN: 978-1-4612-5533-8. DOI: [10.1007/978-1-4612-5533-8](https://doi.org/10.1007/978-1-4612-5533-8).
 - (1996). *Creativity In Context: Update To The Social Psychology Of Creativity*. Routledge, New York. 336 pp. ISBN: 978-0-8133-3034-1. DOI: [10.4324/9780429501234](https://doi.org/10.4324/9780429501234).

References

- Amabile, T. M. (2012). *Componential Theory of Creativity*. Vol. 12-096. Harvard Business School.
- American Psychological Association (2023). In: *APA Dictionary of Psychology*. URL: <https://dictionary.apa.org/> (visited on 12/06/2023).
- Anderson, J. R., M. Matessa, and C. Lebiere (1997). ACT-R: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction* 12:4, pp. 439–462. ISSN: 1532-7051. DOI: [10.1207/s15327051hci1204_5](https://doi.org/10.1207/s15327051hci1204_5).
- Arora, S. and P. Doshi (18, 2020). *A Survey of Inverse Reinforcement Learning: Challenges, Methods and Progress*.
- Asprino, L., A. G. Nuzzolese, A. Russo, A. Gangemi, V. Presutti, and S. Nolfi (2017). An Ontology Design Pattern for supporting behaviour arbitration in cognitive agents. In: *Advances in Ontology Design and Patterns*. Vol. 32. Studies on the Semantic Web. IOS Press, p. 10. DOI: [10.3233/978-1-61499-826-6-85](https://doi.org/10.3233/978-1-61499-826-6-85).
- Augello, A., I. Infantino, A. Lieto, G. Pilato, R. Rizzo, and F. Vella (1, 2015). Artwork creation by a cognitive architecture integrating computational creativity and dual process approaches. *Biologically Inspired Cognitive Architectures* 15. DOI: [10.1016/j.bica.2015.09.007](https://doi.org/10.1016/j.bica.2015.09.007).
- Ayadi, A., A. Samet, F. d. B. de Beuvron, and C. Zanni-Merk (1, 2019). Ontology population with deep learning-based NLP: a case study on the Biomolecular Network Ontology. *Procedia Computer Science*. Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 23rd International Conference KES2019 159, pp. 572–581. ISSN: 1877-0509. DOI: [10.1016/j.procs.2019.09.212](https://doi.org/10.1016/j.procs.2019.09.212).
- Badre, D. (1, 2008). Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends in Cognitive Sciences* 12:5, pp. 193–200. ISSN: 1364-6613, 1879-307X. DOI: [10.1016/j.tics.2008.02.004](https://doi.org/10.1016/j.tics.2008.02.004).
- Badreddine, S., A. d. Garcez, L. Serafini, and M. Spranger (16, 2021). *Logic Tensor Networks*.
- Barnabé, S., L. Denet, M. Manrique, D. Menon, E. Pascual, M. Romero, and T. Viéville (4, 2020). *A low-cost tabletop game to collect learning analytics during computational thinking using unplugged or tangible activities*. Research report RR-9379. Inria.
- Basantia, T. K. (11, 2017). Culture, Value and Personality: Three Flowering Agents of Creativity Development Process. *American Journal of Applied Psychology* 5:1, pp. 1–6. ISSN: 2333-472X. DOI: [10.12691/ajap-5-1-1](https://doi.org/10.12691/ajap-5-1-1).
- Beaty, R. E., M. Benedek, P. J. Silvia, and D. L. Schacter (2016). Creative Cognition and Brain Network Dynamics. *Trends in Cognitive Sciences* 20:2, pp. 87–95. ISSN: 1879-307X. DOI: [10.1016/j.tics.2015.10.004](https://doi.org/10.1016/j.tics.2015.10.004).
- Bekolay, T., J. Bergstra, E. Hunsberger, T. DeWolf, T. C. Stewart, D. Rasmussen, X. Choo, A. R. Voelker, and C. Eliasmith (2014). Nengo: a Python tool for building large-scale functional brain models. *Frontiers in Neuroinformatics* 7. ISSN: 1662-5196. DOI: [10.3389/fninf.2013.00048](https://doi.org/10.3389/fninf.2013.00048).
- Bernard, P., B. Hate, and M. Laval (27, 2023). *Symboling : utiliser des structures symboliques dotées d'une métrique*. Research Report RR-9499. Inria & Labri, Univ. Bordeaux, p. 18.

- Besançon, M., B. Barbot, and T. Lubart (1, 2011). Évolution de l'évaluation de la créativité chez l'enfant de Binet à nos jours. *Recherches & éducations* 5, pp. 215–226. ISSN: 1969-0622, 1760-7760. DOI: [10.4000/rechercheseducations.840](https://doi.org/10.4000/rechercheseducations.840).
- Betzler, R. and D. Bassett (31, 2016). Multi-scale brain networks. *NeuroImage* 160. DOI: [10.1016/j.neuroimage.2016.11.006](https://doi.org/10.1016/j.neuroimage.2016.11.006).
- Boden, M. A. (1, 1998). Creativity and artificial intelligence. *Artificial Intelligence*. Artificial Intelligence 40 years later 103:1, pp. 347–356. ISSN: 0004-3702. DOI: [10.1016/S0004-3702\(98\)00055-1](https://doi.org/10.1016/S0004-3702(98)00055-1).
- (2004). *The Creative Mind: Myths and Mechanisms*. Psychology Press. 368 pp. ISBN: 978-0-415-31452-7.
- Bogacz, R. (2007). Optimal decision-making theories: linking neurobiology with behaviour. *Trends in Cognitive Sciences* 11:3, pp. 118–125. ISSN: 13646613. DOI: [10.1016/j.tics.2006.12.006](https://doi.org/10.1016/j.tics.2006.12.006).
- Boutilier, C. and V. Beche (1, 1995). Abduction as belief revision. *Artificial Intelligence* 77:1, pp. 43–94. ISSN: 0004-3702. DOI: [10.1016/0004-3702\(94\)00025-V](https://doi.org/10.1016/0004-3702(94)00025-V).
- Bowers, J. (1, 2016). Psychology, Not Educational Neuroscience, Is the Way Forward for Improving Educational Outcomes for All Children: Reply to Gabrieli (2016) and Howard-Jones et al. (2016). *Psychological Review* 123, pp. 628–635. DOI: [10.1037/rev0000043](https://doi.org/10.1037/rev0000043).
- Brand-Gruwel, S., I. Wopereis, and Y. Vermetten (2005). Information problem solving by experts and novices: analysis of a complex cognitive skill. *Computers in Human Behavior* 21:3, pp. 487–508. ISSN: 07475632. DOI: [10.1016/j.chb.2004.10.005](https://doi.org/10.1016/j.chb.2004.10.005).
- Buckner, R. L. (2010). The role of the hippocampus in prediction and imagination. *Annual Review of Psychology* 61, pp. 27–48, C1–8. ISSN: 1545-2085. DOI: [10.1146/annurev.psych.60.110707.163508](https://doi.org/10.1146/annurev.psych.60.110707.163508).
- Bull, S. and J. Kay (2010). Open Learner Models. In: *Advances in Intelligent Tutoring Systems*. Ed. by R. Nkambou, J. Bourdeau, and R. Mizoguchi. Studies in Computational Intelligence. Springer, Berlin, Heidelberg, pp. 301–322. ISBN: 978-3-642-14363-2. DOI: [10.1007/978-3-642-14363-2_15](https://doi.org/10.1007/978-3-642-14363-2_15).
- Bunge, S. A. (1, 2004). How we use rules to select actions: A review of evidence from cognitive neuroscience. *Cognitive, Affective, & Behavioral Neuroscience* 4:4, pp. 564–579. ISSN: 1531-135X. DOI: [10.3758/CABN.4.4.564](https://doi.org/10.3758/CABN.4.4.564).
- Callan, G. L., L. D. Rubenstein, L. M. Ridgley, and J. R. McCall (2021). Measuring self-regulated learning during creative problem-solving with SRL microanalysis. *Psychology of Aesthetics, Creativity, and the Arts* 15:1, pp. 136–148. ISSN: 1931-390X. DOI: [10.1037/aca0000238](https://doi.org/10.1037/aca0000238).
- Carnine, D. (2000). *Why Education Experts Resist Effective Practices (And What It Would Take To Make Education More Like Medicine)*. Thomas B.
- Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford University Press. 221 pp.
- Cassone, L., M. Romero, T. Viéville, C. De Smet, and M. Ndiaye (2019). “Co-creativity, robotics and maker education Proceedings”. In: *ANR# CreaMaker workshop*. Nice, France.
- Cessac, B. and T. Viéville (2008). On Dynamics of Integrate-and-Fire Neural Networks with Adaptive Conductances. *Frontiers in Neuroscience* 2:2.

References

- Chi, M. T. H. (2009). Active-Constructive-Interactive: A Conceptual Framework for Differentiating Learning Activities. *Topics in Cognitive Science* 1:1, pp. 73–105. ISSN: 17568757, 17568765. DOI: [10.1111/j.1756-8765.2008.01005.x](https://doi.org/10.1111/j.1756-8765.2008.01005.x).
- Chi, M. T. H. and R. Wylie (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist* 49:4, pp. 219–243. ISSN: 0046-1520, 1532-6985. DOI: [10.1080/00461520.2014.965823](https://doi.org/10.1080/00461520.2014.965823).
- Childs, P., J. Han, L. Chen, P. Jiang, P. Wang, D. Park, Y. Yin, E. Dieckmann, and I. Vilanova (2022). The Creativity Diamond—A Framework to Aid Creativity. *Journal of Intelligence* 10:4, p. 73. ISSN: 2079-3200. DOI: [10.3390/jintelligence10040073](https://doi.org/10.3390/jintelligence10040073).
- Chraïbi Kaadoud, I., A. Bennetot, B. Mawhin, V. Charisi, and N. Díaz-Rodríguez (2022). Explaining Aha! moments in artificial agents through IKE-XAI: Implicit Knowledge Extraction for eXplainable AI. *Neural Networks* 155, pp. 95–118. ISSN: 0893-6080. DOI: [10.1016/j.neunet.2022.08.002](https://doi.org/10.1016/j.neunet.2022.08.002).
- Chui, C. K. and H. N. Mhaskar (2018). Deep Nets for Local Manifold Learning. *Frontiers in Applied Mathematics and Statistics* 4. ISSN: 2297-4687. DOI: [10.3389/fams.2018.00012](https://doi.org/10.3389/fams.2018.00012).
- Clement, B., D. Roy, P.-Y. Oudeyer, and M. Lopes (2014). “Developmental learning for Intelligent Tutoring Systems”. In: *4th International Conference on Development and Learning and on Epigenetic Robotics*. IEEE, Genoa, Italy, p. 10. ISBN: 978-1-4799-7540-2. DOI: [10.1109/DEVLRN.2014.6983019](https://doi.org/10.1109/DEVLRN.2014.6983019).
- Cohen, W. W., F. Yang, and K. R. Mazaitis (2017). TensorLog: Deep Learning Meets Probabilistic DBs. *arXiv:1707.05390 [cs]*.
- Collins, A. and E. Koechlin (2012). Reasoning, Learning, and Creativity: Frontal Lobe Function and Human Decision-Making. *PLoS Biology* 10:3. Ed. by J. P. O’Doherty, e1001293. ISSN: 1545-7885. DOI: [10.1371/journal.pbio.1001293](https://doi.org/10.1371/journal.pbio.1001293).
- Collins, A. G. E. and J. Cockburn (2020). Beyond simple dichotomies in reinforcement learning. *Nature Reviews Neuroscience* 21:10, pp. 576–586. ISSN: 1471-0048. DOI: [10.1038/s41583-020-0355-6](https://doi.org/10.1038/s41583-020-0355-6).
- Conway-Smith, B., R. West, and M. Mylopoulos (2023). *Metacognitive skill: how it is acquired*.
- Cosgrove, A. L., Y. N. Kenett, R. E. Beaty, and M. T. Diaz (2021). Quantifying flexibility in thought: The resiliency of semantic networks differs across the lifespan. *Cognition* 211, p. 104631. ISSN: 0010-0277. DOI: [10.1016/j.cognition.2021.104631](https://doi.org/10.1016/j.cognition.2021.104631).
- Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in brain research* 169, pp. 323–338. ISSN: 0079-6123. DOI: [10.1016/S0079-6123\(07\)00020-9](https://doi.org/10.1016/S0079-6123(07)00020-9).
- (2010). Sensory and Immediate Memory. *Encyclopedia of Consciousness*. DOI: [10.1016/B978-012373873-8.00048-7](https://doi.org/10.1016/B978-012373873-8.00048-7).
- Cox, J. R. and R. A. Griggs (1982). The effects of experience on performance in Wason’s selection task. *Memory & Cognition* 10:5, pp. 496–502. ISSN: 1532-5946. DOI: [10.3758/BF03197653](https://doi.org/10.3758/BF03197653).
- Cramond, B., J. Matthews-Morgan, D. Bandalos, and L. Zuo (2005). A Report on the 40-Year Follow-Up of the Torrance Tests of Creative Thinking: Alive and Well in the

- New Millennium. *Gifted Child Quarterly* 49:4, pp. 283–291. ISSN: 0016-9862. DOI: [10.1177/001698620504900402](https://doi.org/10.1177/001698620504900402).
- Crawford, E., M. Gingerich, and C. Eliasmith (2016). Biologically Plausible, Human-Scale Knowledge Representation. *Cognitive Science* 40:4, pp. 782–821. ISSN: 03640213. DOI: [10.1111/cogs.12261](https://doi.org/10.1111/cogs.12261).
- Daikoku, T., Q. Fang, T. Hamada, Y. Handa, and Y. Nagai (2021). Importance of environmental settings for the temporal dynamics of creativity. *Thinking Skills and Creativity* 41, p. 100911. ISSN: 18711871. DOI: [10.1016/j.tsc.2021.100911](https://doi.org/10.1016/j.tsc.2021.100911).
- De Grave, W. S., H. P. A. Boshuizen, and H. G. Schmidt (1, 1996). Problem based learning: Cognitive and metacognitive processes during problem analysis. *Instructional Science* 24:5, pp. 321–341. ISSN: 1573-1952. DOI: [10.1007/BF00118111](https://doi.org/10.1007/BF00118111).
- De Penning, H. L. H., A. S. D. Garcez, L. C. Lamb, and J.-J. C. Meyer (16, 2011). “A Neural-Symbolic Cognitive Agent for Online Learning and Reasoning”. In: *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence - Volume Volume Two*. IJCAI’11. AAAI Press, Barcelona, Catalonia, Spain, pp. 1653–1658. ISBN: 978-1-57735-514-4.
- De Villiers, T. (2007). Why Peirce matters: the symbol in Deacon’s Symbolic Species. *Language Sciences* 29:1, pp. 88–108. ISSN: 03880001. DOI: [10.1016/j.langsci.2006.07.003](https://doi.org/10.1016/j.langsci.2006.07.003).
- Denervaud, S., A. P. Christensen, Y. N. Kenett, and R. E. Beaty (9, 2021). Education shapes the structure of semantic memory and impacts creative thinking. *npj Science of Learning* 6:1, pp. 1–7. ISSN: 2056-7936. DOI: [10.1038/s41539-021-00113-8](https://doi.org/10.1038/s41539-021-00113-8).
- Denet, L. (18, 2021). “Analyse de la motivation intrinsèque au cours d’une activité de résolution de problèmes.” Master Thesis. Bordeaux, France: Inria & Labri, Université Bordeaux.
- Denceux, T., D. Dubois, and H. Prade (2020). Representations of Uncertainty in AI: Probability and Possibility. In: *A Guided Tour of Artificial Intelligence Research: Volume I: Knowledge Representation, Reasoning and Learning*. Ed. by P. Marquis, O. Papini, and H. Prade. Springer International Publishing, Cham, pp. 69–117. ISBN: 978-3-030-06164-7. DOI: [10.1007/978-3-030-06164-7_3](https://doi.org/10.1007/978-3-030-06164-7_3).
- Desislavov, R., F. Martínez-Plumed, and J. Hernández-Orallo (2023). Compute and Energy Consumption Trends in Deep Learning Inference. *Sustainable Computing: Informatics and Systems* 38, p. 100857. ISSN: 22105379. DOI: [10.1016/j.suscom.2023.100857](https://doi.org/10.1016/j.suscom.2023.100857).
- Diamond, A. and D. S. Ling (2016). Conclusions about interventions, programs, and approaches for improving executive functions that appear justified and those that, despite much hype, do not. *Developmental Cognitive Neuroscience* 18, pp. 34–48. ISSN: 1878-9307. DOI: [10.1016/j.dcn.2015.11.005](https://doi.org/10.1016/j.dcn.2015.11.005).
- Dietrich, A. (2004). The cognitive neuroscience of creativity. *Psychonomic Bulletin & Review* 11:6, pp. 1011–1026. ISSN: 1069-9384. DOI: [10.3758/bf03196731](https://doi.org/10.3758/bf03196731).
- Dietrich, A. and H. Haider (2017). A Neurocognitive Framework for Human Creative Thought. *Frontiers in Psychology* 7, p. 2078. ISSN: 1664-1078. DOI: [10.3389/fpsyg.2016.02078](https://doi.org/10.3389/fpsyg.2016.02078).

References

- Dietrich, A. and R. Kanso (2010). A review of EEG, ERP, and neuroimaging studies of creativity and insight. *Psychological Bulletin* 136:5, pp. 822–848. ISSN: 1939-1455, 0033-2909. DOI: [10.1037/a0019749](https://doi.org/10.1037/a0019749).
- Dillenbourg, P. and J. Self (1992). A Framework for Learner Modelling. *Interactive Learning Environments* 2:2, pp. 111–137. ISSN: 1049-4820, 1744-5191. DOI: [10.1080/1049482920020202](https://doi.org/10.1080/1049482920020202).
- Domenech, P. and E. Koechlin (2015). Executive control and decision-making in the prefrontal cortex. *Current Opinion in Behavioral Sciences* 1, pp. 101–106. ISSN: 23521546. DOI: [10.1016/j.cobeha.2014.10.007](https://doi.org/10.1016/j.cobeha.2014.10.007).
- Domingos, P. (1996). Unifying instance-based and rule-based induction. *Machine Language* 24:2, pp. 141–168. ISSN: 0885-6125. DOI: [10.1023/A:1018006431188](https://doi.org/10.1023/A:1018006431188).
- Donoso, M. (2013). “Le cerveau stratège : les fondements du raisonnement dans le cortex préfrontal humain”. PhD thesis. Paris, France: Université Pierre et Marie Curie. 1 vol. (235 p.)
- Doriath-Döhler, G. (2021). *Implémentation neuronale d’inférences logiques dans une SPA*. Research report in preparation. Inria Bordeaux Sud-Ouest.
- Dubitzky, W., T. Kötter, O. Schmidt, and M. Berthold (1, 2012). Towards Creative Information Exploration Based on Koestler’s Concept of Bisociation. In: *Bisociative Knowledge Discovery*. Vol. 7250. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 11–32. DOI: [10.1007/978-3-642-31830-6_2](https://doi.org/10.1007/978-3-642-31830-6_2).
- Dubois, D. and H. Prade (2015a). Practical Methods for Constructing Possibility Distributions. *International Journal of Intelligent Systems* vol. 31, n° 3, pp. 215–239. DOI: [10.1002/int.21782](https://doi.org/10.1002/int.21782).
- Dubois, D. and H. Prade (2015b). Possibility Theory and Its Applications: Where Do We Stand? In: *Springer Handbook of Computational Intelligence*. Ed. by J. Kacprzyk and W. Pedrycz. Springer Handbooks. Springer, Berlin, Heidelberg, pp. 31–60. ISBN: 978-3-662-43505-2. DOI: [10.1007/978-3-662-43505-2_3](https://doi.org/10.1007/978-3-662-43505-2_3).
- Duncker, L. and M. Sahani (1, 2021). Dynamics on the manifold: Identifying computational dynamical activity from neural population recordings. *Current Opinion in Neurobiology*. Computational Neuroscience 70, pp. 163–170. ISSN: 0959-4388. DOI: [10.1016/j.conb.2021.10.014](https://doi.org/10.1016/j.conb.2021.10.014).
- Duverne, S. and E. Koechlin (2017). Hierarchical Control of Behaviour in Human Prefrontal Cortex. In: *The Wiley Handbook of Cognitive Control*. John Wiley & Sons, Ltd, pp. 207–220. ISBN: 978-1-118-92049-7. DOI: [10.1002/9781118920497.ch12](https://doi.org/10.1002/9781118920497.ch12).
- Ede, F. v., A. G. Board, and A. C. Nobre (29, 2020). Goal-directed and stimulus-driven selection of internal representations. *Proceedings of the National Academy of Sciences* 117:39, pp. 24590–24598. ISSN: 0027-8424, 1091-6490. DOI: [10.1073/pnas.2013432117](https://doi.org/10.1073/pnas.2013432117).
- Eichenbaum, H. (3, 2017). Memory: Organization and Control. *Annual Review of Psychology* 68:1, pp. 19–45. ISSN: 0066-4308, 1545-2085. DOI: [10.1146/annurev-psych-010416-044131](https://doi.org/10.1146/annurev-psych-010416-044131).
- Eidoon, Z., N. Yazdani, and F. Oroumchian (2008). “Ontology Matching Using Vector Space”. In: *Advances in Information Retrieval*. Ed. by C. Macdonald, I. Ounis, V. Plachouras, I. Ruthven, and R. W. White. Lecture Notes in Computer Science. Springer,

- Berlin, Heidelberg, pp. 472–481. ISBN: 978-3-540-78646-7. DOI: [10.1007/978-3-540-78646-7_45](https://doi.org/10.1007/978-3-540-78646-7_45).
- Eliasmith, C. (27, 2013). *How to Build a Brain: A Neural Architecture for Biological Cognition*. OUP USA. 476 pp. ISBN: 978-0-19-979454-6.
- Eliasmith, C. and C. H. Anderson (2002). *Neural Engineering: Computation, Representation, and Dynamics in Neurobiological Systems*. A Bradford Book. The MIT Press.
- Engeness, I. (2020). Developing teachers' digital identity: towards the pedagogic design principles of digital environments to enhance students' learning in the 21st century. *European Journal of Teacher Education* 44:1, pp. 96–114. ISSN: 0261-9768. DOI: [10.1080/02619768.2020.1849129](https://doi.org/10.1080/02619768.2020.1849129).
- Engestrom, Y. (1987). "Learning by Expanding: An Activity Theoretical Approach to Developmental Research." Doctoral Thesis. Helsinki: Orienta-Konsultit.
- Engestrom, Y. (1, 2000). Activity theory as a framework for analyzing and redesigning work. *Ergonomics* 43:7, pp. 960–974. ISSN: 0014-0139, 1366-5847. DOI: [10.1080/001401300409143](https://doi.org/10.1080/001401300409143).
- Eppe, M., C. Gumbsch, M. Kerzel, P. D. H. Nguyen, M. V. Butz, and S. Wermter (2022). Intelligent problem-solving as integrated hierarchical reinforcement learning. *Nature Machine Intelligence* 4:1, pp. 11–20. ISSN: 2522-5839. DOI: [10.1038/s42256-021-00433-9](https://doi.org/10.1038/s42256-021-00433-9).
- Epstein, Z., A. Hertzmann, and THE INVESTIGATORS OF HUMAN CREATIVITY (16, 2023). Art and the science of generative AI. *Science* 380:6650, pp. 1110–1111. DOI: [10.1126/science.adh4451](https://doi.org/10.1126/science.adh4451).
- Ethayarajh, K., D. Duvenaud, and G. Hirst (2019). "Towards Understanding Linear Word Analogies". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Ed. by A. Korhonen, D. Traum, and L. Màrquez. Association for Computational Linguistics, Florence, Italy, pp. 3253–3262. DOI: [10.18653/v1/P19-1315](https://doi.org/10.18653/v1/P19-1315).
- Fauconnier, G. and M. Turner (1, 2003). Conceptual Blending, Form and Meaning. *Recherches en Communication; No 19: Sémiotique cognitive — Cognitive Semiotics; 57-86* 19. DOI: [10.14428/rec.v19i19.48413](https://doi.org/10.14428/rec.v19i19.48413).
- Freksa, C. (2015). "Strong Spatial Cognition". In: *Spatial Information Theory*. Ed. by S. I. Fabrikant, M. Raubal, M. Bertolotto, C. Davies, S. Freundschuh, and S. Bell. Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 65–86. ISBN: 978-3-319-23374-1. DOI: [10.1007/978-3-319-23374-1_4](https://doi.org/10.1007/978-3-319-23374-1_4).
- Friston, K. (2003). Learning and inference in the brain. *Neural Networks: The Official Journal of the International Neural Network Society* 16:9, pp. 1325–1352. ISSN: 0893-6080. DOI: [10.1016/j.neunet.2003.06.005](https://doi.org/10.1016/j.neunet.2003.06.005).
- Fuchs, K. and P. A. Henning (2018). "Cognitive Space Time: A Model for Human-Centered Adaptivity in E-Learning". In: *2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*. IEEE, Stuttgart, pp. 1–9. ISBN: 978-1-5386-1469-3. DOI: [10.1109/ICE.2018.8436283](https://doi.org/10.1109/ICE.2018.8436283).
- Garcez, A. d. and L. C. Lamb (16, 2020). Neurosymbolic AI: The 3rd Wave. *arXiv:2012.05876* [cs].
- Garcez, A. d. and L. C. Lamb (15, 2023). Neurosymbolic AI: the 3rd wave. *Artificial Intelligence Review* 0. ISSN: 1573-7462. DOI: [10.1007/s10462-023-10448-w](https://doi.org/10.1007/s10462-023-10448-w).

References

- Gärdenfors, P. (1, 2004). Conceptual Spaces as a Framework for Knowledge Representation. *Mind and Matter* 2, pp. 9–27.
- Garnelo, M., K. Arulkumaran, and M. Shanahan (1, 2016). Towards Deep Symbolic Reinforcement Learning. *arXiv:1609.05518* [cs].
- Gaut, B. (2010). The Philosophy of Creativity. *Philosophy Compass* 5:12, pp. 1034–1046. ISSN: 1747-9991. DOI: [10.1111/j.1747-9991.2010.00351.x](https://doi.org/10.1111/j.1747-9991.2010.00351.x).
- Gayler, R. (1, 2003). “Vector Symbolic Architectures answer Jackendoff’s challenges for cognitive neuroscience”. In: *Frontiers in Artificial Intelligence and Applications*.
- George, G. and A.M. Lal (1, 2019). Review of ontology-based recommender systems in e-learning. *Computers & Education* 142, p. 103642. ISSN: 0360-1315. DOI: [10.1016/j.compedu.2019.103642](https://doi.org/10.1016/j.compedu.2019.103642).
- Gerver, C. R., J. W. Griffin, N. A. Dennis, and R. E. Beaty (25, 2023). Memory and creativity: A meta-analytic examination of the relationship between memory systems and creative cognition. *Psychonomic Bulletin & Review*. ISSN: 1531-5320. DOI: [10.3758/s13423-023-02303-4](https://doi.org/10.3758/s13423-023-02303-4).
- Gibson, J. J. (1977). The theory of affordances. In: *Perceiving, acting, and knowing: toward an ecological psychology*. Ed. by J. B. Robert E Shaw. Hillsdale, N.J. : Lawrence Erlbaum Associates, pp.67–82.
- (1979). *The Ecological Approach to Visual Perception: Classic Edition*. Psychology Press. 347 pp. ISBN: 978-1-317-57938-0.
- Gilhooly, K. and P. Murphy (1, 2005). Differentiating insight from non-insight problems. *Thinking & Reasoning* 11:3, pp. 279–302. ISSN: 1354-6783. DOI: [10.1080/13546780442000187](https://doi.org/10.1080/13546780442000187).
- Glennerster, A. (15, 2002). Computational theories of vision. *Current biology: CB* 12:20, R682–685. ISSN: 0960-9822. DOI: [10.1016/s0960-9822\(02\)01204-6](https://doi.org/10.1016/s0960-9822(02)01204-6).
- Gobet, F. and G. Sala (2019). How Artificial Intelligence Can Help Us Understand Human Creativity. *Frontiers in Psychology* 10. ISSN: 1664-1078.
- Gosmann, J. and C. Eliasmith (2019). Vector-Derived Transformation Binding: An Improved Binding Operation for Deep Symbol-Like Processing in Neural Networks. *Neural Computation* 31:5, pp. 849–869. ISSN: 0899-7667, 1530-888X. DOI: [10.1162/neco_a_01179](https://doi.org/10.1162/neco_a_01179).
- Grosz, B. N., I. Horrocks, R. Volz, and S. Decker (20, 2003). “Description logic programs: combining logic programs with description logic”. In: *Proceedings of the 12th international conference on World Wide Web. WWW ‘03*. Association for Computing Machinery, New York, NY, USA, pp. 48–57. ISBN: 978-1-58113-680-7. DOI: [10.1145/775152.775160](https://doi.org/10.1145/775152.775160).
- Guerra, J., R. Hosseini, S. Somyurek, and P. Brusilovsky (7, 2016). “An Intelligent Interface for Learning Content: Combining an Open Learner Model and Social Comparison to Support Self-Regulated Learning and Engagement”. In: *Proceedings of the 21st International Conference on Intelligent User Interfaces*. ACM, Sonoma California USA, pp. 152–163. ISBN: 978-1-4503-4137-0. DOI: [10.1145/2856767.2856784](https://doi.org/10.1145/2856767.2856784).
- Guilford, J. (1967). *The nature of human intelligence*. The nature of human intelligence. McGraw-Hill, New York, NY, US.
- Guo, S., Q. Wang, L. Wang, B. Wang, and L. Guo (2016). “Jointly Embedding Knowledge Graphs and Logical Rules”. In: *Proceedings of the 2016 Conference on Empirical Meth-*

- ods in Natural Language Processing*. Association for Computational Linguistics, Austin, Texas, pp. 192–202. doi: [10.18653/v1/D16-1019](https://doi.org/10.18653/v1/D16-1019).
- Han, J., F. Shi, L. Chen, and P. R. Childs (2018). A computational tool for creative idea generation based on analogical reasoning and ontology. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 32:4, pp. 462–477. ISSN: 0890-0604, 1469-1760. doi: [10.1017/S0890060418000082](https://doi.org/10.1017/S0890060418000082).
- Hao, N., Y. Ku, M. Liu, Y. Hu, M. Bodner, R. H. Grabner, and A. Fink (2016). Reflection enhances creativity: Beneficial effects of idea evaluation on idea generation. *Brain and cognition* 103, pp. 30–37.
- Hardcastle, V. G. and K. Hardcastle (2015). Marr’s Levels Revisited: Understanding How Brains Break. *Topics in Cognitive Science* 7:2, pp. 259–273. ISSN: 1756-8765. doi: [10.1111/tops.12130](https://doi.org/10.1111/tops.12130).
- Harnad, S. (1, 1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena* 42:1, pp. 335–346. ISSN: 0167-2789. doi: [10.1016/0167-2789\(90\)90087-6](https://doi.org/10.1016/0167-2789(90)90087-6).
- (11, 2024). Language Writ Large: LLMs, ChatGPT, Grounding, Meaning and Understanding. doi: [10.31234/osf.io/ch2wx](https://doi.org/10.31234/osf.io/ch2wx).
- Hay, L., A. H. Duffy, C. McTeague, L. M. Pidgeon, T. Vuletic, and M. Grealy (2017). A systematic review of protocol studies on conceptual design cognition: Design as search and exploration. *Design Science* 3.
- Heiser, L., M. Romero, and V. Prokofieva Nelson (10, 2022). *Analyse des activités technocréatives au collège en contexte d’éducation FabLab*. Université Côte d’Azur.
- Hennessey, B. A. and T. M. Amabile (2010). Creativity. *Annual Review of Psychology* 61:1, pp. 569–598. ISSN: 0066-4308, 1545-2085. doi: [10.1146/annurev.psych.093008.100416](https://doi.org/10.1146/annurev.psych.093008.100416).
- Hohenecker, P. and T. Lukasiewicz (7, 2020). Ontology Reasoning with Deep Neural Networks. *Journal of Artificial Intelligence Research* 68, pp. 503–540. ISSN: 1076-9757. doi: [10.1613/jair.1.11661](https://doi.org/10.1613/jair.1.11661).
- Hommel, B., J. Müsseler, G. Aschersleben, and W. Prinz (1, 2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *The Behavioral and brain sciences* 24, 849–78, discussion 878. doi: [10.1017/S0140525X01000103](https://doi.org/10.1017/S0140525X01000103).
- Hommel, B. and R. W. Wiers (1, 2017). Towards a Unitary Approach to Human Action Control. *Trends in Cognitive Sciences* 21:12, pp. 940–949. ISSN: 1364-6613. doi: [10.1016/j.tics.2017.09.009](https://doi.org/10.1016/j.tics.2017.09.009).
- Horridge, M. (1, 2011). *A Practical Guide To Building OWL Ontologies Using Protégé 4 and CO-ODE Tools Edition 1.3*.
- Horvath, J. C. and G. M. Donoghue (2016). A Bridge Too Far – Revisited: Reframing Bruer’s Neuroeducation Argument for Modern Science of Learning Practitioners. *Frontiers in Psychology* 7. ISSN: 1664-1078.
- Howard, T. J., S. J. Culley, and E. Dekoninck (1, 2008). Describing the creative design process by the integration of engineering design and cognitive psychology literature. *Design Studies* 29:2, pp. 160–180. ISSN: 0142-694X. doi: [10.1016/j.destud.2008.01.001](https://doi.org/10.1016/j.destud.2008.01.001).
- Huang, W. and C.-K. Looi (2, 2021). A critical review of literature on “unplugged” pedagogies in K-12 computer science and computational thinking education. *Computer Science Education* 31:1, pp. 1–29. ISSN: 0899-3408. doi: [10.1080/08993408.2020.1789411](https://doi.org/10.1080/08993408.2020.1789411).

References

- Jacob, S. R. and M. Warschauer (25, 2018). Computational Thinking and Literacy. 1:1, p. 1. issn: 2574-108X. doi: [10.26716/jcsi.2018.01.1.1](https://doi.org/10.26716/jcsi.2018.01.1.1).
- Jaeger, H. (22, 2017). *Controlling Recurrent Neural Networks by Conceptors*. doi: [10.48550/arXiv.1403.3369](https://doi.org/10.48550/arXiv.1403.3369).
- Jamone, L., E. Ugur, A. Cangelosi, L. Fadiga, A. Bernardino, J. Piater, and J. Santos-Victor (2016). Affordances in Psychology, Neuroscience, and Robotics: A Survey. *IEEE Transactions on Cognitive and Developmental Systems* 10:1, pp. 4–25. issn: 2379-8920, 2379-8939. doi: [10.1109/TCDS.2016.2594134](https://doi.org/10.1109/TCDS.2016.2594134).
- Janowicz, K. and M. Compton (7, 2010). “The Stimulus-Sensor-Observation Ontology Design Pattern and its Integration into the Semantic Sensor Network Ontology”. In: *Proceedings of the 3rd International Conference on Semantic Sensor Networks - Volume 668*. SSN’10. CEUR-WS.org, Aachen, DEU, pp. 64–78.
- Jeffries, K. (1, 2012). Amabile’s consensual assessment technique: Why has it not been used more in design creativity research? *ICDC 2012 - 2nd International Conference on Design Creativity, Proceedings 1*, pp. 211–220.
- Jiménez, A., B. Elizalde, and B. Raj (2018). “Sound event classification using ontology-based neural networks”. In: *Proceedings of the Annual Conference on Neural Information Processing Systems*, p. 9.
- Johnson-Laird, P. N., P. Legrenzi, V. Girotto, M. S. Legrenzi, and J. P. Caverni (1999). Naive probability: a mental model theory of extensional reasoning. *Psychological Review* 106:1, pp. 62–88. issn: 0033-295X. doi: [10.1037/0033-295x.106.1.62](https://doi.org/10.1037/0033-295x.106.1.62).
- Johnson-Laird, P. and M. Ragni (2019). Possibilities as the foundation of reasoning. *Cognition* 193, p. 103950. issn: 00100277. doi: [10.1016/j.cognition.2019.04.019](https://doi.org/10.1016/j.cognition.2019.04.019).
- Johnson-Laird, P. N. (1989). Analogy and the exercise of creativity. In: *Similarity and analogical reasoning*. Cambridge University Press, New York, NY, US, pp. 313–331. isbn: 978-0-521-36295-5 978-0-521-38935-8. doi: [10.1017/CB09780511529863.015](https://doi.org/10.1017/CB09780511529863.015).
- Jung, R., B. Mead, J. Carrasco, and R. Flores (2013). The structure of creative cognition in the human brain. *Frontiers in Human Neuroscience* 7, p. 330. issn: 1662-5161. doi: [10.3389/fnhum.2013.00330](https://doi.org/10.3389/fnhum.2013.00330).
- Jung-Beeman, M., E. M. Bowden, J. Haberman, J. L. Frymiare, S. Arambel-Liu, R. Greenblatt, P. J. Reber, and J. Kounios (13, 2004). Neural Activity When People Solve Verbal Problems with Insight. *PLOS Biology* 2:4, e97. issn: 1545-7885. doi: [10.1371/journal.pbio.0020097](https://doi.org/10.1371/journal.pbio.0020097).
- Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.
- Kajić, I., J. Gosmann, T. C. Stewart, T. Wennekers, and C. Eliasmith (2017). A Spiking Neuron Model of Word Associations for the Remote Associates Test. *Frontiers in Psychology* 8, p. 99. issn: 1664-1078. doi: [10.3389/fpsyg.2017.00099](https://doi.org/10.3389/fpsyg.2017.00099).
- Kalmpourtzis, G. and M. Romero (2020). “Artifactual Affordances in Playful Robotics”. In: *International Conference on Games and Learning Alliance*. Ed. by I. Marfisi-Schottman, F. Bellotti, L. Hamon, and R. Klemke. Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 316–325. isbn: 978-3-030-63464-3. doi: [10.1007/978-3-030-63464-3_30](https://doi.org/10.1007/978-3-030-63464-3_30).

- Kang, S. H. K. (1, 2016). Spaced Repetition Promotes Efficient and Effective Learning: Policy Implications for Instruction. *Policy Insights from the Behavioral and Brain Sciences* 3:1, pp. 12–19. ISSN: 2372-7322. DOI: [10.1177/2372732215624708](https://doi.org/10.1177/2372732215624708).
- Kaplan, A. and H. Flum (1, 2010). Achievement goal orientations and identity formation styles. *Educational Research Review* 5:1, pp. 50–67. ISSN: 1747-938X. DOI: [10.1016/j.edurev.2009.06.004](https://doi.org/10.1016/j.edurev.2009.06.004).
- Kaufman, J. C. and R. A. Beghetto (2009). Beyond big and little: The four C model of creativity. *Review of general psychology* 13:1, p. 1.
- Keefe, L. A. and M. J. Landau (2016). Metaphor and analogy in everyday problem solving. *WIREs Cognitive Science* 7:6, pp. 394–405. ISSN: 1939-5086. DOI: [10.1002/wcs.1407](https://doi.org/10.1002/wcs.1407).
- Kelley, P. and T. Watson (2013). Making long-term memories in minutes: a spaced learning pattern from memory research in education. *Frontiers in Human Neuroscience* 7. ISSN: 1662-5161.
- Kenett, Y. N. and M. Faust (1, 2019). A Semantic Network Cartography of the Creative Mind. *Trends in Cognitive Sciences* 23:4, pp. 271–274. ISSN: 1364-6613. DOI: [10.1016/j.tics.2019.01.007](https://doi.org/10.1016/j.tics.2019.01.007).
- Khalil, R. and A. A. Moustafa (1, 2022). A neurocomputational model of creative processes. *Neuroscience & Biobehavioral Reviews* 137, p. 104656. ISSN: 0149-7634. DOI: [10.1016/j.neubiorev.2022.104656](https://doi.org/10.1016/j.neubiorev.2022.104656).
- Khemlani, S. S., A. K. Barbey, and P. N. Johnson-Laird (28, 2014). Causal reasoning with mental models. *Frontiers in Human Neuroscience* 8. ISSN: 1662-5161. DOI: [10.3389/fnhum.2014.00849](https://doi.org/10.3389/fnhum.2014.00849).
- Khemlani, S. S., M. Lotstein, and P. N. Johnson-Laird (2015). Naive probability: Model-based estimates of unique events. *Cognitive Science* 39:6, pp. 1216–1258. ISSN: 1551-6709. DOI: [10.1111/cogs.12193](https://doi.org/10.1111/cogs.12193).
- Kim, K. H. (1, 2006). Can We Trust Creativity Tests? A Review of the Torrance Tests of Creative Thinking (TTCT). *Creativity Research Journal* 18:1, pp. 3–14. ISSN: 1040-0419, 1532-6934. DOI: [10.1207/s15326934crj1801_2](https://doi.org/10.1207/s15326934crj1801_2).
- Kitcher, P. (1988). Marr’s Computational Theory of Vision. *Philosophy of Science* 55:1, pp. 1–24. ISSN: 0031-8248.
- Knorr, M., P. Hitzler, and F. Maier (2012). “Reconciling OWL and Non-monotonic Rules for the Semantic Web”. In: *ECAI 2012 - 20th European Conference on Artificial Intelligence*. Montpellier, France. DOI: <http://dx.doi.org/10.3233/978-1-61499-098-7-474>.
- Kohler, L. and M. Romero (2023). “Assessment of Divergent Thinking in a Social and Modular Robotics Task: An Edit Distance Approach at the Configuration Level”. In: *Learning and Collaboration Technologies*. Ed. by P. Zaphiris and A. Ioannou. Lecture Notes in Computer Science. Springer Nature Switzerland, Cham, pp. 336–349. ISBN: 978-3-031-34550-0. DOI: [10.1007/978-3-031-34550-0_24](https://doi.org/10.1007/978-3-031-34550-0_24).
- Köhler, L., M. Romero, and P. Reynaud-Bouret (15, 2021). “Graph models for learning”. Workshop. Understanding creative problem solving through a computational learning sciences perspective. ANR CreaMaker workshop. INSPE Nice, France.

References

- Komer, B., T. C. Stewart, A. R. Voelker, and C. Eliasmith (2019). "A neural representation of continuous space using fractional binding". In: *41st Annual Meeting of the Cognitive Science Society*. Cognitive Science Society, Montreal, Canada, p. 6.
- Kounios, J. and M. Beeman (3, 2014). The Cognitive Neuroscience of Insight. *Annual review of psychology* 65, pp. 71–93. DOI: [10.1146/annurev-psych-010213-115154](https://doi.org/10.1146/annurev-psych-010213-115154).
- Kupers, E., A. Lehmann-Wermser, G. McPherson, and P. van Geert (1, 2019). Children's Creativity: A Theoretical Framework and Systematic Review. *Review of Educational Research* 89:1, pp. 93–124. ISSN: 0034-6543. DOI: [10.3102/0034654318815707](https://doi.org/10.3102/0034654318815707).
- Laird, J., P. Rosenbloom, and A. Newell (1986). Soar—A General Problem-Solving Architecture. In: *Universal Subgoaling and Chunking: The Automatic Generation and Learning of Goal Hierarchies*. Ed. by J. Laird, P. Rosenbloom, and A. Newell. The Kluwer International Series in Engineering and Computer Science. Springer US, Boston, MA, pp. 286–288. ISBN: 978-1-4613-2277-1. DOI: [10.1007/978-1-4613-2277-1_15](https://doi.org/10.1007/978-1-4613-2277-1_15).
- Laird, J. E., C. B. Congdon, M. Assanie, N. Derbinsky, and J. Xu (2, 2017). "The Soar User's Manual Version 9.6.0".
- Lakkaraju, S. K. and Y. Zhang (11, 2000). "Rule Based Abduction". In: *Proceedings of the 12th International Symposium on Foundations of Intelligent Systems*. ISMIS '00. Springer-Verlag, Berlin, Heidelberg, pp. 525–533. ISBN: 978-3-540-41094-2. DOI: https://doi.org/10.1007/3-540-39963-1_55.
- Lallement, Y., M. Hilario, and F. Alexandre (1995). "Neurosymbolic Integration: Cognitive Grounds and Computational Strategies". In: *Proceedings World Conference on the Fundamentals of Artificial Intelligence*, p. 12.
- Langdon, C., M. Genkin, and T. A. Engel (2023). A unifying perspective on neural manifolds and circuits for cognition. *Nature Reviews Neuroscience* 24:6, pp. 363–377. ISSN: 1471-0048. DOI: [10.1038/s41583-023-00693-x](https://doi.org/10.1038/s41583-023-00693-x).
- Le Meudec, J., A. Bezerianos, C. Mercier, and A. Prouzeau (2024). "Learning with immersive technologies : a VR task to study collaborative learning". In: *IHM'24 - 35e Conférence Internationale Francophone sur l'Interaction Humain-Machine*. Vol. IHM'24 : Actes étendus de la 35ème conférence Francophone sur l'Interaction Humain-Machine. AFIHM and Sorbonne Université, Paris, France.
- Lefebvre, L. (6, 2013). Brains, innovations, tools and cultural transmission in birds, non-human primates, and fossil hominins. *Frontiers in Human Neuroscience* 7, p. 245. ISSN: 1662-5161. DOI: [10.3389/fnhum.2013.00245](https://doi.org/10.3389/fnhum.2013.00245).
- Leroy, A., M. Romero, and L. Cassone (31, 2021). Interactivity and materiality matter in creativity: educational robotics for the assessment of divergent thinking. *Interactive Learning Environments* 0:0, pp. 1–12. ISSN: 1049-4820. DOI: [10.1080/10494820.2021.1875005](https://doi.org/10.1080/10494820.2021.1875005).
- Lespiau, F. and A. Tricot (1, 2022). Primary vs. secondary knowledge contents in reasoning: Motivated and efficient vs. overburdened. *Acta Psychologica* 227, p. 103610. ISSN: 0001-6918. DOI: [10.1016/j.actpsy.2022.103610](https://doi.org/10.1016/j.actpsy.2022.103610).
- Levesque, H. J. (1, 1986). Knowledge Representation and Reasoning. *Annual Review of Computer Science* 1:1, pp. 255–287. ISSN: 8756-7016. DOI: [10.1146/annurev.cs.01.060186.001351](https://doi.org/10.1146/annurev.cs.01.060186.001351).

- Levy, D. J. and P. W. Glimcher (2012). The root of all value: a neural common currency for choice. *Current opinion in neurobiology* 22:6, pp. 1027–1038. ISSN: 0959-4388. DOI: [10.1016/j.conb.2012.06.001](https://doi.org/10.1016/j.conb.2012.06.001).
- Levy, S. D. and R. Gayler (2008). “Vector Symbolic Architectures: A New Building Material for Artificial General Intelligence”. In: *Frontiers in Artificial Intelligence and Applications*, p. 6.
- Lieto, A., F. Perrone, G. L. Pozzato, and E. Chiodino (1, 2019). Beyond subgoalting: A dynamic knowledge generation framework for creative problem solving in cognitive architectures. *Cognitive Systems Research* 58, pp. 305–316. ISSN: 1389-0417. DOI: [10.1016/j.cogsys.2019.08.005](https://doi.org/10.1016/j.cogsys.2019.08.005).
- Lieto, A. and G. L. Pozzato (2, 2020). A description logic framework for commonsense conceptual combination integrating typicality, probabilities and cognitive heuristics. *Journal of Experimental & Theoretical Artificial Intelligence* 32:5, pp. 769–804. ISSN: 0952-813X. DOI: [10.1080/0952813X.2019.1672799](https://doi.org/10.1080/0952813X.2019.1672799).
- Liu, D., K. Jiang, C. E. Shalley, S. Keem, and J. Zhou (1, 2016). Motivational mechanisms of employee creativity: A meta-analytic examination and theoretical extension of the creativity literature. *Organizational Behavior and Human Decision Processes* 137, pp. 236–263. ISSN: 0749-5978. DOI: [10.1016/j.obhdp.2016.08.001](https://doi.org/10.1016/j.obhdp.2016.08.001).
- Lodge, J. M., S. S. J. Alhadad, M. J. Lewis, and D. Gašević (1, 2017). Inferring Learning from Big Data: The Importance of a Transdisciplinary and Multidimensional Approach. *Technology, Knowledge and Learning* 22:3, pp. 385–400. ISSN: 2211-1670. DOI: [10.1007/s10758-017-9330-3](https://doi.org/10.1007/s10758-017-9330-3).
- Lodi, M. (3, 2020). “Introducing Computational Thinking in K-12 Education: Historical, Epistemological, Pedagogical, Cognitive, and Affective Aspects”. PhD thesis. Dipartimento di Informatica - Scienza e Ingegneria, Alma Mater Studiorum - Università di Bologna.
- Lopez-Persem, A., S. Moreno-Rodriguez, M. Ovando-Tellez, T. Bieth, S. Guiet, J. Brochard, and E. Volle (14, 2023). How Subjective Idea Valuation Energizes and Guides Creative Idea Generation. *The American psychologist*. DOI: [10.1037/amp0001165](https://doi.org/10.1037/amp0001165).
- Lubart, T. I. and R. J. Sternberg (1995). An Investment Approach to Creativity: Theory and Data. In: *The Creative Cognition Approach*. The MIT Press, Cambridge, pp. 269–302.
- Lubart, T. I. (28, 1998). Creativity Across Cultures. In: *The Handbook of Creativity*. 1st ed. Cambridge University Press, pp. 339–350. ISBN: 978-0-511-80791-6. DOI: [10.1017/CB09780511807916.019](https://doi.org/10.1017/CB09780511807916.019).
- Ma, Z., Y. Zhuang, P. Weng, D. Li, K. Shao, W. Liu, H. H. Zhuo, and J. Hao (28, 2020). “Interpretable Reinforcement Learning With Neural Symbolic Logic”. In: *ICLR 2021 Ninth International Conference on Learning Representations*.
- Mandler, G. (2002). Origins of the cognitive (r)evolution. *Journal of the History of the Behavioral Sciences* 38:4, pp. 339–353. ISSN: 1520-6696. DOI: [10.1002/jhbs.10066](https://doi.org/10.1002/jhbs.10066).
- (2011). From Association to Organization. *Current Directions in Psychological Science* 20:4, pp. 232–235. ISSN: 0963-7214, 1467-8721. DOI: [10.1177/09637214111414656](https://doi.org/10.1177/09637214111414656).

References

- Mandler, J. (1, 2010). The spatial foundations of the conceptual system. *Language and Cognition* 2. DOI: [10.1515/LANGCOG.2010.002](https://doi.org/10.1515/LANGCOG.2010.002).
- Mandler, J.M. and G. Mandler (1964). *Thinking: from association to Gestalt*. New York, Wiley. 326 pp. ISBN: 978-0-471-56696-0.
- Mangan, J. (1978). Piaget's Theory and Cultural Differences: The Case for Value-Based Modes of Cognition. *Human Development* 21:3, pp. 170–189. ISSN: 0018-716X.
- Mansfield, R. S., T. V. Busse, and E. J. Krepelka (1, 1978). The Effectiveness of Creativity Training. *Review of Educational Research* 48:4, pp. 517–536. ISSN: 0034-6543. DOI: [10.3102/00346543048004517](https://doi.org/10.3102/00346543048004517).
- Martin, D., A. Serrano, A. W. Bergman, G. Wetzstein, and B. Masia (2022). ScanGAN360: A Generative Model of Realistic Scanpaths for 360° Images. *IEEE Transactions on Visualization and Computer Graphics* 28:5, pp. 2003–2013. ISSN: 1941-0506. DOI: [10.1109/TVCG.2022.3150502](https://doi.org/10.1109/TVCG.2022.3150502).
- Masolo, C., S. Borgo, A. Gangemi, N. Guarino, and A. Oltramari (2003). *Wonderweb deliverable d18: Ontology library*. Technical report IST-2001-33052 del-18. Trento, Italy: Laboratory For Applied Ontology - ISTC-CNR.
- Mathisen, G. E., S. Einarsen, K. Jørstad, and K. S. Brønnick (2004). Climate for work group creativity and innovation: Norwegian validation of the team climate inventory (TCI). *Scandinavian Journal of Psychology* 45:5, pp. 383–392. ISSN: 1467-9450. DOI: [10.1111/j.1467-9450.2004.00420.x](https://doi.org/10.1111/j.1467-9450.2004.00420.x).
- Mayseless, N., A. Eran, and S. G. Shamay-Tsoory (2015). Generating original ideas: The neural underpinning of originality. *NeuroImage* 116, pp. 232–239. ISSN: 10538119. DOI: [10.1016/j.neuroimage.2015.05.030](https://doi.org/10.1016/j.neuroimage.2015.05.030).
- McClelland, J. L. and T. T. Rogers (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience* 4:4, pp. 310–322. ISSN: 1471-003X, 1471-0048. DOI: [10.1038/nrn1076](https://doi.org/10.1038/nrn1076).
- McCulloch, W. S. and W. Pitts (1, 1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics* 5:4, pp. 115–133. ISSN: 1522-9602. DOI: [10.1007/BF02478259](https://doi.org/10.1007/BF02478259).
- McNerney, T. S. (1, 2004). From turtles to Tangible Programming Bricks: explorations in physical language design. *Personal and Ubiquitous Computing* 8:5, pp. 326–337. ISSN: 1617-4917. DOI: [10.1007/s00779-004-0295-6](https://doi.org/10.1007/s00779-004-0295-6).
- Mednick, S. A. (1968). The Remote Associates Test*. *The Journal of Creative Behavior* 2:3, pp. 213–214. ISSN: 2162-6057. DOI: [10.1002/j.2162-6057.1968.tb00104.x](https://doi.org/10.1002/j.2162-6057.1968.tb00104.x).
- Mekern, V., B. Hommel, and Z. Sjoerds (1, 2019). Computational models of creativity: a review of single-process and multi-process recent approaches to demystify creative cognition. *Current Opinion in Behavioral Sciences*. *Creativity* 27, pp. 47–54. ISSN: 2352-1546. DOI: [10.1016/j.cobeha.2018.09.008](https://doi.org/10.1016/j.cobeha.2018.09.008).
- Mercier, C. (2023). An Ontology to Formalize a Creative Problem Solving Activity. *IEEE Transactions on Cognitive and Developmental Systems* 15:4, pp. 1891–1904. ISSN: 2379-8939. DOI: [10.1109/TCDS.2022.3210234](https://doi.org/10.1109/TCDS.2022.3210234).
- Mercier, C., L. Roux, M. Romero, F. Alexandre, and T. Vieville (23, 2021). “Formalizing Problem Solving in Computational Thinking : an Ontology approach”. In: *2021 IEEE*

- International Conference on Development and Learning (ICDL)*. IEEE, Beijing, China / Virtual, pp. 1–8. ISBN: 978-1-72816-242-3. DOI: [10.1109/ICDL49984.2021.9515660](https://doi.org/10.1109/ICDL49984.2021.9515660).
- Mercier, C., F. Alexandre, and T. Viéville (2021). “Reinforcement Symbolic Learning”. In: *Artificial Neural Networks and Machine Learning – ICANN 2021*. Ed. by I. Farkaš, P. Masulli, S. Otte, and S. Wermter. Lecture Notes in Computer Science. Springer International Publishing, Bratislava, Slovakia / Virtual, pp. 608–612. ISBN: 978-3-030-86380-7. DOI: [10.1007/978-3-030-86380-7_49](https://doi.org/10.1007/978-3-030-86380-7_49).
- (31, 2022). “Ontology as manifold: towards symbolic and numerical artificial embedding”. Webinar. Workshop on VSA and hyperdimensional computing - Winter 2022 session, Luleå University and UC Berkeley. Lulea, Sweden / Virtual.
- Mercier, C., H. Chateau-Laurent, F. Alexandre, and T. Viéville (11, 2021). “Ontology as neuronal-space manifold: towards symbolic and numerical artificial embedding”. In: *KRHCAI-21@KR2021*. Hanoi, Vietnam / Virtual.
- Mercier, C., M. Romero, F. Alexandre, and T. Viéville (2, 2022). “Apprentissage par renforcement sur des connaissances symboliques structurées : modéliser la résolution créative de problèmes”. Poster. CJC-SCo 2022 - Colloque des jeunes chercheur-se-s en sciences cognitives. ENS, Paris.
- Mercier, C. and T. Vieville (2023). Where is Aristotle in our brain? On biologically plausible reasoning embedded in neuronal computation. *Cognitive Neurodynamics* 0.
- Metcalfe, J. and D. Wiebe (1, 1987). Intuition in insight and noninsight problem solving. *Memory & Cognition* 15:3, pp. 238–246. ISSN: 1532-5946. DOI: [10.3758/BF03197722](https://doi.org/10.3758/BF03197722).
- Miller, K. J., E. A. Ludvig, G. Pezzulo, and A. Shenhav (1, 2018). Realigning Models of Habitual and Goal-Directed Decision-Making. In: *Goal-Directed Decision Making*. Ed. by R. Morris, A. Bornstein, and A. Shenhav. Academic Press, pp. 407–428. ISBN: 978-0-12-812098-9. DOI: [10.1016/B978-0-12-812098-9.00018-8](https://doi.org/10.1016/B978-0-12-812098-9.00018-8).
- Mnih, V., K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis (26, 2015). Human-level control through deep reinforcement learning. *Nature* 518:7540, pp. 529–533. ISSN: 0028-0836, 1476-4687. DOI: [10.1038/nature14236](https://doi.org/10.1038/nature14236).
- Moerland, T., J. Broekens, and C. Jonker (1, 2018). Emotion in Reinforcement Learning Agents and Robots: A Survey. *Machine Learning* 107. DOI: [10.1007/s10994-017-5666-0](https://doi.org/10.1007/s10994-017-5666-0).
- Molinaro, G. and A. G. E. Collins (9, 2023). A goal-centric outlook on learning. *Trends in Cognitive Sciences*. ISSN: 1364-6613. DOI: [10.1016/j.tics.2023.08.011](https://doi.org/10.1016/j.tics.2023.08.011).
- Murray, M. D. (25, 2023). *Generative AI Art: Copyright Infringement and Fair Use*. Rochester, NY. DOI: [10.2139/ssrn.4483539](https://doi.org/10.2139/ssrn.4483539).
- Musial, M. and A. Tricot (2020). *Précis d'ingénierie pédagogique*. 1st. PEDAGOGIES EN DEVELOPPEMENT. De Boeck Supérieur. 352 pp. ISBN: 978-2-8073-2419-0.
- Mygind, E. and M. Bølling (2022). Pupils’ Well-Being, Mental and Social Health. In: *High-Quality Outdoor Learning: Evidence-based Education Outside the Classroom for Children, Teachers and Society*. Ed. by R. Jucker and J. von Au. Springer International Publishing, Cham, pp. 153–168. ISBN: 978-3-031-04108-2. DOI: [10.1007/978-3-031-04108-2_8](https://doi.org/10.1007/978-3-031-04108-2_8).

References

- Nemiro, J., C. Larriva, and M. Jawaharlal (2017). Developing Creative Behavior in Elementary School Students with Robotics. *The Journal of Creative Behavior* 51:1, pp. 70–90. ISSN: 2162-6057. DOI: [10.1002/jocb.87](https://doi.org/10.1002/jocb.87).
- Ness, D. and S. J. Farenga (2007). *Knowledge under construction: The importance of play in developing children's spatial and geometric thinking*. Knowledge under construction: The importance of play in developing children's spatial and geometric thinking. Rowman & Littlefield, Lanham, MD, US. xxiii, 257. ISBN: 978-0-7425-4788-9.
- Newell, A. (1981). Reasoning, Problem Solving, and Decision Processes: The Problem Space as a Fundamental Category. In: *Attention and Performance VIII*. Psychology Press, p. 26. ISBN: 978-1-315-80296-1.
- Newell, A. and H. A. Simon (1972). *Human problem solving*. Prentice-Hall, Englewood Cliffs, N.J.
- Ni, M., L. Yang, J. Chen, H. Chen, and X. Li (31, 2014). How to Improve Divergent Thinking Capability by Information Technology and Extenics. *Procedia Computer Science* 31, pp. 158–164. DOI: [10.1016/j.procs.2014.05.256](https://doi.org/10.1016/j.procs.2014.05.256).
- Nielsen, J. A., B. A. Zielinski, M. A. Ferguson, J. E. Lainhart, and J. S. Anderson (2013). An evaluation of the left-brain vs. right-brain hypothesis with resting state functional connectivity magnetic resonance imaging. *PLoS One* 8:8, e71275. ISSN: 1932-6203. DOI: [10.1371/journal.pone.0071275](https://doi.org/10.1371/journal.pone.0071275).
- Nijstad, B., C. De Dreu, E. Rietzschel, and M. Baas (1, 2010). The Dual Pathway to Creativity Model: Creative Ideation as a Function of Flexibility and Persistence. *European Review of Social Psychology* 21, pp. 34–77. DOI: [10.1080/10463281003765323](https://doi.org/10.1080/10463281003765323).
- Nonaka, I. (1994). A Dynamic Theory of Organizational Knowledge Creation. *Organization Science* 5:1, pp. 14–37. ISSN: 1047-7039, 1526-5455. DOI: [10.1287/orsc.5.1.14](https://doi.org/10.1287/orsc.5.1.14).
- Norman, D. (5, 2013). *Design of Everyday Things*. Revised and expanded edition. Hachette Book Group USA, Cambridge (Mass.) 347 pp. ISBN: 978-0-465-05065-9.
- Nouira, A., L. Cheniti-Belcadhi, and R. Braham (2019). An ontology-based framework of assessment analytics for massive learning. *Computer Applications in Engineering Education* 27:6, pp. 1343–1360. ISSN: 1099-0542. DOI: <https://doi.org/10.1002/cae.22155>.
- Noy, N. and D. L. McGuinness (2001). *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford University.
- O'Reilly, J. X., S. Jbabdi, and T. E. J. Behrens (2012). How can a Bayesian approach inform neuroscience? *The European Journal of Neuroscience* 35:7, pp. 1169–1179. ISSN: 1460-9568. DOI: [10.1111/j.1460-9568.2012.08010.x](https://doi.org/10.1111/j.1460-9568.2012.08010.x).
- O'Reilly, R. C. (2010). The What and How of Prefrontal Cortical Organization. *Trends in Neurosciences* 33:8, pp. 355–361. ISSN: 0166-2236. DOI: [10.1016/j.tins.2010.05.002](https://doi.org/10.1016/j.tins.2010.05.002).
- O'Reilly, R. C., R. Bhattacharyya, M. D. Howard, and N. Ketz (2014). Complementary Learning Systems. *Cognitive Science* 38:6, pp. 1229–1248. ISSN: 1551-6709. DOI: [10.1111/j.1551-6709.2011.01214.x](https://doi.org/10.1111/j.1551-6709.2011.01214.x).
- Oaksford, M. and N. Chater (2009). Précis of Bayesian Rationality: The Probabilistic Approach to Human Reasoning. *Behavioral and Brain Sciences* 32:1, pp. 69–84. ISSN: 0140-525X, 1469-1825. DOI: [10.1017/S0140525X09000284](https://doi.org/10.1017/S0140525X09000284).

- OECD (2012). PISA 2012 Problem-Solving Framework. In: *PISA 2012 assessment and analytical framework*. PISA. Organisation for Economic Co-operation and Development.
- (21, 2017). *PISA 2015 Results: Collaborative Problem Solving*. Vol. V. PISA. Organisation for Economic Co-operation and Development. ISBN: 978-92-64-28552-1. DOI: [10.1787/9789264285521-en](https://doi.org/10.1787/9789264285521-en).
- (26, 2019). PISA 2018 Well-being Framework. In: *PISA 2018 Assessment and Analytical Framework*. PISA. Organisation for Economic Co-operation and Development, Paris, pp. 257–298. DOI: [10.1787/38a34353-en](https://doi.org/10.1787/38a34353-en).
- (31, 2023). PISA 2022 Creative Thinking Framework. In: *PISA 2022 Assessment and Analytical Framework*. PISA. Organisation for Economic Co-operation and Development, Paris. DOI: [10.1787/471ae22e-en](https://doi.org/10.1787/471ae22e-en).
- Olszewski, W. and A. Sandroni (2011). Falsifiability. *American Economic Review* 101:2, pp. 788–818. ISSN: 0002-8282. DOI: [10.1257/aer.101.2.788](https://doi.org/10.1257/aer.101.2.788).
- Olteteanu, A.-M. (2020). *Cognition and the Creative Machine: Cognitive AI for Creative Problem Solving*. Springer International Publishing, Cham. ISBN: 978-3-030-30322-8. DOI: [10.1007/978-3-030-30322-8](https://doi.org/10.1007/978-3-030-30322-8).
- Ouangraoua, A. and P. Ferraro (2009). A Constrained Edit Distance Algorithm Between Semi-ordered Trees. *Theoretical Computer Science* 410:8, pp. 837–846. DOI: [10.1016/j.tcs.2008.11.022](https://doi.org/10.1016/j.tcs.2008.11.022).
- Palau, A., C. Mercier, and M. Romero (2, 2023). “Représenter l’apprentissage humain dans des problèmes ouverts”. Presentation. Dataquitaine 2023 - IA, Recherche Opérationnelle & Data Science. Kedge Business School Bordeaux, France.
- Paschos, V. T. (9, 2007). *An overview on polynomial approximation of NP-hard problems*.
- Petrucci, G., C. Ghidini, and M. Rospocher (2016). “Ontology Learning in the Deep”. In: *Knowledge Engineering and Knowledge Management*. Ed. by E. Blomqvist, P. Ciancarini, F. Poggi, and F. Vitali. Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 480–495. ISBN: 978-3-319-49004-5. DOI: [10.1007/978-3-319-49004-5_31](https://doi.org/10.1007/978-3-319-49004-5_31).
- Phan, N., D. Dou, H. Wang, D. Kil, and B. Piniewski (1, 2017). Ontology-based deep learning for human behavior prediction with explanations in health social networks. *Information Sciences* 384, pp. 298–313. ISSN: 0020-0255. DOI: [10.1016/j.ins.2016.08.038](https://doi.org/10.1016/j.ins.2016.08.038).
- Piaget, J. (1997). *Le Langage et la Pensée chez l’enfant*. 10e éd. édition. Delachaux & Niestlé, Neuchâtel. 213 pp. ISBN: 978-2-603-00707-5.
- Pintrich, P. R. (2000). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology* 92:3, pp. 544–555. ISSN: 1939-2176(Electronic), 0022-0663(Print). DOI: [10.1037/0022-0663.92.3.544](https://doi.org/10.1037/0022-0663.92.3.544).
- Plate, T. (1995). Holographic reduced representations. *IEEE Transactions on Neural Networks* 6:3, pp. 623–641. ISSN: 10459227. DOI: [10.1109/72.377968](https://doi.org/10.1109/72.377968).
- Poldrack, R., A. Kittur, D. Kalar, C. Seppa, Y. Gil, D. Parker, F. Sabb, and R. Bilder (6, 2011). The Cognitive Atlas: Toward a Knowledge Foundation for Cognitive Neuroscience. *Frontiers in neuroinformatics* 5, p. 17. DOI: [10.3389/fninf.2011.00017](https://doi.org/10.3389/fninf.2011.00017).

References

- Polya, G. (1945). *How to Solve It: A New Aspect of Mathematical Method*. In collab. with J. H. Conway. Expanded Princeton Science Library ed (2004). Princeton Science Library. Princeton University Press, Princeton [N.J.] 253 pp. ISBN: 978-0-691-11966-3.
- Poortvliet, P.M. (2016). Mastery Goals. In: *Encyclopedia of Personality and Individual Differences*. Ed. by V. Zeigler-Hill and T. K. Shackelford. Springer International Publishing, Cham, pp. 1–4. ISBN: 978-3-319-28099-8. DOI: [10.1007/978-3-319-28099-8_533-1](https://doi.org/10.1007/978-3-319-28099-8_533-1).
- Pulvermüller, F. (1, 2013). How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences* 17:9, pp. 458–470. ISSN: 1364-6613. DOI: [10.1016/j.tics.2013.06.004](https://doi.org/10.1016/j.tics.2013.06.004).
- Purves, D., G. J. Augustine, D. Fitzpatrick, L. C. Katz, A.-S. LaMantia, J. O. McNamara, and S. M. Williams (2001). *The Interplay of Emotion and Reason. Neuroscience. 2nd edition*.
- Quillien, J.-B. (2019). “The Roles of Convergent, Divergent Thinking, and Contextual Focus during Scientific Reasoning: Birth of the “Z” Model”. PhD thesis. University of Minnesota.
- Rabinovich, M., C. Bick, and P. Varona (11, 2023). Beyond neurons and spikes: cognon, the hierarchical dynamical unit of thought. *Cognitive Neurodynamics*. ISSN: 1871-4099. DOI: [10.1007/s11571-023-09987-3](https://doi.org/10.1007/s11571-023-09987-3).
- Raczaszek-Leonardi, J. and T.W. Deacon (2018). “Ungrounding symbols in language development: implications for modeling emergent symbolic communication in artificial systems”. In: *2018 Joint IEEE 8th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, pp. 232–237. DOI: [10.1109/DEVLRN.2018.8761016](https://doi.org/10.1109/DEVLRN.2018.8761016).
- Radel, R., K. Davranche, M. Fournier, and A. Dietrich (2015). The role of (dis)inhibition in creativity: decreased inhibition improves idea generation. *Cognition* 134, pp. 110–120. ISSN: 1873-7838. DOI: [10.1016/j.cognition.2014.09.001](https://doi.org/10.1016/j.cognition.2014.09.001).
- Radji, W., C. Léger, and L. Bardisbanian (23, 2023). *Early Empirical Results on Reinforcement Symbolic Learning*. Research Report RR-9509. Bordeaux, France: Inria & Labri, Univ. Bordeaux, ENSEIRB Bordeaux INP, p. 9.
- Ragni, M. and P.N. Johnson-Laird (2018). Reasoning about possibilities: human reasoning violates all normal modal logics. *CogSci*, p. 6.
- Raufaste, E., R. da Silva Neves, and C. Mariné (1, 2003). Testing the descriptive validity of possibility theory in human judgments of uncertainty. *Artificial Intelligence. Fuzzy Set and Possibility Theory-Based Methods in Artificial Intelligence* 148:1, pp. 197–218. ISSN: 0004-3702. DOI: [10.1016/S0004-3702\(03\)00021-3](https://doi.org/10.1016/S0004-3702(03)00021-3).
- Reingold, E. M. and C. A. Ray (2006). Implicit Cognition. In: *Encyclopedia of Cognitive Science*. John Wiley & Sons, Ltd. ISBN: 978-0-470-01886-6.
- Rhodes, M. (1961). An Analysis of Creativity. *The Phi Delta Kappan* 42:7, pp. 305–310. ISSN: 0031-7217.
- Riegel, R., A. Gray, F. Luus, N. Khan, N. Makondo, I. Y. Akhalwaya, H. Qian, R. Fagin, F. Barahona, U. Sharma, S. Ikbali, H. Karanam, S. Neelam, A. Likhyani, and S. Srivastava (23, 2020). *Logical Neural Networks*. DOI: [10.48550/arXiv.2006.13155](https://doi.org/10.48550/arXiv.2006.13155).

- Ritchie, G. (1, 2012). "A closer look at creativity as search". In: *Proceedings of the Third International Conference on Computational Creativity*.
- Rmus, M., S. D. McDougle, and A. G. Collins (1, 2021). The role of executive function in shaping reinforcement learning. *Current Opinion in Behavioral Sciences* 38, pp. 66–73. ISSN: 2352-1546. DOI: [10.1016/j.cobeha.2020.10.003](https://doi.org/10.1016/j.cobeha.2020.10.003).
- Rodrigues, C. T. (1, 2011). The Method of Scientific Discovery in Peirce's Philosophy: Deduction, Induction, and Abduction. *Logica Universalis* 5:1, pp. 127–164. ISSN: 1661-8300. DOI: [10.1007/s11787-011-0026-5](https://doi.org/10.1007/s11787-011-0026-5).
- Romero, M. (2017). "CreaCube, analyse de la résolution créative de problèmes par le biais d'une tâche de robotique modulaire". In: *Journées Nationales de la Recherche en Robotique, JNRR 2017*. Femto-st, p. 1.
- (2019). "Analyzing Cognitive Flexibility in Older Adults Through Playing with Robotic Cubes". In: *Human Aspects of IT for the Aged Population. Social Media, Games and Assistive Environments*. Ed. by J. Zhou and G. Salvendy. Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 545–553. ISBN: 978-3-030-22015-0. DOI: [10.1007/978-3-030-22015-0_42](https://doi.org/10.1007/978-3-030-22015-0_42).
- (28, 2022). *Creative intention is not enough ! The role of regulatory processes in creative problem solving (CreaComp, 2022)*. DOI: [10.13140/RG.2.2.28349.33761](https://doi.org/10.13140/RG.2.2.28349.33761).
- (29, 2023). "#CreaCube: a task, a platform, and a transdisciplinary community for the study of creative problem solving". Annual Meeting of the NeuroMod Institute, Antibes 28th – 29th June 2023. Antibes, France. DOI: [10.13140/RG.2.2.34692.50566](https://doi.org/10.13140/RG.2.2.34692.50566).
- Romero, M. and S. Barma (10, 2022). Analysing an Interactive Problem-Solving Task Through the Lens of Double Stimulation. *Canadian Journal of Learning and Technology* 48:1. ISSN: 1499-6685. DOI: [10.21432/cjlt28170](https://doi.org/10.21432/cjlt28170).
- Romero, M., D. David, and B. Lille (2019). CreaCube, a Playful Activity with Modular Robotics. In: *Games and Learning Alliance*. Ed. by M. Gentile, M. Allegra, and H. Söbke. Vol. 11385. Springer International Publishing, Cham, pp. 397–405. ISBN: 978-3-030-11548-7. DOI: [10.1007/978-3-030-11548-7_37](https://doi.org/10.1007/978-3-030-11548-7_37).
- Romero, M., L. DeBlois, and P. Abadjyan (8, 2018). CréaCube, comparaison de la résolution créative de problèmes, chez des enfants et des adultes, par le biais d'une tâche de robotique modulaire. *MathemaTICE* 61, Les nouvelles technologies pour l'enseignement des mathématiques.
- Romero, M., T. Laferrière, and T. M. Power (2016). The Move is On! From the Passive Multimedia Learner to the Engaged Co-creator. *eLearn Magazine, an ACM Publication*.
- Romero, M., A. Lepage, and B. Lille (2017). Computational thinking development through creative programming in higher education. *International Journal of Educational Technology in Higher Education* 14:1, pp. 1–15. ISSN: 2365-9440. DOI: [10.1186/s41239-017-0080-z](https://doi.org/10.1186/s41239-017-0080-z).
- Romero, M., A. Palaude, C. Mercier, T. Vieville, and F. Alexandre (2024). *Deciphering Cognitive Processes in Creative Problem Solving: A Computational Approach*.
- Romero, M., T. Vieville, and L. Heiser (2021). Analyse d'activités d'apprentissage méditatisées en robotique pédagogique. In: *Traité de méthodologie de la recherche en Sciences de l'Éducation et de la Formation*. B. Alberto and J. Thievenaz.

References

- Rong, X. (5, 2016). *word2vec Parameter Learning Explained*. DOI: [10.48550/arXiv.1411.2738](https://doi.org/10.48550/arXiv.1411.2738).
- Roussel, S., D. Joulia, A. Tricot, and J. Sweller (1, 2017). Learning subject content through a foreign language should not ignore human cognitive architecture: A cognitive load theory approach. *Learning and Instruction* 52, pp. 69–79. ISSN: 0959-4752. DOI: [10.1016/j.learninstruc.2017.04.007](https://doi.org/10.1016/j.learninstruc.2017.04.007).
- Roux, L., M. Romero, F. Alexandre, T. Viéville, and C. Mercier (2020). *Développement d'une ontologie pour l'analyse d'observables de l'apprenant dans le contexte d'une tâche avec des robots modulaires*. Research Report RR-9376. Inria, p. 48.
- Ruiz, C., H. Ren, K. Huang, and J. Leskovec (2, 2023). "High dimensional, tabular deep learning with an auxiliary knowledge graph". In: Thirty-seventh Conference on Neural Information Processing Systems.
- Rusawuk, A. L. (8, 2018). *Possibility and Necessity: An Introduction to Modality*. 1000-Word Philosophy: An Introductory Anthology.
- Ryan, R. M., J. J. Duineveld, S. I. Di Domenico, W. S. Ryan, B. A. Steward, and E. L. Bradshaw (2022). We know this much is (meta-analytically) true: A meta-review of meta-analytic findings evaluating self-determination theory. *Psychological Bulletin* 148:11, pp. 813–842. ISSN: 1939-1455. DOI: [10.1037/bul0000385](https://doi.org/10.1037/bul0000385).
- Sajjad, H. P., A. Docherty, and Y. Tyshetskiy (22, 2019). Efficient Representation Learning Using Random Walks for Dynamic Graphs. *arXiv:1901.01346 [cs, stat]*.
- Sander, E., H. Gros, K. Gvozdic, and C. Scheibling-Seve (2018). *Les neurosciences en éducation*. Mythes et réalités. Retz. 160 pp. ISBN: 978-2-7256-3583-5.
- Sannino, A. and Y. Engeström (2018). Cultural-historical activity theory: founding insights and new challenges. *Cultural-Historical Psychology* 14:3, pp. 43–56. ISSN: 1816-5435, 2224-8935. DOI: [10.17759/chp.2018140304](https://doi.org/10.17759/chp.2018140304).
- Sannino, A. and A. Laitinen (1, 2015). Double stimulation in the waiting experiment: Testing a Vygotskian model of the emergence of volitional action. *Learning, Culture and Social Interaction*. The emergence of transformative agency and double stimulation: Activity-based studies in the Vygotskian tradition 4, pp. 4–18. ISSN: 2210-6561. DOI: [10.1016/j.lcsi.2014.07.002](https://doi.org/10.1016/j.lcsi.2014.07.002).
- Sauerwald, T. and L. Zanetti (4, 2019). Random Walks on Dynamic Graphs: Mixing Times, Hitting Times, and Return Probabilities. *arXiv:1903.01342 [cs, math]*.
- Savic, M. (2016). Mathematical Problem-Solving via Wallas' Four Stages of Creativity: Implications for the Undergraduate Classroom. *The Mathematics Enthusiast* 13:3, p. 255. DOI: [10.54870/1551-3440.1377](https://doi.org/10.54870/1551-3440.1377).
- Schaeffer, R., M. Khona, and I. R. Fiete (16, 2022). "No Free Lunch from Deep Learning in Neuroscience: A Case Study through Models of the Entorhinal-Hippocampal Circuit". In: *Proceedings NeurIPS 2022*.
- Schatz, J. and J. Laird (2018). "An Architecture Approach to Modeling the Remote Associates Test". In: *Proceedings of the 16th international conference on cognitive modelling (iccm)*. P. 6.
- Schlegel, K., P. Neubert, and P. Protzel (11, 2020). A comparison of Vector Symbolic Architectures. *arXiv:2001.11797 [cs]*.

- Schlichting, M. L. and A. R. Preston (2015). Memory integration: neural mechanisms and implications for behavior. *Current Opinion in Behavioral Sciences* 1, pp. 1–8. ISSN: 2352-1546. DOI: [10.1016/j.cobeha.2014.07.005](https://doi.org/10.1016/j.cobeha.2014.07.005).
- Schraw, G., M. E. Dunkle, and L. D. Bendixen (1995). Cognitive processes in well-defined and ill-defined problem solving. *Applied Cognitive Psychology* 9:6, pp. 523–538. ISSN: 1099-0720. DOI: <https://doi.org/10.1002/acp.2350090605>.
- Seijts, G. H. and G. P. Latham (2005). Learning versus performance goals: When should each be used? *Academy of Management Perspectives* 19:1, pp. 124–131. ISSN: 1558-9080. DOI: [10.5465/ame.2005.15841964](https://doi.org/10.5465/ame.2005.15841964).
- Serafini, L. and A. d. Garcez (7, 2016). *Logic Tensor Networks: Deep Learning and Logical Reasoning from Data and Knowledge*. DOI: [10.48550/arXiv.1606.04422](https://doi.org/10.48550/arXiv.1606.04422).
- Shanahan, M. (1, 2000). An abductive event calculus planner. *The Journal of Logic Programming* 44:1, pp. 207–240. ISSN: 0743-1066. DOI: [10.1016/S0743-1066\(99\)00077-1](https://doi.org/10.1016/S0743-1066(99)00077-1).
- Shi, S., H. Chen, W. Ma, J. Mao, M. Zhang, and Y. Zhang (19, 2020). “Neural Logic Reasoning”. In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. CIKM '20. Association for Computing Machinery, New York, NY, USA, pp. 1365–1374. ISBN: 978-1-4503-6859-9. DOI: [10.1145/3340531.3411949](https://doi.org/10.1145/3340531.3411949).
- Singh, S., R. L. Lewis, A. G. Barto, and J. Sorg (2010). Intrinsically Motivated Reinforcement Learning: An Evolutionary Perspective. *IEEE Transactions on Autonomous Mental Development* 2:2, pp. 70–82. ISSN: 1943-0604, 1943-0612. DOI: [10.1109/TAMD.2010.2051031](https://doi.org/10.1109/TAMD.2010.2051031).
- Smith, L. (1, 1994). The development of modal understanding: Piaget’s possibility and necessity. *New Ideas in Psychology* 12:1, pp. 73–87. ISSN: 0732-118X. DOI: [10.1016/0732-118X\(94\)90059-0](https://doi.org/10.1016/0732-118X(94)90059-0).
- Smithson, M. (1, 1970). “Human Judgment And Imprecise Probabilities”.
- Spens, E. and N. Burgess (21, 2023). *A Generative Model of Memory Construction and Consolidation*. DOI: [10.1101/2023.01.19.524711](https://doi.org/10.1101/2023.01.19.524711).
- Stachenfeld, K. L., M. M. Botvinick, and S. J. Gershman (2017). The hippocampus as a predictive map. *Nature Neuroscience* 20:11, pp. 1643–1653. ISSN: 1546-1726. DOI: [10.1038/nn.4650](https://doi.org/10.1038/nn.4650).
- Stancin, K., P. Posic, and D. Jaksic (1, 2020). Ontologies in education – state of the art. *Education and Information Technologies* 25:6, pp. 5301–5320. ISSN: 1573-7608. DOI: [10.1007/s10639-020-10226-z](https://doi.org/10.1007/s10639-020-10226-z).
- Stanovich, K. E. and R. F. West (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences* 23:5, pp. 645–665. ISSN: 1469-1825. DOI: [10.1017/S0140525X00003435](https://doi.org/10.1017/S0140525X00003435).
- Stewart, T. and C. Eliasmith (2011). Neural Cognitive Modelling: A Biologically Constrained Spiking Neuron Model of the Tower of Hanoi Task. *Proceedings of the Annual Meeting of the Cognitive Science Society* 33:33. ISSN: 1069-7977.
- Stewart, T. C., Y. Tang, and C. Eliasmith (2011). A biologically realistic cleanup memory: Autoassociation in spiking neurons. *Cognitive Systems Research*. The 9th International Conference on Cognitive Modeling. Manchester, UK, July 2009 12:2, pp. 84–92. ISSN: 1389-0417. DOI: [10.1016/j.cogsys.2010.06.006](https://doi.org/10.1016/j.cogsys.2010.06.006).

References

- Stinkeste, C., A. Napala, and M. Romero (2021). Impact of Team Climate on Co-Creativity in STEAM Education in Primary Education. *Creative Education* 12:8, pp. 1977–1994. DOI: [10.4236/ce.2021.128151](https://doi.org/10.4236/ce.2021.128151).
- Sutton, R. S. and A. G. Barto (2018). *Reinforcement learning: An introduction, 2nd ed.* Reinforcement learning: An introduction, 2nd ed. The MIT Press, Cambridge, MA, US. xxii, 526. ISBN: 978-0-262-03924-6.
- Sweller, J., J. J. G. van Merriënboer, and F. Paas (1, 2019). Cognitive Architecture and Instructional Design: 20 Years Later. *Educational Psychology Review* 31:2, pp. 261–292. ISSN: 1573-336X. DOI: [10.1007/s10648-019-09465-5](https://doi.org/10.1007/s10648-019-09465-5).
- Taddeo, M. and L. Floridi (1, 2005). Solving the Symbol Grounding Problem: A Critical Review of Fifteen Years of Research. *Journal of Experimental and Theoretical Artificial Intelligence* 17. DOI: [10.1080/09528130500284053](https://doi.org/10.1080/09528130500284053).
- Tang, S. and M. Hanneghan (2011). “Game Content Model: An Ontology for Documenting Serious Game Design”. In: *2011 Developments in E-systems Engineering*, pp. 431–436. DOI: [10.1109/DeSE.2011.68](https://doi.org/10.1109/DeSE.2011.68).
- Tettamanzi, A., C. F. Zucker, and F. Gandon (2017). Possibilistic testing of OWL axioms against RDF data. *International Journal of Approximate Reasoning* 91, pp. 114–130. DOI: [10.1016/j.ijar.2017.08.012](https://doi.org/10.1016/j.ijar.2017.08.012).
- Thurlings, M., A. T. Evers, and M. Vermeulen (1, 2015). Toward a Model of Explaining Teachers’ Innovative Behavior: A Literature Review. *Review of Educational Research* 85:3, pp. 430–471. ISSN: 0034-6543. DOI: [10.3102/0034654314557949](https://doi.org/10.3102/0034654314557949).
- Torrance, E. P. and K. Goff (1989). A quiet revolution. *The journal of creative behavior*.
- Torrance, E. P. (1974). *Torrance Tests of Creative Thinking: Norms-technical Manual, Verbal Tests, Forms A and B, Figural Tests, Forms A and B*. Xerox, Personnel Press, Ginn, Xerox Education Company. book.
- (1990). *Torrance tests of creative thinking: Manual for scoring and interpreting results*. Scholastic Testing Service, Bensenville, Ill. 75 pp.
- Torrente, J., P. Moreno-Ger, I. Martínez-Ortiz, and B. Fernandez-Manjon (2009). Integration and Deployment of Educational Games in e-Learning Environments: The Learning Object Model Meets Educational Gaming. *Journal of Educational Technology & Society* 12:4, pp. 359–371. ISSN: 1176-3647.
- Tous, R. and J. Delgado (2006). “A Vector Space Model for Semantic Similarity Calculation and OWL Ontology Alignment”. In: *Database and Expert Systems Applications*. Ed. by S. Bressan, J. Küng, and R. Wagner. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 307–316. ISBN: 978-3-540-37872-3. DOI: [10.1007/11827405_30](https://doi.org/10.1007/11827405_30).
- Tubb, R. and S. Dixon (2014). “A Four Strategy Model of Creative Parameter Space Interaction”. In: *ICCC*.
- Turkle, S. and S. Papert (1992). Epistemological Pluralism and the Revaluation of the Concrete. *The Journal of Mathematical Behavior* 11:1, pp. 3–33.
- Turner, J. A. and A. R. Laird (2012). The cognitive paradigm ontology: design and application. *Neuroinformatics* 10:1, pp. 57–66. ISSN: 1559-0089. DOI: [10.1007/s12021-011-9126-x](https://doi.org/10.1007/s12021-011-9126-x).

- Uddin, L. Q. (2015). Salience processing and insular cortical function and dysfunction. *Nature Reviews Neuroscience* 16:1, pp. 55–61. ISSN: 1471-003X, 1471-0048. DOI: [10.1038/nrn3857](https://doi.org/10.1038/nrn3857).
- UNESCO (1996). Education for the 21st century: learning to learn - UNESCO Digital Library. *The UNESCO Courier* 49:4.
- Vallaëys, T. (9, 2021). *Généraliser les possibilités-nécessités pour l'apprentissage profond*. Research Report RR-9422. Inria, p. 1.
- Varadarajan, K. M. and M. Vincze (2013). "AfNet: The Affordance Network". In: *Computer Vision – ACCV 2012*. Ed. by K. M. Lee, Y. Matsushita, J. M. Rehg, and Z. Hu. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 512–523. ISBN: 978-3-642-37331-2. DOI: [10.1007/978-3-642-37331-2_39](https://doi.org/10.1007/978-3-642-37331-2_39).
- Viéville, T. (2002). *An improved biologically plausible trajectory generator*. Research report RR-4539. Inria.
- Viéville, T., D. Lingrand, and F. Gaspard (20, 2001). Implementing a multi-model estimation method. *International Journal of Computer Vision* 44, pp. 41–64.
- Viéville, T. and C. Mercier (6, 2022). *Representation of belief in relation to randomness*. Research Report RR-9493. Inria & Labri, Univ. Bordeaux, p. 20.
- Villars, A., A. Solignac, V. Piton, L. Pereira, B. Miège, C. Guillaud, and T. Cauquil (2023). *Une activité de robotique éducative transposée dans un environnement virtuel*. ENSEIRB-MATMECA, Bordeaux-INP.
- Voelker, A., E. Crawford, and C. Eliasmith (1, 2014). "Learning large-scale heteroassociative memories in spiking neurons". Poster. Unconventional Computation and Natural Computation. Centre for Theoretical Neuroscience, Waterloo, Canada. DOI: [10.13140/RG.2.1.1382.1688](https://doi.org/10.13140/RG.2.1.1382.1688).
- Voogt, J. and N. P. Roblin (1, 2012). A comparative analysis of international frameworks for 21st century competences: Implications for national curriculum policies. *Journal of Curriculum Studies* 44:3, pp. 299–321. ISSN: 0022-0272. DOI: [10.1080/00220272.2012.668938](https://doi.org/10.1080/00220272.2012.668938).
- Vygotski, L. S., L. Sève, Y. Clot, and F. Sève (1934). *Pensée et langage*. 3e édition. La Dispute, Paris. 536 pp. ISBN: 978-2-84303-004-8.
- Vygotsky, L. S. and A. Luria (15, 1980). Tool and symbol in child development. In: *Mind in Society: Development of Higher Psychological Processes*. Harvard University Press. ISBN: 978-0-674-07668-6 978-0-674-57628-5. DOI: [10.2307/j.ctvjf9vz4](https://doi.org/10.2307/j.ctvjf9vz4).
- Wallas, G. (1926). *The art of thought*. The art of thought. London, J. Cape. Pp. 320.
- Wang, M., L. Qiu, and X. Wang (2021). A Survey on Knowledge Graph Embeddings for Link Prediction. *Symmetry* 13:3, p. 485. DOI: [10.3390/sym13030485](https://doi.org/10.3390/sym13030485).
- Wason, P. C. (1, 1968). Reasoning about a Rule. *Quarterly Journal of Experimental Psychology* 20:3, pp. 273–281. ISSN: 0033-555X. DOI: [10.1080/14640746808400161](https://doi.org/10.1080/14640746808400161).
- Wertheimer, M. (1, 2000). *A Brief History Of Psychology*. 4th edition. Harcourt Brace Jovanovich, Fort Worth, TX. 192 pp. ISBN: 978-0-15-507997-7.
- Wiggins, G. A. (1, 2006a). A preliminary framework for description, analysis and comparison of creative systems. *Knowledge-Based Systems*. Creative Systems 19:7, pp. 449–458. ISSN: 0950-7051. DOI: [10.1016/j.knosys.2006.04.009](https://doi.org/10.1016/j.knosys.2006.04.009).

References

- Wiggins, G. A. (1, 2006b). Searching for computational creativity. *New Generation Computing* 24:3, pp. 209–222. issn: 1882-7055. doi: [10.1007/BF03037332](https://doi.org/10.1007/BF03037332).
- (1, 2020). Creativity, information, and consciousness: The information dynamics of thinking. *Physics of Life Reviews* 34–35, pp. 1–39. issn: 1571-0645. doi: [10.1016/j.plrev.2018.05.001](https://doi.org/10.1016/j.plrev.2018.05.001).
- Wing, J. M. (1, 2006). Computational thinking. *Communications of the ACM* 49:3, pp. 33–35. issn: 0001-0782. doi: [10.1145/1118178.1118215](https://doi.org/10.1145/1118178.1118215).
- Xiao, H., M. Huang, and X. Zhu (2015). From one point to a manifold: Knowledge graph embedding for precise link prediction. *arXiv preprint arXiv:1512.04792*.
- Zander, T., M. Öllinger, and K. G. Volz (2016). Intuition and Insight: Two Processes That Build on Each Other or Fundamentally Differ? *Frontiers in Psychology* 7, p. 1395. issn: 1664-1078. doi: [10.3389/fpsyg.2016.01395](https://doi.org/10.3389/fpsyg.2016.01395).
- Zelazo, P. D., A. Carter, J. S. Reznick, and D. Frye (1997). Early Development of Executive Function: A Problem-Solving Framework. *Review of General Psychology* 1:2, pp. 198–226. issn: 1089-2680, 1939-1552. doi: [10.1037/1089-2680.1.2.198](https://doi.org/10.1037/1089-2680.1.2.198).
- Zemla, J. C. and J. L. Austerweil (1, 2018). Estimating Semantic Networks of Groups and Individuals from Fluency Data. *Computational Brain & Behavior* 1:1, pp. 36–58. issn: 2522-087X. doi: [10.1007/s42113-018-0003-7](https://doi.org/10.1007/s42113-018-0003-7).
- Zhang, K. (1, 1996). A constrained edit distance between unordered labeled trees. *Algorithmica* 15:3, pp. 205–222. issn: 1432-0541. doi: [10.1007/BF01975866](https://doi.org/10.1007/BF01975866).
- Zhang, L., P. A. Kirschner, W. W. Cobern, and J. Sweller (1, 2022). There is an Evidence Crisis in Science Educational Policy. *Educational Psychology Review* 34:2, pp. 1157–1176. issn: 1573-336X. doi: [10.1007/s10648-021-09646-1](https://doi.org/10.1007/s10648-021-09646-1).
- Zhang, W., Z. Sjoerds, and B. Hommel (15, 2020). Metacognition of human creativity: The neurocognitive mechanisms of convergent and divergent thinking. *NeuroImage* 210, p. 116572. issn: 1053-8119. doi: [10.1016/j.neuroimage.2020.116572](https://doi.org/10.1016/j.neuroimage.2020.116572).
- Zhang, Z., Y. Lei, and H. Li (2016). Approaching the Distinction between Intuition and Insight. *Frontiers in Psychology* 7, p. 1195. issn: 1664-1078. doi: [10.3389/fpsyg.2016.01195](https://doi.org/10.3389/fpsyg.2016.01195).
- Zhu, W., Q. Qiu, J. Huang, R. Calderbank, G. Sapiro, and I. Daubechies (16, 2017). LDM-Net: Low Dimensional Manifold Regularized Neural Networks. *arXiv:1711.06246 [cs]*.