



**HAL**  
open science

# Modeling Creative Problem-Solving Tasks from a Computational and Neuroeducational Approach

Frédéric Alexandre, Chloé Mercier, Axel Palaude, Margarida Romero

► **To cite this version:**

Frédéric Alexandre, Chloé Mercier, Axel Palaude, Margarida Romero. Modeling Creative Problem-Solving Tasks from a Computational and Neuroeducational Approach. RR-9556, Inria & Labri, Univ. Bordeaux. 2024, pp.27. hal-04691371

**HAL Id: hal-04691371**

**<https://inria.hal.science/hal-04691371v1>**

Submitted on 18 Sep 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



# Modeling Creative Problem-Solving Tasks from a Computational and Neuroeducational Approach

Frédéric Alexandre, Chloé Mercier,  
Axel Palaude, Margarida Romero

**RESEARCH  
REPORT**

**N° 9556**

Septembre 2024

Project-Team *Mnemosyne*

ISSN 0249-6399





## Modeling Creative Problem-Solving Tasks from a Computational and Neuroeducational Approach

Frédéric Alexandre<sup>1</sup>, Chloé Mercier<sup>2</sup>, Axel Palaude<sup>3</sup>, Margarida Romero<sup>4</sup>  
Project-Teams Mnemosyne

Research Report N° 9556 — Septembre 2024 — 27 pages.

**Abstract:** Creativity is a complex process that has been studied in different fields with a high level of diversity in relation to the types of tasks, contexts, and assessment methodologies. In this study, we focus on ill-defined individual creative problem-solving (CPS) tasks with the objective of creating a computational model based on the regulatory process of CPS, inspired by the neuroscientific knowledge of the related cognitive processes, and artificial cognitive architectures. The model operationalization considers the emergent character of the path engaged by the learner when solving an ill-defined task and the geometrization of this path within a problem space describing the task. Stimulus-based and goal-directed creative behavior are then distinguished on the computational processes underlying creativity. Through a computational and neuroeducational approach, the study introduces a model of creative problem-solving tasks and provides an operational geometric definition of problem-solving tasks, emphasizing the challenges associated with ill-defined problems. We finish discussing creativity as a semantic grounding process with a focus on data representation, as well as symbolic data manipulation using inference and metric space algorithms.

**Key-words:** problem solving, creativity, computational creativity, creative problem solving, divergent thinking.

---

<sup>1</sup> Frédéric Alexandre Inria Mnemosyne Team – frederic.alexandre@inria.fr

<sup>2</sup> Chloé Mercier Inria Mnemosyne Team – chloe.mercier@inria.fr

<sup>3</sup> Axel Palaude Inria Mnemosyne Team – axel.palaude@inria.fr

<sup>4</sup> Margarida Romero LINE laboratory, U. Côte-d'Azur – margarida.romero@univ-cotedazur.fr

**RESEARCH CENTRE  
BORDEAUX - SUD-OUEST**

351 Cours de la Libération  
Bâtiment A29  
33405 Talence Cedex France

## Modélisation de tâches créatives de résolution de problèmes à partir d'une approche computationnelle et neuroéducative

**Résumé :** La créativité est un processus complexe qui a été étudié dans différents domaines avec un niveau élevé de diversité par rapport aux types de tâches, aux contextes et aux méthodologies d'évaluation. Dans cette étude, nous nous concentrons sur les tâches individuelles de résolution créative de problèmes (CPS) mal définies dans le but de créer un modèle informatique basé sur le processus de régulation du CPS, inspiré par les connaissances neuroscientifiques des processus cognitifs associés et des architectures cognitives artificielles. L'opérationnalisation du modèle considère le caractère émergent du chemin emprunté par l'apprenant lors de la résolution d'une tâche mal définie et la géométrisation de ce chemin au sein d'un espace de problèmes décrivant la tâche. Les comportements créatifs basés sur des stimuli et orientés vers des objectifs sont ensuite distingués sur les processus informatiques sous-jacents à la créativité. Grâce à une approche informatique et neuroéducative, l'étude introduit un modèle de tâches de résolution créative de problèmes et fournit une définition géométrique opérationnelle des tâches de résolution de problèmes, en mettant l'accent sur les défis associés aux problèmes mal définis. Nous terminons en discutant de la créativité en tant que processus de fondement sémantique en mettant l'accent sur la représentation des données, ainsi que sur la manipulation symbolique des données à l'aide d'algorithmes d'inférence et d'espace métrique.

**Mots clés :** résolution de problèmes, créativité, créativité informatique, résolution créative de problèmes, pensée divergente.

---

|  |    |
|--|----|
| 1. Problem-Solving Task: An Operational Geometric Definition               | 6  |
| 2. Considering Ill-Defined Problem-Solving Tasks                           | 8  |
| 3. Creativity as a Brain Functionality                                     | 9  |
| 3.1 Perceive, Decide, and Behave   | 9  |
| 3.2 From Input Stimulus to Internal Representation and Action              | 10 |
| 3.3 From Perception to Action  | 10 |
| 3.4 Episodic Memory for the Past and the Future                            | 11 |
| 3.5 The Prefrontal Cortex as a Biasing Machine                             | 12 |
| 3.6 Interaction between Episodic and Semantic Memory                       | 13 |
| 3.7 The Neural Network Basis of Creativity                                 | 13 |
| 4. Creativity as a Set of Computational Processes                          | 15 |
| 4.1 Creativity as a Semantic Grounding Process: Data Representation        | 15 |
| 4.2 Manipulating Symbolic data: Inference and Metric space algorithms      | 18 |
| 4.3 Application to ill-defined tasks and divergent mechanisms              | 19 |
| 4.4 Creativity as Intentional Processes: State Evaluation and Goal Choices | 20 |
| 5. Discussion  | 22 |
| Bibliography   | 24 |

## 1. Problem-Solving Task: An Operational Geometric Definition

Creative problem-solving (CPS) tasks are a type of problem-solving tasks requiring the exploration of the different strategies to develop emergent intermediate solutions through divergent thinking (DT) approaches, but also the evaluation of these strategies through a convergent thinking (CT) process. For more well-defined problem-solving tasks, the solution can be developed using a specific procedure rather than engaging in a variety of intermediate solutions. Moreover, such problem-solving tasks have been modeled in a way in which there is an optimal path for achieving a solution: Newell (1981) considered problem-solving the “fundamental organizational unit of all human goal-oriented symbolic activity” (p. 696), and he defined problem-solving in a problem space as “a set of initial states, a set of goal states, and a set of path constraints” The key idea is to “find a path through the space that starts at any initial state, passes only along paths that satisfy the path constraints, and ends at any goal state” (p. 695). Let us now explain how such geometric formalization extends to ill-defined open-ended problem-solving tasks requiring a large part of creativity.

This first definition of problem solving is computational in the sense that each element—the initial state, the final targeted state (i.e., the goal), and the generated trajectory from one to another (i.e. the path)—is precisely formalized. This is the case in several computational domains, such as robotic trajectory generation and optimization<sup>5</sup>, transportation theory<sup>6</sup>, and operation research<sup>7</sup>. This idea will be enriched below.

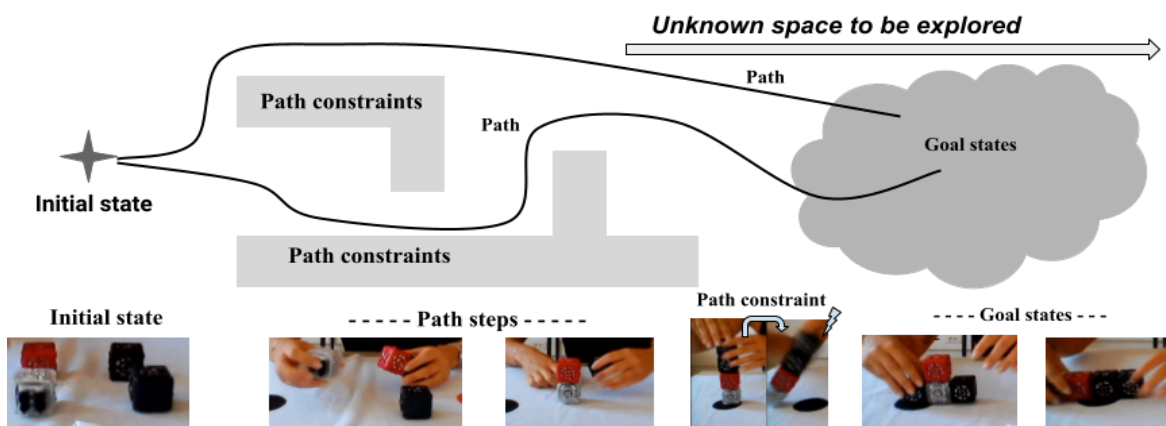


FIGURE 1: Formalizing Problem Solving as a Trajectory Generation Problem in an Abstract State Space., see Text.

A computational model can be operationalized in different ways. In this study, we consider a geometrical approach as a way to “spatialize” the model of the problem-solving process. Following this approach, we

<sup>5</sup> [https://en.wikipedia.org/wiki/Trajectory\\_optimization](https://en.wikipedia.org/wiki/Trajectory_optimization). In this pluri-disciplinary paper, despite not being a common practice in scientific papers, we provide the reader with encyclopedic links regarding obvious but not necessarily universally known elements.

<sup>6</sup> [https://en.wikipedia.org/wiki/Transportation\\_theory\\_\(mathematics\)](https://en.wikipedia.org/wiki/Transportation_theory_(mathematics))

<sup>7</sup> [https://en.wikipedia.org/wiki/Operations\\_research](https://en.wikipedia.org/wiki/Operations_research)

define problem solving as the elaboration of a problem state space and the selection of a trajectory within that space, as illustrated in Figure 1. Interestingly, the state space can be abstract and can include contextual information not only about the physical space but also about the agent, including the internal (e.g., emotional) state of the learner. Selecting a trajectory corresponds to deciding, as a step in the problem-solving process, to modify some characteristics of the state 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, 2021). Our representation of problem solving in Figure 1 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.

Given, for instance, for the CreaCube robotic open-ended problem-solving task (Romero et al., 2018), the initial state corresponds to the initial setup, path states to exploration or partial construction of the solution, path constraint to inappropriate configuration (here the cube tower falls down) and goal state to some (not unique) solution, while several paths are possible regarding such an open problem. Since the problem is ill-defined, the state space is not known from the beginning but discovered when solving it. Each point in this abstract state space is not numeric but corresponds to a symbolic description of the scene, as illustrated in Figure 1. Here, the definition is thus extended to complex ill-defined problem solving and allows us to propose an operational specification of creativity in this context, as developed in the sequel.

The initial present state is a point in this state space resulting from the problem-posing activity of the learner initiating 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 is constrained by certain requirements; furthermore, each path has a valuation (e.g., its cost in the wide sense or advantage), and valuations are partially ordered.

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).

When revisiting the literature on creative problem solving (CPS), we relate the different concepts to this geometric formulation to create a model that can be considered from an effective computational perspective. Furthermore, to exemplify our developments, we have two different tasks in mind: (i) the Tower of Hanoi problem<sup>8</sup>, as an a contrario example of a well-defined reference task, and (ii) the open CreaCube task (Romero et al., 2018), as the prototypal ill-defined CPS task that we want to understand better, as made explicit now.

---

<sup>8</sup> [https://en.wikipedia.org/wiki/Operations\\_research](https://en.wikipedia.org/wiki/Operations_research)



## 2. Considering Ill-Defined Problem-Solving 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 and colleagues 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 state space), yielding hierarchical decompositions.

Beyond complex problem-solving tasks, we are interested here in ill-defined problems for which several 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<sup>9</sup>;
- 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);
- it is thus not possible to apply a predefined 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.

For instance, working to create a simple Lego house and obtaining a large castle in the end is an example of the final production being specified progressively along the way and clearly differing from the initial views. Solving the Tower of Hanoi problem is a counterexample if the way to proceed is known from the beginning. Of course, there is a continuum between well-defined and ill-defined problems and a continuum between more and less complex problem solving.

We hypothesize that solving such an ill-defined problem requires more creativity than well-defined tasks. What we mean by creativity will be defined precisely in the sequel, and we will show that this corresponds to what is considered creativity in cognitive and computational neuroscience, computer science as well as learning science studies ((Boden, 2007); Dietrich, 2004; Tubb & Dixon, 2014). We also consider the review of computational creativity by Oltețeanu (2020) and the more specific cognitive neuroscience mechanisms described by Alexandre (2020) and Dietrich (2004). Here, we mainly consider “personal” creativity in the

---

<sup>9</sup> ... but can also be much more ill defined if the goal is to be found out (e.g., “only asking to move the tower to its correct place,” neither specifying where nor making explicit what the tower corresponds to) and the object manipulation is not stated (e.g., not defining the valid manipulation rules but only providing negative feedback if broken).

sense of Boden (1998) rather than “historical” creativity found in culturally significant achievements, also called Big-C creativity by some authors (Kaufman & Beghetto, 2009).

### 3. Creativity as a Brain Functionality

Kupers et al. (2019) mentioned that, whereas children’s creativity is mainly conceptualized today 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. 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 & Moustafa, 2022) for a recent review.

How the Brain Is Involved in Creative Problem Solving? Recent progresses in Artificial Intelligence are deeply due to analyses and interpretations, at different levels of description, of neuroscientific knowledge and data (Hassabis et al., 2017). This is particularly true for the description of how higher cognitive functions, like creativity and problem solving, might be mapped in the brain circuitry. We borrow from these interpretations to propose a functional sketch below, used to better understand how creativity occurs during a problem-solving task within an educational context.

#### 3.1 Perceive, Decide, and Behave

Let us first present the landscape of the brain mechanisms involved in a creative problem-solving task, based on the brain architecture schematized in Figure 2.

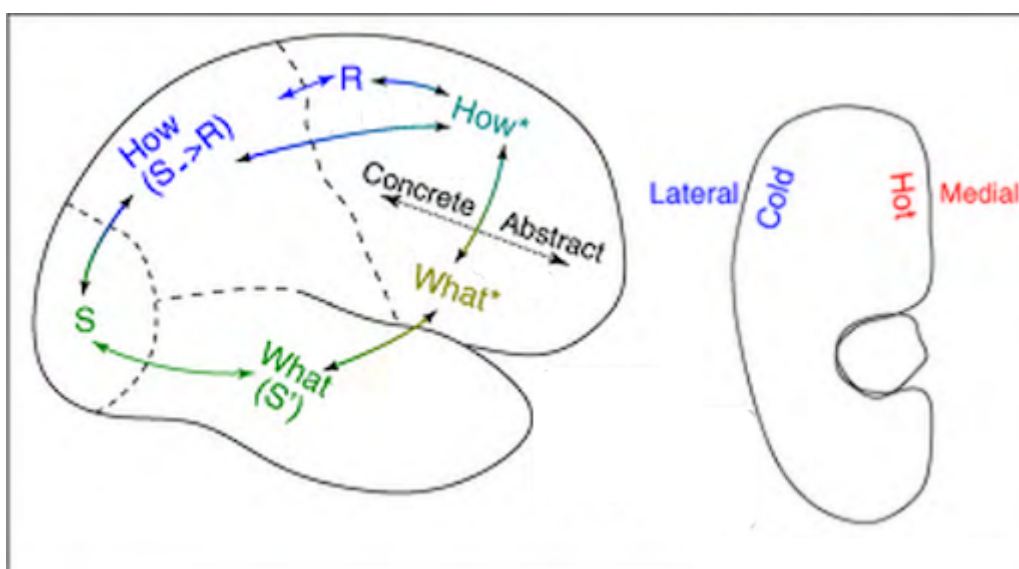


FIGURE 2: A General View of the Broad Functional Organization Discussed Here. Adopted from O'Reilly (2010) and Adapted to Our Purpose, See the text for details.

The What stimulus-driven (S) semantic representation (S') corresponds to the ventral posterior pathway, while the How stimulus (S) to response (R) mapping, that is, perception to action, corresponds to the dorsal

posterior pathway. In the anterior part of the brain, the prefrontal cortex (PFC) organization is multi-dimensional, as detailed by Alexandre (2021), with a What-Why-Where-How goal-directed organization, a What concrete-abstract gradient regarding representation abstraction, and a How concrete-abstract gradient regarding rule complexity. Taking a step further, the lateral-medial axis corresponds to cold, cognitive calculation versus hot, emotional value representation.

### **3.2 From Input Stimulus to Internal Representation and Action**

Information about the situation and the problem posed is received and activates sensory areas, including exteroception (the perception of the situation by, for example, the visual or auditory cortical areas) and interoception (the sensation of pain, pleasure, and emotions in the insular cortex).

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. Associative learning in the sensory cortex takes care of the multimodal sensory dimensions of situations and learns to associate the most frequently linked (for example, seeing a dog is often associated with hearing barking). Building on this, the sensory cortex is associated with the elaboration of semantic 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).

### **3.3 From Perception to Action**

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 orbitofrontal cortex (for giving an emotional value to a situation and making a decision accordingly). More precisely, 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 subgoal, 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 orbitofrontal 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, these learnings involve loops associating the cortex with the basal ganglia. Such a learning process can be very long, but, in the end, it 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. It can be remarked that these behaviors, which are very frequent in a stable and predictable world, are exactly the opposite to what is sought in a creative process. Conversely, two other kinds of memory, presented below, are particularly useful when access to specific cases and flexibility are needed.

### ***3.4 Episodic Memory for the Past and the Future***

Episodic memory involves a cerebral region called the hippocampus (or more generally the hippocampal formation) and corresponds to memorizing episodes, with the spatio-temporal context of the specific personal experiences of situations 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.

Later, when a similar episode is experienced (or part of it, like in the question “what did you have for breakfast yesterday?”), the hippocampus is able to recall the full initial episode and to reactivate the corresponding cortical regions, making the person re-experiment with the various sensations of 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, if this binding is not yet represented in the semantic memory, these episodes will be sent back to the cortex (in what is called a replay mechanism) 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.

Moreover, the hippocampus is not only able to recall a full initial episode or an episode with common features but also able to replay a partial episode or composite episodes made up of several pieces, in other words to predict unrealized episodes through simulation. This capability provides another way to decide which behavior to select in a certain situation, apart from the dominant behavior, remembering that, in similar cases, other decisions were made, being able to anticipate the subsequent outcomes and deciding which case is the most interesting. The capability to decide which behavior to select is essential in creative behavior. When subjects have a creative intention and aim to engage in creative behavior, they should develop a metacognitive judgment of candidate ideas based on their creative properties (mainly novelty and usefulness) and select those that can lead to an efficient and creative process or outcome. This yields a regulatory process to behave in a creative way instead of reusing existing habitual behaviors, especially when these become inappropriate.

### **3.5 The Prefrontal Cortex as a Biasing Machine**

The prefrontal region is another 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).

The medial part monitors errors, suggesting that the dominant behavior might not be adapted in certain circumstances, and aims to predict when these errors might occur to inhibit the dominant behavior accordingly. This is especially important in creative behavior regulation to ensure that certain ideas and actions are inhibited and those that could better respond to the creative intention of the subject are selected.

The lateral part learns new contextual rules, which are better adapted to new specific cases, based on interactions with episodic memory, as previously reported.

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.

Furthermore, as in the posterior sensory cortex, the distinction between the ventral (What, semantic representation) and the dorsal (How, perception for action) representation is valid here (Bunge, 2004). 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.

In addition, on a posterior to anterior axis of the lateral prefrontal cortex, at the How level, it has been observed (Badre, 2008) 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 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).

### **3.6 Interaction between Episodic and Semantic Memory**

This fundamental biasing role of the prefrontal cortex is not only directed toward the sensory cortex but also applies to the hippocampus since these two structures are also interconnected (Eichenbaum, 2017). More concretely, in one direction, concerning the definition of the contexts in which the dominant behavior should be inhibited and replaced by specific rules, this context can be sophisticated and defined by the binding of arbitrary cues by the hippocampus and provided to the prefrontal cortex. 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. The metacognitive judgment of the creative properties of the memories will allow us to judge which ones will be better candidates in the creative process regulation. Altogether, this forms the basis of what is called prospective memory or imagination (Buckner, 2010). More precisely, the hippocampus can be controlled step by step to produce a virtual (not really experienced) trajectory within the mapping of previous episodes. As proposed by Miller & Cohen (2001) with the analogy of a railroad, the hippocampus can lay new tracks and the prefrontal cortex can switch between them flexibly according to contextual rules. Such properties will have a major effect on creative processes, as we will discuss below.

### **3.7 The Neural Network Basis of Creativity**

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 the creative process (Beaty et al., 2016):

- The default network corresponds to brain regions that are active during spontaneous thought and mind wandering. It includes the insular cortex and the hippocampus, and a primary function of this network is episodic memory retrieval.
- The control network includes the prefrontal cortex and is activated for the control of attention.
- The salience network consists of the insular cortex and the prefrontal cortex and monitors the salience of sensory features.

Beaty et al. (2016) reported a major association between the default and the control network during creativity, in which the default network proposes candidate ideas (divergent phase) and the control network stands for the evaluation of their appropriateness (convergent phase).

In a more spontaneous mode, new ideas are generated from the default network and episodic memory and from the coarser semantic coding in the right hemisphere (Kounios & Beeman, 2014) and subsequently evaluated by the control network. In a more deliberate mode, the salience and control network can take into account the characteristics of the task and orient the default network toward a more systematic exploration of possible ideas. Dietrich (2004) related the first mode to intuition and discussed this dual process of imposing constraints to be appropriate and relaxing them to be creative, through ages, with a more spontaneous (and creative) approach in childhood. Kounios & Beeman (2014) also reported a neural inhibition of visual inputs by the control network to reduce distracting inputs, facilitate the retrieval of weak solutions, and evoke actors influencing insight, like mood (positive affects and reduced anxiety favoring a long-term view and broadened semantic processing; cf. also Diamond & Ling (2016) on these topics).

---

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. Concerning the control of the appropriateness of the idea, Dietrich (2004) reported that the dorsal part of the lateral prefrontal cortex is involved in its syntactic aspect, whereas the dorsal part of the medial prefrontal cortex is for the control and inhibition of common ideas (Maysless et al., 2015) and the ventral part of the lateral prefrontal cortex is involved in the semantic evaluation. Altogether, as mentioned by Schlichting & Preston (2015), it is interesting to observe that creativity and imagination are achieved through the recombination of prior memories and are consequently not so “new.” The way in which the recombination is carried out requires further analysis to identify the point at which the combination is performed through a regulatory deliberate process at a certain moment and through insights (spontaneous processes) at other moments.

### **3.8 Stimulus-Based versus Goal-Directed Creative Behavior**

To describe this latter concept, let us highlight two dual behaviors of creativity. From Isaac Newton’s falling apple incident<sup>10</sup>, leading to some insights regarding the gravity laws, to his deliberate scientific works driven by religious goals, we have two extreme examples of different creative behavior regulation. One flow is bottom-up and the other is top-down; the former is more spontaneous while the latter is more deliberate. However, we must enter into more details.

Stimulus-based behavior can have internal or external sources and must be described by both a bottom-up and a top-down process<sup>11</sup>. The bottom-up process is driving, but it is concomitant with a modulatory top-down process that allows the focus of attention and the injection of prior information, as in the case of Newton’s apple fall. It is thus not possible to restrain stimulus-based mechanisms to a simple bottom-up process because perceptual stimulus processing in the brain is, from the beginning, modulated by top-down mechanisms.

Goal-directed behavior is itself a dual mechanism at the goal-related level and the action-related level, as explained by O’Reilly et al. (2014) brain functional cognitive model with a What–How axis, enriched by the

---

<sup>10</sup> Indeed, we refer not to Newton’s imaginary apple head hit but to his witness statement about how his observation of a falling apple “enlightened” his vision of the gravity laws because his attention was focused on his problem and the external observation was biased by prior scientific knowledge.

<sup>11</sup> In the parieto-temporo-occipito, say posterior, part of the brain cortex, the top-down versus bottom-up processing mainly concerns sensory and to a smaller extent sensory-motor processing. The bottom-up processing drives the information and processing sensory input, while the top-down processing is modulatory and related to both the injection of prior information and the attentional mechanisms in the process (Friston, 2003). At this level, the top-down process makes the link between deliberate explicit symbolic information and spontaneous implicit numerical processing, thus mapping the former information onto the latter. This view is coherent with the corresponding part in the Tubb and Dixon (2014) model, which points out the monitoring and learning role of this feedback connectivity, while the driving link “feeding” the explicit processing by implicit information is omitted in their model.

In the prefrontal and frontal parts, say the anterior part, of the brain cortex, the dual process is qualitatively different and made explicit, for example, by Alexandre (2021), corresponding to the concrete versus abstract axis of the O’Reilly (2010) brain functional cognitive process of the anterior brain structures.

systemic framework proposed by Alexandre (2021) of What and Why versus Where and How, as discussed previously. In our formulation, the What and Why levels correspond to the goal-related level of creativity, proposing new intermediate goals, when not adapting or changing the final goal. We may call this the “semantic” level in the sense that it helps to “make sense” of the behavior. The What level allows us to analyze the present state situation and evaluate the potential goal, and the Why level drives the motivational part of the process.

Furthermore, following Alexandre (2021), the distinction proposed between goal-directed (voluntary) and stimulus-driven (involuntary) 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. Following this track, in line with Zhang et al. (2020) 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).

## 4. Creativity as a Set of Computational Processes

At the computational level, to implement the idea generation, that is, the divergent thinking process (DT), and evaluation, that is, the convergent thinking process (CT), the first challenge is how to represent the data used by such processes. The key point is that for a non-trivial task, as the one mentioned in Figure 1, symbolic information has to be taken into account. We propose to represent the information related to these concepts using a minimal and unified data structure, which is related to the geometric view of cognitive processes described earlier: an element is a point in this space and a concept a region of this abstract space. Our specification allows us to define the notion of a concept region as well as a distance between two concepts, thus turning any set of concepts into a structured concept metric space; that is to say, equipped with a distance to quantify how similar or different any two elements are. Practically, such a metric is useful for evaluation as it provides a basis for deciding the similarity between two concepts. Besides, this notion of metric space provides generation mechanisms, such as the projection of a concept onto a given region; for instance, given the concept “cat,” it can project this animal onto the region corresponding to the concept of “bird” and create a bird-cat chimera.

### **4.1 Creativity as a Semantic Grounding Process: Data Representation**

The effective modeling and implementation of creative processes require a choice of data representation at the proper level of abstraction: data are indeed represented in the brain by neuron activities and connection weights, but we need another level of representation allowing us to account for creativity at the geometric modeling level chosen here. We thus propose a symbolic representation that is compatible with a distributed (rather than a localist) view, as suggested by McClelland & Rogers (2003), in line with what is known about how information is processed in the brain through activations of multiple (rather than single) neurons, reviewed in the previous section.

The structure of semantic memory, thus data representation, plays an important role in creativity (Dietrich, 2004; Jung et al., 2013) as a function of both educational experience (Denervaud et al., 2021) and lifespan (Cosgrove et al., 2021). In a semantic network approach, that is, modeling semantic knowledge as a network of concepts, high connectivity and low modularity allow for more flexibility (Kenett & Faust, 2019) and foster associative processes (Mednick, 1962). Such computational models also provide a metric to assess some dimensions of creativity. However, they do not account for other organizational mechanisms beyond associativity (Mandler, 2011), such as hierarchical relationships and different levels of abstraction.



Following, for example, (Gärdenfors, 2004), we start with the simple common idea that a concept can be defined by “quality dimensions” that we will simply call features, that is, attributes with some typed value. In line with the mild abstraction approach proposed by Freksa (2015) and Oltețeanu (2020), we anchor these features in feature spaces built around different sensorimotor dimensions. We add hierarchical taxonomic relationships such that, if a concept is subsumed by another concept, it will inherit its features but can also override some (a few) of these features and add new ones. For example, a bird can be partially defined as being able to fly and sing and having feathers and a beak, and so on, and a penguin, which is a type of bird, will inherit those features with a few exceptions, such as being unable to fly but being able to swim, as described in Figure 3.

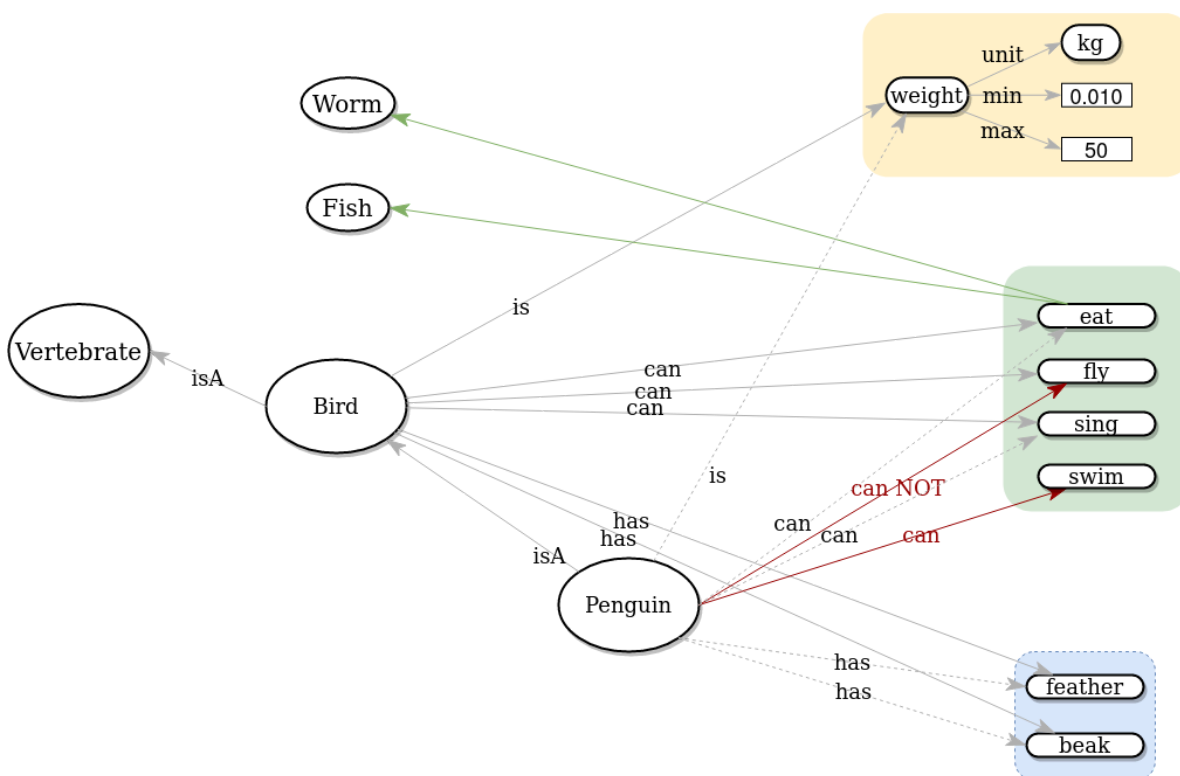


FIGURE 3: Symbolic Representation of Knowledge: a Semantic Graph View of What Is Implicitly or Explicitly Considered. See the text for details.

Feature values are anchored in sensori-motor feature spaces (colored regions) corresponding, for example, to different sensory modalities. Inherited features (e.g., the penguin “is-a” bird and thus inherits the features of a bird) are in dotted lines, while red lines represent overwritten values (e.g., a penguin can also swim but cannot fly). Green arrows point toward concepts that are themselves attributes of other concept features, accounting for inter-concept relationships.

Features thus correspond to sensorimotor information in the brain: some are properties—which can be qualitative (e.g., having a beak) or quantitative (e.g., weight)—while others are relationships. We follow the general approach of (McClelland & Rogers, 2003) using only four different predicates: hierarchical taxonomy (is-a), capability (can), extrinsic features (has), and intrinsic features (is). Beyond this, we need to introduce additional meta-information into the data structure specification to interpret the feature values. At this level,

features can be parametrized (e.g., weight unit, min or max).

A concept is a bipartite notion corresponding to:

- a prototype for this concept (e.g. in our example, the node "bird" defines a typical bird) defining "default" values for undefined features; it is thus a point in this abstract space;
- it is also a region of the state space, corresponding to all individuals having features that are compatible with a certain specification (called schema, as developed below).

This is aligned with the view of Gärdenfors (2004), which hypothesizes that a concept is a convex<sup>12</sup> region around prototypical examples, while we can also introduce the notion of exception at this geometric level, for example a penguin is a bird that cannot fly, meaning that the penguin concept region is almost included in (i.e., in a largely overlapping intersection with) the bird region, except for the flying capability.

For instance, a "bird" is any "vertebrate" for which there is a capability feature "fly" and so on. In other words, defining a concept using features is equivalent to defining a "thing" constrained by some equalities. 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). Such a notion of schema<sup>13</sup> is relevant in our context because divergent thinking might be implemented as relaxing some of these constraints while convergent thinking could be implemented as verifying that the resulting new data are still compliant with what could be expected to attain the desired goal.

Another important aspect is that we consider being in an open world, that is, all that is not specified is unknown. This allows us to define 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 (2006). At the implementation level, this representation is voluntarily very standard and corresponds<sup>14</sup>, up to a certain dialect variant, to most of the usual hierarchical data structure, allowing it to be easily readable and editable even by a non-specialist, either at a syntactic level or using any visual interface. It is also easily associated with a basic memory model, that is, representation in tables as in any database, and is a biologically plausible structure (Eichenbaum, 2017). The proposal is to create a table, i.e., an associative memory, for each relationship type, i.e., "is-a," "can," "is," and "has", and store references like:

---

<sup>12</sup> Convex means here that if two values are, e.g. birds, then intermediate values at a minimal distance between them (i.e., on a geodesic) are still birds.

<sup>13</sup> In the sense of XML-schema, in which concepts are re-used when considering JSON schema. At the programming level, this corresponds to the value of a given type.

<sup>14</sup> For instance, the JSON syntax, or even the weak\* wJSON syntax, allows us to write "birds can eat worms and fish" as "bird: { can: { eat: { worm, fish } } }" or "birds can eat" (i.e. "birds can eat something" as "bird: { can: eat }" or equivalently "bird: { can: { eat: thing } }".

(\*) <https://line.gitlabpages.inria.fr/aide-group/wjson>

"can [bird] = eat \* (thing + worm + ...) + fly + ... ."

This "equality" means that, in the "can" table, the "bird" entry contains "eat + fly" while attributes like "worm" are also bound to the "eat" capability. Interestingly, this mechanism corresponds to very standard data structures (e.g., the dictionary in Python) but also to a biologically plausible implementation of symbolic representation using a vector symbolic architecture, as developed, for example, by Mercier, Chateau-Laurent, et al. (2021).

#### 4.2 Manipulating Symbolic data: Inference and Metric space algorithms

This representation is interesting because it is effective in the sense that it can be manipulated by symbolic algorithms, including reasoners. Such a specification corresponds to a set of triple statements of the form:

<subject> <predicate> <object>

(e.g., <bird> <is-a> <vertebrate>) and is equivalent<sup>15</sup> to a standard knowledge graph representation, available for logical reasoning and allowing the enrichment of the representation. Such a representation was, for instance, used by Han et al. (2018) to implement computer-assisted analogy reasoning using an ontology. It is also possible to perform associative thinking, that is, noticing that "bird" and "fly" are related without explicitizing how in order to jump from one concept to another, modeled by a random walk in a semantic network, as developed by Zemla & Austerweil (2018).

A step further, such a semantic data structure also allows us to define behavioral rules, as schematized in Figure 4, which are of common use at both the cognitive modeling level (e.g., Collins & Koechlin, 2012) and the computational level (e.g., Laird et al., 2017).

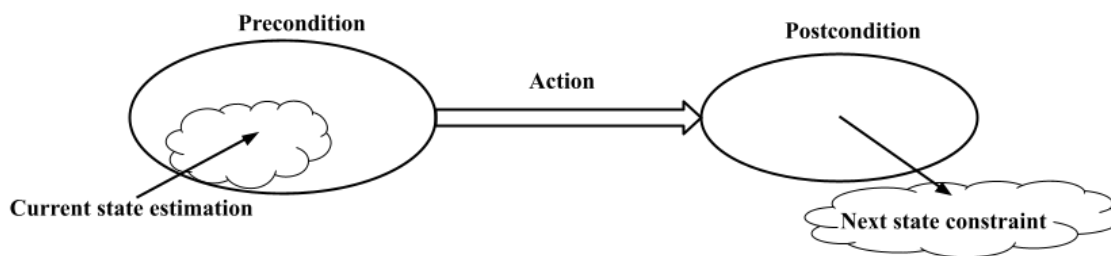


FIGURE 4: The Notion of a Behavioral Rule To Formalize an Action in a Problem-Solving Task.  
See text for details..

<sup>15</sup> On the one hand, it is straightforward to "flatten" the hierarchical structure into a set of triplets, as proposed for instance in the Resource Description Framework (RDF) specification, while the subject references are defined by local Internationalized Resource Identifiers (IRIs), to preserve the location of the element in the hierarchy. On the other hand, given a set of triplets with IRI indicating the location in the hierarchical structure, this structure can easily be reconstructed, see <https://line.gitlabpages.inria.fr/aide-group/wjson/turtoise.pdf> for a description at the implementation level.

As illustrated in Figure 4, a precondition is to be satisfied for the rule to be applicable. The precondition is defined as a region of the state space and the current state space estimation as another region, and the rule is applicable if the latter is included in the former. The result of the action corresponds to a postcondition, defined as a region of the state space, and the next state space is included in this region.

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 (Olteteanu, 2020) with problem templates structured around the concepts that are anchored in sensorimotor features.

As introduced before, we equip the state space with an edit distance, as proposed in this context by Mercier, Alexandre, et al. (2021). Very simply, given two concepts represented by their features, we consider the editing change, insertion, or deletion of a feature to transform the former into the latter and attribute a cost to each elementary operation. We consider the minimum cumulative cost among all the feasible operation sequences and define it as a distance. It appears that this notion<sup>16</sup> is well defined and rather efficient to calculate. Furthermore, such a distance can be adapted to a given applicative field because the cost of each operation can be weighted to balance the importance of different features. Another important aspect is that, beyond the distance value, we can explicitize the sequence of editing operations, allowing us to transform one concept into another, more simply called the editing sequence. This sequence defines a path<sup>17</sup> from the former to the latter concept. For example, we can change some features of a dog to turn it into a cat, with an editing sequence defining a path from the concept of dog to the concept of cat, with any point along that path corresponding to some kind of cat-dog chimera. Similarly, we can define the distance between two state values, each of them being defined over a set of concepts. In this case, editing operations will operate over concepts instead of their features.

As the basis of goal-directed behavior, such a distance is also to be used with respect to the task completion monitoring, considering the present estimation of the distance to the goal (which, in an ill-defined problem, is hypothesized during the preparation phase and adjusted after the evaluation phase), as made explicit by O'Reilly et al. (2014) and Pezzulo & Castelfranchi (2009). On the one hand, this allows us to integrate the notion of the learning progress (related to intrinsic motivation), as proposed by (Oudeyer et al., 2016), as the decrease of the distance to the goal. On the other hand, it is an important element if a backward problem-solving strategy is used (see, e.g., (Amidu et al., 2019) for a concrete example in a valuation problem and (Kapoor & Bahl, 2016) for a more general presentation) because we need to start the problem-solving process from the goal and thus require an estimation of the distance (which comes with a path to it, one of the short-distance paths, when using and editing the distance).

### **4.3 Application to ill-defined tasks and divergent mechanisms**

This formalization allows to precisely 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. Beyond these, all the previous concepts described in a geometric manner have an effective

---

<sup>16</sup> It is easy to verify that it defines a metric in the mathematical sense, and it appears that, with some precaution, the computation cost is polynomial with respect to the data size and is thus still tractable, as reviewed and discussed by (Mercier, Alexandre, et al. (2021) and implemented and discussed at the implementation level here <https://line.gitlabpages.inria.fr/aide-group/symboling>

<sup>17</sup> More precisely a geodesic in the sense that on this path between two points A and B we obtain for any point C, for the distance:  $d(A, B) = D(A, C) + d(C, B)$  instead of the triangular inequality.

instantiation and implementation, including<sup>18</sup> a biologically plausible one (Kajić et al., 2017; Mercier, Chateau-Laurent, et al., 2021).

This proximity is directly applicable to specify a process of reasoning by analogy, defined as a mapping between a source domain and a target domain, to predict some relationships between concepts, as formalized by Han et al. (2018), at a computational level, and revisited in Figure 5.

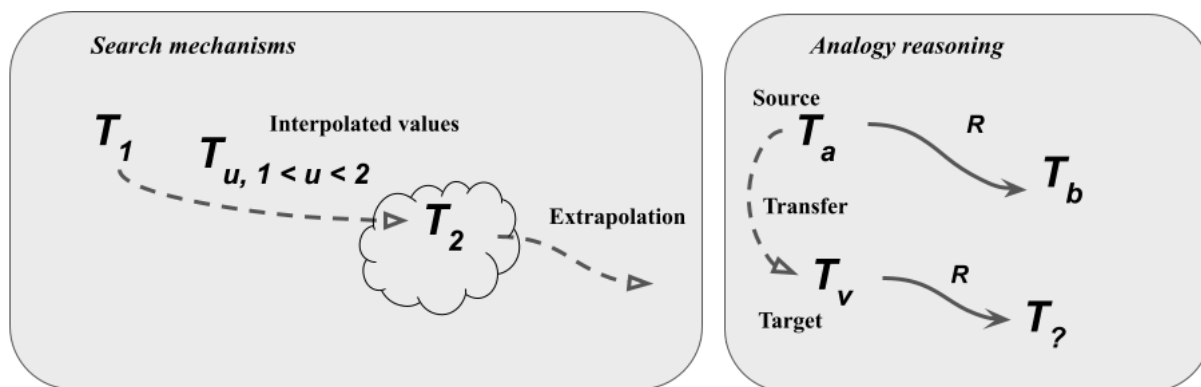


FIGURE 5: A Schematic Representation of Two Idea Generation Mechanisms.

Left: Exploratory search in the sense of Wiggins (2006) and Ritchie (2012).

Right: Reasoning by analogy, as defined in Han et al. (2018).

See the text for details.

#### 4.4 Creativity as Intentional Processes: State Evaluation and Goal Choices

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. 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). Each state may be assigned some valuations called rewards, which can be either extrinsic<sup>19</sup> (external and typically tangible, such as food or money, but also social, such as fame or praise) or intrinsic (i.e. fulfilling an intrinsic need, such as curiosity, e.g., the learning progress). 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 a priori versus a posteriori probability, for example in Bayesian approaches, implement such ideas with a numeric data representation. Such a computing mechanism, as made explicit by Viéville et al. (2001), for example, corresponds to a

<sup>18</sup> It appears that our data structure is a particular case of Resource Description Framework (RDF) representations, which can be mapped using a vector symbolic architecture (VSA) onto biologically plausible spiking networks, as proposed and discussed by Mercier, Chateau-Laurent, et al. (2021).

<sup>19</sup> According to Amabile (1983), extrinsic rewards tend to hinder creativity, unlike intrinsic motivation, which tends to foster it.

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.

Here, we also propose to consider the notion of partial knowledge and its related degree of belief, that is, the estimation of the available knowledge by a subject, which seems to be closer to cognitive representations. Motivated by these aspects, we implement this notion of belief considering necessity and possibility, “distinguishing what is plausible from what is less plausible, what is the normal course of things from what is not, what is surprising from what is expected” (Dencœux et al., 2020). At the data representation level, we simply add an attribute with a belief value<sup>20</sup>, while the chosen theory provides rules to combine different belief values<sup>21</sup>. This is supported by the fact that, for the last century and even slightly before it (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:

- (i) 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);
- (ii) 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.

In the study by (Raufaste et al., 2003), the 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. Furthermore, the notion of the development of modal understanding has been well studied since Piaget (Smith, 1994). Another key aspect (Dubois & Prade, 2015) 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. All together, this simply means that the estimated value is now weighted by a level of belief combining the belief with a priori information and the level of confidence in the stimulus.

These elements allow us to offer a complete operational description of how a partially defined and partially observable state value is estimated and how to evaluate the extent to which the targeted goal has been attained or not. We now have to discuss not only how we evaluate but also how we manage the subgoals and goals. Choosing and keeping track of goals is a key point in regulation processes. Goal setting takes into account both external and internal stimuli, which are considered with respect to extrinsic motivations (e.g. physiological needs) and intrinsic motivations, as discussed in the previous section. In problem-solving tasks, there may be concurrent goals, for example directed either toward performance (solving the task efficiently) or toward better mastery of the problem (developing task-related self-improvement, which corresponds to intrinsic motivation) (Poortvliet et al., 2007).

Following our preliminary proposal Mercier, Alexandre, et al. (2021), 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,

---

<sup>20</sup> As pointed out, for example, by Mercier, Chateau-Laurent, et al. (2021), the value stands between -1 and +1, where

-1 corresponds to a false value, i.e., impossible,

0 corresponds to a unknown value, i.e., fully possible but totally unnecessary,

+1 corresponds to a true value, i.e., fully necessary and thus fully possible

while negative values correspond to partially possible but unnecessary values and positive values to fully possible and more or less necessary values.

<sup>21</sup> This modeling of imprecise knowledge was used, for instance, by Tettamanzi et al. (2017), who linked possibility theory and ontology-based representations and proposed an example of effective combining rules, while Dubois & Prade (2015) discussed how such combining rules are constrained by the theory.

the lower the potential cost, that is, that editing 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<sup>22</sup>. Such a common currency is directly related to emotion, as explained by Alexandre (2020) and studied by Dietrich (2004) and as reviewed previously. 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<sup>23</sup> is available.

This generates an internal subgoal, which is to minimize 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.

## 5. Discussion

This study aims to develop a computational model of ill-defined CPS which takes into account creativity not only from the perspective of the regulatory processes in the learning sciences, but also the neuroscientific models allowing to understand the cognitive processes supporting the CPS process. To target computational modeling, 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. Based on this formulation, we reviewed both the learning science and the cognitive neuroscience literature to make an inventory of the different notions allowing us to explain and shed light on creativity processes in humans. Based on these concepts, and considering biologically plausible computational approaches, we proposed an effective computational specification, with a particular emphasis on data representation, which is often a scientific barrier in CPS modeling for ill-defined tasks.

In this paper, creativity unfolds from the socio-cultural context to the neuroscientific level. While we considered a certain degree of invariability at the neuroscientific level, the socio-cultural context can show important variability not only at the interindividual level but also considering the domain specificity of tasks for the assessment of creative problem solving. We also acknowledge that creativity research has been undertaken mostly among privileged participants and that further studies should be developed in under-privileged populations to ensure a more reliable study of human creative problem solving.

---

<sup>22</sup> It is not obvious, but still rather easy to see, that there is an optimal value when considering our geometric distance and suitable weighting function since the cost is a negative increasing function of the distance and the reward a positive increasing function of the distance. To exemplify, let us write:

$$\text{Common currency} = \text{Reward} - \text{Cost} = d - d^2,$$

considering the simplest case of a linear reward and quadratic cost, with a maximal value for  $d = 1/2$ . This is indeed only a toy example showing that the cost increase must dominate the reward increase to obtain such an optimum.

<sup>23</sup> Given 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.

We claim that this proposal is not only computationally well-defined, yielding a rigorously formalized computational framework, but also corresponds to the neuro-cognitive modeling as reviewed in detail here. Furthermore, it is rather “simple to explain” and can be used as a concrete pedagogical model, allowing the teacher and learner to better represent such a learning task, thus introducing meta-learning in such tasks and possibly improving learning performances. We also attempt to explain some abstruse computational concepts regarding creativity and ill-defined complex problem-solving based on this geometric “metaphor”, aiming to make them approachable to colleagues of other disciplines. Taking a step further, our major contribution is to explain creativity at the level of brain processes and provide enough material for non-specialists to understand better what happens in the brain when engaged in such creative behaviors.



## Bibliography

- Alexandre, F. (2020). Creativity explained by Computational Cognitive Neuroscience. ICC20 - International Conference on Computational Creativity, 4. <https://hal.inria.fr/hal-02891491>
- Alexandre, F. (2021). A global framework for a systemic view of brain modeling. *Brain Informatics*, 8(1), 22. <https://doi.org/10.1186/s40708-021-00126-4>
- Amabile, T. (1983). *The Social Psychology of Creativity*. New York city. Springer Verlag.
- Amidu, A.-R., Boyd, D., & Gobet, F. (2019). A protocol analysis of use of forward and backward reasoning during valuation problem solving. *Property Management*, 37, 638-661. <https://doi.org/10.1108/PM-10-2018-0056>
- Badre, D. (2008). Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends in Cognitive Sciences*, 12(5), 193-200. <https://doi.org/10.1016/j.tics.2008.02.004>
- Beaty, R. E., Benedek, M., Silvia, P. J., & Schacter, D. L. (2016). Creative Cognition and Brain Network Dynamics. *Trends in Cognitive Sciences*, 20(2), 87-95. <https://doi.org/10.1016/j.tics.2015.10.004>
- Boden, M. A. (1998). Creativity and artificial intelligence. *Artificial Intelligence*, 103(1), 347-356. [https://doi.org/10.1016/S0004-3702\(98\)00055-1](https://doi.org/10.1016/S0004-3702(98)00055-1)
- Boden, M. A. (2007). Creativity in a nutshell. *Think*, 5(15), 83-96. <https://doi.org/10.1017/S147717560000230X>
- Buckner, R. L. (2010). The role of the hippocampus in prediction and imagination. *Annual Review of Psychology*, 61, 27-48, C1-8. <https://doi.org/10.1146/annurev.psych.60.110707.163508>
- Bunge, S. A. (2004). How we use rules to select actions : A review of evidence from cognitive neuroscience. *Cognitive, Affective, & Behavioral Neuroscience*, 4(4), 564-579.
- Collins, A., & Koechlin, E. (2012). Reasoning, Learning, and Creativity: Frontal Lobe Function and Human Decision-Making. *PLoS Biology*, 10(3), e1001293. <https://doi.org/10.1371/journal.pbio.1001293>
- Cosgrove, A. L., Kenett, Y. N., Beaty, R. E., & Diaz, M. T. (2021). Quantifying flexibility in thought : The resiliency of semantic networks differs across the lifespan. *Cognition*, 211, 104631. <https://doi.org/10.1016/j.cognition.2021.104631>
- Denervaud, S., Christensen, A. P., Kenett, Y. N., & Beaty, R. E. (2021). Education shapes the structure of semantic memory and impacts creative thinking. *Npj Science of Learning*, 6(1), 1-7. <https://doi.org/10.1038/s41539-021-00113-8>
- Denœux, T., Dubois, D., & Prade, H. (2020). Representations of Uncertainty in AI: Probability and Possibility. In P. Marquis, O. Papini, & H. Prade (Éds.), *A Guided Tour of Artificial Intelligence Research : Volume I: Knowledge Representation, Reasoning and Learning* (p. 69-117). Springer International Publishing. [https://doi.org/10.1007/978-3-030-06164-7\\_3](https://doi.org/10.1007/978-3-030-06164-7_3)
- Diamond, A., & Ling, D. S. (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, 34-48. <https://doi.org/10.1016/j.dcn.2015.11.005>
- Dietrich, A. (2004). The cognitive neuroscience of creativity. *Psychonomic Bulletin & Review*, 11(6), 1011-1026. <https://doi.org/10.3758/bf03196731>
- Dubois, D., & Prade, H. (2015). Possibility Theory and Its Applications : Where Do We Stand? In J. Kacprzyk & W. Pedrycz (Éds.), *Springer Handbook of Computational Intelligence* (p. 31-60). Springer. [https://doi.org/10.1007/978-3-662-43505-2\\_3](https://doi.org/10.1007/978-3-662-43505-2_3)
- Eichenbaum, H. (2017). Memory: Organization and Control. *Annual Review of Psychology*, 68(1), 19-45. <https://doi.org/10.1146/annurev-psych-010416-044131>
- Eppe, M., Gumbsch, C., Kerzel, M., Nguyen, P. D. H., Butz, M. V., & Wermter, S. (2022). Intelligent problem-solving as integrated hierarchical reinforcement learning. *Nature Machine Intelligence*, 4(1), Article 1. <https://doi.org/10.1038/s42256-021-00433-9>

- Freksa, C. (2015). Strong Spatial Cognition. In S. I. Fabrikant, M. Raubal, M. Bertolotto, C. Davies, S. Friendschuh, & S. Bell (Éds.), *Spatial Information Theory* (p. 65-86). Springer International Publishing. [https://doi.org/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), 1325-1352. <https://doi.org/10.1016/j.neunet.2003.06.005>
- Gärdenfors, P. (2004). Conceptual Spaces as a Framework for Knowledge Representation. *Mind and Matter*, 2, 9-27.
- Gibson, J. J. (1977). The theory of affordances. In J. B. Robert E Shaw (Éd.), *Perceiving, acting, and knowing: Toward an ecological psychology* (p. pp.67-82). Hillsdale, N.J.: Lawrence Erlbaum Associates. <https://hal.archives-ouvertes.fr/hal-00692033>
- Han, J., Shi, F., Chen, L., & Childs, P. R. (2018). A computational tool for creative idea generation based on analogical reasoning and ontology. *AI EDAM*, 32(4), 462-477.
- Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. M. (2017). Neuroscience-Inspired Artificial Intelligence. *Neuron*, 95, 245-258. <https://doi.org/10.1016/j.neuron.2017.06.011>
- Hay, L., Duffy, A. H. B., McTeague, C., Pidgeon, L. M., Vuletic, T., & Grealy, M. (2017). A systematic review of protocol studies on conceptual design cognition: Design as search and exploration. *Design Science*, 3. <https://doi.org/10.1017/dsj.2017.11>
- Jung, R., Mead, B., Carrasco, J., & Flores, R. (2013). The structure of creative cognition in the human brain. *Frontiers in Human Neuroscience*, 7, 330. <https://doi.org/10.3389/fnhum.2013.00330>
- Kajić, I., Gosmann, J., Stewart, T. C., Wennekers, T., & Eliasmith, C. (2017). A Spiking Neuron Model of Word Associations for the Remote Associates Test. *Frontiers in Psychology*, 8, 99. <https://doi.org/10.3389/fpsyg.2017.00099>
- Kapoor, N., & Bahl, N. (2016). Comparative Study Of Forward And Backward Chaining In Artificial Intelligence. *International Journal Of Engineering And Computer Science*. <https://doi.org/10.18535/ijecs/v5i4.32>
- Kaufman, J. C., & Beghetto, R. A. (2009). Beyond big and little: The four C model of creativity. *Review of general psychology*, 13(1), 1.
- Kenett, Y. N., & Faust, M. (2019). A Semantic Network Cartography of the Creative Mind. *Trends in Cognitive Sciences*, 23(4), 271-274. <https://doi.org/10.1016/j.tics.2019.01.007>
- Khalil, R., & Moustafa, A. A. (2022). A neurocomputational model of creative processes. *Neuroscience & Biobehavioral Reviews*, 137, 104656. <https://doi.org/10.1016/j.neubiorev.2022.104656>
- Kounios, J., & Beeman, M. (2014). The Cognitive Neuroscience of Insight. *Annual review of psychology*, 65, 71-93. <https://doi.org/10.1146/annurev-psych-010213-115154>
- Kupers, E., Lehmann-Wermser, A., McPherson, G., & van Geert, P. (2019). Children's Creativity: A Theoretical Framework and Systematic Review. *Review of Educational Research*, 89(1), 93-124. <https://doi.org/10.3102/0034654318815707>
- Laird, J. E., Lebiere, C., & Rosenbloom, P. S. (2017). A Standard Model of the Mind: Toward a Common Computational Framework across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics. *AI Magazine*, 38(4), 13-26. <https://doi.org/10.1609/aimag.v38i4.2744>
- Mandler, G. (2011). From Association to Organization. *Current Directions in Psychological Science*, 20(4), 232-235. <https://doi.org/10.1177/096372141141414656>
- Mayselless, N., Eran, A., & Shamay-Tsoory, S. G. (2015). Generating original ideas: The neural underpinning of originality. *NeuroImage*, 116, 232-239. <https://doi.org/10.1016/j.neuroimage.2015.05.030>
- McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience*, 4(4), 310-322. <https://doi.org/10.1038/nrn1076>
- Mednick, S. (1962). The associative basis of the creative process. *Psychological review*. <https://doi.org/10.1037/H0048850>
- Mercier, C., Alexandre, F., & Viéville, T. (2021). Reinforcement Symbolic Learning. In I. Farkaš, P. Masulli, S. Otte, & S. Wermter (Éds.), *Artificial Neural Networks and Machine Learning – ICANN 2021* (p. 608-612). Springer International Publishing. [https://doi.org/10.1007/978-3-030-86380-7\\_49](https://doi.org/10.1007/978-3-030-86380-7_49)
- Mercier, C., Chateau-Laurent, H., Alexandre, F., & Viéville, T. (2021, mars 11). Ontology as neuronal-space manifold: Towards symbolic and numerical artificial embedding. KRHCAI-21@KR2021. Workshop on Knowledge Representation for Hybrid & Compositional AI.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.*, 24, 167-202.
- Newell, A. (1981). The Knowledge Level: Presidential Address. *AI Magazine*, 2(2), Article 2. <https://doi.org/10.1609/aimag.v2i2.99>

- Oltețeanu, A.-M. (2020). *Cognition and the Creative Machine: Cognitive AI for Creative Problem Solving*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-30322-8>
- O'Reilly, R. C. (2010). The What and How of Prefrontal Cortical Organization. *Trends in neurosciences*, 33(8), 355-361. <https://doi.org/10.1016/j.tins.2010.05.002>
- O'Reilly, R. C., Hazy, T. E., Mollick, J., Mackie, P., & Herd, S. (2014). *Goal-Driven Cognition in the Brain: A Computational Framework*. arXiv:1404.7591 [q-bio]. <http://arxiv.org/abs/1404.7591>
- Oudeyer, P.-Y., Gottlieb, J., & Lopes, M. (2016). Intrinsic motivation, curiosity, and learning: Theory and applications in educational technologies. In *Progress in Brain Research* (Vol. 229, p. 257-284). Elsevier. <https://doi.org/10.1016/bs.pbr.2016.05.005>
- Pezzulo, G., & Castelfranchi, C. (2009). Thinking as the control of imagination: A conceptual framework for goal-directed systems. *Psychological Research PRPF*, 73(4), 559-577. <https://doi.org/10.1007/s00426-009-0237-z>
- Poortvliet, P. M., Janssen, O., Yperen, N. W. V., & Vliert, E. V. de. (2007). Achievement Goals and Interpersonal Behavior: How Mastery and Performance Goals Shape Information Exchange. *Personality and Social Psychology Bulletin*, 33(10), 1435-1447. <https://doi.org/10.1177/0146167207305536>
- Raufaste, E., da Silva Neves, R., & Mariné, C. (2003). Testing the descriptive validity of possibility theory in human judgments of uncertainty. *Artificial Intelligence*, 148(1), 197-218. [https://doi.org/10.1016/S0004-3702\(03\)00021-3](https://doi.org/10.1016/S0004-3702(03)00021-3)
- Ritchie, G. (2012, mai 1). A closer look at creativity as search. *Proceedings of the Third International Conference on Computational Creativity*. [https://www.academia.edu/5626724/A\\_closer\\_look\\_at\\_creativity\\_as\\_search](https://www.academia.edu/5626724/A_closer_look_at_creativity_as_search)
- Romero, M., David, D., & Lille, B. (2018). CreaCube, a playful activity with modular robotics.
- Schlichting, M. L., & Preston, A. R. (2015). Memory integration: Neural mechanisms and implications for behavior. *Current Opinion in Behavioral Sciences*, 1, 1-8. <https://doi.org/10.1016/j.cobeha.2014.07.005>
- Smith, L. (1994). The development of modal understanding: Piaget's possibility and necessity. *New Ideas in Psychology*, 12(1), 73-87. [https://doi.org/10.1016/0732-118X\(94\)90059-0](https://doi.org/10.1016/0732-118X(94)90059-0)
- Smithson, M. (1970). *Human Judgment And Imprecise Probabilities*.
- Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2017). The hippocampus as a predictive map. *Nature Neuroscience*, 20(11), 1643-1653. <https://doi.org/10.1038/nn.4650>
- Tettamanzi, A., Zucker, C. F., & Gandon, F. (2017). Possibilistic testing of OWL axioms against RDF data. *International Journal of Approximate Reasoning*. <https://doi.org/10.1016/j.ijar.2017.08.012>
- Tubb, R., & Dixon, S. (2014). A Four Strategy Model of Creative Parameter Space Interaction. ICCC.
- Viéville, T. (2002). An improved biologically plausible trajectory generator (Research Report RR-4539). Inria. <https://hal.inria.fr/inria-00072049>
- Viéville, T., Lingrand, D., & Gaspard, F. (2001). Implementing a multi-model estimation method. *International Journal of Computer Vision*. <https://hal.inria.fr/inria-00000172>
- Wiggins, G. A. (2006). Searching for computational creativity. *New Generation Computing*, 24(3), 209-222. <https://doi.org/10.1007/BF03037332>
- Zemla, J. C., & Austerweil, J. L. (2018). Estimating Semantic Networks of Groups and Individuals from Fluency Data. *Computational Brain & Behavior*, 1(1), 36-58. <https://doi.org/10.1007/s42113-018-0003-7>
- Zhang, W., Sjoerds, Z., & Hommel, B. (2020). Metacontrol of human creativity: The neurocognitive mechanisms of convergent and divergent thinking. *NeuroImage*, 210, 116572. <https://doi.org/10.1016/j.neuroimage.2020.116572>



**RESEARCH CENTRE  
BORDEAUX - SUD-OUEST**  
351 Cours de la Libération  
Bâtiment A29  
33405 Talence Cedex France

Publisher  
Inria  
Domaine de Voluceau - Rocquencourt  
BP 105 - 78153 Le Chesnay Cedex  
[inria.fr](http://inria.fr)  
ISSN 0249-6399