



Creativity as potentially valuable improbable constructions

Mark Fedyk¹  · Fei Xu²

Received: 12 November 2019 / Accepted: 11 December 2020 / Published online: 13 January 2021
© Springer Nature B.V. 2021

Abstract

We argue that creative ideas are potentially valuable improbable constructions. We arrive at this formulation of creativity after considering several problems that arise for the theories that suggest that creativity is novelty, originality, or usefulness. Our theory avoids these problems. But since we also derive our theory of creativity from the scientific commitments of a more general theory of cognitive development, a theory called *rational constructivism*, our theory is unique insofar as it explains creativity in both adults and children through reference to a set of computational mechanisms that have been posited on the basis of independently plausible experimental research.

Keywords Creativity · Rationality · Rational constructivism · Developmental psychology · Naturalism · Normativity · Learning

1 Introduction

More than a few psychological theorists of creativity have said that creativity requires ideas that, in combination with one another, are both original and have utility (Runco and Jaeger 2012; Simonton 2018). Can a psychological theory of creativity exemplify these two virtues, and thereby exemplify what it aims to explain?

This article belongs to the Topical Collection: Creativity in Art, Science & Mind
Guest Editors: Creativity in Art, Science & Mind

✉ Mark Fedyk
mfedyk@ucdavis.edu

Fei Xu
fei_xu@berkeley.edu

¹ Betty Irene Moore School of Nursing & School of Medicine (Bioethics), University of California, Davis, 2450 48th St #2652, Sacramento, CA 95817, USA

² Department of Psychology, University of California, Berkeley, 2121 Berkeley Way, Berkeley, CA 94704, USA

In this paper, we try. We first set for ourselves four intellectual hurdles that, if cleared by our theory, are evidence of its scientific utility. The originality of our theory then comes from how we avoid a series of problems that confront the thesis that either novelty, utility, originality, or usefulness is sufficient for creativity [c.f. (Diedrich et al. 2015; Simonton 2013, 2018; Weisberg 2015)]. Something else is required.

But saying *what* else creativity consists of forces us to dig into some deep, and in places messy, philosophical dirt. This digging nevertheless yields a theory of creativity that combines aspects of past psychological theories with some new ideas of our own. We argue that the minds of children and adults both form creative ideas by computing an idea's posterior probability and projecting its potential value. Put another way, our theory of creativity holds that creative ideas are *potentially valuable improbable constructions*, and that sometimes occurrences of creativity are a sui generis kind of instrumental rationality. But the rationality of creativity is not linked just to the role it can play in building increasingly complex patterns of reliable inference or belief. According to our theory, the rationality of creativity is grounded in its ability to both *facilitate* and *express* the by-products of increasingly sophisticated patterns of learning.¹

1.1 What is the point?

Why do we need a new psychological theory of creativity? First, many psychological theories of creativity lack important philosophical virtues that are characteristic of good scientific theories. For example, even though psychologists have become skilled at measuring creativity, they have not developed theories of what is being measured that have the theoretical plausibility which matches their often impressive experimental sophistication.² There are several hundred articles reporting the results of versions of the Alternative Uses Test.³ This test typically asks experimental participants to generate useful, original, or novel uses of an everyday object like a box of tissues or a wooden block. Data from experiments based upon this test give the impression that creativity can be safely operationalized as novelty or originality or usefulness. As we will argue below, this is a mistake. But it is a mistake that emerges clearly only when we try to situate the thesis that creativity is (that is, is adequately operationalized as) novelty, or originality, or usefulness, among a network of independently plausible scientific, philosophical, and psychological commitments. The point, then, is that many existing psychological theories of creativity lack the specific philosophical virtue of being theoretically plausible. Taken on their own, existing theories of creativity frequently seem attractive; but when we try to situate them into a psychological framework that

¹ Here, we follow Bouwmeester, who distinguishes between two types of instrumental rationality: generative and expressive. Cognitive processes are generative if they produce more accurate representation; expressive if they find a way to use these representations for some independently valuable end (Bouwmeester 2017, p. 36).

² Counterexamples to this generalization are Boden and Simonton (Boden 2004; Simonton 2013, 2018). Key elements of our theory are inspired by the respective work of Boden and Simonton.

³ The alternative uses task was discussed by Guilford in a connection with his factor analysis of human intelligence — see, e.g., “Utility Test (flexibility)” in (Guilford and Hoepfner 1966), extending his earlier theoretical work (Guilford 1950). Since then it has been widely adopted by psychologists as either the “alternative uses task”, the “alternative uses test”, or generic “tests of divergent thinking”. Unfortunately we cannot find a review of historical study of this experimental lineage.

includes, e.g., a commitment to the computational theory of mind or the fact that children and adults are both creative, problems arise.

A focus on variations of a small set of standardized experimental protocols that aim to measure creativity has largely failed to illuminate the connection between creativity and learning. Creativity is one of the primary drivers of learning. Learning is a messy, long, social, piecemeal, and rarely linear process. Learning does not easily lend itself to experimental study, and the kinds of creativity that are easy to study experimentally are not the kinds of creativity that are most valuable to real-world learners. We hope to add some clarity to the connection between learning and creativity by stepping back from the details of experimental studies of creativity and asking, conceptually: can we make sense of the idea that *creativity can both express and facilitate learning*? Again, as per our remarks above, one of the virtues of our theory is that it answers this question in the affirmative.

We also think it is important that philosophical theories of mental phenomena like creativity — and also knowledge, reasoning, feelings, etc. — have secure psychological foundations. By this we mean more than ensuring that the concepts in a theory are logically consistent with some formulation of physicalism or another. Instead, the stronger standard — the one that we try to achieve in this paper — is to derive novel concepts and definitions from an independently plausible scientific theory that, when assembled, support substantive, original philosophical conclusions. This of course does not show that these conclusions are *true*. But it does mean that the relevant philosophical conclusions have a non-trivial amount of scientific plausibility. Accordingly, we use a theory in developmental psychology called rational constructivism to provide the scientific grounding for our theory of creativity.

Let us introduce rational constructivism, then. Rational constructivism (Xu 2011, 2019; Xu and Kushnir 2012, 2013) is a theory in developmental psychology that charts a middle path between nativism/rationalism and connectionism/empiricism. It does this by hypothesizing that all learning is caused by three cognitive mechanisms that are computational in nature: language acquisition and symbolic/lexicalized cognition, Bayesian induction and hypothesis selection, and constructive thinking. The psychological substrates of these mechanisms are more or less innate [cf. (Fedyk and Xu 2019)], but the causal powers of these mechanisms transform dramatically over developmental time. Even at the neonatal stage they are profoundly influenced by both the learner's environments and the people in these environments. Furthermore, as a child grows, she gains increasingly impactful ways of intentionally shaping and influencing the operation of the relevant mechanisms; one of the most distinctive axioms of rational constructivism, thus, is that the intentional agency of learners is an important cause of learning and related cognitive behaviour (Fedyk and Xu 2018). The function of the mechanisms may change over developmental time, but according to rational constructivism, the same set of foundational cognitive mechanisms generates cognition in infants, children, teenagers, and adults.

Rational constructivism is committed to a computational mental ontology (meaning, the mind is an information processing system, and that cognitive processes are therefore defined in terms of their informational inputs and outputs), and rational constructivism is also, as we said, a theory of cognitive development. These two features of rational constructivism allow our theory of creativity to offer two philosophical innovations relative to many other psychological theories of creativity: compatibility with both the

computational mental ontology and consistency with the basic developmental fact that children and adults are both genuinely creative. Or another way of stating the problem we see for some popular psychological theories of creativity that we noted above is that these theories are not usually grounded in a simultaneously computational *and* developmental framework. This paper can be read as an (original, and hopefully useful) effort to work out what a computationally and developmentally plausible theory of the psychology of creativity can look like.

We therefore take children's creativity as given (cf. Mottweiler & Taylor 2014). More specifically: we assume that creativity is cognitively general in children: anything that children can think about (by the age of four), they can think creatively about. However, we do not also hold the position that the creativity of children is the same as the creativity of adults. While we think that any good psychological theory should account for as many facts of human cognitive development as possible, and do so by positing the fewest number of underlying cognitive mechanisms as possible, this does not mean that the cognitive mechanisms always will operate in a fixed and non-malleable fashion over the course of cognitive development. We think that this happens in creativity: as children age, creativity expands from being directed towards the acquisition of new knowledge to being directed towards both the acquisition of new knowledge and the expression of acquired expertise.⁴ Thus, another philosophical virtue of our theory of creativity is that it can offer an explanation of both children's and adults' creativity that does not require positing different cognitive mechanisms to explain child and adult creativity respectively.

Finally, we aim to address one last lacuna in some existing psychological theories of creativity: these theories frequently leave the normativity of creativity unexplained. One of the benefits of grounding our theory of creativity in rational constructivism is that we can offer a computational, developmentally plausible analysis of the normativity of creativity. As we mentioned above, we do this by analyzing how creativity can sometimes exemplify a *sui generis* form of instrumental rationality which we, uncreatively, term "creative rationality".

1.2 A map of the paper

Here is how our efforts to accomplish these various aims are organized. In section 2 we introduce the four philosophical hurdles—or, as we will now start calling them, principles—that we believe any good psychological theory should satisfy. Then, in section 3, we investigate the possibility of reducing creativity to originality, novelty, or usefulness (either individually or in some combination), showing that any such reduction is incompatible with the principles we introduce in section 2. This investigation also motivates our attempt to formulate a theory of creativity that can satisfy the four philosophical principles. We spell out the details

⁴ The claim we are making here is easily confused with Alison Gopnik's analysis of play and exploration in early childhood (Gopnik 2020). We all agree that there is a process of refinement that characterizes most people's learning over developmental time. Gopnik offers a computational analysis of this refinement, as do we. But she does not share our efforts to extend this computational analysis into an epistemology of creativity, by linking the specific cognitive mechanisms (see below) that we posit in order to explain creativity with an explanation of the normativity of creativity (again, see below).

of our new theory in sections 4 and 5. While there are many moving parts, ultimately the story we tell boils down to showing how a theory of creativity that is able to satisfy the four principles is a theory that can be derived from a computational theory of mind, namely rational constructivism. The upshot, as we say, is what aims to be an original, useful theory of creativity — and, for that, one that is scientifically plausible, philosophically innovative, computationally tractable, accounts for the normativity of creativity and in so doing helps to clarify the connection between creativity and learning, and is consistent with some basic facts about human cognitive development.⁵

With apologies, then, the details of our theory, its motivations, and its supporting arguments are complicated—but, so far as we know, no one has yet argued that creativity requires simplicity.⁶

2 Four Jamesian principles

When looking for criteria able to demarcate good theory from bad in any scientific discipline, it can be helpful to look to the discipline's philosophical foundations. We begin therefore by extracting three principles from William James's *Principles of Psychology* that, when taken together, capture some of his views about what makes for a good psychological theory. We will also propose a fourth principle that is Jamesian in its spirit, though not letter.

The first principle is that a good psychological theory should offer a causal explanation of whatever mental phenomena are most characteristic of the cognitive life of humans, and this explanation must situate the relevant mental phenomena somewhere in the natural world. James begins his *Principles* by apologizing for the length and complexity of his treatise, and he then explains that what little concision *Principles* has comes from his commitment to sticking to “the point of view of natural science” for most of his book. For James, the point of view of natural science consists of the following three interlocking axioms: “Psychology, the science of finite individual minds, assumes as its data (1) *thoughts and feelings*, and (2) *a physical world* in time and space with which they coexist and which (3) *they know*.” (James 1918a) James then argues that the goal of psychological science is to explain how, through causal interactions with the physical world, “thoughts and feelings” emerge and then eventually crystalize into knowledge-bearing entities. Crucially, James requires that these theories account for both the causes and effects of “thoughts and feelings”: he writes, “mental phenomena are not only conditioned *a parte ante* by bodily processes; but they lead to them *a parte post*... Our psychology must therefore take account not only of the conditions antecedent to mental states, but of their resultant consequences as well.”

⁵ Briegel and colleagues have developed a theory of creativity that is similar in its intended scope and computational tractability — see (Briegel 2012; Briegel & De las Cuevas, 2012; Hangl et al. 2017). Briegel et al. also found themselves positing a strong concept of agency in order to explain creativity, which is one of the ways our theory is deeply similar to their earlier work

⁶ Frank Barron (Barron 1963) reports a positive association between creativity and a preference for complexity. If this is right, our theory should appeal to our most creative readers.

But, if it is interpreted as a *sui generis* category, “thoughts and feelings” is too vague and abstract to be useful as a starting point for empirical, scientific inquiry. It does not name a single category of mental stuff that can be individuated according to a distinctive causal pattern. James consequently breaks the category down, initially into “feelings, desires, cognitions, reasonings, decisions, and the like” (James 1918a, p. 1), and as he gets deeper into *Principles*, relatively smaller categories of analysis like “will” (James 1918a, p. 128), “concepts” (James 1918a, p. 266) and “abstraction” (James 1918a, p. 470) emerge. We think “creativity” should also be included to the list,⁷ and if you will concede as much, then when applied to psychological analyses of creativity, James’ first principle is the first of our four principles: a good scientific explanation of creativity should be a causal theory which explains how creativity is situated both amongst other “thoughts and feelings” and also within the physical world.

The remaining two Jamesian principles are more demanding. James insists in several places that a good psychological theory should be consistent with the fact that variations in the behavior of most mental phenomena of interest to scientific psychology are common enough that the existence of these variations should be consistent with any acceptable scientific psychological theory. Consider what James has to say about memory:

Why should this [...] faculty retain so much better the events of yesterday than those of last year, and, best of all, those of an hour ago? Why, again, in old age should its grasp of childhood’s events seem firmest? Why should illness and exhaustion enfeeble it? Why should repeating an experience strengthen our recollection of it? (James 1918a, pp. 2–3)

The second Jamesian principle is that a good psychological theory should be consistent with—and, ideally, *explain*—the variability in the mental phenomena that the theory aims to be about. An important corollary of this principle is that psychological theories should not require the faculties and mechanisms that they posit to operate *optimally* (or rationally, or appropriately, or ideally, etc.) *all* or even *most* of the time. We cannot explain memory, for instance, by saying it is the faculty that retrieves stored information that, upon retrieval, is as accurate as when it was first stored; since memory too frequently fails to work even remotely close to this ideal, constructing an explanation of memory as the faculty which satisfies this ideal would leave too much of memory’s variable nature a scientific mystery. Analogously, any theory of creativity should be consistent with the fact that the mechanisms that produce it exhibit a non-trivial amount of variability in their respective behavior.⁸

⁷ James does not mention creativity, and he uses the word “creative” only once across both volumes of his *Principles*. In speaking of the phenomenology of coming to a decision made on the basis of deliberation, James writes “... we feel, in deciding, as if we ourselves by our own wilful act inclined the beam; in the former case by adding our living effort to the weight of the logical reason which, taken alone, seems powerless to make the act discharge; in the latter by a kind of creative contribution of something instead of a reason which does a reason’s work.” (James 1918b, p. 534)

⁸ This principle places an important constraint on computational theories of psychological mechanisms: computational theories must account *in computational terms* for all of the observed behaviour of a given mental process of faculty, and not just claim that a computational theory has explained the relevant phenomena when only some (indeed, quite possibly a very small fraction of) the relevant phenomena can be described as approximating certain algorithms. In practical terms, this likely means that algorithmic explanations need to be supplemented by different kinds of causal explanations.

These remarks also show why it is important not to confuse variability with normativity, which is the subject of our third Jamesian principle.⁹ James believed that it was important for psychology to account for the mind's capacity for normativity—that is: the ability to perform cognitive processes (or tasks, operations, inferences et cetera) *correctly* (or appropriately, rightly, or badly, rationally, et cetera). For example, here is James' gloss of what it is to reason: “To reason [...] we must be able to extract characters,—not any characters, but the *right* characters for our conclusion.” (James 1918b, p. 343) (emphasis ours) He continues to offer an account of the mind's ability to learn to pick out the “right”—that is: normatively appropriate—characters. Elsewhere, James also takes up the question of the cause of a “good” or “moral” character, and much of James' discussion of instinct and habit reflects a concern with deep philosophical questions about the foundation of normative behavior.

2Jamesian scientific psychology does therefore not deny that aspects of the mind exhibit normativity. This distinguishes Jamesian psychology from more recent flavors of reductionism about the cognitive system (e.g., behaviorism, connectionism) that usually deny, or downplay, the significance of the fact that cognition is normative. By comparison, James has a more ambitious proposal: the normativity of the mind must be part of the complex causal story we tell about mind-world-body interactions, rather than eliminated through choosing a simpler story, namely, one referring to only world-body interactions.

Our fourth and final principle has no direct connection with the text of *Principles*, even though we like to imagine that James would have endorsed it had he access to the bulk of 20th and twenty-first century psychological science.¹⁰ This is the principle that a psychological theory should be *developmentally parsimonious*: if a mental ability or psychological capacity can be observed in both young children and adults, then, all things being equal, it is best to explain the capacity with reference to only one set of underlying computational mechanisms, as opposed to positing two or more sets of mechanisms in order to explain any developmental differences. This is a challenging principle, as developmental differences in the capacities of the mind are the rule, not the exception. And since creativity is amongst the earliest complex cognitive abilities to come ‘on-line’ in development, a theory of creativity that is consistent with this

⁹ We believe that it is likely that a confusion between variability and normativity is endemic in contemporary cognitive science, and that this explains why epistemology and cognitive psychology have not converged, as Quine predicted, over the last half-century (Antony 2018; Quine 1969). To a good first approximation, psychologists, being experimentalists, are interested in studying non-overlapping distributions – and so are inherently interested in variations. Philosophers, by contrast, are interested in only a very small range of the total possible variability in the operation of any cognitive mechanism – the space in which the mechanism is operating *rationally*, or as close to ideally as possible. If this is right, then there is a sense in which psychologists and philosophers might be talking about the same cognitive mechanisms, but that impression largely misses the point. The philosophers are interested in phenomena that usually occur several (3? 10?) standard deviations rightward of a central tendency, while psychologists are interested in phenomena tightly clustered around the central tendency. If this is right, then there is little focal overlap between the two disciplines. However, since there is variability even when the mechanisms are performing optimally, there is no reason why there cannot be experimental work which studies variability of optimal (or rational) performance; likewise, there is no reason why there cannot be causal explanations of how it is that the mind produces its rationality. What may seem like a fundamental conceptual canyon could really be a by-product of a canalized history of choices about tasks and sample populations that is not favorable to studying the causal basis of optimal or rational cognition.

¹⁰ James on the moral wisdom of children: “Could the young but realize how soon they will become mere walking bundles of habits, they would give more heed to their conduct while in the plastic state.” (James 1918a, p. 127)

principle should, by referring to a single set of underlying causal mechanisms, offer a unified explanation of the nature of creativity in both adults and children.

Thus, collecting them all together, we have the following list of Jamesian principles:

1. Psychological science should aim to produce *causal* explanations of the mental phenomena most characteristic of human life, where these explanations situate the relevant phenomena in the physical world.
2. Psychological explanations should take into account the natural variability of mental phenomena, as opposed to ignoring it, or analyzing it away.
3. Psychological explanations should account for the normativity of mental phenomena, if the relevant phenomena characteristically exhibit normativity.
4. Psychological explanations should be constrained by the principle of developmental parsimony: if children have an ability that adults also have, it is better, all things being equal, to explain both sets of abilities by reference to a single underlying cluster of mechanisms, as opposed to two or more clusters of substantially different clusters of mechanisms.

These are the four principles that we aim to satisfy with our theory of creativity. Readers will also soon see that a commitment to the computational theory of mind (Chomsky 2009; Fodor 1990; Turing 1950) functions in our argument like a fifth principle; we do think that any good psychological theory should have a computational ontology. However, it would be anachronistic to attribute such a commitment to James.

3 Against the thesis that novelty, originality, or usefulness are definitive of creativity

Before turning the details of what we think creativity is, we will argue that a sequence of problems imperils any attempt to reduce creativity to novelty, originality, or usefulness. [cf. (Albert and Runco 1999; Gaut and Kieran 2018)]. Our argument is that any such attempt is not compatible with the first and the fourth Jamesian principles, plus the computational theory of mind.

This investigation is relevant for two reasons. As we've mentioned, it helps to set up our theory. But it also helps clarify what can and cannot follow from the practice of studying creativity by measuring novelty, usefulness, and/or originality. Indeed, many psychologists can be interpreted as operationalizing creativity as thinking that has one or more of the following properties:

Novelty¹¹ and/or Usefulness

(Amabile 1996; Feist 2006; Guilford 1950; Kaufman and Baer 2004; MacKinnon 1970; Runco 2004; Simonton 2008;

¹¹ Many experimentalists distinguish between “novelty” and “originality” in something like the following way. Novelty is often measured by the novel uses test which takes something that has a known use (a hammer, a Kleenex) and asks people to put it to new (i.e. nonstandard) uses. Originality means “brand new” as far as that person is concerned, as per Boden’s concept of *psychological creativity* (see below). So novel uses are often original uses of non-original things. But sometimes what we want to talk about are original *ideas* (i.e. ideas that have not occurred before to that person). These are “original” ideas, not “novel” ideas in this schema, because being new ideas, they can’t be nonstandard applications of prior beliefs.

	Sternberg 1988; Szen-Ziemiańska et al. 2017)
Originality	(Abraham 2013, 2014; Acar et al. 2017; Runco 1988, 2004; Runco et al. 2005; Runco and Charles 1993)
Feeling of insight or feeling an “aha” moment, or surprisingness	(Simonton 2018; Thagard 2002, 2019; Thagard and Stewart 2011)

While we agree that occurrences of creative thought frequently can be characterized by these properties, we aim to show that the thesis that novelty, usefulness, or originality — either individually or in some logical combination — is sufficient for creativity is false, *if* the computational theory of mind is also assumed, children’s creativity is taken as a given, and the first and fourth of our Jamesian principles apply. (Because we will not argue for this claim, we hope you will grant that feelings of “aha” are not individually sufficient for creativity.)

We will illustrate the problems by examining novelty first.¹² The question of how determination of novelty can be caused by psychological processes, for example, how agents realize that some nonstandard use of a familiar object *is novel*, is ultimately the question of how cognitive processes could compute that some mental representation is novel. Here, the question is how someone’s mind could produce a novel thought that is not only novel as far as that agent is concerned – or “*to*” or “*for*” that agent, or the computation process that subserves her cognition – but how they can correctly categorize (compute) it as such. Novelty is a relational property. Let w be the content of some mental representation. Then, w is novel only relative to a set of mental representations the content of which is past examples of things more or less similar to w . Given that, it is hard to see how a child could compute that a representation of hers is novel without having somewhere in her mind a sufficiently large set of past comparators. That is: computing that some w is novel cannot be done without having access to a rich store of previously acquired knowledge that serves as input for the relevant algorithms. But this pre-existing knowledge of comparators cannot plausibly be attributed to young children, which means we lack a computational explanation of their creativity.

So, the “novelty-compared-to-someone’s-existing-beliefs” definition of novelty does not apply to children. Perhaps a simpler definition of novelty is called for. We could instead say that a mental representation is novel only if exactly the same content does not already exist in a person’s memory. To keep things clear, call this second definition of novelty the “not-exactly-the-same-as-something-else” conception of novelty (“NETSASE novelty”). Checking for NETSASE novelty does not require the mind to compute whether something is like, but not the same, as each member of a set of other things. The mind’s algorithms only must check for exact sameness. However, if NETSASE novelty is sufficient for creativity, then a random idea generator is creative, and perhaps maximally so. Or, to put the same point another way, there is no difference in the creativity of a completely unexpected one-page solution to the Riemann hypothesis and a page of paper filled with exactly the same number of randomly generated words and numbers.

The reason why a random idea generator is not creative — and thus why NETSASE novelty is not creativity — is that there is a hue of practical rationality to creativity.

¹² Since our project is one in Jamesian psychology not metaphysics, we are not concerned with novelty simpliciter but how novelty emerges in the mind and is categorized accurately by both children and adults.

Occurrences of creativity frequently have a point or a purpose; they make sense in terms of personal experience, either past or present; or if not that, they make sense in the context or social setting in which they occur. For example, a person rapidly and wildly guessing answers to a technical problem seems to exhibit much less creativity than the person who thinks quietly for a few hours and then comes up with a brilliant solution. These facets of creativity would be rendered mysterious if creativity were just randomness. Again NETSASE novelty is not creativity.

At this point, someone may be tempted to think that these problems can be avoided by saying that it is not, for instance, novelty *alone* that is sufficient for creativity. Instead, it is novelty *plus* some additional criterion that makes creativity.

That does not help. It is true that many definitions of creativity define it as the joint instantiation of a set of properties. However, in that case we still must account for how it is that a joint instantiation of novelty and x could be caused by psychological mechanisms that are computational in nature and which do not depend for their operation on mental representations that children most likely lack. It therefore must also be the case that either there is an algorithm which can compute *jointly* the presence of novelty + x , or, if not that, then an algorithm which computes novelty and another which computes x , and some further process that fuses the output of these two algorithms together to form a creative idea or judgment. Either way, we do not avoid the problem of imputing to children's mind a computation task that it lacks the resources to accomplish. The proposal that creativity is a complex property with novelty as a necessary component does not help us avoid the question of how novelty can be computed in the mind of a young child.

The same difficulties apply to the proposal that originality is sufficient for creativity. When it is discussed in the psychological literature, originality is operationalized as being stronger than novelty: not merely a novel use of a familiar object, but involving a thought or action entirely brand new as far as that person is concerned. And that is why the same criticism as before applies: originality is also a relational property, so judgements of originality must be based upon some comparison with pre-existing knowledge, and originality is therefore something that a person can accurately detect only after a non-trivial amount of learning. Thus, if originality is sufficient for creativity, then creativity should not be observable in people without the relevant histories of learning. But again, creativity is easy to find in the minds of young children—children, just like adults, can have creative thoughts about topics that they know very little to nothing about.

As before, someone may be tempted to say that there is NETSASE originality, which consists simply in checking whether or not a new mental representation is the same as any existing representation in memory. As before, this would make a random idea generator a creative idea generator — but that is implausible for the same reasons as before. Alternatively, someone may want to claim that creativity is a complex property of which originality is an essential logical component. As before, that still means that a mind still must compute judgments of originality in order to manifest its creativity: a child's mind would still have to compute whether or not some w is original in order to manifest a creative inference or expression.

Finally, the almost same sequence of problems applies to the hypothesis that usefulness is sufficient for creativity. According to this definition, creativity is not the *observation* that something is useful made only after something has been put to use.

Creativity is here the *ex-ante* insight that something has a particular use: the ability to find one particular use out of the (if you think about it) infinite range of options that exist. The explanandum is the ability to see (before trying) that a stick can be used (not as if, but literally as) a sword. As before, we face the difficulty of working out how a mind without a rich store of prior learning can determine *ex ante* that something is (or is not) useful. If usefulness is sufficient for creativity, then many of the seemingly most creative ideas of young children are not really creative, because children cannot be expected to know what, if any, the potential uses of their ideas are. And this is true even recognizing that much of a child's undirected, seemingly random, is "useful" for her learning — in the sense that, strictly speaking, exploration contributes causally to learning.

If that line of reasoning is unconvincing, then there are other, further objections that can be made against the hypothesis that usefulness is sufficient for creativity. For instance: making usefulness individually sufficient for creativity excludes from the ambit of creative thought most instances of mind wandering (Gray et al. 2019). Furthermore, the history of science provides many examples of scientific discoveries that at the time of their invention did not seem particularly useful but were still considered creative—e.g., Riemannian geometry or the drug Sildenafil (Tiefer 2006; Wiredu 1970). Obviously many creative ideas are useful, but it seems to be a mistake to think that usefulness is sufficient for creativity.

Margaret Boden distinguishes between personal creativity and historical creativity, which she calls "P-creativity" and "H-creativity". A H-creative idea is "one which (so far as is known) no one had ever thought of before" (Boden 2004, p. 233), while a P-creative idea is one that its thinker has never thought of before, whether or not it is H-creative. All H-creative ideas are P-creative, but most P-creative ideas are not H-creative. Do the objections above apply to Boden's distinction? They do. Because all NETSASE ideas would be P-Creative ideas, equating creativity with P-creativity, which is the more basic of Boden's two concepts, fails to account for the practical rationality of creativity (see section 5.4 for more on this).^{13,14}

It seems to be a mistake to think that either novelty, originality, or usefulness are sufficient for creativity. These conceptions of creativity are not compatible with the first and fourth Jamesian principles plus the computational theory of mind. That said, it may be that the four principles we have set for ourselves are ultimately undesirable for any number of philosophical or scientific reasons. We certainly do not have the space here to say that psychological science *must* follow our Jamesian principles. Still, it is our impression that many psychologists are practicing their science using methods and techniques that are consistent with James' first principle [cf. (Harre 2002; Stillings et al. 1995; Thagard 2005)¹⁵], and the principle of developmental parsimony should be

¹³ We nevertheless think that the distinction between P-creativity and H-creativity is important: according to our theory of creativity, an expert is someone whose creativity is frequently H-creative. According to our view, the expert's creativity isn't creative *because* it has never been thought of before. But we still try to capture the underlying insight that inspires Boden's distinction with our distinction between knowledge-seeking and expertise-dependent creativity.

¹⁴ Depending on the combinatorics of ideas, the relationship between NETSASE ideas and P-Creative ideas may be much more complicated. As one editor has noted, several very finely grained NETSASE ideas can be lumped together as a single P-Creative idea.

¹⁵ Philosophical discussion of the methods of cognitive science are, of course, idealizations. For a reality check, see (Bakker et al. 2012)

attractive to anyone who recognizes the difficulty of establishing any claims about the nature of human cognition using experimental methods. Or, to put this brief defense of the fourth principle differently: one of the reasons to care about how children think is that experimental studies of their cognition ought to be very deeply relevant to understanding the nature of adult cognition because of the high baseline probability that adult cognition is driven by the same underlying computational mechanisms as children's cognition. So, one of the reasons to think that developmental psychology should be in the vanguard of the cognitive sciences is that the relevant computational mechanisms are likely easier — though, not *easy* — to investigate experimentally in children as opposed to adults.

Indeed, the various strands of our criticism are strengthened if we ignore children's creativity for a second. To measure adults' creativity, NETSASE criteria are also not enough, since random number generators produce output that is novel and original, and usefulness, in the case of adults, is too general of a category to be specific to creativity. The deeper point here, linking our discussion of children's creativity with this mention of adults' creativity is that any measure of creativity needs to be a consequence of a computational theory of mind that characterizes our knowledge and skills. We need a theory of creativity that does not presume too much knowledge of children, nor too little knowledge from adults.

So, any temptation to reduce creativity to novelty, originality, or usefulness generates tension with the precepts of our Jamesian methodology. At this point, the options from here moving forward seem to be:

- Reject some or all of the Jamesian principles — presumably at least the 4th principle.
- Reject the computational theory of mind.
- Reject psychological theories of creativity that hold that creativity is generated by cognitive processes that compute the originality, novelty, or usefulness of an idea.

The remainder of this paper consists of an exploration of the third option. We think that the problems surveyed above are avoided by not requiring that the mind *compute* that certain mental representations are either novel, original, or useful — instead, we argue that the minds of children and adults both form creative ideas by computing an idea's posterior probability and projecting its potential value. By exploring the details of how this could happen, we will also show how our theory of creativity can satisfy the second and third Jamesian principles along with the first, the fourth, and the computational theory of mind.

4 Knowledge-seeking creativity and expertise-dependent creativity

However, we have one last preliminary issue to deal with. We hypothesize that the creativity of children is caused by qualitatively very similar psychological mechanisms as which cause creativity in adults. But we nevertheless think that connection between knowledge and creativity does exhibit an important developmental pattern: our theory of creativity needs a way of integrating the impact of the fact that adults just know more than children.

We think it is therefore important to distinguish between *knowledge-seeking creativity* and *expertise-dependent creativity*. The idea we are trying to express with this distinction is that creativity both facilitates *and* expresses learning. Expertise-dependent creativity is defined by cases in which either experience or an accumulation of knowledge causes increases in creativity (creativity *expresses* learning). Knowledge-seeking creativity is defined by cases in which creativity has the potential to cause learning (creativity *facilitates* learning).

The distinction—which, please note, is not between mutually exclusive or exhaustive categories; some occurrences of creativity are both—is grounded in common sense. For instance, here are some everyday examples of knowledge-seeking creativity:

- Asking a series of questions which are not linked by any underlying logic but which nevertheless generate a new avenue of inquiry.
- Persisting with an extended complex counterfactual train of thought, in order to search for and hopefully uncover previously unexplored pathways for learning.
- Assuming that a historically trusted teacher or interlocutor is mistaken about a new piece of information.
- Constructing hypotheses about what ideas or arguments have not yet been considered without carrying out an exhaustive, deterministic search of the available data.
- Inventing a new term, strategy, idea, or hypothesis that can scaffold any of the above occurrences or outcomes.

At a more general level, knowledge-seeking creativity can emerge in contexts where there is no easy route for a learner to acquire knowledge: the learner cannot, for instance, just reason in a stepwise fashion from existing beliefs, or ask an expert, or even just remind herself of a past lesson.

Expertise-dependent creativity is probably what many of the psychologists we listed in the previous section have in mind when they are thinking about creativity. Expertise-dependent creativity consists of taking something that is known and deploying it in something other than a mechanical, rote, or literal form. Examples of expertise-dependent creativity likely should include:

- The performance of a complex musical masterpiece that is truly original in its manifestation.
- The formulation of a simpler procedure for performing an extremely complicated task — for instance, designing a new surgical technique that accomplishes the same outcome as prior techniques in half the time with half the risk and half the costs.
- The condensation of hundreds of humanistic insights into a single coherent body of writing, sculpture, or play – or condensing the complex history of a country or era into a captivating piece of written fiction, visual art, or song.
- Alternatively, the condensation of hundreds of scientific facts into a single coherent body of writing, graph, textbook, or lecture series.
- The crafting of any artwork which almost perfectly balances form with content.
- The insight that a complex network of equations can be replaced by a single equation.

- The compounding effects on learning that expertise-dependent creativity can produce. Experts often become *more* creative with more and more practice in their domain of activity.

The first item on this list—musical performance—illustrates rather nicely the distinction between expertise-dependent creativity and knowledge-seeking creativity. Think of the difference between someone learning what fundamental frequencies count as notes by playing around with a piano (and probably annoying nearby listeners) and someone ensuring by listening that their fingers produce the first C# minor chords of a certain Rachmaninov prelude. And then, there are all of the expressions of creativity that fall between these two extremes—expressions of creativity that both are an expression of acquired knowledge and which can generate new knowledge. Plausibly, even Vladimir Horowitz learned something new each time he played *Prelude in C#*.

Is there an additional analytic criterion that can be used to better refine our definition expertise-dependent creativity? Yes, possibly — we can repurpose a distinction from an earlier chapter in the history of cognitive science, the competence/performance distinction. We think that a plausible sufficient condition for expertise-dependent creativity is when the distance between competence and performance on some task is shrunk past either some absolute or performance-relative minimum. As most readers know, the competence/performance distinction comes from Chomsky's early work (Chomsky 1965): "competence" denotes a person's total understanding of the grammar of any natural languages they can use with minimal fluency, while "performance" denotes acts of putting this knowledge to use by speaking or interpreting other's uses of a shared natural language. In most real-world cases, performances are imperfect expressions of the underlying competence: we may know that "'i' comes before 'e' except after 'c'" holds for most idioms of English, but it does not follow that all of our written "performances" of words with 'i' and 'e' will follow this rule. There is usually a gap between performance and competence, and because of that, we should not assume that flawed performances reflect impartial or degraded underlying competency. Our suggestion is that when some underlying epistemic competence and some occurrent performance are so close to one another that the "gap" between competence and performance is almost eliminated, this is a sufficient condition for the presence of expertise-dependent creativity. This explains why expertise-dependent creativity can take a lot of practice to express, and why it may therefore be the rarer form of creativity.

5 An original, useful definition of creativity

Let us now survey all of the desiderata that we have put on the table. Our theory of creativity must be consistent with the four Jamesian principles. It also must avoid the problems noted above for theories of creativity which make novelty, usefulness, or originality sufficient for creativity. It must be able to explain how knowledge-seeking creativity and expertise-dependent creativity can arise from the same set of underlying cognitive mechanisms. The remainder of this paper

shows how the resources of rational constructivism can generate a theory of creativity that satisfies all these demands.¹⁶

The simplest place to start is with our rational constructivist theory of creativity. To wit:

A creative idea is a mental representation caused by a creative psychological process, and

A psychological process is creative when it constructs potentially valuable improbable mental representations.

In cases of knowledge-seeking creativity, what makes the constructions valuable is that they can facilitate learning.

In cases of expertise-dependent creativity, what makes the constructions valuable is that the constructions are virtually “gapless” performances of some underlying expertise.¹⁷

As you can see, three analytic concepts constitute the core of our theory: *construction*, *valuable*, and *improbable*. The next four subsections use the resources of rational constructivism to explicate each of these concepts, showing thereby how our definition of creativity can be derived from rational constructivism.

5.1 Rational constructivism

Recall from the introduction that, according to rational constructivism, there are three cognitive mechanisms that are the causal basis of all learning: language acquisition and symbolic/lexicalized cognition, Bayesian induction and hypothesis selection, and constructive thinking.

Language acquisition provides cognition with a constantly growing library of symbolic structures and templates that can be used to encode information. Language also places a constraint on the repertoire of symbolic representations that are available to the mind, ensuring that two speakers of the same language share many of the same thoughts and ideas.

Bayesian learning refers to a cognitive mechanism that can support drawing increasingly reliable inductive conclusions from evidence provided by experience. Bayesian learning is hypothesized by rational constructivists to play an important role in belief fixation, such that the cognitive representations with a sufficiently high posterior

¹⁶ A clarification for the philosophers reading this paper: we are trying to develop a philosophically sophisticated scientific explanation of creativity. So, our theory illustrates one way of bringing together a computational theory of cognition with an explanation of creativity — this is hardly the only way of uniting the two. Philosophers will be able to imagine alternatives. The key question in evaluating these alternatives is whether or not they are entirely hypothetical, or plausible conditional upon at least some experimental data.

¹⁷ A reviewer offers the following counterexample that is helpful to consider as a way of clarifying our proposal: “Consider someone who practices mental arithmetic until they are at world-leading levels at it. The person has shrunk the gap to a minimum, but they are not in the least creative: they are outstandingly good at a rote activity. The competence / performance condition may be sufficient for some types of expertise, but it is not sufficient for creative expertise.” While we are grateful for the example, we do not think that this is a counterexample. The mathematician’s ability is not a case of creativity because the expression of world-leading arithmetical skill is *not* improbable *because* it is rote.

probability conditional upon available evidence typically become beliefs. The function of the Bayesian cognitive mechanism, then, is to compute posterior probability, and importantly for what follows, it also can be used by the cognitive system to screen for highly improbable representations.

Finally, there are the mechanisms which constitute the mind's faculties of constructive thinking. Bayesian reasoning can, by hypothesis, only sort and filter cognitive representations in terms of their relative probability—it cannot be a source of new information. Constructive thinking provides this function, as the capacity for constructive thought refers to the ability to formulate novel explanations, imagine alternative causes and effects, draw analogies and create metaphors *ex nihilo*, engage in deeply counterfactual thought experiments and, generally speaking, enlarge the mind's cognitive repertoire by creating new information that can form the basis of any number of different kinds of mental representation (Gendler, 2000; Gentner & Hoyos, 2017; Lombrozo, 2012, 2020).

It is important not to confuse any operation whatsoever of the mind's constructive faculties with creativity. Returning to an idea mentioned in section 3, the production of NETSASE novel or NETSASE original mental representations is not *ipso facto* a manifestation of creativity. As we have said, we are sceptical that a random concept or hypothesis generator that could produce a limitless amount of NETSASE novel or original concepts is thereby creative. The problem is that a random new information generator would not be able to produce output that is often sufficiently practically rational.

Indeed, one of the strengths of our account of creativity is that it provides a simple, computational explanation of creativity's normativity. But before turning to this explanation, we want to first provide our analysis of how improbable constructions can enter the mind.

5.2 The origins of improbable constructions

The most basic way that the improbable constructions that eventually form a person's creative ideas can come into existence is by way of operations joining sufficiently improbable combinations of existing ideas. The process has two steps. Existing ideas are combined in imagination to form new ideas, which are then evaluated relative to some specification of improbability by the Bayesian mechanisms.¹⁸ One or more ideas then emerge as the relevant improbable idea(s).

But it would be a mistake, however, to think that combinations are the only source of improbable ideas. Rational constructivism allows there to be many, many other routes by which information can be added to the cognitive system: learning new words and concepts; mental processes like simplification, reframing, and abstraction; sustained mind-wandering, whether in isolation or with other people; constructing a paracosm; engaging in wordplay; imperfect imitation; and so on. It is now easy to think about how ideas and thoughts that have varying amounts of relatively low posterior probability beliefs at a time can nevertheless become fixtures in a person's thinking for

¹⁸ Here, our theory of creativity overlaps with "first-generate-then-evaluate" theories of creativity. However, we think that there are different kinds of creativity (there is knowledge-seeking and expertise-dependent creativity, at least), as well as a broader range of mental representation constructing processes.

a time (Denison and Xu 2019; Kushnir et al. 2010; Schulz 2012; Tenenbaum et al. 2011; Xu 2011; Xu and Kushnir 2012; Xu and Tenenbaum 2007).

However, one of the ways that rational constructivism distinguishes itself from other theories of cognitive development is that it makes room for NETSASE novel mental representations (even though we also think that NETSASE novel mental representations are not *ipso facto* creative). These are ideas that are not constructed by *entirely* computational processes—since, plausibly, computational processes must, at bottom, implement logical rules, and this means computational processes must ultimately consist in transformations of existing information structures into new information structures where, crucially, the output structures share at least some information with the input structures. But it does not follow that even proponents of the computational theory of mind cannot allow that NETSASE novel information either enters or constructs the mind; it is just that the mechanisms which produce such mental representations cannot take computational input. The relevant mechanisms are best thought of as transducers (mechanisms that convert signals from one format to another), not processors (mechanisms that generate new constructs by performing logical operations on a set of information).

This is not as odd of a proposal as it may at first seem. Here, an analogy with the random number generator found in most digital computers is helpful, since the function of these devices is to, quite literally, inject information into computational systems that the systems cannot produce themselves. Because computers are deterministic machines, they cannot be programmed to generate random sequences of numerals. Random number generators overcome the problem of providing computers with random input by “harvesting” stochastic information and converting it into electrical signals that are then treated as digital information by a computer. RAND’s famous book *A Million Random Digits with 100,000 Normal Deviates*, first printed in 1955, was produced by an “electronic roulette wheel” that operated this way:

A random frequency pulse source, providing on the average of about 100,000 pulses per second, was gated about once per second by a constant frequency pulse. Pulse standardization circuits passed the pulses through a 5-place binary counter. [...] A binary-to-decimal converter was used which converted 20 of the 32 numbers (the other twelve were discarded) and retained only the final digit of two-digit numbers; this final digit was fed into an IBM punch to produce finally a punched card table of random digits.¹⁹ (*A Million Random Digits with 100,000 Normal Deviates* 1955)

Likewise, it is possible that novel information enters the cognitive systems by way of transducers the function of which is to convert the stochastic properties of dynamic perceptual, physiological, or neurophysiological processes into NETSASE novel content for the mind’s computational processes to subsequently act upon.

¹⁹ RAND has placed the book online. In the foreword to the digital edition, they offer the following commentary to their readers: “A humorous sidelight: The New York Public Library originally indexed this book under the heading ‘Psychology.’”

The possibility of these strict-novelty-supplying transducers also shows that we need to make explicit a distinction that has been implicit in some of our arguments above. We need to distinguish between mental representations (constructions) that are NETSASE novel (or original) because they are not caused to become part of the inventory of a mind by the computational transformation of either perceptual data, interoceptive data, or existing cognitively embedded information, and mental representations that are novel (or original) because they are produced by an algorithm that searches existing cognitive information, and then constructs a new mental representation on the basis of some comparison with the contents of the stored information. Above, we argued that the latter kind of novelty (or originality) is hard to impute to children; the former does not face this problem. Our contention is that neither kind of novelty or originality is sufficient for creativity.

This distinction is worth pausing to consider a bit more, as it could explain why psychologists have previously concluded there is an important connection between creativity and novelty or originality. It could be that the root of most creative ideas are mental representations that first enter the mind as NETSASE novel constructions—but that is a very different commitment than taking on the thesis that novelty or originality, strict or not, is sufficient for creativity.

So much, then, for the metaphysics of the construction of mental representations. We should now say more about the role that a Bayesian inferential mechanism could play in generating creativity. A cognitive device that is able to implement something like Bayes's rule to calculate the posterior probability of a proposition (or the probability of each of an ordered n -tuple of propositions) automatically generates information that can be used to sort highly-probable propositions from low-probability propositions [cf. (Denison and Xu 2019)]. As we have said, rational constructivism holds that the propositions with the highest posterior probability tend to become beliefs. But it does not follow that the low-probability propositions have no use in the economy of the mind. A proposition that has a sufficiently low posteriori probability conditional only upon a person's existing beliefs is a belief that is likely to be useful for either further validating existing beliefs or, what may be better in the long run, adjusting the probability assigned to certain beliefs by placing them into a broader web of probability relations.

It is easiest to first link these observations with knowledge-seeking creativity. This creativity helps a mind avoid getting trapped in local maxima, helps a mind ensure that it has a rich library of alternative explanations to draw upon, and helps a mind ensure that it is not relying on exactly the same strategy time and again in order to generate learning. But the same observations also can be connected with expertise-dependent creativity as well. The ability to construct an improbable representation that is nevertheless an expression of an underlying expertise is a way of ensuring that the eventual performance is truly original—and this can have benefits that are analogous to those that we have just noted can flow from knowledge-seeking creativity. Out of these combinations can come things that to common sense seem to be most paradigms of creativity: unexpected but clever design for an experimental protocol, an awe-inspiring performance of Bach's chaconne in D, or a new interpretation of pad thai that help groups avoid gustatory cul-de-sacs.

So, creativity may start with NETSASE novel mental representations—but, crucially, in our analysis, a person's cognitive system does not have to categorize them as

novel. Instead, all that is required is that the cognitive system be sensitive to the fact of their occurrence, and thereafter be able to make a projective inference about whether a new mental representation is potentially valuable. This inference could be a simple inductive inference that reflects memories of the earlier beneficial effects of taking new ideas seriously.

5.3 Creative rationality

We have now shown how rational constructivism can explain *construction* and *improbable*. We turn now to *valuable*.

Here, our analysis of the normative value of creativity is instrumental. Something is instrumentally valuable if it can cause or influence the probability of the occurrence of something else that is valuable. Accordingly, we think that creative ideas are valuable to the extent that they either facilitate learning (which itself is valuable; a claim we take can be entered without further argument) or facilitate the expression of some kind of cognitive output that has some other kind of value (aesthetic, mathematical, pragmatic, hedonic). A creative process is therefore valuable to the extent that it produces either, or both, of these two kinds of outcomes.

There is one particularly valuable by-product of creative thinking that we want to highlight. We have termed this by-product “cognitive agency” (Fedyk and Xu 2018); it is what can happen when knowledge-seeking creativity and expertise-dependent creativity converge and subsequently reinforce one another. In such cases, a learner’s past learning about how to learn has reached such a state of refinement and potential that the person is able to use their knowledge to frame unlikely ideas that nevertheless facilitate future learning. When a person knows enough about learning that they are able to take control of their own learning and thereby *increase* the efficacy of their subsequent learning, they have cognitive agency. Put more simply: someone with cognitive agency has enough experience “learning to learn” that they *know* how to learn.²⁰

Another way of putting the idea here is this. As expertise in a domain grows, the learner also gains expertise in how to grow their expertise effectively and productively. Familiar examples are how pianists’ performances of a particular piece over many concerts increases the efficiency of subsequent practicing, or how past experimental failures can lead biologists to more and more ingenious experiments. One could say that ‘exploration’ becomes much more directed, intentional, and effective in experts, compared to children, and this results in deeper and more productive creativity, eventually culminating, sometimes, in cognitive agency.

So, creativity is valuable because it can lead to cognitive agency. Nevertheless, it is a mistake to think that that is the only important normative dimension to creativity. Because of our instrumental approach to normativity, it is possible to treat the variability inherent in much of the mind’s creative computational processes as the basis of

²⁰ An important clarification: we do not mean that they know how to learn in a domain-general sense. Someone achieves cognitive agency with respect to playing the piano when they have spent enough time practicing piano (learning to learn to play the piano) that they sooner or later know how to keep learning. The process of learning searches from being exploration-like to being intentional and directed. Crucially, this does *not* mean that learning is thereafter path-dependent: the value of introducing intentionality here is that the learner can recognize that she has reached a local maximum, and can thereby switch strategies or areas of effort by making good (i.e. epistemically productive) choices.

an explanation for how the mind both succeeds and fails to manifest valuable ends of many different kinds.

Here is how the story goes. According to rational constructivism, cognition emerges out of an interaction between three fundamental learning mechanisms. Each of these mechanisms exhibits its own characteristic variations and dynamics, and this fact serves to dramatically amplify the total number of causal interactions that rational constructivism's three mechanisms can produce. But out of all of these possible interactions, only a small fraction will be either conducive to learning or conducive to producing constructs that are valuable for some other reason than their potential to produce knowledge. Moreover, the variability of the operations of each of the three fundamental mechanisms *expands* the total number of possible causal interactions and is therefore a key element in rational constructivism's explanation of the cognitive system's normativity—sometimes the mind is able to achieve rationality because of, not in spite of, variability inherent in the operation of its most fundamental cognitive mechanisms.

So, when we say that creativity requires the production of “potentially valuable” mental representations, by this we mean that the mental representations may either be actually valuable (because they are a realization of cognitive agency, say), or they may lack any value but nevertheless be produced by interactions between the mechanisms posited by rational constructivism that are like the interactions which typically do produce mental representations of value. Not all of van Gogh's paintings are of equivalent aesthetic value; but they all arose out of interactions between the same set of underlying computational mechanisms.

With that clarification in mind, let us return to the idea that, sometimes, knowledge-seeking creativity and expertise-dependent creativity converge, yielding cognitive agency. Metaphysically speaking, cognitive agency is a subset of the total set of possible interactions between the learning mechanisms posited by rational constructivism. But because they are a subset that is different from the subset which defined cognitive agency, they can be characterized as exemplifying its own *sui generis* normative category, that of creative rationality. Put more succinctly: creative rationality is what creativity exhibits when it causes or sustains cognitive agency.

“Rationality” is not an easy term to import into contemporary cognitive science. Amongst philosophers, it can usually mean one of three things. Rationality is *normative* in the sense of defining a set of regulative ideals: certain logical or statistical or inferential principles dictate how one should think, and someone is rational to the extent that their thinking conforms to these ideals. Rationality is *descriptive*: a set of regulative ideals really describes how people actually do think. Rationality is *constitutive* of thought; conformity to some of the regulative ideals of rationality is, at least in part, just what it is to think (Rescorla 2013).

What we want to do is lean on the ‘normative’ and ‘descriptive’ meanings of rationality here and say that some occurrences of creativity exhibit creative rationality to the extent that they satisfy certain regulative ideals related to epistemic development. The relevant regulative ideals, however, are not principles of logical or statistical inference. Instead, the creative rationality manifest by facilitating cognitive agency — and so the patterns of creative thinking most likely to count as patterns of creative rationality will include such things as patience, open-mindedness, attentiveness to detail, tolerance of contradiction and paradox, curiosity, and even playfulness — for these can all cause low-probability and therefore likely initially surprising ideas to

remain in the mind long enough for them to facilitate meaningful advances in learning. Creative rationality therefore seems to conform to very different principles than, for instance, deductive rationality (conformity to the axioms of first-order logic) or statistical rationality (conformity to the Kolmogorov axioms).

There are some interesting payoffs that come from pushing our thinking about creative rationality and rational constructivism into even more speculative areas. We have noted that at some point in their cognitive development, all learners acquire knowledge about how to acquire knowledge—and so, at the limit, there can be individual expertise about learning. When this expertise is manifest in the service of learning, this is, again, cognitive agency—and the virtues expressed by cognitive agency which themselves explain how cognitive agency can causally facilitate subsequent learning can be thought of collectively as creative rationality. Over time the creative rationality of occurrences of cognitive agency may increase substantially and thereby generate very powerful reflexive feedback loops. Possible examples of this are easy to find: Van Gogh's last burst of creative output, the independence of thought and action that solid training in the liberal arts provides, or the ability of seasoned clinicians or detectives to generate either diagnosis or an indictment without any apparent deliberative reasoning. There seems to be an important connection between cognitive agency and the more striking expressions of human creativity — we speculate that it may be important to see developing cognitive agency as an important pedagogical goal.

Stepping back from these speculations, though, we would be remiss to ignore the more basic instrumental role that creative processes play in facilitating learning, which illustrated why psychological Bayesians need a theory of creativity. Learning driven by *only* Bayesian reasoning would be a process of directed selection that would sooner or later become locked into path-dependencies that either substantially degrade the efficiency of learning or make further progress impossible [cf. (Griffiths and Machery 2008)]. The mind therefore needs a source of variation in the representations that it evaluates, acts upon, rejects, or eventually believes—and, ideally, this variation will add new information to the set of ideas that have either been accepted or rejected by the mind. A cognitive mechanism that produced cognitive variety by some kind of random walk may be able to do this work, but far better is a capacity that can ensure that the novel information is, as frequently as possible, valuable in some way or another; there is less risk of cognitive pollution this way. Creativity can therefore explain why Bayesian learning can, over the long run, be effective, and thus instrumentally valuable: some mechanism needs to generate out-of-the-box hypotheses for the Bayesian evaluative mechanisms to evaluate if these mechanisms are going to be able to produce belief structures that, after enough time, converge on the truth.

In sum, because creativity is productive it is also valuable. It is instrumentally valuable because this variability can facilitate such ends as learning. But once enough learning has occurred, (knowledge-seeking) creativity can facilitate the formation of cognitive agency (relative to some domain or another), and then cognitive agency can be a source of yet further creativity (by being both knowledge-seeking and expertise-dependent at the same time). When creativity of either of these two more complex forms facilitates this epistemically virtuous integration, we can speak of these integrations as manifestations of “creative rationality” and thereby refer to a *sui generis* form of epistemic normativity.

6 Conclusion

Can a theory of creativity exemplify the virtues of originality and usefulness, and thereby exemplify what it aims to explain? If the arguments in section 3.0 stand, then the answer is “no”. This paper may be original and useful, but if we are right about the nature of creativity, it does not follow that this paper is creative. And if this is not a creative paper, that is a failing we happily accept. As many scientists know, it is hard to publish truly creative scholarship; the absence of creativity in this paper may explain why it got through peer review and you are now reading it in print.

Jokes aside, we want to conclude by stressing a more important point. A theory of creativity should be integral to any theory of human cognition, given that any learner is capable of creative thought, and given the comparatively smaller utility that purely formal forms of rationality have with respect to learning (no one learns by deductive reasoning alone, after all). As this paper shows, a theory of creativity can find roots in rational constructivism, which is an independently plausible theory of human rationality and cognitive development, and also a theory which provides a unified explanation of the cognition of both children and adults.

However, our theory of creativity is hardly the only theory of creativity that can be derived from any number of the contemporary, independently plausible psychological theories. For instance, we are sceptical of associationism and connectionism, but we find reading about creativity enjoyable. A connectionist theory of creativity could be very creative by the lights of our theory, because it would be (scientifically) improbable and valuable (because theories of creativity are entertaining). We look forward to seeing what other (creative or not) theories of creativity can be germinated in the rich soil of contemporary cognitive science.

Acknowledgments We wish to express our thanks to the excellent comments and criticisms we received from both anonymous reviewers and the editors. We are especially grateful to Dr. Killin for his deep and incisive editing.

Compliance with ethical standards

Conflict of interest None to declare.

Ethical approval n/a

Informed consent n/a

References

- A million random digits with 100,000 normal deviates. (1955). Rand Corporation.
- Abraham, A. (2013). The promises and perils of the neuroscience of creativity. *Frontiers in Human Neuroscience*. <https://doi.org/10.3389/fnhum.2013.00246/abstract>.
- Abraham, A. (2014). Is there an inverted-U relationship between creativity and psychopathology? *Frontiers in Psychology*, 5, 750.
- Acar, S., Burnett, C., & Cabra, J. F. (2017). Ingredients of creativity: originality and more. *Creativity Research Journal*, 29(2), 133–144.
- Albert, R. S., & Runco, M. A. (1999). A history of research on creativity. *Handbook of Creativity*, 2, 16–31.

- Amabile, T. M. (1996). *Creativity in context*. Westview Press.
- Antony, L. M. (2018). Quine as feminist: The radical import of naturalized epistemology. In *A mind of one's own* (pp. 110–153). Routledge.
- Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 7(6), 543–554.
- Barron, F. (1963). *Creativity and psychological health*. 292. <https://psycnet.apa.org/fulltext/1964-06077-000.pdf>
- Boden, M. A. (2004). *The creative mind: Myths and mechanisms* (2nd ed.). Routledge.
- Bouwmeester, O. (2017). *The social construction of rationality: Policy debates and the power of good reasons*. Taylor & Francis.
- Briegel, H. J. (2012). On creative machines and the physical origins of freedom. *Scientific Reports*, 2, 522.
- Briegel, H. J., & De las Cuevas, G. (2012). Projective simulation for artificial intelligence. *Scientific Reports*, 2, 400.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. M.I.T. Press.
- Chomsky, N. (2009). *Cartesian Linguistics: A Chapter in the History of Rationalist Thought*. Cambridge University Press.
- Denison, S., & Xu, F. (2019). Infant statisticians: the origins of reasoning under uncertainty. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 14(4), 499–509.
- Diedrich, J., Benedek, M., Jauk, E., & Neubauer, A. C. (2015). Are creative ideas novel and useful? *Psychology of Aesthetics, Creativity, and the Arts*, 9(1), 35.
- Fedyk, M., & Xu, F. (2018). The epistemology of rational constructivism. *Review of Philosophy and Psychology*, 9(2), 343–362.
- Fedyk, M., & Xu, F. (2019). The metaphysics of the development of the cognitive system: Why the brain cannot replace the mind. In S. Cullen & S.-J. Leslie (Eds.), *Current controversies in philosophy of cognitive science*. Routledge.
- Feist, G. J. (2006). How development and personality influence scientific thought, interest, and achievement. *Review of General Psychology: Journal of Division 1, of the American Psychological Association*, 10(2), 163–182.
- Fodor, J. A. (1990). *A theory of content and other essays*. MIT Press.
- Gaut, B., & Kieran, M. (2018). *Creativity and philosophy*. Routledge.
- Gendler, T. S. (2000). The puzzle of imaginative resistance. *The Journal of Philosophy*, 97(2), 55–81.
- Gentner, D., & Hoyos, C. (2017). Analogy and abstraction. *Topics in Cognitive Science*, 9(3), 672–693.
- Gopnik, A. (2020). Childhood as a solution to explore-exploit tensions. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 375(1803), 20190502.
- Gray, K., Anderson, S., Chen, E. E., Kelly, J. M., Christian, M. S., Patrick, J., Huang, L., Kenett, Y. N., & Lewis, K. (2019). “Forward flow”: a new measure to quantify free thought and predict creativity. *The American Psychologist*, 74(5), 539–554.
- Griffiths, P. E., & Machery, E. (2008). Innateness, canalization, and “biologizing the mind”. *Philosophical Psychology*, 21(3), 397–414.
- Guilford, J. P. (1950). Creativity. In *American Psychologist* (Vol. 5, Issue 9, pp. 444–454). <https://doi.org/10.1037/h0063487>
- Guilford, J. P., & Hoepfner, R. (1966). *Structure-of-intellect factors and their tests*. The Psychological Laboratory of the University of Southern California.
- Hangl, S., Dunjko, V., Briegel, H. J., & Piater, J. (2017). Skill learning by autonomous robotic playing using active learning and creativity. In *arXiv [cs.RO]*. arXiv. <http://arxiv.org/abs/1706.08560>
- Harre, R. (2002). *Cognitive science: A philosophical introduction*. SAGE.
- James, W. (1918a). *The principles of psychology: Authorized Ed., Unabridged* (Reprint edition). Dover Publications.
- James, W. (1918b). *The principles of psychology: Authorized Ed., Unabridged* (Revised ed. edition). Dover Publications.
- Kaufman, J. C., & Baer, J. (2004). Hawking’s Haiku, Madonna’s Math: Why it is hard to be creative in every room of the house. In R. J. Sternberg (Ed.), *Creativity: From potential to realization*, (Vol. 226, pp. 3–19). American Psychological Association, x.
- Kushnir, T., Xu, F., & Wellman, H. M. (2010). Young children use statistical sampling to infer the preferences of other people. *Psychological Science*, 21(8), 1134–1140.
- Lombrozo, T. (2012). Explanation and abductive inference. *The Oxford Handbook of Thinking and Reasoning*, 836, 260–276.
- Lombrozo, T. (2020). “Learning by thinking” in science and in everyday life. In *The Scientific Imagination*. Oxford University Press.
- MacKinnon, D. W. (1970). Creativity: A multi-faceted phenomenon. *Creativity Research Journal*, 19–32.

- Mottweiler, C. M., & Taylor, M. (2014). Elaborated role play and creativity in preschool age children. *Psychology of Aesthetics, Creativity, and the Arts*, 8(3), 277–286
- Quine, W. V. O. (1969). Epistemology naturalized. In W. V. O. Quine (Ed.), *Ontological relativity and other essays*. Columbia University Press.
- Rescorla, M. (2013). Rationality as a constitutive ideal. In E. Lepore & K. Ludwig (Eds.), *A companion to Donald Davidson* (Vol. 578, pp. 472–488). John Wiley & Sons, Inc.
- Runco, M. A. (1988). Creativity research: originality, utility, and integration. *Creativity Research Journal*, 1(1), 1–7.
- Runco, M. A. (2004). Everyone has creative potential. In R. J. Sternberg (Ed.), *Creativity: From potential to realization*, (Vol. 226, pp. 21–30). American Psychological Association, x.
- Runco, M. A., & Charles, R. E. (1993). Judgments of originality and appropriateness as predictors of creativity. *Personality and Individual Differences*, 15(5), 537–546.
- Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal*, 24(1), 92–96.
- Runco, M. A., Illies, J. J., & Eisenman, R. (2005). Creativity, originality, and appropriateness: what do explicit instructions tell us about their relationships? *The Journal of Creative Behavior*, 39(2), 137–148.
- Schulz, L. (2012). Finding new facts; thinking new thoughts. *Advances in Child Development and Behavior*, 43, 269–294.
- Simonton, D. K. (2008). Scientific talent, training, and performance: intellect, personality, and genetic endowment. *Review of General Psychology: Journal of Division 1, of the American Psychological Association*, 12(1), 28–46.
- Simonton, D. K. (2013). What is a creative idea? Little-c versus Big-C creativity. *Handbook of Research on Creativity*, 2, 69–83.
- Simonton, D. K. (2018). Defining creativity: don't we also need to define what is not creative? *The Journal of Creative Behavior*, 52(1), 80–90.
- Sternberg, R. J. (1988). A three-facet model of creativity. *The Nature of Creativity*, 125–147.
- Stillings, N. A., Chase, C. H., Weisler, S. E., Feinstein, M. H., Garfield, J. L., & Rissland, E. L. (1995). *Cognitive science: An introduction*. MIT Press.
- Szen-Ziemiańska, J., Lebuda, I., & Karwowski, M. (2017). Mix and match. In *The Cambridge handbook of creativity across Domains* (pp. 18–40). Cambridge University Press.
- Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: statistics, structure, and abstraction. *Science*, 331(6022), 1279–1285.
- Thagard, P. (2002). Curing cancer? Patrick Lee's path to the reovirus treatment. *International Studies in the Philosophy of Science*, 16(1), 79–93.
- Thagard, P. (2005). *Mind: Introduction to cognitive science (Vol. 17)*. Cambridge: MIT Press.
- Thagard, P. (2019). Creativity. In *Brain-Mind* (pp. 223–251). <https://doi.org/10.1093/oso/9780190678715.003.0011>.
- Thagard, P., & Stewart, T. C. (2011). The AHA! Experience: creativity through emergent binding in neural networks. *Cognitive Science*, 35(1), 1–33.
- Tiefer, L. (2006). The Viagra phenomenon. *Sexualities*, 9(3), 273–294.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind; a Quarterly Review of Psychology and Philosophy*, 59(236), 433–460.
- Weisberg, R. W. (2015). On the usefulness of “value” in the definition of creativity. *Creativity Research Journal*, 27(2), 111–124.
- Wiredu, J. E. (1970). KANT'S synthetic a priori in geometry and the rise of non-EUCLIDEAN geometries. *Kant-Studien*, 61(1-4), 5–27.
- Xu, F. (2011). Rational constructivism, statistical inference, and core cognition. *The Behavioral and Brain Sciences*, 34(03), 151–152.
- Xu, F. (2019). Towards a rational constructivist theory of cognitive development. *Psychological Review*, 126(6), 841–864.
- Xu, F., & Kushnir, T. (2012). *Rational constructivism in cognitive development* (Vol. 43). Academic Press.
- Xu, F., & Kushnir, T. (2013). Infants are rational constructivist learners. *Current Directions in Psychological Science*, 22(1), 28–32.
- Xu, F., & Tenenbaum, J. B. (2007). Sensitivity to sampling in Bayesian word learning. *Developmental Science*, 10(3), 288–297.