Emerging Ideas: How, if at all, Can Creativity Be Captured Computationally?

By
Bailey Fernandez

Submitted to Central European University
Department of Philosophy

In partial fulfillment of the requirements for a degree of Master of Arts in Philosophy

Supervisor: Tim Crane

Vienna, Austria
2025

=:

Author's declaration

I, the undersigned, **Bailey Fernandez**, candidate for the MA degree in Philosophy declare herewith that the present thesis is exclusively my own work, based on my research and only such external information as properly credited in notes and bibliography. I declare that no unidentified and illegitimate use was made of the work of others, and no part of the thesis infringes on any person's or institution's copyright.
I also declare that no part of the thesis has been submitted in this form to any other institution of higher education for an academic degree.

This is a true copy of the thesis, including final revisions.

Vienna, 27 January 2025

Bailey Fernandez
Signature

Copyright Notice

ABSTRACT

Creativity has recently received significant attention in the brain sciences and in artificial intelligence. However, the literature on creativity in contemporary philosophy — especially in the philosophy of these fields — has been comparatively scant. This thesis aims to take steps to address this gap by arguing for an account in which computers can be creative, and then reflecting on the significance of its results. I first synthesize diverse literature in the cognitive sciences, A.I., and philosophy which aims to derive additional necessary conditions for creativity beyond the standard definition of novelty and usefulness. Noting a resulting puzzle, I then examine existing cognitive and computational models of creativity, with a particular focus on a computational model of creative problem solving known as CreaCogs. I argue that the synthesis initially described can best be captured by the metaphysical notion of "emergence," and argue that this condition *is* compatible with computational creativity and that the model examined exhibits it. Finally, I use these results to offer an account of why "computational" cases of creativity might still seem like edge cases. Namely, emergence is often associated not only with a *metaphysical* gap between the base and emergent properties, but also an *epistemological* one regarding the explanation for how the emergent properties are produced from the base. By specifying a procedure which produces the creative/emergent results in explicit formal terms, we erode the epistemological gap even if we retain the metaphysical one. This observation brings forward a reflective sketch on the future of philosophy in discussing computational creativity.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

This thesis was written *in absentia* after having already begun a doctoral program at another university. These unusual circumstances of composition have in turn left me with an unusual debt to the kindness of others in facilitating its completion.

I would first like to thank my supervisor Tim Crane, and my second reader Maria Kronfeldner, for their time and generosity. Since I was not a student in residence at the time of composition, their (remote) commentary was given out of a remarkable good will towards me. It goes without saying that my effort was significantly improved through their feedback.

Second, I would like to thank the following individuals for their time and commentary on the thesis at various stages in its composition. I presented on papers relating to the "Blind Variation and Selective Retention" theory of creativity for two weeks in a row at the Massachusetts Institute of Technology's Philosophy of Cognitive Science reading group. I am particularly indebted to the comments of Selina Guter, Juan Vargas, Xin Hui Jong, and Mathias Michel. I am also grateful for the comments of friends on the thesis itself. I would like to thank Reid Kurashige, Gabriel Levc, Ida Miczke, Rohit Sanatani, and Irmena Tsankova for these.

I owe a particularly heartfelt thank you to Martyna Meyer. Our weekly ZOOM conversations about the literature on "problem solving" in cognitive psychology were instrumental in working through the tangled formal and empirical literature on the subject. Other conversations regarding emergence in the cognitive sciences helped me refine the positions expressed in the third chapter. In lieu of a traditional thesis seminar, these meetings provided a much-needed regularity and point-of-conversation for my thinking around this topic. It is difficult to imagine what the months writing the thesis would have been like were it not for these conversations.

Thank you, finally, to the Central European University, for considering my candidacy for a Master of Arts in philosophy.

INTRODUCTION

Nearly fifteen years ago, Berys Gaut (2010) noted a discrepancy in the literature on creativity in philosophy and the brain sciences. On the one hand, creativity was a core issue in psychology and cognitive science, with multiple high-profile journals dedicated to it.[1] The same could not (and still can not) be said for contemporary philosophy. There, treatment of creativity has been marginal. Given that the discussion of creativity (or "genius") plays a significant role in the *history* of philosophy –for instance, in Plato and Kant[2] – this is a surprising fact. Anecdotally, however, I can confirm that many of my philosophical colleagues find my deep interest in this topic inexplicable.

Gaut's article preceded a minor surge in philosophical publications on creativity. However, it is equally worth noting that the vast majority of these publications did not come from authors in the philosophy of mind or the philosophy of cognitive science. Rather, they came from aesthetics, value theory, and the philosophy of science more broadly. The most influential writer from the former subfields – Margaret Boden – was further already nearing the end of the active phase of her career by 2010. I by no means aim to disparage either the articles published or the named subfields by pointing out the discrepancy. Rather, it is to highlight a possible sociological reason for why these publications focused mainly on the social reception of creativity, did not frequently comment in-depth on the empirical literature, and typically advocated for a perspective on creativity which emphasized its relationship to human agency, institutions, and values.

The dominance of this perspective in the already-barren philosophical literature on creativity has left a remarkable gap for philosophical comment on creativity in the brain sciences and in artificial intelligence. The gap is all-the-more-notable given the sea change in

[1] See for instance the Creativity Research Journal and the Journal for the Psychology of Creativity, Aesthetics, and the Arts.

[2] See the *Ion* for Plato and the *Critique of the Power of Judgment* for Kant

the A.I. landscape which has taken place since Boden's most involved commentary in the late 90s and early 2000s, and especially in the past five years. Further, the *use* of the term "creativity" in the brain sciences and in A.I. is *clearly* problematic. "Creative A.I." is canvassed by startups and tech leaders with little clarification on its meaning, and "creativity" has also completely resisted real theoretical unification in the brain sciences despite the multiple journals devoted to it.

This gap is clearly beyond the scope of a master's thesis to resolve. What I hope to achieve instead is to take a small step towards solving it by focusing on a particularly salient part of the gap. Namely, due to the aforementioned inclinations of many of the philosophers writing on creativity in the 2010s, most existing literature in the philosophy of creativity rules out the very possibility of computational creativity *a priori* (see Mark Runco, 2023; Dustin Stokes and Elliot Paul, 2021, for stark examples). Such a categorical denial leaves no room for philosophical *comment* on computational creativity, and this leaves the gap between the philosophy and A.I. literature on creativity unbridgeable.

In this thesis, I construct an account of creativity which renders computational creativity *possible*. In doing so, however, I uncover certain aspects of computational creativity's *significance*. The resulting perspective is deflationary. This perspective, however, can still dialogue with and comment upon the efforts to capture creativity computationally. My argument is structured as follows.

In **Chapter 1,** I will examine the literature in philosophy and psychology regarding proposals for an additional condition for creativity beyond the "standard definition" of novelty and usefulness (see Mark Runco and Garret J. Jaeger, 2012). I consider proposals of two kinds. First, I discuss proposals including "surprise" (Margaret Boden, 2003; 1998), "non-obviousness" (D.K. Simonton, 2022), and "spontaneity" (Maria Kronfeldner, 2009; C.E. Hausman, 1975). Second, I discuss conceptions involving "agency" (Runco, 2023;

Stokes and Paul, 2021), which would rule out computational creativity *a priori*. I address these proposals and argue that they fail, leaving an agency requirement inconclusive and computational creativity theoretically possible. I instead present a formulation of "spontaneity" inspired by Kant (Kant, 2000; see Matthew Kieran, 2015, p. 169 for Kant's third critique as an account of creativity) in terms of "independence from a rule" as a candidate third condition. I argue this conception has explanatory power over the other two conditions presented in the first group. Doing so, however, opens up a conceptual puzzle regarding computational creativity. How can something which requires independence from rules be captured by an algorithm, which consists of nothing but rules made formally precise?

Chapter 2 aims to provide conceptual resources to address the problem through investigating a computational model of insight problem solving. Investigation into this model was chosen due to its proximity to standard methods for investigating human creativity in empirical psychology. I first outline the history of the literature on problem solving, which began in artificial intelligence before eventually migrating into cognitive psychology (Stuart Russell and Peter Norvig, 2010; Allan Newell and Herbert Simon, 1972). I then note that some cognitive scientists separate methods for problem solving into "analytic" approaches, and approaches requiring "insight" (Jessica Fleck and Robert Weisberg, 2013). Intuitively, "insight" problems are problems in which the solution appears "all at once," triggering an "a-ha" moment. Problem designs aim to correlate to historical cases of creativity, such as the solving of a difficult theorem or the creation of an artwork. I outline the CreaCogs model for insight problem solving presented in Ana-Maria Olteteanu (2020) and Arpit Bahety and Olteteanu (2019).

In **Chapter 3**, I articulate a formulation of "independence from rule" which aims to address the issue *metaphysically* through "emergence." I first note affinities that discussions of emergence have with discussions of creativity, including associations with novelty

(Richard Spencer-Smith 1995) and surprise (David Chalmers, 2006). I then distinguish the formulation required for my approach from existing proposals for emergence in creativity such as those in Ruth Richards and Terri Goslin-Jones (2018) as well as Zachary Estes and Thomas B. Ward (2002) through my focus on what Jason Winning and William Bechtel (2019) consider "pattern emergence" rather than "being emergence." I then argue that the notion of insight discussed in the previous chapter can be considered an instance of just this sort of pattern emergence. Since emergent phenomena are simultaneously *dependent* and *distinct* (Tim Crane, 2001) from their base, pattern emergence in relation to a *rule* provides a conceptual resolution for the puzzle presented in **Chapter 1**. However, I then follow Crane (2001) in arguing for a subjective, epistemological interpretation of the emergent phenomena in relation to the basal rule. Namely, the phenomena appear creative insofar as we cannot determine when they will emerge from the application of a rule. This in turn explains intuitions regarding why computational creativity often feels like an edge case. By specifying the formal methodology which will give rise to phenomena, even if it is an emergent, one plays a part in explaining it. Since the notion of creativity contains within it a conceptual uncertainty regarding the rule's outcome, these cases will not feel as or as fully creative as when the conditions are not thus specified.

I conclude by reflecting on the thesis's deflationary implications. I reject the view that there is a phenomenon *called* creativity, which exists in human minds, that computational creativity aims to capture. I instead note that the implications of the above line of reasoning are that creativity is a property which adheres to human (and machine) reasoning *in general*, one where novel, valuable, and emergent outcomes occur from algorithms or cognitive processes. Computational creativity should *not* be understood as a specific technical milestone to be achieved in A.I. research, but the process of achieving new milestones in mechanical reasoning itself. The role of a philosophy of creativity in artificial

intelligence and cognitive science is thereby therapeutic. We look for cases in which creativity is attributed to a *term* in current research, and show instead that it is a *property* of the term. How and when this property is instantiated, however, can be the source of genuine empirical, technical, and philosophical interest.

CHAPTER 1: FINDING A THIRD CONDITION FOR CREATIVITY

Introduction

In this chapter, I will argue that a third condition for creativity can best be understood as an independence from rule or procedure, which I — partially following C.E. Hausman (1975) and Maria Kronfeldner (2009)— will term "spontaneity." This formulation — however — will give rise to a conceptual puzzle regarding computational creativity. If we do not wish to reject the possibility of computational creativity *in principle*, then we need to find a method to resolve a tension to which this formulation seems to give rise. Namely, the fact that any generative system will be partially constituted by a set of substitution or transformation rules.[33] Producing an account on which it is conceptually possible for programs to be creative will be the focus of the subsequent chapters.

I will mostly keep both the empirical literature and the existing computational models to the side for this chapter. This move is not motivated by any general position regarding the use of this kind of evidence in philosophy. Indeed, both kinds of evidence will feature prominently in the second chapter. Rather, my motivation for doing so comes from the current state of the literature *in creativity* more specifically. There is good reason to be skeptical that creativity in the brain sciences can be identified one-to-one with any existing neurological correlate (Arne Dietrich, 2024; Dietrich 2019a) or cognitive process (Maria Kronfeldner, 2018).[4] Thus, we have no reason to assume creativity can be captured directly

[3] Conceiving computational creativity in terms of rule-based formalisms of course ignores the sea change in artificial intelligence towards data driven approaches over the past 5-10 years. However, rule-based formalisms at least have the advantage of being explicit, while in the current landscape of machine learning the granular operation of algorithms is mysterious even to their developers. It may be possible that *these* systems are creative as well, but they would not be so in virtue of their formal characteristics.

[4] It may seem surprising that any researchers would have ever thought that there would be a "creativity neural correlate." However, quite a bit of discussion of the Default Mode Network takes on this tenor. Likewise, many psychologists have identified creativity one-to-one with "divergent thinking" as tested by tasks such as the Alternative Uses Task (Guilford, 1967) or the Remote Associates Test (Mednick,1959; Mednick, 1962). However, as Dietrich points out, both cases are instances of false category formation. Plenty of tasks we consider "creative" are associated with other neural networks, and plenty of creative tasks are associated just

— in Marrian terms — on either the implementational or algorithmic level (David Marr, 1982). Attempts in the existing literature to segregate creativity into types (Dietrich, 2019b) must thereby find some other conceptual method by which to define creativity such that it be clear what "types of creativity" be types *of*. This chapter will aim at getting closer to a framework for doing so by means of conceptual analysis without specific commitment to any particular branch of philosophy. In the third chapter, the relevant condition(s) will be stipulated in explicitly metaphysical terms. These formulations will motivate my eventual, quietist, conclusion.

Section 1.1: Outlining Novelty, Usefulness, and Conceptual Spaces

A standard definition for creativity includes "novelty" and "usefulness" as necessary conditions (Runco, 2023; Runco and Jaegar, 2012).[5] I by no means aim to suggest that even these two conditions are without philosophical issues. However, in the brain sciences literature, there is at least a broad agreement that *these* are the conditions which need to be met. For the additional conditions which have been proposed, not even this can be said. While, as we will see, there is significant conceptual overlap between the different alternatives proposed for the third condition, *which* of these proposals has explanatory priority is in need of conceptual clarification. First, however, we should acquire some understanding of novelty and usefulness. For the sake of parsimony, I will present the literature around these conditions as unified even where it is not.

Novelty can be construed in either objective or subjective terms. My use of "objective" and "subjective" novelty here are terms of art deriving from Margaret Boden. An idea is subjectively novel if it has not been previously produced at some point in the agent's

as much with "convergent" thinking as they are with divergent thinking.

[5] Other authors call these the "Originality Condition" and "Adequacy Condition" (Kronfeldner, 2009). I take it that in both authors the meaning is the same.

history; objectively novel if it has not been previously produced in the agent's community. This community may, but need not, extend to the whole of human history. Call an idea which is subjectively novel "psychologically creative" (or "P-Creative"). Call an idea which is objectively novel "historically creative" (or "H-Creative") (Boden, 1998). Other authors have already argued that the set of H-Creative ideas is a proper subset of the set of P-Creative ideas (Kronfeldner, 2009). I will treat this as a claim already proved. This move will allow me to navigate away from the thorny questions involving the social reception of creativity and its role in the philosophy of science, to focus instead on the issues treated here.

The standard definition stipulates that to be creative, an idea must not only be novel, but also *useful*.[6] This intuition can be pumped by noting the literature's focus on creative *problem solving*. There, a creative solution must not only be novel, but also must help *solve the problem*. A related, but distinct, intuition is that to be creative, an idea must be *valuable*. If you do not feel this intuition, consider the credit which is often given to creative people. People who formulate important scientific theories, or generate great works of art, are the recipients of social *praise*. Further, these scientists and artists are typically praised *on account of the fact that they were creative*. This social phenomenon would be difficult to explain were creativity not something which we value (Gaut, 2018). Note that these two claims are distinct: it is possible that something be valuable without being useful. Both the standard definition and standard practice in the empirical literature on psychology focus on *usefulness* or *utility* as the relevant formulation of the condition. However, this analysis – particularly in respect to works of art – is quite strained. For this reason, I will henceforth construe the "usefulness" in the standard condition as *value*.

Novel and valuable variations may be produced in a variety of ways. Margaret Boden (2003; Boden, 1998) gives us three broad categories of increasing creative "significance":

[6]A small minority of authors question this condition. See for instance Robert Weisberg (2015).

combinatorial, exploratory, and transformational creativity. Each of these are defined in respect to the role they play in respect to a "conceptual space." A "conceptual space" may be thought of as a generative system (Boden, 2003). In combinatorial creativity, we use the rules of the conceptual space to generate novel and useful combinations of ideas. An immediate example might come from the perspective on creativity advocated by Noam Chomsky and Michel Foucault (1971) in relation to his formulation of generative grammars.[77] Exploratory creativity involves using and following *transformation* rules of the conceptual space, which operate on as well as substitute content. Transformational creativity, finally, consists of a modification of the rules of the conceptual space itself. Other authors have proposed some revisions to this framework. Jaakko Hintikka (1997) has noted that Boden's formulation is problematic in cases of mathematical creativity. There, any theorem — no matter how creative — will be produced within a set of generative rules. In response, he argues that transformational creativity should be understood in relation to proof *strategy*. Likewise, both Hintikka (1997) and Geraint Wiggins (2006) note an ambiguity in Boden's notion of a "conceptual space." Namely, the rules *constituting* the conceptual space are conflated with the rules for *traversing* it. Hintikka takes the *strategic* rules above to solve the issue, while Wiggins introduces *both* a distinction between constitutive and transformational rules *and* an additional set of preference rules to characterize the agent's evaluation of the space.

For our purposes, it is not important to legislate between Hintikka and Wiggins' proposed solutions. It is only important to note, as of yet, we have no other conditions beyond

[7] Boden (1998) notes that generative grammars as first conceived by Noam Chomsky meet the conditions for combinatorial creativity, in that the rule set (or "conceptual space") allows for – to use Chomsky´s term– "discretely infinite" combinations of atomic elements in new sentences (Chomsky,1991). We might then use other areas of linguistics – such as pragmatics or semantics–to rule out sentences produced by the system which are grammatical but meaningless such as "Colorless green ideas sleep furiously," "Caesar is a prime number," and the like. However, as Boden notes, to call such a system sufficient for creativity in a narrower sense appears limiting. To account for this, Boden introduces the other notions of exploratory and transformational creativity. I will add to Boden´s claim only that this precise lack of variation accounts for why – as Chomsky notes –his use of the term "creative" is "admittedly idiosyncratic" (Chomsky and Foucault,1971)

novelty and value.

Section 1.2: Kant on Genius

Before I discuss competing accounts for another condition, I will discuss something which at first may seem unrelated. However, it will play a central role in my argument going forward.

Immanuel Kant's *Critique of the Power of Judgment* contains an account of "Genius." Matthew Kieran (2015, p. 169) notes that Kant's formulation of "Genius" has served as a historical touchstone for philosophers interested in creativity. Even without this historical context, there is clear overlap between Kant's formulation of Genius and the standard definition. Kant proposes first that works of "Genius" must have "the foremost property…of originality," and second that "they must be exemplary…as a standard or rule by which to judge" (Kant 2000, Section 46). In the first, we see a transparent stipulation of *novelty*. In the second, we arguably see an account of *value*. Genius is to play the role of a *standard* for *aesthetic judgment*, and an *aesthetic judgment* is a kind of valuation.

Kant also presents a third condition, namely that "Genius itself cannot describe…or indicate scientifically how it brings about its products." (Kant 2000, Section 46). That is, Genius cannot clarify its *method* for bringing about products.

Let's dissect this further. Kant argues elsewhere in the Third Critique that the rules which give standards for aesthetics — i.e., the rules which are exemplary —cannot be generated from an empirical concept (Kant 2000, Section 46). Nevertheless, there are standards for aesthetic assessment. We are at least capable of *disputing* taste in conversation. Further, Kant proposes that "genius" cannot have a standard in mind as it is creating.

Otherwise, their product will be defined by an empirical concept, namely the standard. But Kant has already argued that this is impossible.

Instead, Kant proposes that "the rule to art…must be abstracted from what the artist

has done, i.e. from the product, which others have used to test their own talent" (Kant 2000, Section 47). I take Kant to mean that the rules for generating the creative product are understood by the agent and by others *retrospectively*. If an agent comes to retrospectively understand the rules by which they generated the product, then such a product could be *creative for that agent*. Such cases are the cases of P-Creativity discussed earlier. A smaller subset of procedures so produced will be retrospectively understood as a novel standard along with the agent. These are cases of H-Creativity.

Recall our earlier formulation of conceptual spaces. An additional characterization of these rules might now be given. In addition to producing novel products and containing a mechanism for evaluation, we might then have a clue regarding how to approach the third condition. Namely, *the application of the rules themselves* must only be clear retrospectively. The creator must somehow act independently from them, or give rise to them, or be unaware of them.

Section 1.3: A First Group of Proposals for Additional Conditions

D.K. Simonton (2022) notes that "it's necessary to go a step further [than the Standard Definition] because the standard definition provides only the necessary, but not the sufficient, conditions for judging personal creativity. (Simonton 2018b)" (in-text citations Simonton's). I do not think any of the proposals for additional conditions, including Simonton's, are plausibly *sufficient* for creativity. However, there should be additional constraints on creativity beyond *novelty* and *value* in the brain sciences.

Simonton first proffers Boden's (2003) suggestion that a third criterion for creativity might be *surprise*. He then offers his own suggestion "non-obviousness," a categorization used by the U.S. Patent Office.

His analysis of the third condition stems from his statistical model of the Blind Variation and Selective Retention Theory for creativity. The theory stems from work

originally presented by Donald Campbell (1960). Like my investigation, BVSR does not aim to identify creativity with either a neural correlate or cognitive process. Rather, it aims to account for creativity through a statistical analysis which can be applied to the evaluation of empirical tasks. BVSR first segregates creative idea generation has two aspects: first, a variation of an existing stock of ideas; second, the agent's selective retainment of the ideas which meet its goals. By itself, this formulation is really only an elaboration on the standard definition. The variation corresponds to novelty; the selective retainment to value. What makes BVSR unique is its commitment that the variation phase occurs *blindly*, i.e. that the generation of a new idea is statistically uninfluenced by its value. Often, and controversially, BVSR is construed in analogy with natural selection.[8] The "blind" variation is supposed to reflect the "undirectedness" of mutations in the genome, which are then "selectively retained" through natural selection. The two versions of a "third condition" thus far presented – surprise and non-obviousness – aim to reflect this supposed gap between the agent's prior knowledge and the ideas' novelty and value.

Boden (2003) observes that "to be surprised is to find that some of one's previous expectations did not fit the case." This formulation, if one wants, could easily be formulated in light of BVSR. An agent's "expectations" are their prior knowledge, so being "surprised" would just amount to being blind. We might also consider "surprise" phenomenologically. Simonton's BVSR conceives surprise as a *statistical* irregularity, but we might just as well conceive it as the "a-ha" moment which accompanies a sudden insight (Dustin Stokes, 2024). As we will see in **Chapter 2**, this formulation is particularly salient in respect to "insight problems."

[8] Scathing criticism of this formulation can be found in Kronfeldner (2011) and Kronfeldner (2010). The essence of the critique is that the analogy breaks down because it is clear for various reasons that our prior knowledge *does* have statistical influence on idea generation. This leads to a "partial coupling" between the agent's prior knowledge and the generation probability of an idea, which is not present in natural selection. Simonton (2022) partially recants any analogy between BVSR and Darwinism without any explicit reference to, but potentially inspired by, this criticism.

The U.S. Patent Office (and presumably Simonton) construe "non-obviousness" in relation to "skill in the art." Here we also find a relationship to prior expertise. Namely, an option not being obvious to someone "with prior skill in the art" is equivalent to it "not being obvious" to someone with *prior expertise*. It is less clear how this characterization might be linked to the previous discussion. We might define obviousness in terms of complexity in a conceptual space. A solution is "obvious" if it requires only a few transitions within the space, obvious if it requires many. However, many things I do every day – including all actions that require long term planning – are quite complex. But we would be hard-pressed to characterize actions like planning a family visit over the holidays *creative*. Since I have reneged on the requirement that the conditions specified be sufficient, and only require them to be necessary, this may be alright. However, we often *don't* associate creativity with complexity, but rather an unexpected *simplicity*. In the aforementioned cases of insight, we often experience a solution as appearing phenomenally "all-at-once." The kind of "non-obviousness" at issue, then, cannot be captured purely in terms of complexity. An alternative formulation is that a "non-obvious" solution is one that is not predictable, likely, or expected.[9] This formulation has conceptual links to surprise, as I will discuss shortly.

I will present one more option. We might construe a third condition for creativity in terms of "spontaneity." Spontaneity was originally introduced to literature by Hausman 1975, but was codified in relationship to the conditions for creativity in Kronfeldner (2018) and Kronfeldner (2009). The condition can be variously specified in ways which do or do not include agency. Since I aim to handle the conditions for creativity requiring agency in the next subsection, I will only focus on a formulation which does *not* require it here. Following

[9] My intention in this passage is to describe these conditions in such a way that it would be easily convertible into the Bayesian language of the predictive processing framework. Although there is an existing computational model of creativity which has been developed with predictive processing in mind (Wiggins 2020), no one to my knowledge has specifically tried to convert Boden's work into this framework. In light of the current popularity of the framework, especially in neuroscience, this seems like a worthwhile task. For an overview of predictive processing for philosophers, see Andy Clark (2013) and Wanja Wiese and Thomas Metzinger (2017).

Kronfeldner, we might think of spontaneity as "partial causal independence from a method or routine." For Kronfeldner, the method is related to prior expertise, so partial causal independence from method is partial causal independence from one's prior knowledge. This formulation has the advantage of explanatory power over the other two. If something is independent from a method, then it will not be obvious. Further, if it is not in one's prior knowledge base, it will be surprising.

We are now in the position to propose a third necessary condition for creativity. I will do so by unifying the discussion of Kant in the previous section with the prior observation.

Recall that Kant's Genius invents and discover the rules which gave a standard to its activity *retrospectively*. They thus exhibit some kind of causal independence from a *rule*. Likewise, Kronfeldner posits independence from a *method*, and this had explanatory power over surprise and non-obviousness. But independence from rule and independence from method, as I will argue in more detail later, are the same. I will thus propose here that the third condition should be understood as an *independence from rules*. This formulation, in addition to capturing both prints of Kronfeldner's definition and retaining its explanatory power over the other definitions, also connects it directly back to Boden's formulation of creativity in terms of conceptual spaces. We noted before that the third condition was not captured in Wiggin's (2006) formalization of Boden's framework: we now have a conceptual reason why it might have seemed elusive: the third condition is nothing but a sort of independence from such formalizations.

Section 1.4: Conditions for Creativity Involving "Agency"

I have argued in the previous section that one group of proposals for additional conditions for creativity can simultaneously be captured with the condition "independence from a rule." However, there is a second cluster of conditions of which I would like to treat. All formulations of these conditions imply that creativity requires agency or a related property

like “authenticity.” I take it, with these authors, to be obvious that machines are not agents in the strong sense: i.e. they do not act for and because of reasons, and they have no “self” to express. An agency requirement for creativity will thus rule out computational creativity *a priori*. The purpose of this section is not to demonstrate that there *could* not be an agency requirement. Rather, I only aim to argue that the existing arguments are not persuasive. Since the question remains open, we might have some reason to pursue the how it might be theoretically *possible* for artificial intelligence to be creative.

Gaut (2010) first expresses that creativity must require agency. These arguments have been developed more specifically in relationship to computers and artificial intelligence by Runco (2023) and Stokes and Paul (2021). I find both arguments rather weak, and consider each of them in turn. I then present what I think is a stronger version of this line of argumentation, which comes from the notion of “aesthetic responsibility” introduced in Susan Wolf (2016). I note that a version of “aesthetic responsibility” modified to fit the literature on creativity has a great deal of explanatory power regarding the social phenomenon of creativity. However, an asymmetry between the aforementioned H-Creativity and P-Creativity allows for cases of P- Creativity which don’t require social reception, and provide resources to argue against the position that this kind of responsibility for creative production is a *necessary* condition for creativity. I then handle an objection presented to me from the thesis advisor.

Runco (2023) argues that we should add “intentionality” and “authenticity” conditions to the standard definition in order to counteract the “artificial creativity” of A.I. I take both conditions to be related to agency. He gives examples in which novel results might provide uncreative conclusions, such as in divergent thinking. However, he does not mention the value condition in this section. He further states that intentionality – conceived as intrinsic motivation, and authenticity – conceived as the expression of emotions, are instantiated in

humans but not A.I. This is obviously true. But it does not follow that A.I. could not be creative without these things. Runco says repeatedly that he is trying to revise the standard definition to counteract the possibility that A.I. might be creative. No argument is given that this should be done. Hence, the paper fails to give any *reason* to reject the possibility of computational creativity whatever.

Stokes and Paul (2021) likewise argue for an agency condition, but with a different example. Their example comes from snowflakes: snowflakes arrange themselves in novel and valuable (i.e. "beautiful") constellations. However, the processes which produce them are not judged to be "creative." From this, they reason to the idea that creativity must be the product of agency, as snowflakes are not agents. The argument is a failed *modus tollens*. Here is a plausible reconstruction of its form:

1. If something is novel and valuable, then it is creative. (Premise) [A—>B]
2. Snowflakes are not creative (Premise). [~B]
3. Snowflakes are novel and valuable (Premise)[A]
4. Snowflakes are creative (Modus Ponens on 1 and 3). [B]
5. Contradiction on 2 and 4 (Reductio) [Falsum]
6. It is not sufficient that if something is novel and valuable, then it is creative (Negation Introduction on 5) [~(A—>B)]

….???????Creativity requires agency.

The argument only shows that additional conditions for creativity beyond the standard definition are required. This is something to which each side of the debate already agreed. It in no way demonstrates that we need agency to meet the conditions. Thus, the argument does not demonstrate why we could not meet this demand with the kind of proposal discussed in the previous section.

A more sophisticated version of the view can be found in Susan Wolf (2016). Wolf pumps our intuition that there is a distinctive form of *responsibility* which accompanies works of art. However, her conception can easily be extended to scientific and other achievements as well. First, we follow Gary Watson (1996) in distinguishing between two

"faces" of responsibility: accountability and attributability. Accountability is the sense of (moral) responsibility in which an agent is deemed liable to sanction or praise for their actions. "Attributability" refers to the degree to which an agent's activity can be attributed to their personal characteristics. These can come apart from each other, as well as from merely causal responsibility. Let's consider three cases in which someone shoves someone else while trying to board a busy train. In the first case, the shove is purely accidental. The agent is then only causally responsible for the shove: they did in fact physically move someone else, but in no way which engages moral consideration. Now assume that they did so intentionally, with a minor amount of force, with the intention of boarding the train. Perhaps this reveals certain personality characteristics about the person — their impulsiveness, aggressiveness — and so forth. But one might think that the action reveals all of these things about the person's self without thinking that they are liable for legal, moral, or even social sanction and reproach. It is this "attributability" sense of responsibility which is engaged in Wolf's example. Imagine that you visit the studio of a friend, who is a painter. You look at an interesting abstract composition which they seem to have made. Pointing it out to your friend, they comment that they did not compose it at all: merely spilled a bucket of paint on the canvas by accident. The intuition is that you are liable to *withdraw credit* to the artist for the composition. Wolf argues that the best way to explain this is as a kind of attributive, aesthetic responsibility. You give credit to the artist for the arrangement of forms because you attribute that composition which you value to personal characteristics that artist has.

This argument can easily be extended to other areas besides the arts. Take, for example, scientific achievement. We are liable to give scientific and technological innovators a kind of *credit* for their innovation, whether or not we think of them as strictly *moral* people. For instance, we might give credit to scientific innovators due to their industriousness and intelligence. Moreover, the kind of credit that we give to scientific and artistic achievement is

often given *in virtue of it being creative*. If this argument is right, however, then it seems like we cannot give any such credit unless the agent is a candidate for being held responsible in the attributive sense. This would require the possession of personal characteristics, which in turn requires agency.

However, we might also note a factor of this sort of assessment. In order for a creator to be given positive credit in this framework, it must be done on account of a community which receives their work. Recall the distinction between P-Creativity and H-Creativity in **Section 1.2**. Recall further that H-Creative ideas were a proper subset of P-Creative ideas.

Since aesthetic responsibility specifically designates that a *community* has given credit to a creator for their work, aesthetic responsibility can only occur in cases of H-Creativity. However, we have already established that there are cases of P-Creativity which are not cases of H-Creativity. Thus, there are instances of creativity which do not require social recognition of the act as creative. So aesthetic responsibility cannot occur in response to *all* cases of creativity, and thus agency need not be a necessary condition for creativity as a whole. It need only be such for H-Creativity.

I believe this evidence shows that the idea that agency-like conditions must be *necessary* conditions for creativity is inconclusive. Certainly, it would not be enough to write off the notion that machines can be creative in principle.

Tim Crane, who advised this thesis, gave one more related and very sophisticated objection. There has been some ambiguity thus far in my development of this idea regarding whether I am defending the claim that *computers* can be creative, or that creativity can be *computational*. These two claims can come apart. Notably, they come apart in a manner deeply relevant to the agency condition. Suppose I subscribe to a computational theory of mind. Then I might hold that all human thinkers can engage in creative thought, which would then be algorithmic in virtue of being thought. Now I might hold that agency is a requirement

for creativity and thereby that *computers* cannot be creative, but simultaneously hold that the creativity in *humans* is *computational*.

My first response is to note once more that I don't believe the arguments for an agency condition are themselves currently persuasive for the reasons already given. So there isn't reason yet to consider the situation except as a hypothetical. Second, it seems to me that this sort of functionalism about the mind will interact strangely with the kind of agency and systems of value proposed. This can be illustrated most easily with the aesthetic responsibility framework. If creativity is a property of cognitive processes which are computational, then creativity can occur on a sub personal level. The evaluation of a creative agent, however, will occur on a personal level. So we end up with a picture like this: i). creativity can be sub personal, but ii). it is necessary that a personal level exist in order for sub personal creativity to obtain, and iii). if a creative idea is subject to praise or credit (which I have previously argued are the strongest arguments for an agency condition), then that credit is given on the personal level for a sub personal process. These premises are not contradictory, but certainly jointly counterintuitive. If it is the sub personal *process* which is creative, then it seems odd that the *credit* for the process be given to the *person*.

A more palatable position is given by denying the necessity of the agency condition. Then one holds that I). Creativity can be subpersonal, but need not be, 2). Agency is not required if creativity is subpersonal, and 3). Creativity can only be subject to credit if it is personal and therefore comes from an agent. This gets us out of the counterintuitive idea that creativity is computational but requires unrelated, non-computational social practices of aesthetic responsibility attribution to be considered to be such. It seems to me that this would be a better way to track a "computational" creativity understood in light of an individual agent's cognitive processes. And this would not *require* agency.

Section 1.5: Some Tensions with Computational Creativity

Having addressed the second group of additional conditions for creativity, which rule out computational creativity in principle, let us return to the first. I have given a formulation of the third condition in terms of "independence from rule." I have argued that this formulation has explanatory power over alternatives, including surprise, non-obviousness, and independence from routine and expertise. Further, it theoretically connects the additional condition to the notion of conceptual spaces also widely discussed in the literature.

Recall that I also set out to leave the question open regarding computational creativity. However, my formulation also renders computational creativity problematic. Namely, any program — including ones capable of doing things we would ordinarily call "creative" if done by a human — is nothing but the implementation of an algorithm. An algorithm, as we know from the Church-Turing thesis, is nothing but a formalization of our intuitive notion of a rule or procedure. So, algorithms are nothing but a special case of rules, and any instance of computational creativity will be nothing but the implementation of algorithms. If we have specified that a *necessary* condition for creativity is an "independence from rules," then these two definitions seem at best to be in tension and at worst outright contradictory.

The challenge for the rest of this thesis will be to formulate this condition of "independence from rules" in a manner compatible with computational creativity.

CHAPTER 2: INSIGHT PROBLEMS AND COMPUTATIONAL CREATIVITY

Introduction

In the previous chapter I argued that we should adopt a third necessary condition for "creativity" in terms of an "independence from a rule." I then noted a puzzle for computational creativity which this proposal brought into view. How could computational creativity be possible, given that this is a condition?

This chapter aims to provide conceptual resources to resolve this puzzle through a consideration of the literature in problem solving. I will first link the literature on problem solving, and particularly in what are known as "insight problems," to creativity research.[1010] These steps will be taken in **Section 2.1** and **Section 2.2**. In **Section 2.3**, I will then discuss a computational model of insight problems known as CreaCogs and its related formalism OROC in detail. This will show that a prominent area of human creativity research can plausibly be captured computationally. **Section 2.4** will discuss the use of CreaCogs to solve the classical empirical tasks the Remote Associates Task [RAT] and Alternative Uses Task [AUT].

Section 2.1: Problem Solving

"Problem solving" is a field of research codified by the publication of Human Problem Solving (Allen Newell and Herbert Simon, 1972). Their approach aimed to define research in artificial intelligence with a focus on the ways in which humans solved problems (Stuart

[10] Results from Beaty et al (2014) suggest that there is no clear relation between solution rates in insight problems and other creativity tasks. Preliminarily, this is not an issue for this thesis's formulation, as I have already renounced commitment that creativity is captured by a single cognitive process. It is thus not problematic that people often exhibit creativity on some tasks (like insight problems) and not in others. Indeed, this is reflected by intuition: people may be very creative in mathematics but hopeless in art. A sophisticated objection stems from the use of the CreaCogs framework in **Section 2.3** and its explicit theoretical linkage between insight problem solving, the alternative uses task, and the remote associates task. If *that* model is an accurate one, then performance on these tasks should be linked. I treat the divide between these two parts of the literature as an open empirical question. If the evidence sides with the results in Beaty et al. (2014), then I respond by noting that the use of CreaCogs is that of an illustrative example. My argument would stand or fall depending on whether the replacement models exhibited some instance of what will later be termed "algorithmic emergence," whether or not they involved insight.

Russell and Peter Norvig, 2010). The solution to a problem for a human agent would not come into being through the development of a specific algorithm. Rather, it would come through an analysis of the way in which all of that agent's cognitive processes worked in consort. This approach required two innovations: 1). A theoretical framework regarding how generalized problem solving occurs, and 2). An empirical paradigm which could be used to investigate human behavior regarding problem solving in the task environment.

The standard theory of problem solving is hallmark classical representationalism, although non-representationalist formulations of problem solving exist (see David Kirsh, 2009; Sune Vork Steffensen, 2013). The relationship between the problem solver and the problem is conceived as a "problem space" (Benjamin Angerer, 2024). There is some controversy as to whether agents explicitly represent the problem space or not. Some authors argue that this is the case, others that it merely models human problem solving (Benjamin Angerer and Cornell Schreiber, 2019). In either case, the problem space consists of a graph in which vertices represent individual states in the problem, and edges represent transitions between states. The problem "begins" at the initial state, and transitions between problem states are governed by transformation rules. The accepting state at the end of the graph can be understood as a goal state. The *complexity* of a solution to a problem is measured by the number of states visited on the way to the solution state. *Heuristics* are strategies for reducing the complexity of a problem solution (Angerer, 2024). Note that Margaret Boden's "conceptual spaces" discussed in the previous chapter may also be understood as problem spaces.

Let us now consider an archetypal example of a problem in the problem solving literature: the Tower of Hanoi (Ana-Maria Olteteanu, 2020). The Tower of Hanoi was initially developed by the French mathematician Édouard Lucas (Eric Weisstein, 2022). Owing to its use in the problem solving literature, finding the solution to the puzzle has been

used to test many cognitive capacities such as the capacity to store information recursively (Mauricio Martins et al., 2015). The game typically consists of a starting position in which there are three pegs and *n* disks of increasing size placed on top of the leftmost peg. Solving the puzzle in the shortest number of moves requires a recursive strategy, in which one iteratively moves pieces to the end of the board and then back to the beginning with the aim of unloading all disks from the first peg and then stacking them in the final peg in reverse order. The shortest solution will be given in 2n-1 steps, where n is the number of disks (Weisstein, 2022).

Now let's understand this puzzle as a problem space. The initial state in the graph will be given as the state in which all disks are stacked onto the first peg. The rules govern possible state transitions from this initial state and beyond it according to the game rules. Namely, only one piece can be moved at a time and a larger piece cannot be placed on top of a smaller piece. The number of states in the problem space is theoretically infinite, but will be limited in practice by memory. So heuristics for reaching the solution state – in which all disks are stacked from smallest to largest on the third peg – are needed. These heuristics can be found via a classification of rule *sequences*. Due to the recursive nature of the solution, a rule sequence can be followed until the solution state is reached.

The ability to solve problems like these are indicators of general intelligence. Empirical investigation into how these kinds of problems are solved might thus inform us about problem solving in general.

Section 2.2: Insight Problem Solving

Call problems like the Tower of Hanoi "analytical" problems. Analytical problems are noteworthy for their stepwise nature. The structure of the problem is clear from the outset, as are the moves which are legal. Only the *strategy* for traversing the problem space, or

"heuristic," is uncertain.

The Gestalt psychologists initially proposed that another conception of problems involved "insight" (Jessica Fleck and Robert Weisberg, 2013). In contrast to analytic problem solving, the problem solver does not experience an initially clear problem space for which one then requires a heuristic. Instead, the problem solver initially feels "stuck" and then suddenly sees the solution as a whole. This sudden recognition of the solution is what we call an "insight." Phenomenally, one might think of an insight as an "a-ha" moment one encounters while working on a difficult problem, and suddenly discovering a solution. It is associated with a phenomenal feeling of *surprise* (Stokes, 2024). An investigation of insight is thus intimately related to the study of creativity, and in particular the family of conditions discussion in **Section 1.3**. There is thus reason to investigate them in accordance with our previously mentioned condition.

There are several ways of testing insight problem solving in the empirical task environment. One popular kind of "insight problem" are "matchstick problems" (Guenther Knoblich et al., 2001; Olteteanu, 2020). Consider a group of matchsticks which each depict unequal equations. The challenge is for the problem solver to find a way to make the equation equal while moving only a single matchstick. The catch is that the solution for the insight problem is given by moving one of the sticks composing an *operand* within the equation, as opposed to one of the number values. The case models the specific kind of difficulty found in insight problems. The solution is, in terms of complexity, trivial.

However, agents frequently have a difficult time "seeing" that they can actually move one of the sticks in the matchstick problem. Thus, the problem takes participants a long time to solve, and many participants do not solve it.

A popular theory for modelling insight problem solving is called "representational change theory" or RCT (Stellan Ohlsson, 1992). RCT and other models of insight problem

aim to account for the fact that insight problems seem "difficult" to solve even if the steps required to solve them are trivial. RCT breaks the stages of insight problem solving into three stages. First, the agent explicitly represents the problem space of the insight problem. In the problem space, however, there is no path from the initial state to the goal state.

Second, the agent reaches "impasse," when they notice that they cannot find a way to the goal state using the rules they are representing. Third, the agent changes their representation of the problem space (hence "representational change") so as to allow a path to the solution.[11][11]The "a-ha" moment, or phenomenal feeling of surprise, comes from the agent experiencing a sudden change in their problem space representation. The task of the empirical literature on insight problems is then to investigate how these changes occur.

There are many intuitive connections between insight problems and creativity, beyond the notion of surprise previously discussed. The impasse seems to parallel many of the cases of actual creativity which we often encounter in daily life. Imagine that you are writing a philosophy paper. You have certainly come to some point in the paper at which you can no longer draw the conclusion you aim to draw from the premises which you have originally included. These periods often require breaks: a long walk or a few days off. In the literature on creativity, such impasses are frequently called "incubation periods" (Simonton, 2022; stemming from Wallas, 1926). The "Eureka" moment comes when you have figured out how to reframe your argument so that the desired conclusion is reached. Since we cannot observe such instances of creativity in the wild, insight problems provide an attempt to model what this process must be like for problem solvers in microcosm.

Section 2.3: CreaCogs

With the literature on insight problems having been discussed, I can now turn to a proposed

[11] Guenther Knoblich et al. (2001) discuss that this approach replaces an earlier theory of "ill-structured problems," in which the representation was merely vague (C.A. Kaplan and Herbert Simon, 1990).

computational model for them. This model will be discussed in detail to provide a case study for how the puzzle discussed in **Section 1.5** might be resolved. A discussion of the model in light of my proposed solution can be found in **Section 3.2**.

CreaCogs is a theoretically ambitious computational description of insight problem solving (Bahety and Olteteanu, 2019; Olteteanu, 2020). In addition to providing a database and procedure for the generation of insight problems from non insight problems, the underlying architecture has also been used in two famous creativity tasks: the Remote Associates Task [RAT] and the Alternative Uses Task [AUT] (Olteteanu, 2020). This theoretical extension makes it an especially interesting example of computational creativity, as it thereby appears to be a promising approach for the theoretical unification of several distinct areas in the creativity literature. The model might thereby be particularly useful for examining how computational creativity could evade the dilemma discussed in **Section 1.5**.

At the core of CreaCogs and its descendent "Object Replacement and Object Recomposition" [OROC] formalism is the notion of an "affordance." Originally drawn from J.J. Gibson (1979) and theoretically extended in the ecological psychology literature, an "affordance" refers to the action possibilities which a given object has relative to an agent. For instance, a given tree may or may not "afford" climbing depending on your size, shape, and dexterity. If a tree "affords" climbing, we call it a climbing "affordance." None of the rest of the theoretical framework of ecological psychology is necessary to appreciate CreaCogs. However, it is central to the framework that concepts are stored in relation to their "uses," and in CreaCogs this will occur at multiple levels of the formalism.

The formal ontology for CreaCogs consists of three layers. All of the layers are stored in the knowledge base [KB] of an agent, but they are hierarchically grounded. That is, the first layer is taken as primitive (although it may be extended through interaction with the environment), the second layer is grounded on it, and the third layer is grounded on the

second layer. My presentation of this material will be summative and pseudo-formal. However, a more detailed description can be found in Chapters 7-8 of Olteteanu (2020).

On the first layer, the agent interacts with visuospatial features of the environment. This level is called F and consists of the basic properties of color, shape, and so forth. In interacting with these elements of the environment, an agent might come to note that these items have distinctive "motor affordances." These are strictly affordances which are embodied, and do not involve high level planning or conceptualization.

On the second level, the Knowledge Base contains a set of Concepts [C]. As discussed previously, these concepts are grounded on the first layer of Features. The agent binds together features in the environment, motor affordances they have, and "semantic tags" (i.e. names given to these bundles). Any member of the set which constitutes a concept can also be disjunctive. To use an example from Oltetenau (2020), we might define an apple as the concept C_1, where C_1 is the set {red v green, round(ish), to eat, "apple"}. C (the set of all concepts) will be a proper subset of the Cartesian product of the power sets of features, motor affordances, and linguistic tags [$C \subset (P(F) \times P(A) \times P(S))$].[12]

On the third level, the elements of set C will be augmented with different relations [R], actions [H], and solution states [sol]. Possibly following Anthony Chemero (2003), the relations in R are conceived as affordances. Actions in H are also two place relations, for instance Place Upon (Plate, Table). The solution affordance A_{sol} for a Problem Template is also a two place relation, and is the affordance for the action which will directly produce the solution state at the penultimate point in the problem. Elements of C, R, H, PT_{sol}, and A_{sol} are

[12] In the formalization, Olteteanu (2020, p. 93) writes that C is *equal* to the set described on the right hand side [C=P(F) x P(A) x P(S)]. However, I believe that this is a typo. First, Olteteanu *does* write that C is a *proper* subset of the product of the power sets. Second, it seems obvious that there will be elements in the set on the right hand side which will not be in the agent's set of concepts. For example, the set of the right hand side will include a set where the linguistic tag "apple" is paired with the motor affordance "to drink."

grouped into "Problem Templates." The set of all "Problem Templates" is called PT. A Problem Template is a known problem space where an agent will not have to spend any time searching for a solution. For instance, if I know how to cook Pasta Bolognese, I will have the Problem Template for Pasta Bolognese stored (Olteteanu, 2020, Chapter 8). We thus have the machinery for a compositional conception of the actions involved in problem spaces and the different uses which objects have within them. Further, we can always add uses to objects on the conceptual level by going back down to the feature level and investigating their motor affordances. We can do this because the feature level grounds the upper levels.

Now how do we generate insight problems using this framework? The key is to take an existing problem template in KB, and then decompose it so that it is difficult to find A_{sol}. Call the procedures for decomposing objects OROC, as discussed previously. There are several procedures for obscuring A_{sol} in a problem template. One might decompose the object which possesses A_{sol} for the agent such that it has to get put back together. For instance, Bahety and Oltetenau (2019) generate an insight problem entitled the "Blown Away Teddy" problem as follows: your child's teddy bear has blown away into your neighbor's yard, which is behind a wooden fence. You cannot climb the fence.[13] In your yard is a clothesline hooked to a wall, which you can use as a fishing hook to hook the teddy bear out of the neighbor's yard. This will likely not be the salient use of this object for the agent, however. To model the insight issue, the clothesline is not formalized as a hook but decomposed into a rod (whose affordance is to hold up clothes) and hook (whose affordance is to hold up the line). The agent must decompose the objects in this Problem Template of drying the clothes, and recompose them into the Problem Template for fishing the bear.

How might an agent go about doing this? In Olteteanu (2020), upon not being able to

[13] Problem Templates also allow for "constraints," or forbidden actions, which can be formalized into the problem.

find A_{sol} (in RCT terms, "reaching an impasse"), the agent can search through Problem Templates looking for a solution. These searches are weighted by the salient affordances of the objects in the environment. An "upwards" search will look for a Problem Template in which the objects in the environment can be used so as to produce the solution. A "downwards" search will move back to the Feature Level, looking for motor affordances to add to the concepts in the agent's Knowledge Base in order to solve the problem. A sideways search can use compositional procedures to install elements from other Problem Templates into the current one. Oltetenau refers to the appearance of a new Problem Template which successfully allows for a path to A_{sol} a "restructured representation" (Olteteanu 2020, Chapter 11). The affinities with "representational change" in RCT should be clear. The "a- ha" moment is explained as the sudden appearance of a path to the solution state of the current problem.

Existing insight problems have been generated through OROC. For example, the candle problem first proposed in Duncker (1937) has been generated (Bahety and Olteteanu, 2019). Strategies for composing insight problems in OROC beyond decomposing the solution affordance might also include adding distracting within the problem space. These objects can trigger other Problem Templates (i.e. useless problem representations) through presenting useless salient affordances. Another method for doing so is for the problem to contain a solution affordance involving an object not typically used for that task.

Section 2.4: CreaCogs and OROC for AUT and RAT

I mentioned that Olteteanu (2020) develops the CreaCogs framework discussed into two other creativity tasks. The theoretical links between these frameworks provides a great deal of CreaCog's theoretical interest.

The Alternative Uses Task [AUT] developed in J.P. Guilford (1967) in the paradigm

divergent thinking task. A patient is given an object, stereotypically a brick. They are then asked to write as many "uses" of the brick as they can think of within a specified time span. These answers may then be scored according to qualities such as their novelty, likeability, and usefulness. Two of these qualities should immediately strike the reader as evoking the standard definition discussed in the first chapter. However, CreaCogs and OROC give us another way of thinking about the "usefulness" of alternative uses.

Namely, a highly novel "alternative use" for an object might just be an affordance which is not salient. When stuck with an insight problem, we might go back to the Feature Level to investigate motor affordances of that object not contained within existing Problem Templates. One might recognize this procedure from **Section 3.3** as the downwards search procedure.

Likewise, the Remote Associates Task [RAT] is a task in which an agent thinks of an item which links three seemingly unusual words. For instance, if one lists the words cottage, cheese, and bank, one might find that the "associate" for these three words is "Swiss." Oltetenau (2020) notes that following B.R. Worthen and C.M. Clark (1971), we can distinguish "structural" and "functional" associates. "Structural" associates are ones like the one previously listed. The "association" has to do with frequent expressions, and often changes in translation. However, "functional" associates exist "in more than a language relationship." Olteteanu (2020, Chapter 9) develop a procedure for solving both kinds of associates, but only the "functional" associates are relevant to CreaCogs. There, a procedure is generated in which each of the concepts listed in the RAT trigger salient Problem Templates (Olteteanu, 2020, p. 117). The Template which is most salient across all concepts is selected as the Remote Associate. This procedure replicates the procedure of "upwards search" described in **Section 3.2**.

CreaCogs provides an exceptional theoretical unification of many areas in the creativity literature. If we can show how CreaCogs resolves the puzzle posed in **Section 1.5**, we may have a plausible case for illustrating how creativity can be captured generally in computational terms.

CHAPTER 3: EMERGENCE AS THE THIRD CONDITION

Introduction

In the previous chapter, I presented one computational model which may seem to have at least a plausible case of capturing creativity. In this chapter, I propose a solution to the puzzle in **Section 1.5**. Namely, I propose that the puzzle can be resolved if creative ideas can *emerge* from the application of the rule. I do *not* argue in this section that this is the only way that this condition can be met, only that it is at least one plausible candidate for doing so.

To defend this view, I first note in **Section 3.1** the requirements that emergent phenomena exhibit both "distinctness and dependence" (Tim Crane, 2001) from their bases. This formulation provides a groundwork for arguing that phenomena can be "independent" from a rule while still being determined by it. Although emergence is often construed as a metaphysical relationship between properties, I use work by Winning and Bechtel (2019) on "pattern emergence" to argue that emergence can also hold in relationship to a rule. I name this phenomenon "algorithmic emergence." In **Section 3.2**, I argue that the "insights' generated by the CreaCogs framework plausibly meet the formulation of algorithmic emergence. I defend this claim on account of the fact that the "re-structured representation" *emerges* from a search procedure. In **Section 3.3**, I comment on resulting epistemological issues, and how these might factor into an understanding of computational creativity.

Namely, epistemological interpretations of emergence might illuminate why cases of computational creativity still often seem like edge cases of creativity in general, even if they satisfy the more metaphysical interpretation defended elsewhere in the chapter.

Section 3.1: The Third Condition for Creativity as "Algorithmic Emergence"

At first, the proposed solution of "emergence" might hardly seem a novel one in creativity

studies. Some authors (Ruth Richards and Terri Goslin-Jones, 2018; Zachary Estes and Thomas B. Ward, 2002) have already argued that creativity might require emergence. Runco (2023) likewise spends significant time arguing that Large Language Models are in fact *not emergent*, in light of these authors' suggestion that emergence is an indicator of creativity in artificial intelligence. If there is innovation in my proposal, it comes from my direct relation of emergence to the standard definition and conditions discussed in **Section 1.3**. Emergent phenomena have a been associated with "novel" properties occurring in relationship to their base (Spencer-Smith, 1995; although as stated in Crane, 2001, Spencer-Smith does not endorse the view) as well as with "surprise" (Chalmers, 2006). However, a direct relation to an additional condition for creativity has to my knowledge not yet been proposed.

The focus of my argument, however, is to resolve the previously mentioned puzzle. I will do so by commenting on how the simultaneous "distinctness" and "dependence" conditions for emergence allow us to do so on a specific conceptualization thereof. The relation to algorithms must first be explained, not least because of the unclarity of the conception of emergence itself irrespective of my particular use of it.[14] I thus aim to sketch a very specific sense of the word emergence, which I aim to argue can meet the condition of being "independent from a rule" and yet be compatible with computational creativity. In these general terms, I have two desiderata. First, I must show how it can be the case that emergent phenomena can be independent from the base while still requiring that base.

Second, I must demonstrate how the base can be understood in terms of a rule or algorithm.

According to Tim Crane (2001), emergent phenomena exhibit the capacity for "dependence" and "distinctness" from their bases. The standard case examined in the

[14] As Jaegwon Kim (2006, p. 548) notes, "emergence is very much a term of philosophical trade; it can pretty much mean whatever you want it to mean, the only condition being that you had better be reasonably clear about what you mean, and that your concept turns out to be something interesting and theoretically useful."

philosophy of mind is typically the relation between mental and physical properties, so let us first consider it from that perspective. According to an emergentist point of view, mental properties depend on physical substances for their existence. Counterfactually, if the brain did not exist, nor would any of its mental properties. However, these properties are *distinct* from physical properties in that they have different characteristics. One way of expressing dependence and distinctness is through the metaphysical notion of supervenience. In case of a supervenient phenomenon, we imagine a base layer and a non-base layer which "supervenes" on it, and where changes on the supervenient layer only occur when there are changes in the base layer. The supervenient layer is dependent on the base in that it only exists and changes in light of it. It is distinct from it, however, due to a different sortal hierarchy in kinds of properties.

However, this relation does not seem sufficient for emergence. Jaegwon Kim (2006) argues that in addition to this supervenience relation, we must also consider other factors in individuating emergent phenomena. One candidate might be "predictability": emergent phenomena seem to behave in ways which are not *understandable* in virtue of the base layer. For example, one might contrast the relationship that the molecular structure of water has to its wetness. In the philosophy of mind, I might also contrast the relationship my neural configurations have to my conscious mental states.

In the first case, while there are multiple levels of description between the molecular structure of the water and its wetness, we can directly explain the wetness in terms of its chemical composition. We can also say definitively which chemical compositions will give rise to wetness, and which will not. The same cannot be said for connectionism. Neural configurations have no one-to-one correlation between any particular conscious mental state, as far as we know. At the very least, emergence in contrast to supervenience seems not to possess a *linear causal* relationship between base and emergent properties.

The picture sketched so far is one of an *ontological* which obtains between different kinds of properties. For my purposes, however, this kind of relationship will not be sufficient. In order to properly describe something as emergent from an algorithm, I must define a different sort of metaphysical relationship. Namely, I must define a structural one.

To produce a notion of emergence which will be useful for these purposes, I will consider a distinction which Jason Winning and William Bechtel (2019) draw between "being-emergence" and "pattern-emergence." "Being-emergence" is an ontological relationship of the kind just discussed. "Pattern-emergence" is a distinct phenomenon, however, and is just as often discussed in the philosophy of science as it is in metaphysics. When we model natural phenomena in the sciences, we will do so by tracking statistical regularities in the data which we are gathering. Viewed on their own, however, these "patterns" in the data might exhibit behavior which is not adequately captured in the description of the pattern. In the words of the authors, we "come to treat some patterns as emergent when one cannot account for their generation in the same manner as patterns regarded as more basic." The Game *Life* is often used as an illustrative example of emergent patterns (Francesco Berto and Jacopo Taglilabue, 2023; Winning and Bechtel, 2019; Chalmers, 2006).[15] There, a very simple set of rules can be used to define the behavior of an m x n dimensional grid. Assume some elements are "on" at the initial state. Transitions between each subsequent state in the game will be defined by whether the "on" cells are surrounded by a sufficient and non-excessive amount of other "on" cells. If the cell is surrounded by either not enough or too many other cells, it dies. If it is surrounded by a good mean of other cells, it will survive or regenerate (depending on whether it was on or off in the previous state). These simple rules give rise to extraordinarily complex rules and visual behavior with many different initial positions.

[15] In philosophy, it was famously used by Daniel Dennett (1992) to argue for an idiosyncratic view on the reality of folk psychological mental states such as intentions, desires, and beliefs (Berto and Tagliabue, 2023)

Winning and Bechtel (2019) give three families of this kind of pattern emergence in scientific phenomena. First, we might consider the issue from a systems-theoretic point of view, in which the application of one rule for pattern generation *bifurcates* such that another pattern becomes visible in the data. Second, one might think of *non-linearity*, in which the output of a given set of input data exhibits no regularities whatsoever. These irregularities might also be called emergent. Third, one might think of pattern emergence in terms of "rate-independence." To use the article's example (Winning and Bechtel, 2019, p. 140) , "The switching of a light switch provides a simple example of a rate-independent phenomenon. Flipping the switch requires the application of a certain threshold level of energy to the switch, but once it is flipped, the light is turned on or off independently of the energy applied to the switch" The flipping of the switch is a discrete event which occurs in the application of a continuous pattern of increasing force.

The visual phenomena produced in the *Game of Life* are produced from the iteration of an algorithm and a classified set of initial states. This state of affairs provides an important bit of evidence for us. In these cases of emergence, we need not only speak of "Pattern" emergence, but also "algorithmic" emergence. Cellular automata are formal automata and, as such, can be converted formally into graphs like the problem spaces discussed in **Section 2.1.** The transition function for the automata functions like the transformation rules of the problem space. So it seems to be *possible* that we conceive "algorithmic emergence" as either a kind of or equivalent to the "pattern" emergence discussed in scientific modeling.

Now we must show that these phenomena exhibit the characteristics of "distinctness" and "dependence" from the underlying formal description. Counterfactually, we might note that they are *dependent* on the formal description. If the formalism were not to be implemented, than the patterns in the *Life* would not be generated. We might note that they are "distinct" in light of the "rate independence" criterion given above. The visual behavior

which often appear in the *Life* occur at arbitrary configurations in the computational history of the game. They are not predicted by the initial state and state rules in advance. Thus, the game meets the conditions of distinctness and dependence and thereby illustrates a case of algorithmic emergence.

Recall that in **Section 1.5** the puzzle I posed involved how a computational creative system could simultaneously require independence from a rule and still be constituted by them. I present here the theory that the *distinctness* condition involved in algorithmic emergence can meet the independence condition that I previously specified. As such, we have one possible method in which the plausible explanation of the family of conditions in **Section 1.3** might still be compatible with research into computational creativity.

Section 3.2: CreaCogs as a Case of Algorithmic Emergence

Given that CreaCogs provides a particularly promising method for unifying various creativity frameworks in the empirical literature, it would be particularly exciting if the condition for algorithmic emergence specified in **Section 3.1** could be satisfied by the model. I argue here that it can be.

Before I turn to CreaCogs specifically, I will note that there is something of emergence at work in the conception of insight problems more generally. Insight is a *gestalt* notion. According to representational change theory (Angerer, 2024; Ohlsson, 1992), an insight occurs all-at-once after a patient experiences an impasse. That is, from the process of searching an inadequate problem space, an adequate one takes shape through the problem's re-representation.. We can thus turn to two phenomena: the process of *searching* the old problem space, and the *appearance* of the new problem space all-at-once at a sudden point in the search. The new problem space is *distinct* from the old one in virtue of the fact it contains a path to the solution. Further, it is *dependent* on the procedure of search for its generation,

and the searching of the old space is *distinct* from the appearance of the new space. Since the insight occurs as a Gestalt "all-at-once," it is also rate independent from the linear search procedure.CreaCogs and OROC model this structure computationally. Recall that upon reaching an impasse in a problem, the agent will begin heuristic searches across multiple layers of the ontology. The procedure of *searching* Problem Templates either downwards, upwards, or sideways is thus clearly *distinct* from the procedure of *executing* the Problem Template.

Further, the newly composed Problem Template will be generated *suddenly* according to a chain of linking affordances either on the Problem Level or on the Feature Level. This will make the procedure rate-independent and give further evidence to its distinctness. Finally, the newly generated Problem Template would not exist counterfactually, were the previous problem template to work and the search procedures thereby not initiated. As such, insight also depends on the impasse. It thus manifests *dependence*, which is sufficient for algorithmic emergence. Since we have linked algorithmic emergence to insight, we have also thus linked it to the phenomenology of "surprise," as discussed in **Section 2.2**. This fact gives further evidence that our framework links to the family of conditions discussed in **Section 1.3.** Since the AUT and RAT can be captured computationally as applications of search procedures in CreaCogs, as discussed in **Section 2.4**, these creativity tasks can also be captured as algorithmic emergent phenomena.

Section 3.3: Emergence as an Epistemological Issue

The argumentative strategy of the previous sections likely seemed to give a clean, metaphysical answer to the question I initially intended to address. Namely, it seems that I have identified a strategy for defending the view that computational creativity is possible. I have not yet, however, reflected on the significance of my strategy for doing so. Further, my

approach does nothing to explain away the following strong intuitions. Namely, that it doesn't really seem that we have a clear, unambiguous case of computational creativity that *any* observer would accept, irrespective of professional or technical training. In examining Crane's (2001) view of emergence specifically, I have a tentative suggestion for accommodating this intuition.

I discussed in **Section 3.1** that Kim (2006) distinguishes emergence from supervenience in virtue of the *unpredictability* of emergent phenomena. Crane argues a similar point by distinguishing emergent phenomena from supervenient ones in terms of their *explainability*. Emergent phenomena are dependent on their base for their existence, but are not *explained* by them causally or otherwise. For Crane, the result in the metaphysics of mind is that we should treat it as an open empirical question whether there are physical correlates to mental phenomena such as consciousness.

The phenomenology and significance of a *pattern* which is not *explained* in virtue of its base is first felt principally in its discovery. It is seeing something which we could not have expected to find. The pattern still *remains* emergent in the metaphysical sense even after we have discovered that we can generate it. We say this because the base pattern plays no *explanatory* role in relationship to the emergent pattern. Following the argument in **Section 3.1**, this explanatory gap can be seen in relationship to *Life* and other algorithmic- emergent phenomena. However, upon noticing that these insights can emerge through the application of this procedure, we will no longer be *surprised* by them. They will be metaphysically distinct from their bases. But from our point of view, they will no longer feel epistemologically alien.[16]Something of the significance which I aim to put forward here has

[16] Chalmers (2006) uses examples involving cellular automata such as Conway's Game of Life as an example of weak emergence. There is some conceptual relation between this sort of weak emergence and the "pattern emergence" described here, which this metaphor suggests via Daniel Dennett's discussion of mental phenomena in "Real Patterns" (1992). For Dennett, "folk psychological" terms are not reducible to neurological substrates but are rather like the patterns we observe in *Life*.

already been suggested by several of the previously discussed authors. Bahety and Olteteanu (2019) write that "One psychometric limitation of insight problems is that, once the participant has solved a problem, this will most likely not produce insight anymore, as the solution path has already been trodden by the participant." Likewise, Kronfeldner (2009) compares the "spontaneity" condition to the Meno Paradox in a philosophical discussion of the relationship between the conditions of its partial independence from routine and its partial independence from an agent's prior knowledge.[17]. Once an insight problem has been resolved in the CreaCogs framework, it can easily be assimilated into the agent's existing knowledge base of Problem Templates. This is what both authors mean by stating that the problem will "no longer produce insight anymore." The implications of this problem are beyond the scope of this document. They are the province of future work to resolve. Then, why should I discuss this claim at all?

Namely, what we are doing in specifying something in terms of an algorithm is precisely to fully explain it in terms of the procedure that would generate it. Even if the phenomena which the system outputs are in fact emergent, the programmer gives an explanation or procedure for their generation. Thus, even if the output generated is emergent from the code, the procedure for generating them seems repeatable. This provides some explanation for why computational cases of creativity feel like edge cases.

It also provides a heuristic for the perspective on computational creativity which I wish to offer. I rest the true philosophical work of this thesis in the prior sections. What comes forward should only be thought of as an intuitive sketch.

I do not think the project of "creative A.I." or computational creativity is one which

[17] This touches briefly on the issue of whether creative idea generation is "blind" or "guided," discussed earlier. The formulation here is compatible with the results in Kronfeldner (2010): there is a *partial coupling* between the Problem Templates generated and their usefulness, as their salience will be connected to past usefulness in a variety of situations. The coupling, however, will not be absolute, otherwise there would be no need for a search procedure. The relative creativity of a problem template could be quantified with Simonton (2022)'s formalization.

can be completed in a manner satisfying to us. In generating algorithmically creative phenomena, we are also specifying the results of that procedure to ourselves. That procedure will now not feel creative to us, because it will be incorporated into a prior stock of procedures to generate those emergent results. In touting a program as creative A.I. to multiple audiences, we are essentially trying to do the same insight problem twice.
As stated above, the metaphysical thesis has not been completely forgotten.

Whether or not a phenomenon is "algorithm emergent" *is* a question which stands outside of innovation history. But its significance for *us*, and its link to the phenomenology of insight and surprise, is instead related to the *discovery* rather than the *fact* of this algorithm emergence. This is the deflationary perspective which I wish to offer.

CONCLUSION

In the introduction, I noted the dearth of research in the contemporary philosophy of creativity. This gap is particularly pronounced in the case of the philosophy of cognitive science and the philosophy of artificial intelligence. I noted as well that the distant intention of the conceptual analysis provided here was to provide a framework which might form the basis for future research in this area. I would like to sketch how I see this work developing.

The thesis as it has preceded so far has given a metaphysical answer to the question of computational creativity which is quietist in spirit. Namely, that one of the conditions on which computational creativity is most plausible entails a lack of explanation between the algorithm and its behavior. This lack of explanation has certain associations of surprise for us, such that when a phenomenon is sufficiently predictable by a method, it no longer feels creative to us. This explained why instances of computational creativity often feel like edge cases.

I would like to reflect briefly on what I think this might mean for philosophical comment on computational creativity going forward. A task which was beyond the scope of this thesis, but may very well be worth pursuing, is to examine cases of "Creative A.I." more specifically to aim to identify where they seem to exhibit emergent behavior. Simultaneously, however, such investigations should emphasize that both the development of "creative A.I" and this sort of analysis is just that – a developmental process – rather than an explicit phenomenon which software engineers should aim to capture. This conception entails a *diagnostic* attitude regarding "creative A.I." currently. According to the framework I have defended: if developers argue that one should look for creativity in A.I. through A.I.'s ability to perform a specific task, philosophers should resist this conception.

An unexplored corollary of this process so far is a connection between emergence and evolution. Kronfeldner (2011) and Kronfeldner (2010) has developed severe critiques of an

oft-made comparison between creativity and evolution both in terms of BVSR and elsewhere. Likewise, there is a triad of associations between emergence and evolution, emergence and cellular automata (Chalmers, 2006), evolution and automata (Stephen Wolfram, 2002), evolution and creativity (Kronfeldner, 2011; Kronfeldner, 2010), and emergence and creativity (this thesis). Additional work in this area should disentangle these conceptual relations to make the distinctions between creativity, emergence, and natural selection precise.

In addition to computational creativity, I think the thesis leaves room for further discussion regarding agency and creativity as discussed in **Section 1.4.** I noted that while I do not this family of conditions to be plausible for creativity *in general*, the framework developed in Wolf (2016) seems so in response to H-Creativity. We thus require a more developed conceptual link between the P-Creative processes which *are* necessary for creativity and the systems of assessment and value which are necessary for H-Creativity.

I hope to continue to work on these conceptual threads in future research, both in and outside of philosophy.

BIBLIOGRAPHY

Angerer, B. (2024). *Representational Change and the Use of Metaphors in Problem Solving: The Winding Roots of Understanding*. Routledge.

Angerer, B., & Schreiber, C. (2019). Representational Dynamics in the Domain of Iterated Mental Paper Folding. *Cognitive Systems Research, 54*, 217-231. https://doi.org/10.1016/j.cogsys.2018.11.011

Bahety, A., & Olteteanu, A-M. (2019). An Approach to Computational Creation of Insight Problems Using CreaCogs Principles. In *Proceedings of the 7th International Workshop on Artificial Intelligence and Cognition* (pp.22-38).

Beaty R. E., Nusbaum, E. C., & Silvia, P. J. (2014). Does Insight Problem Solving Predict Real-World Creativity? *Psychology of Aesthetics, Creativity, and the Arts*, *8*(3), 287-292. https://doi.org/10.1037/a0035727

Berto, F., & Tagliabue, J. (2023). Cellular Automata. in E.N. Zalta, & U. Nodelman (Eds.). *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/win2023/entries/cellular-automata/

Boden, M.A. (2003). *The Creative Mind: Myths and Mechanisms* (2nd ed.). Routledge. https://doi.org/10.4324/9780203508527

Boden, M. A. (1998). What is Creativity?. In M. A. Boden (Ed.), *Dimensions of creativity* (pp. 75–117). The MIT Press

Campbell, D. T. (1960). Blind Variation and Selective Retention in Creative Thought as in Other Knowledge Processes. *Psychological Review*, *67*(6), 380–400. https://doi.org/10.1037/h0040373

Clark, A. (2013). Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science. *The Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Chalmers, D.J. (2006). Strong and Weak Emergence. In Philip Clayton & Paul Davies (Eds.), *The Re-Emergence of Emergence: the Emergentist Hypothesis from Science to Religion.* Oxford University Press.

Chemero, A. (2003). An Outline of a Theory of Affordances. *Ecological Psychology, 15*(2), 181–195. https://doi.org/10.1207/S15326969ECO1502_5

Chomsky, N. (1991). Linguistics and Cognitive Science: Problems and Mysteries. In Asa Kasher (Ed.), *The Chomskyan Turn* (pp. 26-53). Oxford University Press.

Chomsky, N. and M. Foucault (1971). *Debate between Noam Chomsky and Michel Foucault:On Human Nature* [Video]. Youtube. https://www.youtube.com/watch?v=3wfNl2L0Gf8

Crane, T. (2001). The Significance of Emergence. In Barry Loewer & Grant Gillett (Eds.), *Physicalism and its Discontents* (pp. 207-224). Cambridge University Press.

Dennett, D.C. (1991). Real Patterns. *Journal of Philosophy*, 88 (1), 27-51

Dietrich, A. (2024). Where in the Brain is Creativity: The Fallacy of a Creative Faculty in the Brain. *Frontiers in Psychology, 15*. https://doi.org/10.3389/fpsyg.2024.1373299

Dietrich, A. (2019a). Where in the Brain is Creativity: a Brief account of a Wild-Goose Chase *Current Opinions in the Behavioral Sciences, 27*, 36–39. https://doi.org/10.1016/j.cobeha.2018.09.001.

Dietrich, A. (2019b). Types of Creativity. *Psychological Bulletin Review, 26*, 1–12. https://doi.org/10.3758/s13423-018-1517-7

Duncker, K. (1935). *Zur Psychologie des produktiven Denkens*. Verlag Julius Springer.

Estes, Z., & Ward, T. B. (2002). The Emergence of Novel Attributes in Concept Modification. *Creativity Research Journal*, *14*(2), 149–156. https://doi.org/10.1207/S15326934CRJ1402_2

Fleck, J., & Weisberg, R. (2013). Insight versus Analysis: Evidence for diverse methods in problem solving. *Journal of Cognitive Psychology*, *25*(4), 436-463. https://doi.org/10.1080/20445911.2013.779248

Gaut, B. (2018). The Value of Creativity. In Berys Gaut & Matthew Kieran (Eds.), *Routledge Handbook on Creativity and Philosophy* (pp. 213-29). Routledge.

Gaut, B (2010). The Philosophy of Creativity. *Philosophy Compass, 5* (12), 1034-1046. https://doi.org/10.1111/j.1747-9991.2010.00351.x

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception: Classic Edition.* Houghton Mifflin.

Guilford, J. P. (1967). *The Nature of Human Intelligence*. MacGraw Hill.

Hausman, C.R. (1975). *A Discourse on Novelty and Creation*. Markus Nijhoff.

Hintikka, J. (1997). On Creativity in Reasoning. In M.A. Boden (Ed.), *The Complexity of Creativity* (pp. 67-78). Springer.

Kant, I. (2000). *Critique of the Power of Judgment*. Paul Guyer (Ed.), Paul Guyer and Erin Matthews (Trans.). Cambridge University Press.

Kaplan, C. A., & Simon, H. A. (1990). In Search of Insight. *Cognitive Psychology*, *22*(3), 374–419. https://doi.org/10.1016/0010-0285(90)90008-R

Kieran, M (2015). Creativity. In Noel Carroll and John Gibson (Eds.) , *The Routledge Companion to the Philosophy of Literature*. Routledge.

Kim, J. (2006). Emergence: Core ideas and issues. *Synthese, 151* (3), 547-559. https://doi.org/10.1007/s11229-006-9025-0

Kirsh, D. (2009). Problem Solving and Situated Cognition. In Phillip Robbins (Ed.),

The Cambridge Handbook of Situated Cognition (pp.264-306). Cambridge University Press.

Knoblich, G., Ohlsson, S. & Raney, G.E (2001). An Eye Movement Study of Insight Problem Solving. *Memory & Cognition*, *29*, 1000–1009. https://doi.org/10.3758/bf03195762

Kronfeldner, M. (2018). Explaining Creativity. In Berys Gaut & Matthew Kieran (Eds.), *Routledge Handbook on Creativity and Philosophy* (pp. 213-29). Routledge.

Kronfeldner, M (2011). *Darwinian Creativity and Memetics*. Acumen Publishing.

Kronfeldner, M. (2010). Darwinian 'Blind' Hypothesis Formation Revisited. *Synthese, 175* (2), 193--218. https://doi.org/10.1007/s11229-009-9498-8

Kronfeldner, M. (2009). Creativity Naturalized. *Philosophical Quarterly, 59* (237), 577-592. https://doi.org/10.1111/j.1467-9213.2009.637.x

Marr, D. (1982). Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. The MIT Press.

Martins M.D., Laaha S, Freiberger E.M., Choi S, & Fitch W.T. (2014). How Children Perceive Fractals: Hierarchical Self-Similarity and Cognitive Development. *Cognition, 133*(1), 10-24. https://doi.org/10.1016/j.cognition.2014.05.010

Mednick, S.A. & Mednick, M.T. (1959, 1962). Remote Associates Test, college and adult form.

Newell, A., & Simon, H. A. (1972). *Human Problem Solving.* Prentice-Hall.

Ohlsson, S. (1992). Information-Processing Explanations of Insight and Related Phenomena. In K.J. Gilhooley (Ed.), *Advances in the Psychology of Thinking* (pp. 1-44). Harvester-Wheatsheaf.

Olteteanu, A-M. (2020). *Cognition and the Creative Machine: Cognitive AI for Creative Problem Solving*. Springer.

Richards, R., & Goslin-Jones, T. (2018). Everyday Creativity Challenges for Self and World— Six Questions. In R. J. Sternberg & J. C. Kaufman (Eds.), *The Nature of Human Creativity* (pp. 224–245). Cambridge University Press. https://doi.org/10.1017/9781108185936.017

Runco, M.A. (2023). Updating the Standard Definition of Creativity to Account for the Artificial Creativity of AI. *Creativity Research Journal*, *37*(1), 1-5. https://doi.org/10.1080/10400419.2023.2257977

Runco, M. A., & Jaeger, G. J. (2012). The Standard Definition of Creativity. *Creativity Research Journal*, *24*(1), 92–96. https://doi.org/10.1080/10400419.2012.650092

Russell, S. and Norvig, P. (2010). *Artificial Intelligence: A Modern Approach*. 3rd Edition, Prentice-Hall.

Simonton, D.K. (2022). The Blind-Variation and Selective-Retention Theory of Creativity: Recent Developments and Current Status of BVSR, *Creativity Research Journal, 35* (3), 304-323. https://doi.org/10.1080/10400419.2022.2059919

Spencer-Smith, R. (1995). Reductionism and Emergent Properties. *Proceedings of the Aristotelian Society*, *95*, 113–129. https://doi.org/10.1093/aristotelian/95.1.113

Steffensen, S.V. (2013). Human Interactivity: Problem Solving, Solution Probing, and Verbal Patterns in the Wild. In Stephen J. Cowley & Frédéric Vallée-Tourangeau (Eds.), *Cognition Beyond the Brain: Interactivity, Computation and Human Artifice* (pp. 85- 114). Springer Verlag.

Stokes, D. (2024). Empirical Treatments of Imagination and Creativity. In Amy Kind & Julia Langkau (Eds.), *Oxford Handbook of Philosophy of Imagination and Creativity*. Oxford University Press.

Stokes, D. & Paul, E.S. (2021, November 11th). Computer Creativity is a Matter of

Agency. Institute of Arts and Ideas. https://iai.tv/articles/ai-will-never-rival-picasso-auid-1971?_auid=2020

Wallas, G. (1926). *The Art of Thought.* Harcourt, Brace and Company.Watson, G. (1996). Two Faces of Responsibility. *Philosophical Topics*, *24*(2), 227–248. https://doi.org/10.5840/philtopics199624222

Weisberg, R. W. (2015). On the Usefulness of "Value" in the Definition of Creativity. *Creativity Research Journal*, *27*(2), 111–124. https://doi.org/10.1080/10400419.2015.1030320

Weisstein, E.W. (2022). "Tower of Hanoi." MathWorld--A Wolfram Web Resource. https://mathworld.wolfram.com/TowerofHanoi.html

Wiese, W. & Metzinger, T. (2017). Vanilla PP for Philosophers: A Primer on Predictive Processing. Philosophy and Predictive Processing. http://dx.doi.org/10.15502/9783958573024

Wiggins, G.A. (2020). Creativity, Information, and Consciousness: The Information Dynamics of Thinking. *Physics of Life Reviews, 34–35*, 1–39. https://doi.org/10.1016/j.plrev.2018.05.001

Wiggins, G.A. (2006). A Preliminary Framework for Description, Analysis, and Comparison of Creative Systems. *Knowledge-Based Systems, 19*, 449–458. https://doi.org/10.1016/j.knosys.2006.04.009

Winning, J. & Bechtel, W. (2018). Being Emergence vs. Pattern Emergence: Complexity, Control, and Goal-Directedness in Biological Systems. In Sophie Gibb, Robin Findlay, Hendry & Tom Lancaster (Eds.), *The Routledge Handbook of Philosophy of Emergence* (pp. 134-144). Routledge.

Wolf, S. (2016). *Aesthetic Responsibility* [Conference Presentation]. The Amherst Lectures in Philosophy 11. Amherst College, Amherst, MA, United States.

https://www.amherstlecture.org/wolf2016/wolf2016_ALP.pdf

Wolfram, S. (2002). A New Kind of Science. Wolfram Media Inc.

Worthen, B. R., & Clark, P. M. (1971). Toward an Improved Measure of Remote Associational Ability. *Journal of Educational Measurement*, *8*(2), 113-123. https://doi.org/10.1111/j.1745-3984.1971.tb00914.x