# Toward a Formal and Computational Theory of Creativity: Extending Boden’s Framework via Projection-Space Search

# Outline

### **1. Introduction**

* The need for a unified, testable theory of creativity
* The paradox: novelty and intelligibility
* Prior approaches (cognitive psychology, AI, philosophy)
* The role of search in understanding creativity
* Our contributions:
  + Formal definition of creativity
  + Operationalization of creativity types
  + Empirical analysis across synthetic environments
  + Extension of Boden’s model

### **2. Background**

#### **2.1. Boden’s Theory of Creativity**

* Combinatorial, exploratory, and transformational creativity
* Emphasis on conceptual spaces and structured representations
* Lack of formalization or simulation framework

#### **2.2. Creativity in AI and Cognitive Science**

* Search-based creativity
* Monte Carlo and evolutionary approaches
* Recent work on novelty search, curiosity, and generative AI

### **3. A Formal Framework of Creativity**

#### **3.1. Search, Semantics, and Projection**

* Overview of projection-sampling framework
* Representational context as projection P
* Sampling as exploration under P

#### **3.2. Formal Definition of Creativity**

* Full mathematical definition as discussed
* Semantic continuity + structural divergence + functional viability
* Relation to compression, abstraction, and generalization

**Definition**: Creativity is the strategic transformation or compression of the search space to yield outputs that are:

1. Intelligible in light of prior knowledge (semantic continuity),
2. Novel enough to lack direct traceability to known inputs (structural divergence), and
3. Useful or meaningful within a context (functional validity).

Creativity is the capacity to strategically reshape the search process in order to produce novel outputs that are intelligible through indirect connections to prior knowledge, and functionally viable within a given context.

### **Computationally**

Let S be a search space, K the prior knowledge context, and f:S→O a generative process.  
 A creative agent transforms S′⊂S such that:

* f(S′)∩K=∅ (novel output),
* but ∃ϕ:f(S′)→K (interpretable mapping),
* and Eval(f(S′))>δ (threshold utility).

Where:

* ϕ is a projection to the space of known semantic interpretations,
* Eval is a task- or domain-specific evaluator,
* δ is a minimal threshold for relevance/usefulness.

#### **3.3. Creativity as Intelligent Search-Space Transformation**

* Search space S→S′
* Mapping to known semantics
* Non-triviality constraint

### **4. Typology of Creative Agents**

* Mapping of creativity types to agent behaviors

|  |  |  |
| --- | --- | --- |
| **Type** | **Description** | **Agent Name** |
| **Exploratory** | Biased search toward underexplored or sparse areas | Explorabot |
| **Combinatorial** | Reuses and recombines past trajectories | ComboBot |
| **Transformational** | Alters representational coordinates or grammar | TransBot |
| **Inferential** | Learns the structure of problem and predicts good regions | InferoBot |
| **Reframing** | Alters objective, constraints, or evaluation space | (Future Work) |
| **Constructive** | Composes partial solutions to build new ones | (Merged into ComboBot) |

### **5. Simulation Environments**

#### **5.1. Symbolic Search: Word Discovery**

* Setup: random and real dictionary targets
* Agents compared: MCTS, TransBot
* Results: TransBot excels via meaningful transformations

#### **5.2. Spatial Search: 2D Maze**

* Setup: discrete spatial environment with random obstacles
* Agents compared: all
* Results: MCTS prevails in no-feedback setting

#### **5.3. Topological Search: Surface Maze with Gaussian Mixtures**

* Setup: smooth landscape with implicit gradient
* Agents compared: all
* Results: InferoBot outperforms via structure-following

### **6. Creativity–Problem Fit Matrix**

* Mapping creativity types to problem structures
* Summary of performance patterns
* When is creativity useful vs not?
* Relationship to inductive bias and alignment with problem topology

### **7. Discussion**

* Creativity as generalization mechanism
* Creativity vs brute force exploration
* Importance of projections and interpretability
* Implications for AI, art, science, and education
* Connections to grounded cognition and active inference

### **8. Conclusion**

* Summary of formal definition, simulations, and insights
* Future work: richer environments, compositional tasks, agent hybrids
* Potential for unifying creativity research under testable principles

### **References**

* Boden, M. A. (2004). *The Creative Mind: Myths and Mechanisms*
* Schmidhuber, J. (2009). *Formal Theory of Fun, Curiosity and Creativity*
* Lehman & Stanley (2011). *Abandoning Objectives: Novelty Search*
* Open-ended evolution, generative agents, etc.

## TODO 2

## **1. Fix the citation errors**

* **~~AlphaGo / Move 37~~**
  + Replace *Silver et al., 2017* with *Silver et al., 2016, Nature 529* for the Lee Sedol match.
  + Optionally keep *Silver et al., 2017* if you add a “follow-on” note about AlphaGo Zero.
* **~~Suspension bridge example~~**
  + Either clearly frame Witcher (2022) as a *historical/cultural* reference or swap/augment with a technical civil engineering history source that supports the “invention of structural grammar” claim.
* **~~Duplicate Boden entries~~**
  + Merge into a single citation in the References.
* Do a quick **pass for style compliance**: punctuation, capitalisation in titles, journal names, etc., so you don’t get tripped up in formatting checks.

## **2. ~~Add Agent pseudocodes in Appendix~~**

* **Appendix title**: *Appendix A – Creative Agent Implementations*
* Provide **one short pseudocode block per agent type**:
  + **Inputs**: domain parameters, projection *PθP\_\theta*Pθ , policy π, utility function U.
  + **Core loop**: sampling, projection transforms, evaluation.
  + **Output**: best candidate(s) and performance metrics.
* Keep each ≤ 15 lines; match style across agents so readers can compare easily.
* Include **hyperparameters**: step limits, exploration biases, transformation operators.
* Add a **note on reproducibility**: RNG seed policy, environment library, and availability of code.

## **3. ~~Enlarge literature review and comment on contrast~~**

* Insert a **new subsection in Section 2** (after 2.3) titled *Comparison to Related Formal Frameworks*.
* Cover and contrast with:
  + **~~Wiggins (2006) Creative Systems Framework~~** ~~– rule spaces vs. your projection–sampling.~~
  + **~~Ritchie (2007)~~** ~~– empirical criteria (novelty, value, typicality) vs. your Intelligible / Novel / Valuable.~~
  + **~~Jordanous (2012) SPECS~~** ~~– evaluation dimensions vs. your operationalization.~~
  + **~~Quality-Diversity (MAP-Elites)~~** ~~– exploratory coverage vs. ExploreBot’s novelty bias.~~
  + **~~Gärdenfors (2000) Conceptual Spaces~~** ~~– geometric spaces vs. projection Sθ.~~
  + **~~Colton et al. FACE/IDEA models~~** ~~– framing shifts vs. ReframeBot.~~
* For each, add a **contrast paragraph**:
  + One-sentence summary of their focus.
  + One-sentence note on overlap.
  + One-sentence on what your framework adds or changes.
* Use a **comparison table** for visual clarity:

## TODO 3

## **~~1. Frame the Gap as a “Missing Mechanism” Problem~~**

* **Opening move:**
  + State plainly: *“Despite the wide influence of Boden’s typology, there is no operational mechanism to generate, predict, or compare the three types of creativity.”*
  + Add: *“This absence has limited the development of testable, general theories.”*
* Use **strong contrast**: Boden (descriptive) vs. Your work (formal + operational + testable).
* Explicitly claim: *“We fill this gap by providing a mathematically grounded, simulation-ready formalism that unifies typology and mechanism.”*

## **~~2. Declare the Framework as General & Domain-Independen~~t**

* Somewhere early (end of Intro):
  + *“The projection–sampling formalism is independent of domain content; it applies equally to symbolic, spatial, visual, linguistic, and scientific problem spaces.”*
* Add **examples from multiple domains** (you already have Cubism, Relativity, Suspension Bridge, Carroll, AlphaGo — keep them as proof of universality).

## **~~3. Name and Define the Contribution~~**

* Give the framework a **distinct, citable label** — e.g., *The Projection–Sampling Creativity Framework (PSCF)* or *Projectional Creativity Model (PCM)*.
* This helps others refer to it and marks it as a stand-alone concept.

## **~~4. Position Relative to Landmark Theories~~**

* Add a **“Related Formal Frameworks”** subsection with side-by-side contrasts:
  + **Wiggins CSF** – rule/space search vs. your P–π–U triad.
  + **Ritchie criteria** – novelty/value/typicality vs. intelligibility/novelty/value.
  + **Jordanous SPECS** – evaluation dimensions vs. your operational thresholds (δ, ε, τ).
  + **Conceptual Spaces** – geometry of Sθ vs. your projection-based observer relativity.
  + **MAP-Elites/QD** – diversity search vs. your exploratory agent architecture.
  + **FACE/IDEA** – framing shifts vs. ReframeBot.
* Use a **contrast table** so the reader sees clearly where yours extends and unifies.

## **5. Show It Generates a Research Program**

* Explicitly state: *“This framework defines a generative design space for creative agents, enabling systematic exploration of projection transformations, sampling policies, and evaluation functions.”*
* Suggest **future experiments**:
  + Testing new agent types.
  + Mapping creativity–problem fit across domains.
  + Applying the framework to human–AI co-creativity.

## **~~6. Emphasize Observer-Relativity as a Novelty~~**

* Dedicate a short sub-subsection to the idea that creativity is **frame-dependent**.
* Give a sharp example: how the same move or invention is trivial in one projection but transformative in another.
* This sets your work apart from output-only definitions of creativity.

## **~~7. Strengthen the “Creativity–Problem Fit” Message~~**

* Present it as a **predictive tool**:
  + *“Given the structural properties of a problem space, the framework predicts which creativity strategies will be most effective.”*
* Show this in a simple **matrix or 2D map** with problem structure on axes and agent type performance.

## **8. Make the Formalism Readable**

* Keep the equations, but ensure each is paired with a plain-language restatement.
* Reviewers should leave thinking: “I get the math, but I also see exactly what it means in words.”

## **~~9. Provide Reproducible Agent Archetypes~~**

* Add the **pseudocode appendix** (as you plan).
* Label each agent with its creativity type, P–π–U modifications, and intended problem-space strengths.

## **10. Close with an Ambitious Vision**

* Conclude with language that signals a shift in the field:
  + *“By grounding creativity in projection–sampling transformations, this work lays the foundation for a unified science of creativity — one that spans human cognition, artificial systems, and their interaction.”*