Creativity

Forthcoming in Stanford Encyclopedia of Philosophy
revised and resubmitted

Few things shape the human experience as profoundly or as pervasively as creativity does. And creativity raises a wealth of philosophical issues. Since art is such a salient domain of creativity, you might assume, at first, that the philosophy of creativity is the philosophy of art or aesthetics, or a branch thereof. But creativity invites questions of its own that go beyond the purview of those other fields.

Note that the adjective ‘creative’ can be applied to three kinds of things: a person (‘Beyoncé is creative’), a process or activity (‘Tell us about your creative process’), or a product, where the latter is taken broadly to include an idea in someone’s mind or an observable performance or artifact (‘That’s a creative design’).

Now suppose you are looking at a creative product, like a painting or sculpture. The philosophy of art may ask, “What makes this a work of art?” and aesthetics may ask, “What makes this beautiful?” By contrast, the philosophy of creativity asks, “What makes this creative? Is it just that it’s new, or must it meet further conditions?” We may ask the same question not just of artworks but of any creative product, whether it be a new scientific theory, a technological invention, a philosophical breakthrough, or a novel solution to a mathematical or logical puzzle. Beyond creative products, we can ask about the creative process: Must it proceed without following rules? Is it conscious, unconscious, or both? Must it be an expression of the creator’s agency, and, if so, must that agency be exercised intentionally? Exactly how does the process manage to produce new things? Can it be explained scientifically? Furthermore, we can ask about creative persons, or more generally, creators. What does it mean for a person to be creative? Is it a virtue to be creative? What capacities and characteristics does a being need to have in order to be creative? Could a computer be creative? These are the kinds of questions animating the literature we’ll survey below.
Some of these questions have an empirical dimension, most obviously those which pertain to how the creative process is actually carried out. Thus, much of the research we’ll canvass falls under the inter-disciplinary umbrella of cognitive science, with contributions not only from philosophers but also from researchers in neighboring fields like psychology, neuroscience, and computer science.

1. The philosophy of creativity: past and present
2. What is creativity?
   2.1. Challenges to the value condition
   2.2. Other proposed conditions
      2.2.1. Surprise
      2.2.2. Originality
      2.2.3. Spontaneity
      2.2.4. Agency
   2.3. Is creativity a virtue?
3. Can creativity be learned?
4. Can creativity be explained?
5. The cognitive science of creativity
   5.1. Preparation
   5.2. Generating ideas
      5.2.1. Blind Variation
      5.2.2. The Default-Mode Network
      5.2.3. Imagination
      5.2.4. Incubation
   5.3. Insight
   5.4. Evaluation
   5.5. Externalization
   5.6. Worries and future directions
6. Creativity and Artificial Intelligence
7. Conclusion
Given the significance creativity has in our lives and the deep philosophical questions it raises, one might expect creativity to be a major topic in philosophy. Curiously, it isn’t.

To be sure, some of the most prominent figures in the history of Western philosophy have been fascinated with creativity – or what we now call ‘creativity’. According to some scholars, the abstract noun for creativity did not appear until the 19th century — but the phenomenon certainly existed and many philosophers took an interest in it (McMahon 2013, Nahm 1956, Penelope 1989, Tatarkiewicz 1980: chapter 8).

To name just few examples: Plato (4th century BCE) had Socrates say, in certain dialogues, that when poets produce truly great poetry, they do it not through knowledge or mastery, but rather by being divinely “inspired” by the Muses, in a state of possession that exhibits a kind of madness (Ion and Phaedrus, Plato 1997). Aristotle (3rd century BCE), in contrast, characterized the work of the poet as a rational, goal-directed activity of making (poeisis), in which the poet employs various means (such as sympathetic characters and plots involving twists of fate) to achieve an end (of eliciting various emotions in the audience (Poetics; cf. Gaut 2012)). Margaret Cavendish (1623-1673) and Émilie du Châtelet (1706-1749) championed the creative use of the imagination to pursue freedom, overcome prejudice, and cultivate natural abilities even despite social and political oppression (Borcherding, forthcoming). Immanuel Kant (1724-1804) conceived of artistic genius as an innate capacity to produce original works through the free play of the imagination, a process which does not consist in following rules, can neither be learned nor taught, and is mysterious even to geniuses themselves. Schopenhauer (1788-1860) stressed that the greatest artists are distinguished not only by the technical skill they employ in the production of art, but also by the capacity to “lose themselves” in the experience of what is beautiful and sublime (Schopenhauer 1969, Vol. I: 184–194 and Vol. II: 376–402). Friedrich Nietzsche (1844–1900) argued that the greatest feats of creativity, which he took to be exemplified by the tragic poetry of ancient Greece, was being born out of a rare cooperation between the “Dionysian” spirit of ecstatic intoxication, which imbues the work with vitality and passion, and the “Apollonian” spirit of sober restraint, which tempers chaos with order and form
(Nietzsche 2000). William James (1842-1910) theorized about creative genius exerts the causal power to change the course of history (Simonton 2018). This is just a glimpse of what each of these philosophers had to say about creativity, and many other figures could be added to their number.

Nevertheless, while some of the topics explored by earlier thinkers have come to occupy a central place in philosophy today—such as freedom, justice, consciousness, and knowledge—creativity is not among them. Indeed, “philosophy of creativity” is still a neologism in most quarters, just as, for example, “philosophy of action” and “philosophy of gender” were not too long ago. However, philosophical work on creativity has been picking up steam over the last two decades (as shown, for example, in a few important collections of essays: Gaut and Livingston 2003; Krausz, Dutton, and Bardsley 2009; Paul and Kaufman 2014; Kieran and Gaut 2018). We’ll now dive into those contributions, along with earlier work, beginning with what is perhaps the most basic question one can ask in this field.

2 What is creativity?

As we noted at the outset, the term ‘creative’ can be applied to three kinds of things: a person, a process, or a product (where a product could be an idea, performance, or physical artifact).

Most definitions focus on the product. According to one common approach, persons or processes are creative to the extent that they produce creative products, and a product is creative if it meets two conditions: in addition to being new it must also be valuable. Many theorists argue that novelty is not sufficient, because something can be new but worthless (e.g. a meaningless string of letters), in which case it doesn’t merit the compliment of being called “creative”. Immanuel Kant is often cited as anticipating this definition of creativity in his discussion of (artistic) genius. According to a common interpretation, Kant defines (artistic) genius as the ability to produce works that are not only “original” – since “there can be original nonsense” – but also “exemplary” (Kant 2000: 43–50). (Hills and Bird (2018) challenge this reading of Kant.) This definition is so widely accepted among psychologists that it has come to be known as “the standard definition” of creativity in psychology. In practice, “creativity is often not defined”
(J.C. Kaufman 2009: 19) in psychological experiments – more on this in §5 below. When psychologists do explicitly adopt a definition, however, they usually say that creative products are not only new, but also valuable in some way, though they variously express the product’s value in terms of its being “useful”, “effective”, “worthwhile”, “fit”, or “appropriate to the task at hand” (Bruner 1962: 18; Cropley 1967: 67; Jackson and Messick 1965: 313; Kneller 1965: 7; Cattell and Butcher 1968; Heinelt 1974; J.C. Kaufman 2009: 19-20; S.B. Kaufman & Gregoire 2016; Stein 1953; Sternberg and Lubart 1999: 3 – for an overview, see Runco and Jaeger 2012).

A few psychologists have suggested that the standard definition doesn’t fully capture the concept of creativity (Amabile 1996; Simonton 2012). As for philosophers, at least one of them defends the standard definition with qualifications (Klausen 2011), but many of them challenge it, as we’ll soon see.

While it is uncontroversial that novelty is required for creativity, philosophers have refined that point. Certain examples may seem, at first, to suggest that novelty isn’t really necessary for creativity. Newton’s discovery of calculus was creative even if, unbeknownst to him at the time, Leibniz got there first – one of many examples of what are called “multiples” in the history of science (Simonton 2004). A beginning student’s idea that freedom is compatible with causal determinism might be creative even if, as she will soon learn, philosophers have been defending such “compatibilist” theories for millennia. However, examples like these do not force us to abandon the novelty requirement, but only to qualify it. Newton’s calculus and the student’s compatibilism were not new in all of history, but they were new to their respective creators, and that is enough for them to count as creative. In the terminology of philosopher Margaret Boden, these ideas are ‘psychologically creative’ (P-creative) even though they are not ‘historically creative’ (H-creative). Notice that P-creativity is more fundamental. Anything that is new in all of history (H-creative) must also be new to its creator (P-creative). Thus, creativity always exhibits psychological novelty, though it doesn’t always exhibit historical novelty.

Again, no one denies that a creative product must be new, at least to its creator. But as we’ll now see, some philosophers depart from the standard definition of creativity by rejecting the value condition (§2.1), or by proposing some further condition(s) (§2.2), or by doing both.
2.1 Challenges to the value condition

Some theorists have argued that although creative things are valuable, we shouldn’t build value into the definition of creativity, because doing so is not informative or explanatory: “Knowing that something is valuable or to be valued does not by itself reveal why or how that thing is. By analogy, being told that a carburetor is useful provides no explanatory insight into the nature of a carburetor: how it works and what it does” (Stokes 2008: 119; Stokes 2011: 675-76). Those who maintain that value is required for creativity might reply that it doesn’t need to be informative or explanatory. Being a man is required for being a bachelor even though it’s not informative or explanatory to say that bachelors are men. Stokes notes that “creative” is a term of praise, and uses this point to argue that what is creative must be produced intentionally (since we don’t rightly praise what is unintentional or accidental) – an idea we’ll return to below. But the same point also seems to imply that what is creative must also have value (since we don’t rightly praise what doesn’t have value). And while the concept “carburetor” is value-neutral, as shown by the fact that a carburetor can be worthless or useless (if it’s broken), “creative”, one might argue, is a value-laden concept, like “progress.” Progress necessarily involves novelty or change, but we don’t praise change as progress unless it’s good change. Likewise, defenders of the value condition urge, creativity necessarily involves novelty, but we don’t praise novelty as creative unless it’s good novelty.

Other critics use counterexamples to argue that value isn’t necessary for creativity, the most prominent cases being ones of immoral creativity. (For a collection of essays by psychologists on the phenomenon of immoral or so-called ‘dark’ creativity’, see Cropley 2010). Putative cases of immoral creativity include creative accounting to cheat investors or creative testimony to mislead jurors, and the stock example in the literature is creative torture or murder. One can imagine novel and well-designed murders, as Thomas De Quincey once did in a satirical essay:

[S]omething more goes to the composition of a fine murder than two blockheads to kill and be killed--a knife--a purse--and a dark lane. Design, gentlemen, grouping, light and shade, poetry, sentiment, are now deemed indispensable to attempts of this nature. Mr. Williams has exalted the ideal of murder to all of us […] Like Æschylus or Milton in
poetry, like Michael Angelo in painting, he has carried his art to a point of colossal sublimity (De Quincey 1827; see also discussion in Battin et al. 1989).

Innovative ways of inflicting needless agony and craftily designed murders are not good (they have no value), and yet they can be creative. If this is right, then it seems to follow that creativity doesn’t require value.

One way of trying to save the value condition is by flatly denying that torture methods can be creative, and by denying more generally that creative things can be bad (Novitz 1999). But such denial seems ad hoc and implausible – “evil creativity” is not a contradiction in terms – and some have argued that this denial faces other problems besides (Livingston 2018).

Other theorists revise or qualify the value condition in order to accommodate examples of immoral creativity. Paisley Livingston (2018) proposes that a creative product only needs to be instrumentally valuable or “effective” as means to its intended end, regardless of whether that end is morally good, bad, or indifferent. Berys Gaut (2018) distinguishes between something’s being good (or good, period) versus being good of its kind. In his view, a new way of wielding blades and pulleys may be creative if it’s a good of its kind – good as a method of torture – even though it isn’t good. In order for something to count as creative, Gaut says, it doesn’t need to be good; it just needs to be good of its kind.

Alison Hills and Alexander Bird (2018) are unconvinced by such qualifications. They contemplate an elaborate torture device that ends up killing its victims immediately, “without enough suffering on the way.” The device may still be creative, they hold, even though “as a method of torture, it’s no good” (p. 98). Indeed, they argue, a creative item needn’t be good in any way at all, not even for its creator. The ineffective torture device just described doesn’t satisfy its creator’s preferences, it doesn’t give him pleasure, it isn’t an achievement, it doesn’t contribute at all to his well-being – and yet, they contend, it may be creative, provided that it’s new and was produced in the right way. Exactly what “the right way” amounts to is the topic we turn to next.
2.2 Other proposed conditions

With or without the value condition, some theorists argue that a product must satisfy one or more further conditions, beyond being new, in order to count as creative. The four most prominent proposals are that the product must be (i) surprising, (ii) original (i.e. not copied), (iii) spontaneous, and/or (iv) agential. Each of these is a condition on the process of creativity. To be clear, we are still concerned with what it means for a product to be creative, but the proposals we’ll now consider say that in order for a product to count as creative, it must be brought about in the right way.

2.2.1 Surprise

Margaret Boden holds that a creative product must be “new, surprising, and valuable” (2004: 1; cf. Boden 2010; 2014; 2018). It is perhaps most natural to assume that being surprising – like being new and valuable – is a feature of a product. But while Boden does think of creative products as surprising, her interest is more fundamentally in the underlying generative process, in how a creator manages to make something surprising. In her view, there are “three types of creativity” – combinatorial, exploratory, and transformative – “which elicit different forms of surprise, [and] are defined by the different kinds of psychological processes that generate the new structures” (2010: 1, italics added).

Combinatorial creativity occurs when old ideas are combined in new ways. Obvious examples include fictional hybrid creatures or chimeras: add wings to a horse (Pegasus), add the tail of a fish to a woman’s head and upper-body (a mermaid), add a lion’s body to a woman’s head and torso (Sphinx), and so on. Other combinations are found in analogies, such as when Niels Bohr compared an atom to the solar system. The term “combination” can refer either to the product of things combined or to the process of combining them, but Boden’s focus is on the process here, on the fact that one way to generate new ideas is to begin with old ideas and combine them in new ways.

To explain her other two kinds of creativity, Boden invokes the notion of a “conceptual space,” which is roughly a system comprising a set of basic elements (e.g. basic ideas or representations).
as well as rules or “constraints” for manipulating or re-combining those elements. A conceptual space is not a painting, song, or poem, for example; it’s a way of creating a painting, song, poem, or theory. The rules or constraints are “the organizing principles that unify and give structure to a given domain of thinking.” And so a conceptual space is “the generative system that underlies that domain and defines a certain range of possibilities: chess moves, or molecular structures, or jazz melodies” (1994: 79). We could think of a conceptual space as not just a set of thoughts but also a style of thinking defined by rules for generating new thoughts.

“Within a given conceptual space,” Boden observes, “many thoughts are possible, only some of which may have been actually thought” (2004: 4). Some conceptual spaces contain more possibilities than others. Consider different games. Tic-tac-toe is such a simple game that all of its possible moves have already been made many times over. The same is not true in chess, by contrast, which allows for a mind-boggling number of possible moves. The range of possible ideas is also practically inexhaustible in literature, music, the visual and performing arts, as well as the various domains of theoretical inquiry. And within those pursuits, there are various “structured styles of thought” – genres, paradigms, methodological orientations – which Boden thinks of as conceptual spaces.

Boden argues that the elements as well as the operating rules of a conceptual space can be, and in some cases have been, captured in computer programs. She has used this point not only to argue that computers can be creative (a topic we’ll return to below in §5), but also to suggest that we should employ the computational model of the mind in order to explain how humans create.

With her notion of conceptual spaces in hand, Boden says that exploratory creativity occurs within a given conceptual space. The new idea that emerges is one that was already possible within that space, because it was permitted by its rules. “When Dickens described Scrooge as ‘a squeezing, wrenching, grasping, scraping, clutching, covetous old sinner,’” Boden writes, “he was exploring the space of English grammar” in which “the rules of grammar allow us to use any number of adjectives before a noun” (Boden 1994: 79). Dickens’s description may strike us somewhat surprising, unexpected, or improbable, but it doesn’t have an air of impossibility about it.
By contrast, Boden argues, another form of creativity does. In this kind of case, the creative result is so surprising that it prompts observers to marvel, “But how could that possibly happen?” (2004: 6). Boden calls this transformational creativity because it cannot happen within a pre-existing conceptual space; the creator has to transform the conceptual space itself, by altering its constitutive rules or constraints. Schoenberg crafted atonal music, Boden says, “by dropping the home-key constraint,” the rule that a piece of music must begin and end in the same key. Lobachevsky and other mathematicians developed non-Euclidean geometry by dropping Euclid’s fifth axiom. Kekulé discovered the ring-structure of the benzene molecule by negating the constraint that a molecule must follow an open curve (Boden 1994: 81-3). In such cases, Boden is fond of saying that the result was “downright impossible” within the previous conceptual space (Boden 2014: 228).

Boden’s definition of creativity has perhaps been most influential among researchers who share her interest in computer creativity (e.g. Halina 2021; Miller 2019: ch. 3; du Sautoy 2019). In a variation of Boden’s account, one philosopher proposes that what makes a mental process creative is not that it actually involves “the recombination of old ideas or the transformation of one’s conceptual space”, but rather that the creator experiences the process as having one of those features (Nanay 2014).

**2.2.2 Originality**

Maria Kronfeldner (2009; 2018) argues that the process of making something creative must exhibit originality. As she uses the term ‘original’, it does not simply mean ‘new’; instead, it has to do with the kind of causal process the creator must employ. She motivates her view by asking why it’s the case that, as we noted earlier, psychological novelty is required for creativity while historical novelty is not. Why is it, for example, that Newton’s invention of calculus was creative even if Leibniz invented it first? The answer, of course, is that it’s because Newton didn’t copy his calculus from Leibniz. Insofar as Newton came up with calculus independently, on his own, then he exhibited originality in his discovery, even though someone else got there first. This originality, Kronfeldner argues, is essential to creativity.
2.2.3 Spontaneity
Kronfeldner (2009; 2018) also argues that *spontaneity* is required for creativity. An idea occurs spontaneously to the extent that it is produced without foresight or intentional control. If you were to foresee the output of the creative process at the beginning of that process, then you wouldn’t need any further process to come up with it. So if an idea is creative, you cannot have fully seen it coming. To that extent, insight comes as a surprise, hence the common phenomenological observation that creative breakthroughs feel like they come unbidden or out of the blue: “Eureka!”, “Aha!”, a lightbulb turns on.

2.2.4 Agency
Some philosophers argue that an item does not count as creative unless it has been produced by an agent. Consider a unique snowflake with an intricate shape, a distinctive sunset with stunning layers of red-orange hues, a novel patterning of dunes across a wind-blown desert. All of these things are aesthetically valuable and new. None of them are creative, however, insofar as they all occurred naturally and were not made by an agent. Gaut uses examples like these to argue that creative things must be created by *agents* (Gaut 2018: 129-30; cf. Gaut 2010, and Gaut 2014) and several other philosophers agree (Carruthers 2005, 2011; Kieran 2014a, 2014b; Stokes 2008; Stokes 2011; Stokes 2014; Paul and Stokes 2018).

Of course, many theists would maintain that everything in nature is the handiwork of an agent – namely, God – and so arguably it would make sense for them to regard a natural phenomenon as creative if it is valuable and new. For theists, the unparalleled beauty of nature is a reason to praise the Creator. But this only supports the conceptual point that creativity, by definition, requires agency. We may coherently regard valuable new things as creative if we attribute them to a creative agent, as the theist does with the natural world; otherwise, we can’t. So again, it seems, creativity requires agency.

This leaves open the question of exactly how a creator’s agency must be exercised in order for the result to count as creative. Some philosophers argue that the agent’s act of creation must be *intentional*. Suppose you are snowboarding on a powder day and, unbeknownst to you, the tracks from your board result in a pleasing new pattern as viewed from high above. The new pattern has
aesthetic value, but it isn’t creative. And that is because you didn’t intend to make it. Underlying this intuition, as well as our intuitions about the natural phenomena above, is the fact that “creative” is a term of praise, and we do not extend praise (or blame) for things that are not done by an agent, or for things that an agent doesn’t do in some sense intentionally.

While a number of philosophers endorse some version of the agency requirement for creativity, many theorists make no mention of it, whether to endorse it or reject it, including all of the psychologists cited above. Further, at least two philosophers are willing to attribute creativity to natural phenomena like trees and evolutionary processes: Arnheim (2001) and, in recent work, Boden (2018). These latter theorists don’t discuss agency as such, but insofar as the natural phenomena they call creative are not the result of agency, their view would imply that agency isn’t required for creativity.

The four proposals we’ve just considered all say that a product must arise from a certain kind of process – a process that exhibits surprise, originality, spontaneity, or agency – in order to count as creative. While there is wide agreement among philosophers that creativity requires some special kind of process, not just a special product, there is no consensus on what is required of the process. Of the four process conditions described here, the agency condition seems to be the one that is explicitly endorsed by the greatest number of philosophers thus far, though even they are still just a handful. And as we’ve seen, the other proposed conditions have serious arguments in their favor as well.

Some philosophers argue that if any process requirement is correct, this has an intriguing corollary for judgements about creativity: Even when we are explicitly judging only that a product is creative, we are implicitly assuming something about the process by which it was made. Suppose, for illustration, that the agency requirement is correct – that being generated through an agential process is built into the very concept of a creative product. Suppose further that you are applying that concept competently. It follows that if you come across a captivating arrangement of stones on the beach and you judge it to be creative, you are at least implicitly assuming that it was created through an agential process. If someone later persuades you that the stones happened to be moved into place by the wind and waves, not by any agent but just by
chance, then you may still regard the result as aesthetically interesting but you would have to rescind your judgement that it is creative. So if the agency condition is correct, whenever you point to some item and say, “This is creative,” what you are saying, in part is, “This resulted from a creative process.” Furthermore, on this view, analogous implications follow if any other process condition is correct (Paul and Stokes 2018).

2.3 Is creativity a virtue?

Having considered what is required for something to count as a creative product, and whether it must be produced by a certain kind of process, we now turn to analysis of the creative person.

Some theorists suggest that creativity, as an attribute of persons, is an ability to perform creative acts or produce creative things (Boden 2004). Others argue, however, that creativity isn’t merely an ability. An ability is something you can possess without ever putting it to use. You might have the ability to learn Swahili, for example, without ever making the effort to learn that language, despite having ample opportunities to do so. Creativity is different in this regard. If someone has the ability to be creative but never uses that ability when given numerous chances to do so, we would not call that person creative. Creative people are not merely able to act creatively. They are, moreover, disposed to exercise that ability, such that they do act creatively, at least some of the time, when the occasion arises. On this view creativity is a disposition, also referred to as a trait (Grant 2012; cf. Gaut 2014, 2018).

Philosophers have long distinguished virtues as a special subclass of dispositions or traits. In Western philosophy, the tradition of theorizing about virtues goes back to the ancient Greeks, and over the last half-century it has enjoyed a renaissance in ethics (Hursthouse and Pettigrove 2016) and, more recently, in epistemology (Turri, Alfano, and Greco 2021) and aesthetics (Lopes 2008, Roberts 2018, Hills 2018). Traditional examples of virtues include wisdom, justice, temperance, and courage. Should creativity be added to the list?

The answer depends, of course, on what it means for a trait to be a virtue. At the very least, a virtue is a trait that is good or valuable. So whether creativity counts as a virtue in this minimal
sense depends on whether creativity is necessarily valuable, a point which is contested, as we saw in the previous section. In fact, those who contend that creativity isn’t necessarily valuable often do so in order to prove that it isn’t a virtue.

But let’s suppose for the sake of argument that creativity is indeed a valuable trait. Is it also a virtue in some more robust sense? Virtue theorists commonly take their cue from Aristotle’s classic discussion in the *Nichomachean Ethics*. Citing justice and temperance as paradigm virtues, Aristotle asserts that a trait must meet at least three conditions to count as a virtue:

> Virtuous acts are not done in a just or temperate way merely because they have a certain quality, but only if the agent acts in a certain state, viz. (1) if he knows what he is doing, (2) if he chooses it, and chooses it for its own sake, and (3) if he does it from a fixed and permanent disposition of character (*EN* II.4, pp. 1105a28–1105a33).

So, for example, if you return something you’ve borrowed, that act exhibits the virtue of justice if and only if (1) you know that you’re returning what you borrowed, (2) you choose to do so because it is the just thing to do, and for no other reason, and (3) you are disposed to do the just thing across the range of circumstances when the opportunity arises. In addition to justice and temperance, Aristotle enumerates other ethical virtues like prudence, generosity, and courage, as well as the intellectual virtue of theoretical wisdom. In his view, each of these traits requires one to meet the three conditions above. While he does not consider whether creativity is a virtue, we may ask whether creativity also has these three criteria. Does one have to meet these three requirements in order to count as creative?

We’ll begin with the third requirement to set it to one side. Does a person’s act count as creative only “if he does it from a fixed and permanent disposition of character”? Examples suggest otherwise. Consider the poet Arthur Rimbaud, who abandoned poetry at the age of 21 to pursue a life of adventure. The fact that he never produced another poem after that does not count against the fact that he was a creative poet in his youth (Gaut 2014). Unlike the Aristotelian virtues, then, creativity does not have to be a permanent disposition.
Even so, it would still be significant if creativity turned out to be like an Aristotelian virtue in meeting the first two requirements. And arguably, creativity does meet the first requirement. A person doesn’t count as doing something creative unless “he knows what he is doing.” This was already implied by the agency condition for creativity discussed earlier.

Where things get interesting is with Aristotle’s second criterion for virtue. In order for your action to count as virtuous, he says, you have to do it “for its own sake” – i.e. you have to do it because you value virtue as an end itself, and not as a means to some external reward like praise, money, status, fame, or winning a competition. Consider the virtue of generosity, for instance. If you give money to someone in need merely because it will make you look good in the eyes of your friends, then you aren’t really being generous. Your act may outwardly look like generosity, but it’s not the real thing. To exhibit real generosity, you have to pursue generosity as an end in itself; you have to help others just for the sake of helping others. Now contrast being generous with being polite. If you compliment your colleague on the good work she’s done, then even if you’re doing this in order to manipulate her, you are being polite to her. You can have an ulterior motive for being polite. So politeness is not a virtue the way generosity is.

Is creativity a virtue in this respect? That is, does being creative require acting creatively for its own sake? Matthew Kieran’s (2014a, 2014b, 2018) answer is a qualified yes. While he grants that you can be motivated by external rewards to exhibit “minimal creativity” in producing valuable new things, he maintains that “exemplary creativity” requires you to be motivated by the value of creativity itself. Thus, in his view, exemplary creativity is a virtue.

To support this claim, Kieran points to a research program in psychology which purports to show that creativity is driven by “intrinsic motivation” rather than “extrinsic motivation.” A classic experiment in this program is “the magic markers study”, in which kids end up producing less creative drawings when they are offered a prize (Lepper et al. 1973). Many other studies have reported similar results, which lead Teresa Amabile to conclude, at first without qualification, that creativity is enhances by intrinsic motivation and hampered by extrinsic motivation (Amabile 1983: 107).
Further research introduced complications. In some studies, subjects were given “immunization techniques” whereby they were first primed or trained to focus on intrinsically motivating factors like the pleasure or aesthetical value of engaging in artistic activities, and it was found that when they engaged in those activities afterward, external rewards actually *enhanced* their creativity. As researchers interpreted these findings, offering reward can support one’s intrinsic motivation, provided that the reward works either to boost one’s sense of agency or to provide useful feedback about what’s working and what isn’t. Intrinsic motivation is still what fuels creativity, on this interpretation; rewards help only indirectly, when they reinforce intrinsic motivation. This lead Amabile to revise her hypothesis as the Intrinsic Motivation Principle (IMP): “Intrinsic motivation is conducive to creativity; controlling extrinsic motivation is detrimental to creativity, but informational or enabling extrinsic motivation can be conducive, particularly if initial levels of intrinsic motivation are high” (1996: 107). Kieran takes this as evidence for his claim that creativity, or at least what he calls exemplary creativity, requires intrinsic motivation and is therefore a virtue in that respect.

Objecting to this proposal, Gaut cites evidence that extrinsic motivation is not always detrimental to creativity. In one study, students in an introductory psychology class came up with more creative short story titles if they were offered a financial reward (Eisenberger and Rhodes 2001). In the studies where immunization techniques were used, proponents of IMP argue that rewards enhance creativity only indirectly, by buttressing intrinsic motivation. But in this case no such techniques were used, and so it seems the prospect of a reward enhanced creativity directly.

Further, Gaut argues that this point coheres with the role that rewards seem to play in so many real-world cases of creative achievement. In their quest to discover the structure of the DNA molecule, Watson and Crick were driven “to imitate Linus Pauling and beat him at his own game” (Watson 1999: 46). Picasso and Matisse were both spurred on by their rivalry with each other (Flam 2003: 37). Paul McCready says he was driven to invent his award-winning human-powered glider in 1977 because he needed the prize-money to pay off his debts: “I felt that I didn’t have the time to mess with such things, but I had this strong economic motivation to take an interest in man-powered flight, so I charged around trying to figure out a way to solve it” (quoted in Sternberg and Lubart 1995: 242). One historian argues that in WWII the Poles beat
the French in cracking the Germans’ Enigma Code because they were more terrified of German invasion (Singh 1999: ch. 4). Gaut quips: “Fear of death is a more powerful motivator than the intrinsic satisfactions of code breaking” (Gaut 2014: 196).

Finally, Gaut points out that even if IMP is true, it is only a causal, probabilistic claim: intrinsic motivation is “conducive” to creativity; extrinsic motivation is “detrimental”. But for a trait to be a virtue, intrinsic motivation must be conceptually necessary for the exercise of that trait. If we learn that someone gave to charity just to enhance his reputation, we conclude that he wasn’t really being generous. By contrast, if we discover that someone created gorgeous artwork just for the fame and glory, we may then lose some of our admiration for her creativity, but we do not deny that she was being creative.

Kieran could remind us that, in his view, intrinsic motivation is not required for all creativity, but only for the special form of it that he calls exemplary creativity. Anticipating this reply, Gaut says that to distinguish between two forms of creativity is just to concede his point. There are not two forms of generosity, one that requires intrinsic motivation and another that does not. If your act of giving isn’t motivated by the right kind of reason, then it doesn’t count as an act of generosity at all. Thus, Gaut argues, to grant the possibility of non-exemplary creativity is to grant that, unlike generosity, creativity isn’t a virtue in the traditional Aristotelian sense.

Another way to examine relations between creativity and virtue is through the lens of virtue epistemology. Linda Zagzebksi defines a virtue “as a deep and enduring acquired excellence of a person, involving a characteristic motivation to produce a certain desired end and reliable success in bringing about that end” (1996: 137, italics added). While there is a lot packed into this definition, what we’ll pinpoint here is the idea that virtue involves reliable success in achieving a desired end, and that the agent who is epistemically virtuous, in particular, is one who is reliably successful in achieving knowledge. Knowledge requires truth, of course, so an epistemic virtue is a trait that is “truth-conducive.” Epistemologists typically regard a process as truth-conducive to the extent that the beliefs it produces are more often true than false. But Zagzebksi proposes that a process or trait may be truth-conducive in a different sense, insofar as it is necessary for advancing knowledge in some area, even if it produces a very small proportion
of true beliefs. Creativity, she claims, is truth-conducive in this sense, and thus it qualifies as an epistemic virtue (p. 182). Also note the emphasis on agency. In contrast to contemporary western epistemology, virtue epistemology identifies the agent (rather than, say her beliefs) as the essential locus of epistemic valence; it is the agent who is epistemically good (or not). This emphasis comports well with the proposal, discussed above, that the creator’s agency is necessary for genuine creative achievement. A virtue-theoretic approach thus illuminates what may (as we will discuss again later) be essential to creativity, namely, a process that non-trivially involves a responsible agent.

We’ve seen that even after we fix a specific referent for the term “creative” – whether it be a person, process, or product – there are lively disagreements about what it means. These debates often seem to presuppose that the term always expresses the same concept, for which we can seek necessary and sufficient conditions. But we’ve also seen that some theorists distinguish between different concepts of creativity, corresponding to different senses of the term ‘creative’. In future work we may see theorists develop such pluralistic approaches in more detail. The trick, though, will be to give principled reasons for multiplying different concepts of creativity so that the analyses do not simply reduce to saying that anything goes.

3 Can creativity be learned?

There is a long tradition of thinkers who answer no to the question above. Two of the most influential are from the 18th century – Edward Young and Immanuel Kant – who were concerned specifically with genius, the capacity for achieving the very highest levels of creativity. In Conjectures on Original Composition (1759/1966), Young says, “An Original may be said to be of a vegetable nature; it rises spontaneously from the vital root of genius; it grows, it is not made . . .” (p. 7). His idea is that originality emerges naturally from something implanted in us by nature, and it can only be hindered by learning. Young seems to think of learning as proceeding either through imitation or through the following of rules, and both, he thinks, are detrimental to originality. Regarding imitation he writes, “Born Originals, how comes it to pass that we die Copies? That meddling ape Imitation. . . destroys all mental individuality . . .” (p. 20). And
insofar as learning is “a great lover of rules,” he warns that it “sets rigid bounds to that liberty, to which genius often owes its supreme glory” (p. 13).

Kant makes similar claims in his *Critique of Judgment* (1790). Like Young, he takes genius to be a natural capacity, though a very rare one: “such a skill cannot be communicated, but is apportioned to each immediately from the hand of nature and dies with him” (p. 309). It certainly cannot be learned through imitation: “genius is entirely opposed to the spirit of imitation. Now since learning is nothing but imitation, even the greatest aptitude for learning, facility for learning (capacity) as such, still does not count as genius” (p. 308). Nor can it be learned through rules, Kant holds, for genius is “the talent (natural gift) that gives the rule to art . . . the inborn predisposition of the mind (*ingenium*) through which nature gives the rule to art” (Kant 2000, p. 307). For Kant, a genius does not follow rules; a genius *invents* the rules, indirectly, by creating exemplary works from which other artists might extract rules and undertake “a methodical instruction in accordance with rules” (p. 318).

Young and Kant are concerned with genius, specifically, but if we extend their reasoning to creativity in general, as Berys Gaut (2014) has noted, we can discern two lines of argument:

**The imitation argument**

1. All learning is a form of imitation.
2. Imitating someone or something is incompatible with being creative.
   

   So, one cannot learn to be creative.

**The rules argument**

1. All learning consists in the following of rules.
2. Following rules is incompatible with being creative.
   

   So, one cannot learn to be creative. (p. 266)

Gaut points out, first of all, that both arguments are invalid. In both cases, what the premises would entail is that learning cannot be creative, that, in other words, you cannot *learn creatively* (a claim about *how* you can learn). But even if that were true, it wouldn’t follow that you cannot *learn to be creative* (a claim about *what* you can learn). If you absorb the advice of a creative
writing manual then this act of learning may not itself be creative. But if the manual is effective – and we’ll see in a moment how it can be – then what you will learn is how to become more creative.

Gaut also challenges the premises of these arguments. To start with the first premise of the imitation argument, it simply isn’t true that all learning proceeds through imitation, as we learn many things through direct experience, trial and error, and many other means.

The second premise is also suspect. Something superficially close to it is true: mere copying is incompatible with being creative. But to the extent that we learn from others by imitating them, this is not merely a matter of copying them. When a child learns to speak the language of those around her, she doesn’t simply parrot the exact same sentences she hears; she absorbs the vocabulary and underlying grammar in a way that enables her to form new sentences of her own devising.

Now for the rules argument. Contrary to the first premise, it cannot be the case that all learning consists in following rules, Gaut argues, because for any given rule there will be hard cases where it is unclear whether or how the rule applies to them, and so an individual still has to use her own judgment in applying the rule.

The second premise is false too. Recall the distinction from §3 above between two kinds of rules. An algorithm serves as an exact plan, specifying both the outcome and the path for getting to it in exact detail. In contrast, a heuristic is a looser ‘rule of thumb’ that leaves room for an agent to exercise her own judgment, choice, and creativity in determining whether, when, and how to follow the rule. While algorithms, in this sense, may preclude creativity, heuristics do not, which is why, as we’ll see below, the teaching of creativity so often takes the form of heuristics.

There is a sense in which the question at hand can be answered empirically: We can show that creativity can be taught simply by pointing to cases where it has been taught. Gaut himself discusses such examples as they occur in mathematics and fiction writing, which we’ll turn to below. But while such cases may suffice to show that creativity can be taught, Gaut further enriches our understanding by explaining how this is possible. He does so partly by articulating
and then debunking the imitation and rules arguments to the contrary. But in addition, he offers the following positive argument to show that creativity can be taught and learned. He calls it ‘the constitutive argument’ because it begins with his view of what constitutes or defines creativity itself.

*The constitutive argument*

1. Creativity is a disposition – involving both the ability and the motivation – to produce things that are new and valuable, and to do so in ways that express one’s agency through “the exercise of choice, evaluation, understanding, and judgment” (Gaut 2015: 273).
2. At least some people can learn to enhance their creative motivation.
3. At least some people can learn to enhance their creative abilities.
   
   So, at least some people can learn to become more creative.

Premise 1 recapitulates the point we’ve already seen Gaut and others defend (in §2,3 above), that creativity is not merely an ability but a disposition or trait, whereby the creative person is disposed or motivated to exercise that ability when given the opportunity.

In support of premise 2, Gaut argues that you can strengthen both your intrinsic motivation to be creative (when you take pleasure in your creative activities), as well as your extrinsic motivation to be creative (when you are rewarded with praise, grades, pay, etc. for your creative efforts).

Defending premise 3, Gaut points out that you can develop your ability to produce valuable new things by practising and strengthening the relevant skills. And this development can be substantially aided by learning certain heuristics.

Heuristics are indeed a staple of education in creative pursuits from mathematics (draw the figure; consider special cases; consider extreme cases; generalize the problem; look for a related problem, etc. – see Pólya 1990; Schoenfeld 1982; 1987a; 1987b) to creative writing (write what you know; be specific and detailed in describing sensory experiences; practice seeing similarities between dissimilar things; show, don’t tell, etc. – see Bell and Magrs 2001; Anderson 2006; Maybury 1967; Kaufman and Kaufman 2009). Gaut also identifies several heuristics that might
be used to foster creativity in philosophy, even among children (cf. M. Gaut 2010; Gaut and Gaut 2011).

With this last theme, Gaut has a kindred spirit in Alan Hájek (2014, 2015, 2017, 2018), who has independently proposed that by using various heuristics, philosophers can enhance their abilities to make valuable contributions to their field, including ideas that are distinctively creative. It has been said that anyone of average talent can become a strong chess player by learning and internalizing certain chess heuristics: “castle early”, “avoid isolated pawns”, etc. Analogously, Hájek suggests, philosophy has a wealth of heuristics — philosophical heuristics — although they have not been as well documented and studied. Sometimes these take the form of useful heuristics for generating counterexamples, such as “check extreme cases.” Sometimes they suggest ways of generating new arguments out of old ones, as in “arguments involving possibility can often be recast as arguments involving time, or space.” Sometimes they provide templates for positive arguments (e.g., ways of showing that something is possible). Hájek offers a catalogue of such philosophical heuristics to show that, contrary to a common assumption, creativity, even in philosophy, can be compatible with, and enhanced by, following rules.

4 Can creativity be explained?

Upon observing the work of creative people, it is natural to wonder: How do they do that? How do people create? The issue we turn to now is whether we could, at least in principle, answer this question scientifically, using the methods of modern empirical psychology and other cognitive and behavioral sciences. Those who take a negative stance on this matter are not merely saying that, in practice, it would be exceedingly difficult for science to explain creativity. They are saying that it’s altogether impossible that science could ever explain creativity.

Hospers (1985) defends this kind of pessimism based on the variety and complexity of creativity, given that creativity occurs not only in art, but in science, theorizing of any sort, engineering, business, medicine, sport, gaming, and so on. At least two worries may follow. First, given the complexity of any one of these individual domains, one might worry that there are simply too many variables to allow for a clear explanation. Art provides a paradigmatic example. Consider
an artwork that you judge to be masterful (a sculpture, a painting, a film). Now imagine attempting to describe or identify all the reasons for which you think it is masterful. Take as much time as you like but, the skeptic will urge, any long description you construct will invariably strike you as woefully incomplete by comparison to the artwork, and the experience thereof. So, if the creative achievements of artists, in all of their complexity, cannot even be adequately described, we have little reason to think that such achievements can be explained.

How can theorists respond to these skeptical worries? Both the complexity and generalizability worries might be partially disarmed by noting analogies between creativity and other phenomena. For instance, consider the range of bodily movement involved in some of the very domains of activities listed above: art, science, engineering, medicine, sport. The kinds of bodily action specific to these domains are complex and vary dramatically: the relevant physical movements of the surgeon are much different from the tennis player. However, it is not plausible that this complexity and variety precludes explanation of bodily action in those domains. It simply implies that some features of the explanation will be context-sensitive, that is, specific to that domain of activity. And further to the analogy: the fact that the long description of, say, the tennis serve is incomplete does not preclude it from being apt and explanatory. If this line of reasoning is sound for bodily action, why not also for creative action?

At this point, one might argue that while complexity and generalizability worries would only show that creativity is difficult to explain in practice, the very nature of creativity implies, more strongly, that it could never be explained, not even in principle. Resources to support this kind of pessimism may be adduced from various past philosophers. We need to tread carefully, however, since most of the figures we are about to consider were writing long before the rise of the relevant sciences, so they could not have made any explicit claim either way as to whether creativity could be explained by those sciences. Nevertheless, some of them did make claims which entail, or seem to entail, that creativity simply isn’t the kind of thing that could be explained through scientific inquiry as we understand it today.

The classic expression of such a view comes from Plato. In his dialogues, Plato features his teacher Socrates as a spokesperson for his own views, and in the Ion he has Socrates argue that
poets do not produce poetry through knowledge or skill. When you exercise a skill (*technē*), you apply techniques, rules, or methods to perform a given activity, like charioteering, fishing, or commanding an army. In principle, one could explain these activities by identifying the techniques they involve, and a student or apprentice could learn these activities by applying and practicing those techniques. But poetry is not like that, in Socrates’ view. A poet can only imitate the application of rules or techniques, mimicking the surface appearance of skill. Voicing an idea that was familiar in Ancient Greek culture, Socrates suggests that poetry emerges instead through divine inspiration, whereby a human being is *inspired* – literally ‘filled with a spirit’, with a god or goddess, with a muse:

You know, none of the epic [or lyric] poets, if they’re good, are masters of their subject; they are inspired, possessed, and that is how they utter all those beautiful poems. … [They] are not in their right minds when they make those beautiful lyrics, but as soon as they sail into harmony and rhythm they are possessed by Bacchic frenzy. […] For a poet is an airy thing, winged and holy, and he is not able to make poetry until he becomes inspired and goes out of his mind and his intellect is no longer in him. As long as a human being has his intellect in his possession he will always lack the power to make poetry or sing prophecy. […] You see, it’s not mastery [*technē*] that enables them to speak those verses, but a divine power. That’s why the god takes their intellect away from them when he uses them as his servants, as he does prophets and godly diviners, so that we who hear should know that *they* are not the ones who speak those verses that are of such high value, for their intellect is not in them: the god himself is the one who speaks, and he gives voice through them to us. In this more than anything, then, I think, the god is showing us, so that we should be in no doubt about it, that these beautiful poems are not human, not even *from* human beings, but are divine and from gods; that poets are nothing but representatives of the gods, possessed by whoever possesses them. (534a-d)

Socrates repeats this view in the *Phaedrus*: “Some of the greatest blessings come by way of madness, indeed madness that is heaven-sent” (244a). He adds that while a poet may have some
kind of skill, anyone who aspires to make poetry purely by skill, without the madness or the muse, will fail (245a).

It’s important to note that ‘madness’, for Plato, is a supernatural affair. From the vantage of contemporary behavioral science, we think of madness – or rather, mental illness – as a pathology arising from some combination of genetic and environmental factors, and those factors can be studied scientifically. So even if creativity is linked to mental illness – a highly controversial proposition – it could still be entirely within the scope of science. However, Plato’s talk of ‘madness’ does not refer to any naturally occurring pathology, but rather to the result of divine intervention: the poet is taken over or “possessed” by the muse and that is precisely why he is “out of his mind”. Plato’s poet suffers divine madness.

According to this story, then, the person we call a poet isn’t really a creator of poetry, but is merely the vessel through which a divine being delivers poetry. If it is literally true that the source of poetry is supernatural, then poetic creativity could never be explained by science, which is limited to the investigation of natural causes. (For more on Plato, see Asmis 1992.)

This kind of supernaturalism has enjoyed a long afterlife in Western thought. In ancient Rome, the Latin term ‘genius’ referred to a guiding spirit that was thought to accompany each person throughout their lives. The genius of an artist would occasionally deliver art through that person in the manner of Platonic inspiration.

Conceptions of the artist take a new turn when the idea of genius is transformed in the 18th century. As we saw above, Immanuel Kant defines genius as a natural capacity that a certain kind of artist possesses innately and which partly constitutes that artist’s identity. So rather than saying that a gifted artist “has a genius”, Kant says that such a person “is a genius”. What distinguishes the genius is fundamentally an imaginative capacity – an ability to engage in a “free play” of imagination to produce artworks of “exemplary originality”. These works are exemplary not only in the sense that they have artistic or aesthetic value, unlike “original nonsense”; they are also exemplary in the more radical sense of providing an exemplar – a new paradigm and precedent – for lesser artists to follow. A work of genius sets a new standard of
artistic value, and, looking to that exemplar, lesser artists may then extract techniques or rules for their own craft. The genius therefore “gives the rule to art”. In creating such works, the genius does not follow any rules or methods. Instead the genius creates art through a “free play of imagination” – where the terms “free” and “play” characterize the nature of an activity unconstrained by any pre-established methods or rules:

Genius is a talent for producing something for which no determinate rule can be given, not a predisposition consisting of a skill for something that can be learned by following some rule or other. (§46)

Kant thought that genius, so conceived, is limited to the fine arts, poetry being chief among them. Meanwhile, in Kant’s view, there is no room for genius in science, for example, where good theories and hypotheses must emerge from the careful application of scientific method, and so he said that even Isaac Newton, “that great man of science”, was not a genius. We’ll soon consider why this view might seem to entail that creativity is inexplicable, but first it will be helpful to bring another figure, Arthur Schopenhauer, who was deeply influenced both by Kant and by Plato.

Like Kant, Schopenhauer thought of genius as a natural capacity that is limited to the fine arts. He also echoes Plato’s sentiments about madness, famously stating that “genius and madness have a side where they touch and even pass over into each other” (The World as Will and Representation, I, 190), and that “Genius lives only one story above madness” (Parerga and Paralipomena). In a state of madness, Schopenhauer’s genius is like Plato’s poet in experiencing a momentary loss of self, but what displaces the self is not any divine being but rather a pure Idea which seizes the author’s being and becomes the object of both his fascination and his artistic expression:

We lose ourselves entirely in this object, to use a pregnant expression; in other words, we forget our individuality, our will, and continue to exist only as pure subject, as clear mirror of the object, so that it is as though the object alone existed without anyone to perceive it, and thus we are no longer able to separate the perceiver from the perception,
but the two have become one, since the entire consciousness is filled and occupied by a single image of perception. (Schopenhauer, *World*, vol. I, pp. 178–179, §34).

With their focus on genius construed as a natural capacity, figures like Kant and Schopenhauer abandon the supernaturalism of the Platonic muse. Nevertheless, they retain the idea that creativity – specifically genius-level creativity in the fine arts – is not a matter of exercising a skill or applying given rules, methods, or techniques.

As we noted earlier, these figures did not and could not have explicitly denied that creativity could be explained by the sciences of the 20th and 21st centuries, but they are commonly taken to represent such a denial (Kronfeldner 2018). Why?

Perhaps figures like Kant and Schopenhauer seem to make creativity, or at least creative genius, inexplicable insofar they suppose it to be innate and as they have no story to tell about how one came to acquire an innate capacity except to say that it was either an accident of chance (which is no explanation at all) or a gift from God (which again is not a scientific explanation). But while these figures seemed to think of artistic genius as being endowed entirely by nature with no contribution from nurture, modern genetic theory rejects that dichotomy. Instead of positing all-or-nothing natural abilities, behavioral scientists today think in terms of genetically inherited predispositions. In order for a genetic predisposition to develop into a trait with an observable phenotype, it needs to be triggered and shaped through a complex interaction between an organism’s genes and certain kinds of stimuli or environmental conditions. There are still open questions about exactly how, and how much, genes and environment feed into the development of any given trait, but it’s misguided to pose the binary nature-versus-nurture question as if the two were mutually exclusive (see Tabery 2014). Many researchers agree that some people have a stronger natural predisposition toward creativity than others, and that genius-level creativity partly stems from such a predisposition. Even so, the predisposition itself can be understood scientifically in terms of genetic heritability. (For a sampling of the relevant studies, see the essays collected in S.B. Kaufman 2013.)
Perhaps creativity seems inexplicable according to these accounts because it doesn’t follow rules or methods. In order to explain how to do something – how to build a boat or lead an army etc. – perhaps I need to be able to identify the rules or methods you should follow in order to practice and apply those skills. How-to explanations are instructions. But scientific explanations needn’t be instructions. A lot of good science explains how something happens – e.g. how heat melts ice or how a bat navigates its environment by echolocation – without explaining how to do it yourself.

Perhaps creativity seems inexplicable according to these accounts because creators themselves do not know how they create. But a scientific explanation needn’t be available through introspection. Most people cannot explain how their own digestive, circulatory, or perceptual systems work, but scientists who study those systems can.

Another line of thought is perhaps implicit in Kant but comes to the fore in Schopenhauer, who says that “the nature of genius consists precisely in the preeminent ability” to “consider things independently of the principle of sufficient reason, in contrast to the way of considering which proceeds in exact accordance with this principle, and is the way of science and experience” (World, vol. I, p. 192, §36). The principle of sufficient reason says that for every fact there is a cause which completely explains that fact. So the defining ability of genius is to see things in a way that transcends the causal order and defies all explanation.

A version of this view is defended more recently by Carl Hausman (1979, 1984, 1985) who frames it in terms of novelty that creativity involves. Hausman asserts that if a product is creative, it must be metaphysically novel (or in his terms, ‘genuinely novel’) in the sense that it cannot be predicted from, or explained by, prior events – not even in principle. Creativity is therefore incompatible with causal determination and causal explanation: “A causal view of explanation sets a framework for ways of denying that there is anything new under the sun” (Hausman 1984: ix). If something can be explained by prior causes, it is not metaphysically novel, and is therefore, in Hausman’s view, not truly creative.
Against Hausman’s skeptical charge, Maria Kronfeldner (2009) argues that creativity is compatible with causal determination. First, causal determinism does not preclude novelty or change. Determinism says the emergence of new kinds of things can at least in principle be predicted in advance. Importantly, though, when this prediction becomes true, then something new is added to the world. Of course, not all novelty instantiates creativity. The question is whether the kind of novelty involved in creativity must be metaphysical novelty, which is by definition incompatible with causal determination. This is doubtful. Notice that, by definition, metaphysical novelty defies natural laws. The production of something metaphysically novel would therefore require supernatural powers. Traditional Western religions conceive of God as performing the miracle of creation *ex nihilo*. But are we positing a miracle every time we describe a human artifact or achievement as creative? Surely not. As noted above, human creativity is manifest in things that are novel relative to the agent producing them or new to human history, but both of those kinds of novelty (psychological and historical) are perfectly compatible with causal determination. As Kronfeldner explains, creativity does not preclude causes in general; it only precludes certain *kinds* of causes. A creative product, she argues, must be *original* – which means that it cannot be produced through a process of copying something prior. And it must be *spontaneous* (not produced through a routine or mechanical procedure) - which means that it is to some extent independent of the agent's intentional control and previously acquired knowledge. (For more on originality and spontaneity, recall §2.2 above). Intuitively, the causes of something creative cannot simply be a matter of copying or following a routine. But it may have causes nonetheless, and cognitive science can investigate those causes, at least in principle. Indeed, as we’ll see next, it is doing so in practice.

5 The cognitive science of creativity

Although creativity has been relatively understudied by contemporary philosophers, as we noted in §1, it has been receiving a great deal of attention from psychologists over the past few decades. In 1950, J. P. Guilford gave a presidential address at the American Psychological Association calling for research on the topic, and the field soon took off with waves of research investigating the traits and dispositions of creative personalities; the cognitive and neurological mechanisms at play in creative thought; the motivational determinants of creative achievement; the range of institutional, educational, and environmental factors that enhance or inhibit
creativity; and more. Today, the blossoming of this field can be seen in the flurry of popular writing on its results; an official division of the American Psychological Association for the psychology of aesthetics, creativity, and the arts (Division 10); numerous academic conferences; dedicated peer-reviewed journals (Psychology of Aesthetics, Creativity and the Arts; Creativity Research Journal; Journal of Creative Behavior; International Journal of Creativity and Problem Solving); special issues of journals (Frontiers in Human Neuroscience 2013 and Current Opinion in Behavioral Sciences 2019); literature surveys (Hennessey and Amabile 2010; Runco and Albert 2010, 2017; Glaveanu 2014; Williams et al. 2016); textbooks (J.C. Kaufman 2009; Sawyer 2012; R. W. Weisberg 1986; 2016); and a comprehensive encyclopedia (Runco and Pritzker 2020). According to one overview, creativity has been studied by nearly all of the most eminent psychologists of the 20th century, and “the field can only be described as explosive” (Runco and Albert 1999). There is also a groundswell of new work on creativity in the fields of computer science, artificial intelligence (AI), and robotics.

The present section surveys empirical work in psychology along with some related work in neuroscience, while the next section (§6) covers research in computing, AI, and robotics. Throughout, we’ll see that philosophers are actively in dialogue with these fields under the broad, interdisciplinary umbrella of cognitive science.

The vast body of empirical research of creativity can be seen as addressing a variety of issues, but the central question that concerns us here is the one we identified above as the challenge for explaining creativity: How are people creative? This question is analogous to a number of other questions in cognitive science: How do people perceive through sense modalities such as vision? How do they form concepts? How do they acquire a language? How do they make inferences? Just as psychologists investigate the psychological and neurological processes, systems, and mechanisms at work in these other mental operations, as well as the internal and external factors that either enhance or hinder these operations, they are doing the same for creativity. There is no pretension to achieving a complete explanation which would include each and every causal factor, and provide the basis for perfectly predicting creative outcomes in advance. But to the extent that we identify some of the relevant causal factors involved in creativity we thereby make progress in explaining creativity, just as we do with other features of the mind.
As we noted in §2, the standard definition of creativity in psychology says that a product (idea or artefact) is creative to the extent that it is both new and valuable (‘effective’, ‘useful’ or ‘appropriate’), and, in turn, people and processes are creative to the extent that they produce new and valuable things. As we also noted, many psychologists do not actually employ this, or any, definition of creativity in conducting their research. In one sampling of studies of creativity published in peer-reviewed psychology journals, only 38% of them included an explicit definition of creativity (Plucker, Beghetto, and Dow 2004), as they rely in one way or another on the assumption that we know it when we see it. For example, many studies use the Consensual Assessment Technique (CAT), whereby experimental subjects produce things that are then rated for how creative they are by a panel of experts in the relevant field; so paintings are rated by professional painters, stories by published authors, etc. Many other research methodologies are used, as we’ll see below.

Empirical research on creativity departs in several ways from the traditional approaches that seemed to place creativity outside the scope of science. For starters, in stark contrast to Plato’s supernaturalism, empirical psychologists take creativity to be a completely natural phenomenon. Creative people may of course be ‘inspired’ in the sense of feeling energized or filled with ideas, but rather than being literally ‘breathed into’ by some god or muse, their thoughts and behaviors are presumed to have causes that are perfectly natural. While it is difficult in practice to identify these causes, they are not in principle beyond the reach of science.

Further, the range of phenomena that contemporary researchers countenance within the ambit of creativity is far broader and more diverse than the traditional focus on poetry and the fine arts, as creativity can be manifest in any kind of art or craft, as well as in the sciences, technology, entrepreneurship, cooking, humor, or indeed in any domain where people come up with ideas or things that are novel and valuable in some way or another. Departing from Kant, genius, the highest echelon of creativity, may be acknowledged in virtually any of these domains, not just in the fine arts. And while a few researchers (e.g. Simonton 1984, 1994, 1997, 2009; Root-Bernstein and Root-Bernstein 1999) venture to examine genius (so-called ‘Big-C’ creativity), most of them focus instead on relatively ordinary creative feats (‘little-c’ creativity) including the
kinds of story-making, drawing, and problem-solving that can be elicited on command from regular people in experimental settings. Some researchers propose that in order to understand how the mind generates new ideas, we should begin with even more rudimentary phenomena. For example, philosopher Jesse Prinz and psychologist Lawrence Barsalou focus on how we form new concepts to categorize the things we perceive, a process which they claim is creative, albeit in a “mundane” rather than “exceptional” way (Prinz and Barsalou 2002; Barsalou and Prinz 1997; cf. Child 2018).

Of course, many feats of human creativity, and the ones that are most interesting, go far beyond the basic formation of concepts. A major step toward explaining those feats is to recognize that what we call “the creative process”, as if it were a single, homogenous phenomenon, is in fact an assembly of multiple stages or operations. The simplest recognition of this fact is the Geneplore model which distinguishes just two stages: generating ideas and exploring ideas (Finke 1995; Smith, Ward, and Finke 1995). This distinction may be seen as echoing one made by philosophers of science in the early 20th century, between the context of discovery and the context of justification (Popper 1934). Other theorists posit up to eight stages of creativity (for a summary of proposals, see Sawyer 2012: 89). But the most influential stage-theory traces back to Henri Poincaré’s lecture, “Mathematical Creation” (1913), in which he identifies four phases in his own innovative work as a mathematician: (1) conscious hard work or preparation, (2) unconscious incubation, (3) illumination, and (4) verification. In his book, The Art of Thought (1926), the psychologist Graham Wallas endorses Poincaré’s four stages with corroborating evidence from the personal reports of other eminent scientists like Hermann von Helmholtz. Wallas’s scheme, as a development of Poincaré’s, is still the one that is most widely cited, and we employ a version of it here with some slightly different terminology and with two more substantive alterations: instead of “incubation”, we identify the second operation more generally as the “generation” of ideas, which may include unconscious incubation but may also occur in conscious, deliberate thought; and we add “externalization” for a total of five operations:

(1) **Preparation** – You invest a great deal of effort learning and practicing in order to acquire the knowledge, skills, and expertise required for work in a given domain.
(2) **Generation** – You produce new ideas, whether through conscious reflection or unconscious incubation.

(3) **Insight** – You consciously experience the emergence of a new idea, which would strike you with a feeling of surprise: “Aha!”, “Eureka!”

(4) **Evaluation** – You assess the idea to determine whether it should be discarded, retained, revised, or amended.

(5) **Externalization** – You express your idea in a concrete, observable form.

Artists provide compelling examples (though not the only ones) of each of these five operations. Such examples can be especially illustrative since they come straight from the artists’ mouths, as they reflect upon, and share, their creative process. The 20th century painter Jacob Lawrence was known for painting in the style of visual narratives. Lawrence developed a system, much like a filmmaker’s storyboard, for the preparation of these paintings. He would lay as many as 60 wood panels on the studio floor, each with individual scenes and sometimes with captions. From these storyboards, Lawrence would generate and evaluate ideas and insights for a visual narrative, culminating in the paintings such as those in his *Migration Series*. (Source: https://whitney.org/www/jacoblawrence/art/painting_method.html). Toni Morrison, the Nobel prize winning novelist, remarks on the labors and sustained effort required at the preparation, generation, evaluation, and externalization stages of a creative writing process. Commenting on her novel *Jazz*, she says, “I thought of myself as like the jazz musician—someone who practices and practices and practices in order to be able to invent and to make his art look effortless and graceful. I was always conscious of the constructed aspect of the writing process, and that art appears natural and elegant only as a result of constant practice and awareness of its formal structures.” She further notes that insight does not always come in a flash, “[I]t’s a sustained thing I have to play with. I always start out with an idea, even a boring idea, that becomes a question I don’t have any answers to.” (Source: https://www.theparisreview.org/interviews/1888/the-art-of-fiction-no-134-toni-morrison). Writer Ishmael Reed claims that insight can come unexpectedly and in various contexts:

One can find inspiration from many sources. The idea of Japanese by Spring originated in a news item that claimed the endowment to a major university was traced to Japanese
mob, the Yakuza. Flight to Canada began as a poem. The Terrible series began when I heard someone at party mention that there was a black figure, Black Peter, in the Dutch Christmas, and by coincidence I was invited to the Netherlands shortly afterwards, where I witnessed the arrival of Saint Nicholas and Peter on a barge that floated into Amsterdam with crowds looking on. I took photos of the ceremony … (Howell 2020: 91).

And with signature profundity, James Baldwin suggested that all elements of the creative artistic process, from preparation to externalization, require a basic enabling condition: being (and willing to be) alone (Baldwin 1962).

As Wallas recognized (1926: 81), and as the above examples suggest, the “stages” of the creative process are not necessarily discrete steps that follow one another in a tidy sequence. Creative work is messy: over time you have numerous ideas, keeping some and abandoning others in multiple rounds of trial-and-error; you incubate new ideas for one problem while you’re busy externalizing your ideas for another; and your moments of insight, evaluation, and externalization trigger further generative processes that send you cycling through these operations many times over. It’s still important to distinguish these operations, however, because, as researchers are confirming, they are enabled and influenced by different causal factors.

Among the additional stages that researchers have posited, one of the most widely discussed is known as problem-finding. Psychologists often conceptualize creative thought in terms of problem-solving: the ideas generated within the creative process are seen as candidate solutions to a given problem – where “problems” are broadly construed to include any creative aim, like that of producing a particular kind of artwork or proving a particular theorem, etc. (Flavell & Draguns 1957: 201; Newell, Shaw, and Simon 1962). But following some early work by Mihalyi Csikszentmihalyi (1965), many researchers came to appreciate that a lot of creative work is done not just in solving problems but in finding the right problem to begin with (Abdulla et al. 2020; Csikszentmihalyi & Getzels 1970; Getzels 1965; Getzels & Csikszentmihalyi 1975). While we agree that problem-finding often plays a key role in creativity, we have not assigned it to a separate stage, for the following reasons. Consider that you might settle on a problem to work on
in either of two ways. On one hand, you might choose a problem to work on from a pre-existing menu of options. In that case, your choice would fall under the evaluation phase; it’s just that the idea you select is a problem that calls for the pursuit of further ideas. If, on the other hand, you develop a new problem, you would thereby be engaging in the generation of a new idea – the new problem – which may emerge in a moment of insight. Einstein and his colleague celebrated the novelty in such problem-finding: “The formulation of a problem is often more essential than its solution, which may be merely a matter of mathematical or experimental skill. To raise new questions, new possibilities, to regard old problems from a new angle, requires creative imagination and marks real advance in science” (Einstein and Infeld 1938: 92). Either way – whether you “find” a problem by picking a pre-existing one or by coming up with a new one yourself – problem-finding, though important, does not need to be seen as an additional operation beyond the five listed above; it’s just a special case of generation, insight, or evaluation.

The next five sub-sections will respectively examine the five operations of creative work. Notice that three of them – preparation, evaluation, and externalization – are uncontroversially ordinary activities that involve no apparent mystery; it’s a challenge to explain them but no one is tempted to regard them as inexplicable or as violating the laws of nature. As we saw in §4, traditional skepticism about the possibility of explaining creativity is really focused on the two remaining phenomena: the generation of new ideas (§5.2) and the experience of insight whereby an idea seems to come out of the blue, as if from a god (§5.3).

5.1 Preparation

It’s myth that outsiders are more creative. To put yourself in a position to create anything of value, you have to spend a great deal of time and effort acquiring the relevant knowledge, skills, and expertise. In what has come to be called “the ten-year rule”, Howard Gardner (1993) found that, on average, people spend about 10 years learning and being immersed in a domain before they make any significant creative contribution to it.
Though a certain amount of rote learning is required, gaining mastery in a field is not simply a matter of passively absorbing information. Much of it involves what Anders Ericsson calls *deliberate practice*, where you focus on tasks which are a little beyond your current abilities, but which you eventually conquer through feedback and repetition. Across a variety of domains – including physics, medicine, programming, dance, and music – Ericsson found that, on average, world-class performance becomes possible for people only after 10,000 hours of deliberate practice in their chosen activity. This finding also converges on the ten-year rule, because if you engage in deliberate practice four hours a day, five days a week, that would add up to 10,000 hours in ten years (Ericsson, Krampe, & Tesch-Römer, 1993; Ericsson et al., 2006).

However, there seems to be a point at which too much formal training can dampen creativity. Simonton (1984c: 70-73) has reported that the relationship between creativity and education level is an inverted-U, as too much schooling can reinforce familiar, pre-established styles of thought. Even so, the point remains that, before you run into diminishing returns, years of preparatory learning and practice are required for exceptional creativity.

### 5.2 Generation

In this section we discuss four kinds of mental capacities or processes that researchers have posited for generating new ideas.

#### 5.2.1 Blind Variation

Some scientific theories of creativity are framed in broadly Darwinian or evolutionary terms. Donald T. Campbell (1960, 1965) proposed this approach, and it has been advanced and extended by Dean Keith Simonton (1999a, 1999b). The model comprises three basic conditions. First, there must be what we might call a *variation-generator* in the cognitive system: a mechanism that, by analogy with Darwinian evolution and genetic mutation and variation, produces variations on ideas and concepts. Second, there must be a *selection-mechanism*. This mechanism, by analogy with the role of natural or sexual selection in biological evolution, selects ideas generated by the variation-generator. Finally, there must be a *retention mechanism*:
by analogy with genetic inheritance, creative ideas are the ones retained and stored by memory and passed on further through communication. Thus, the blind variation and selective retention model (BVSR). The second and third mechanisms of BVSR can be accommodated easily enough by other cognitive mechanisms—some kind of reflection, metacognition, or introspection for the second and memory for the third.

If this is correct, what positing the special mechanism may gain in explanatory power (since it posits a creative capacity for novelty generation), it may lose in parsimony and empirical tractability. Kronfeldner argues that to the extent that the creative process must be “blind”, that just means that it must proceed without complete foresight about its end result, which is just to say that it must be “spontaneous”, in her terms (recall §2.2.3). But to say that the creative process must be spontaneous (or blind), Kronfeldner says, is to make a definitional claim, not an explanatory one. It is still a further matter to explain how the mind implements such a process. Thus, she questions the explanatory value of BVSR (Kronfeldner 2010, 2011, 2018). In light of this and other criticisms, Simonton has revised his theory (2011, 2012, 2022).

5.2.2 The Default-Mode Network

Despite the foregoing criticism of BVSR, recent neuroscientific studies suggest a network of brain activity that may serve the blind variation role. Brain activity doesn’t cease when one is not focusing on a task, when one is at rest, daydreaming, and so on. Following this insight, researchers have used neuroimaging methods to identify what is now called the default mode network (DMN). The precise anatomy of this network is still a matter of investigation, but it is supposed to be less active when one is focused on an external task (say a problem in the real world or in the lab) and more active when one is not so focused (Raichle et al. 2001; Buckner and DiNicola 2019). Notice then, that while this network is not creativity-specific—it is supposed to be active during memory recall, imagining future events, daydreaming, and so on—it does seem especially well-suited for creativity, and particularly for the random idea generation hypothesized by the BVSR (Jung et al. 2013). Creativity researchers in these fields often refer to this more “free” production of ideas as “divergent thinking”, and some argue on the basis of neuroimaging studies that creative thought requires cooperation between this mode of thought as
well as that under “executive control”. As one team puts the point, “In general, we contend that
the default network influences the generation of candidate ideas, but that the control network can
constrain and direct this process to meet task-specific goals via top-down monitoring and
executive control.” (Beaty et al. 2016; see also Mayseless et al. 2015; Beaty et al. 2019;
Chrysikou 2019). Notice how well this comports with both the Geneplore and the BVSР
frameworks, perhaps identifying a way to keep some of the insights of both without commitment
to a special creativity mechanism after all.

5.2.3 Imagination
At least since Kant, theorists have identified an important link between creativity and
imagination; indeed, the two are sometimes unfortunately conflated. Construed broadly,
imagination can take various forms: sensory imagery, propositional imagination, supposition,
free association. Gaut (2003, 2009, 2010) and Stokes (2014, 2016) have both recently argued
that, although imagination and creativity are distinct, imagination is especially well-suited to
creative thought because of its characteristic flexibility. They both agree that imagination is
decoupled from action (Gaut 2003) and “non-truthbound” (Stokes 2014) in the sense that, unlike
belief, imagination is not limited by the proper function of accurately representing (some part of)
the world. This freedom or playfulness of imagination is crucial to generating new ideas, since it
allows one to safely “try out” hypotheses, conceptual combinations, strategies for solutions, and
so on, without epistemic or behavioral commitment.

A series of studies illustrates both the need for non-truthbound capacities in creative thought, as
well as the difficulty of employing them. When people – children and adults alike – are asked to
imagine and draw non-existent houses, people, or animals, they depict things that are strikingly
similar to their familiar counterparts in the real world: imagined people, for example, were
generally drawn with some version of a head, limbs, eyes, and so forth. (Karmiloff-Smith 1990;
1992: 155-61; Cacciari et. al 1997; Ward 1994, 1995). This suggests that we are highly
constrained in our creativity by the concepts we already have. Concepts of existing things are
truth-bound: your concept of an animal, for example, has the proper function of accurately
representing the range of things that are in fact animals. When you try to envision a new,
fictional kind of animal, you begin with a mental image that exemplifies your existing concept of
animal, which is why you are constrained by that concept. You then have to manipulate your initial image, varying its features in ways that abandon the aim of accuracy, using a capacity that isn’t truthbound. Generalizing this point yields the following thesis:

Cognitive manipulation thesis: Creative thought requires cognitive manipulation, which involves thinking in ways that are not bound to the truth. (Stokes 2014)

Plausibly, imagination is the mental capacity which is best suited to serve in this cognitive manipulation role. In the studies just cited, subjects must use their imagination to manipulate their existing concepts so as to form new ideas.

Recent empirical research on visual imagery seems to corroborate this claim. Various studies have identified positive correlations between creative problem solving and visual image generation, image transformation, and vividness of imagery (Finke 1990; Finke 1995; Zemore 1995; Morrison and Wallace 2001; Pérez-Fabello and Campos 2007). A more recent study highlights the importance of image transformation ability—the ability to mentally manipulate a given image—and the ability to achieve high degrees of visual creativity. Further, the results of this study suggest that although vividness negatively correlates with the practicality of images created, vividness positively correlates with novel idea generation (Palmiero et al. 2015). The novelty involved is minimal, but again it appears that imagination, here in the form of imagery, well serves the role of cognitive manipulation.

Stokes observes further that we can voluntarily control imaginative states (in contrast with other non-truthbound states, like desires and wishes). And because imagination connects in important ways with inferential systems, as well as affective systems, the thoughts it produces can often be integrated with knowledge and skills to formulate an innovative strategy or solution to a problem. Finally, this role for imagination in creativity is not exclusive to the rich creativity of artists and scientists, but indeed seems to characterize the minimally creative behavior that we all enjoy. This claim is partly motivated by the empirical research just discussed. Here, as in the more radical cases, instances of novel achievement or learning by subjects requires more than rote memorization; it requires cognitive manipulation of the information in the relevant
conceptual space (e.g. combining concepts about houses and persons). This kind of cognitive activity is best done by using the imagination.

Peter Carruthers has argued that imagination is important to creativity on evolutionary grounds (2002; 2006; see also Carruthers and Picciuto 2014). Like the above analyses, he focuses on the playfulness of imagination. Pretend play typically develops early in childhood in humans. And imagination in adults provides the right mechanisms for generating and exploring ideas (just as required by the Geneplore model). Carruthers argues that imagination evolves under adaptive advantage as a kind of practice for adult creativity—and may have been accordingly selected for, aligning with the putative creativity explosion of 40,000 years ago (Mithen 1998, 1999; Harris 2000). This, he argues, is the most parsimonious explanation of both the emergence and the ubiquity of creativity in the human species. See Gaut (2009) for a critique of Carruthers’ analysis.

5.2.4 Incubation

While we may generate ideas consciously in imagination, we may also do so during a period of unconscious incubation, when we are focused on something else. This point is illustrated by any number of famous stories, though some are probably embellished after years of retelling. Isaac Newton witnessed an apple fall from a tree (on some accounts, falling upon Newton’s head) and thereby found the insight for his laws of gravity. August Kekulé is reported to have discovered the structure of the benzene molecule while daydreaming of a serpent circling upon and seizing its own tail. Henri Poincaré alleged that, while boarding a bus, he enjoyed a needed flash of insight that led to his discovery of non-Euclidian geometry. Richard Feynman, the Nobel prize winning physicist, claimed to find inspiration while sipping soda and doodling at adult clubs. And Einstein reported: “I was sitting in a chair in the patent office at Bern when all of a sudden a thought occurred to me. ‘If a person falls freely he will not feel his own weight.’ I was startled. This simple thought made a deep impression on me. It impelled me toward a theory of gravitation” (Cited in Pais 1982). In each case, someone is suddenly struck with a flash of insight about one thing while engaged with something else entirely. The empirically-minded theorist
rejects the notion that such ideas arise *ex nihilo* or through divine possession. So how are they explained in terms of natural mental phenomena?

Arthur Koestler, partly inspired by the work of Henri Poincaré (1913), hypothesized that during creative thought processing, ideas are combined in novel ways, and this combination is performed largely *unconsciously*, by what Poincaré called the subliminal self (Koestler 1964: 164–5). For Poincaré there are only two ways we might think of the unconscious. One, we might think of the unconscious in Freudian terms, as a self capable of careful and fine discernment and, importantly, distinctions and combinations that the conscious self fails to make. Alternatively (and this is the option favored by both Poincaré and Koestler), we can think of the unconscious as a sub-personal automaton that mechanically runs through various combinations of ideas. Importantly, this unconscious process (or, if one likes, automaton) generates random conceptual associations and ideas. And these can then be further considered, examined, explored, and revised.

These behavioral methods can be combined with contemporary understanding of neural plasticity and the effects of cognitive effort and attention. Neuroscientists have long recognized that the human brain is *plastic*—stable in genetic material but constantly undergoing functional change and development in neural networking in response to external stimuli, with the work of Donald Hebb in the middle of the 20th century being one important early precedent. As Hebb put it, neural cells that “fire together, wire together”. *Cell assemblies* thus form as a result of the synchrony and proximity of the firing of individual cells. “[A]ny two cells or systems of cells that are repeatedly active at the same time will tend to become “associated”, so that activity in one facilitates activity in the other” (Hebb, 1949, p. 70). And continued attention to a problem, what some have called *cerebral effort*, causes changes in the networking of the brain’s cortex (Donald 2001: 175–8). Importantly, these changes can continue to take place, to “reverberate” even after one has removed attention from that problem. This motivates a simple (and somewhat unsurprising) hypothesis: attending to and performing cognitive tasks affects neural networking (Posner et al., 1997; Posner and Raichle, 1994; see also Karni et al., 1995), and those changes can involve strengthening of synaptic connectivity (which correlate with conceptual connections and associations). These changes, again, can occur both when one is attending to a task and after
one has diverted attention elsewhere. And, finally, the latter goes some way to explain a moment of insight after incubation (the so-called incubation effect): when one returns attention to the target problem, new or newly strengthened neural connectivity (as a result of previous cognitive effort) can give rise to a new idea. And because that neural process is not in any sense done by you, the emergence of the new idea can feel like a burst of insight (See Stokes 2007; Thagard et al. 2011; Ritter and Dijksterhuis 2014; and Heilman 2016).

There are also various recent studies on closely related topics: on mindwandering and spontaneous thought (Christoff et al. 2016; Thompson and Irving, forthcoming; Murray et al., forthcoming), on so-called “divergent thinking” (Mekern et al. 2019), and more on the neural basis of insight (Jung-Beeman et al. 2004; Bowden et al. 2005; Limb and Braun 2008; Dietrich and Kanso 2010; Kounios and Beeman 2014).

5.3 Insight

It should be intuitive that creativity often involves solving problems and doing so in interesting or surprising ways. In exceptional cases, the individual identifies a problem solution that perhaps no one (including the creator) anticipated. But there are countless examples of more mundane instances of problem solving, where the solution may be surprising (or especially interesting) to only a few individuals, perhaps even only to the problem solver. One broad, standard experimental method used by researchers thus focuses on insight in problem solving. Some problems (thankfully!) can be solved by straightforward appeal to memory, or by applying some technique or method of calculation in a mechanical way. Solving the problem may still take time and effort, but the solution will come so long as one executes the appropriate strategy or applies the relevant knowledge from memory. An insight problem, by contrast, typically requires something new on the part of the individual, and one must often “change views” of the structure of the very problem. Predictably, there are a variety of definitions or characterizations of “insight” in the literature. Here are two recent, representative examples. Bowden et al. suggest that insight occurs “when a solver breaks free of unwarranted assumptions, or forms novel, task-related connections between existing concepts or skills (Bowden et al. 2005: 322). More recently, Kounios and Beeman write, “we define insight as any sudden comprehension,
realization, or problem solution that involves a reorganization of the elements of a person’s mental representation of a stimulus, situation, or event to yield a nonobvious or nondominant interpretation” (2014: 74).

Some early studies on these phenomena centered around a familiar observation. Consider the tip-of-the-tongue phenomenon, when you know that you know some bit of information (an actor’s name or the title of a song) but, try as you may, you just can’t recall it. It often helps to give up for a moment and allow the memory to surface without effort. Researchers found that the same approach—forgetting about a problem—works well to overcome fixation on ineffective ideas so as to allow the actual solution to pop up. Smith and Blankenship primed two groups of subjects with inappropriate or misleading solutions to problems. They left one group to continue struggling with the same problem, while they distracted the second group with a distinct but cognitively demanding task. The second group thereby overcame fixation and outperformed the first group when returning attention to the original target problem (Smith and Blankenship, 1989; 1991; see also Smith et al., 1995).

In the context of creativity in particular, there is precedent, or at least overlap, in Colin Martindale’s cortical arousal theory. This theory centers around the nature of focuses of attention (Martindale 1977, 1981, 1995, 1999; Martindale and Armstrong 1974; Martindale and Hines 1975). Martindale proposes a multi-stage model of problem solving, which if the right mechanism is possessed, leads to creative thought. In the initial stages, information is gathered, various approaches are taken to the problem, and there is a high level of cortical arousal with a narrow focus of attention. As information increases and the problem remains unsolved, two kinds of responses may occur. The first kind of response is to keep attempting the same solutions to the problem such that the arousal and attention focus stay high and narrow, respectively. Alternatively, some persons experience a decrease in cortical arousal coupled with a wider range of attention focus. Information then enters what Martindale calls primary processing: a kind of subconscious cognition not under the complete control of the agent. It is this kind of processing, and the arousal mechanisms that enable it, that distinguish creative insight or achievement from non-creative ones. The first kind of response typically results in frustration and failure (fixation), while the second often results in creative insight.
There are at least two, separable components of insight thus understood. First, an insight problem requires non-mechanical or non-algorithmic solution, and this in turn requires some kind of conceptual reorganization. A hackneyed phrase may come to mind here: one has to “think outside the box”.

The second element of insight as understood here is subjective or phenomenological. An insightful problem solution is often described as occurring suddenly and with little or no apparent effort. It is an aha moment, even if less dramatic than the traditionally romanticized Eureka moment. One way researchers have tested for this subjective feature is to ask subjects to report nearness or “warmth” relative to solving a problem. They find that for insight problems, by contrast to non-insight problems, subjects report that as they near solution they experience abrupt changes in the sense of warmth for solving the problem (Metcalfe and Wiebe 1987; see also Dominowski 1995, Laukkonen and Tangen 2018). More recently, researchers have begun to employ neuroimaging techniques to study insight and insightful problem solving (Luo and Niki 2003; Mai et al. 2004).

First, researchers have developed methods for using subjective report, where subjects rate whether they felt that they used insight in solving a designated problem (Bowden et al. 2005). And second, and coupled with those report methods, researchers have developed simple problems that can be solved with insight. One such example is the “Compound remote associates problem” (CRA). Here is an example of a CRA problem:

Each of the three words in (a) and (b) below can form a compound word or two-word phrase with the solution word. The solution word can come before or after any of the problem words.

a. french, car, shoe
b. boot, summer, ground

(From Bowden et al. 2005)

Because of their simplicity, these problems can be solved unambiguously and quickly, and with this speed comes better potential for neuroimaging study. In instances where subjects report insight solutions to these kinds of problems, “EEG shows a burst of high-frequency (gamma-band) EEG activity over the right temporal lobe, and fMRI shows a corresponding change in blood flow in the medial aspect of the right anterior superior temporal gyrus (Jung-Beeman et al. 2004)” (Kounios and Beeman 2014: 78). The question for neuroscientists is whether this convergence of evidence is sufficient to establish neural correlates of insight.

5.4 Evaluation

A moment of “insight” can be misleading, as what initially strikes you as a promising idea may ultimately turn out to be a dead end. You may have countless ideas in the course of undertaking a complex creative project, while only a few of them will make the final cut. A crucial part of your creative work therefore consists in evaluating your ideas. For any idea that occurs to you, you might have to ask: Will this work? Is it new? How does it fit in with other parts of your project? Do you have the resources and abilities to bring it to fruition? Is it worth the time and effort?

Much of the research on this phase of the creative process is concerned to identify and categorize the range of factors that people take into consider as they evaluate their ideas (Blair and Mumford 2007; Dailey & Mumford, 2006). Unsurprisingly, those factors vary from one domain to another. New culinary dishes are judged by factors like aroma, taste, texture, color, presentation (Horng and Lin 2009), whereas improved musical performances are judged according to their complexity, originality, and technical virtuosity (Eisenberg and Thompson 2003), and so on. Your understanding of the relevant factors is part of your internalized model of the domain (Bink & Marsh, 2000; Csikszentmihalyi & Sawyer, 1995). And since you acquired

---------------------------

1 Solutions: a. horn b. camp.
and refined that model through years of preparation, your capacity for evaluation is largely a consequence of your efforts from that initial stage.

Somewhat more surprisingly, there is some evidence that people who are good at evaluating ideas are also good at generating them (Runco 2003; Runco and Dow 2004; Runco and Chand 1994; Runco & Vega, 1990; Runco, 1991).

Other studies support what Sawyer calls (2012: 131) calls the productivity theory, which says that the best way to get good ideas is to have lots of ideas and just throw away the bad ones. In historiometric studies, Simonton found that creators who yielded the greatest number of works over their lifetimes were mostly likely to produce works that were significant and stood the test of time. Even more striking, he discovered that, from year to year, the periods when creators were most productive were also the ones in which they were most likely to do exceptional work (Simonton 1988a, 1988b). Linus Pauling, who won the Nobel Prize in Chemistry in 1954 as well as the Nobel Peace Prize in 1962, summed up the productivity theory in a famous remark: “If you want to have good ideas you must have many ideas. Most of them will be wrong, and what you have to learn is which ones to throw away” (quoted by Crick 1995).

5.5 Externalization

The final operation of the creative process – externalizing ideas – may involve any number of disparate activities, which Keith Sawyer sums up as follows:

Creativity research has tended to focus on the early stages of the eight-stage creative process— particularly on the idea-generating stage. But a lot has to happen to make any idea a reality. Successful creators are skilled at executing their ideas, predicting how others might react to them and being prepared to respond, identifying the necessary resources to make them successful, forming plans for implementing the ideas, and improvising to adjust their plans as new information arrives. These activities are important in all creativity, but are likely to be even more important in practical domains
such as technological invention and entrepreneurship (Mumford, 2003; Policastro & Gardner, 1999). (Sawyer 2012: 133-4)

It may be tempting to assume that the real creative work is finished once a new idea emerges in the moment of insight, and that externalization is just the uncreative, mechanical chore of making the idea public. But a closer look at the phenomenon reveals that externalization is often integral to creativity itself.

Vera John-Steiner (1985) interviewed, and examined the notebooks of, over 70 exceptional creators (ranging from author Anaïs Nin to composer Aaron Copland), and consulted the notebook of another 50 eminent historical creators such as Leo Tolstoy and Marie Curie. A recurring theme throughout was that at the beginning of each creative endeavor and continually throughout its development, creators manipulate and build upon their impressions, inklings, and tentative hunches using sketches, outlines, and other external representations.

Perkins (1981) corroborated this finding by analyzing the 61 sketches Picasso made en route to painting his famous work, Guernica, as well as Beethoven’s musical drafts and Darwin’s notebooks. In each case, the artist progressed by engaging with external representations.

Other studies found that people discovered and solved more problem when they used sketches during a task (Verstijnen 1997), and that people come up with better ideas for improving inventions when they work with visual diagrams (Meyer 1989).

One reason externalization is so vital to substantial creative work is because of our limited capacity to consciously hold and manipulate information in our minds. It helps to offload ideas and store them in the form of physical symbols and expressions in order to free up space for the mind to examine those ideas at arm’s length while entertaining new ones. Thus research shows that internal strategies like mental visualization can help with relatively simple tasks, but for more complex projects externalization is key (Finke et al. 1992: 60).
5.6 Worries and future directions

We close our survey of the cognitive science of creativity with a brief discussion of some general worries about current work, and some prescriptions for future research.

Some have worried about the validity of the psychometric measures employed in neuroimaging studies. One such concern regards the confidence that we should have that the tests employed are really tracking creative behavior. This is of course a general problem, partly symptomatic of the challenges that come with defining creativity (like other phenomena) and with the special challenges that attach to features such as insight and incubation. But there are particular challenges that come with using neuroimaging technologies such as fMRI scanning to attempt to study naturally occurring phenomena. Use of this technology is almost invariably ecologically invalid—one cannot run an fMRI in the artist’s studio. And because of the cost and sensitivity of these imaging systems, the correlative behavioral tests are often significantly abbreviated. This may impose constraints on space for occurrence of the target phenomena—novel thinking and insight—during the imaging session. As one researcher worries,

Too often single tests are used—or even single items! This is contrary of psychometric theory in general (where longer tests allow errors to cancel themselves out and are thus more reliable) and true of the research on creativity assessment in particular (where differences among items and even tests are common (Richards, 1976; Runco et al. 2016). Results from any one test will not generalize to other tests. Results from a single item of course have even less generalizability (Runco 2017: 310; see also Abraham 2013).

Another empirical researcher criticizes what he sees as “the wild goose chase” in the neuroscience of creativity. Arne Dietrich (2019) recapitulates the above worries about validity of psychometric measures and their abbreviated and piecemeal application. He further worries about the now dominant emphasis on divergent thinking, and the default mode network (as well as the now mostly abandoned emphasis on notions such as madness, the right brain, and REM sleep). Dietrich’s concern in each case is that the research emphasis is unhelpfully myopic, and
that while the imaging methods are sound and state of the art, the characterization of creativity is not. He decries the temptation to identify what may be a feature of creativity with the whole of the phenomenon. Divergent thinking, he suggests, is likely a cluster of various mental phenomena rather than a singular one, and “there is no effort underway to dissect divergent thinking and link it to the kinds of cognitive processes we use to operationalize all other psychological phenomena, such as working memory, cognitive control, semantic memory, perceptual processes, or executive attention” (2019: 37). Notice, then, that the “wild goose” for Dietrich is to hastily conclude and then center studies around a singular, special creativity mechanism.

Dietrich also offers various prescriptions for remedy. To combat myopia, he suggests (as some have in other disciplines, e.g. Boden 2004) a plurality of types of creativity (and/or features of creativity). He cautions, “Since different types of creativity contain opposing brain mechanisms — focused versus defocused attention, for instance — any all-encompassing claim about creativity in the brain will almost certainly qualify as phrenology” (2019: 39). He pairs this with a prescription for a more interdisciplinary approach to the topic. Others in the field have made the same prescription, advocating a “systems” approach sensitive both to the multi-faceted nature of creativity and the value of theorizing at multiple levels of explanation (Hennessy and Amabile 2010).

These directives for future research seem hard to resist. At the very least, it would seem advantageous to ensure that the full range of empirical method across the behavioral and brain sciences is communicated across the relevant sub-disciplines. This would ideally lead to better collaboration amongst such researchers. What’s interesting is that a cousin to this prescription is not well heeded by the same researchers advancing it here. However little crossover there is between, say, behavioral psychologists and neuroscientists in studies of creativity, there is comparatively even less crossover (almost none) between the psychological sciences and computational approaches to creativity. The next section thus begins by highlighting this “gap”, and identifying some of the potentially fruitful areas for interdisciplinary work on that front. It then continues with a discussion, generally, of research on creativity in the fields of computing science, artificial intelligence, and robotics.
6 Creativity and artificial intelligence

Just as we find in psychology and neuroscience, there is a rich research literature on creativity in artificial intelligence and computer science, with devoted journals, special issues, and conferences (The Journal of Artificial Creativity, The Journal of Creative Music Systems, Digital Creativity, Minds and Machines-Special Issue on Computational Creativity (2010, Vol 20:4), The International Conference on Computational Creativity). The question we focus on here is whether a computer could be creative. As background, it is worth considering how theorists approached the analogous question as to whether a computer could think.

Although theorists of various kinds have asked whether machines can think since at least the early modern period, the most important conceptual innovations on the topic came from Alan Turing, centering around his 1950 paper ‘Computing machinery and intelligence’. Here Turing provided a number of groundbreaking insights. Perhaps most familiar is Turing’s “imitation game”, now commonly known as “the Turing Test”. In brief, the test involved an unknowing interrogator who could ask an open-ended series of questions of both a human and a computer. If the interrogator could not distinguish computer from human, Turing postulated that this would suffice to illustrate genuine intelligence. There is no shortage of controversy regarding the aptness of the test for intelligence, and arguably no computer has yet passed it. (For more thorough discussion of Turing and the Turing test see Hodges 2013, Del Mol 2018, Oppy and Dowe 2021).

Successful performance in Turing’s game would require remarkable behavioral flexibility. And it is highly operational: specify a threshold for imitation, and then simply allow the interrogator to ask questions, then assess performance. If the behavior is sufficiently flexible to fool the interrogator, Turing claimed, the behavior was intelligent and, therefore, the computer intelligent.

With this background in mind, what are some of the cases in AI research lauded as success cases, and how do they align with some of Turing’s criteria?
Many of the familiar success cases are highly specialized. *Deep Blue* defeated chess master Garry Kasparov (Kasparov 2017); some language processing systems managed to navigate social contexts such as ordering from a menu at a restaurant (Schank and Abelson 1977); *AlphaGo* more recently defeated the world champion Go player. This specialization is both a virtue and a limitation. On the one hand, achievement in such a specialized domain implies an exceptional amount of detailed memory and skill. On the other hand, this knowledge and skill does not generalize. Neither *Deep Blue* nor *AlphaGo* could successfully order from a menu, along with countless other basic human tasks. Put in terms of Turing’s imitation game, these systems would fail miserably to fool a human, or even remotely imitate one (except for their performance in a very narrow domain). What about systems such as IBM’s *Watson*, which famously won (against humans) on the television game show *Jeopardy!* This performance is more general, since topics on the show vary widely, and seemed to require both language comprehension and some minimal reasoning skills (see Bringsjord and Govindarajulu 2020 for extended discussion). Even so, Watson’s capabilities are still quite limited: it cannot make fluid conversation “in real time” and is largely insensitive to temporal and other factors that come with context.

Suppose, though, that we interpret these cases charitably. There are many, many more examples of computational systems that display sophisticated behavior, from the highly specialized to the more general. To be clear, these are remarkable achievements, and more are on the way. Such achievement involves substantial complexity (and, it appears in some cases, significant flexibility). Turing highlighted this and accordingly focused on the computability of complex human behaviors. But this also underscores a distinction, often invoked by critics of artificial intelligence research. There is a difference between a computer’s displaying or merely imitating an intelligent behavior, and a computer’s instantiating intelligence through such behavior. And the critic will say, even if a computer behaves *as if* it is intelligent, this is just modeling or simulating intelligence. The greater ambition, though, is “genuine artificial intelligence”, a system that *actually thinks*. John Searle refers to this as the distinction between “weak AI” and “strong AI”, respectively.

**Weak AI**: Could a computer *behave as if* it thinks?

**Strong AI**: Could a computer genuinely think?
The general worry here is that however sophisticated a system’s behavior may appear “from the outside”, for all we know it may just be a “hollow shell” (Haugeland 1981; Clark 2001). The worry has then been fleshed out in various ways by specifying what is missing from the shell, as it were. Here are three standard such candidates. And, again, in each case however sophisticated the computer’s behavior may appear it still may be lacking in any or all of the following. First, the computer may lack consciousness. Second, the computer may lack any understanding of the symbols over which it computes (Searle 1980). Finally, the computer may operate without caring about its own behavior or, as John Haugeland colorfully puts it, without “giving a damn”. In each case, any kind of response from the ambitious AI researcher encounters the substantial challenges that come with theorizing mental phenomena such as consciousness, understanding, linguistic competence, and emotion. (Turing 1950, for instance, recognized but largely eschewed these kinds of topics).

It’s one thing to ask whether computers could think, and another to ask whether they could be creative. And just as the prospect of artificial intelligence or thinking divides into two questions - of weak AI and strong AI - we may distinguish two analogous questions about artificial creativity, which we’ll refer to as the questions of “weak AC” and “strong AC”, respectively. To begin with the former:

**Weak AC**: Could a computer behave as if it’s creative?

Something behaves as if it’s creative if it produces things which are psychologically new (new to that thing) and valuable. Arguably, a number of computers have already done that.

In the 1970s, Harold Cohen began using computational technologies to produce new drawings and paintings. The work of his computer painter, Aaron, has exhibited at galleries such as the Tate and the Victoria and Albert Museum in London. David Cope’s ‘EMI’ (Experiments in Musical Intelligence) has composed musical works in the style of various known composers and styles, even a full-length opera. Some of these works have been recorded and produced by bona fide record labels. Just search “Emily Howell” on Spotify or Apple Music and give it a listen.
Simon Colton’s *The Painting Fool* is an ongoing project, involving a software that abstracts phrases, images, and other items from newspaper articles and creates collage-style pieces. It has also produced portraits, based on images of film characters, of the same individual in different emotional states (see here: http://www.thepaintingfool.com/galleries/amelies_progress/index.html; see Colton 2012 for theoretical discussion). In all of these cases, the relevant outputs of the computer program are new relative to its past productions – so they are psychologically (or behaviorally) novel, which again is all the novelty that creativity requires. And although historical novelty isn’t required for creativity, it’s worth noting that these products appear to be to be new in all of history as well.

What about value? As noted above in §1.1, some theorists reject the value condition, but even if value *is* required for creativity, that too is a condition these computer artworks seem to meet. Assessments of value can be controversial, but that is no less true for the outputs of human creativity. The fact that these works are critically acclaimed, showcased in prestigious galleries, and commissioned by selective record labels testifies to their artistic merit, and viewers find them pleasing, interesting, and appealing, even before being apprised of their unusual origin. So it is reasonable to conclude computer programs like the ones just described exhibit at least weak AC insofar as they produce works of valuable novelty, and one could cite many more examples in the same vein.

Some theorists have noted that, whether or not the original Turing test is a good test for intelligence or thinking, we might adopt an analogous test for creativity: If a computer can fool human observers into thinking that it is a human creator, then it is in fact creative (Colton and Pease 2011; see also Chen 2020 for useful discussion of artificial creativity, including many additional examples of particular cases, and so-called Dartmouth-based Turing tests). If we employ this test, we might find ourselves with an unexpected conclusion: computers can be creative; in fact, some of them already are. But one might reasonably worry that the test is inadequate and the conclusion is too quick (Berrar & Schuster 2013; Bringsjord et al. 2001). From the fact that a computer operates *as if* it’s creative, one might argue, it doesn’t follow that it really is. Which brings us to our next question:
**Strong AC:** Could a computer genuinely be creative?

This obviously returns us to the question of what conditions something must meet in order to count as being genuinely creative. And here we need go beyond the outwardly observable product-features of novelty and value to consider the underlying processes of genuine creativity. As we saw in §1.2, theorists have variously proposed that in order for a process to count as creative, it must be *surprising, original, spontaneous,* and/or *agential.* There is no consensus to appeal to here, but *if* any one of these conditions is indeed required for genuine creativity, then a computer could be genuinely creative only to the extent that it executes processes which satisfy that condition.

The classic statement of skepticism regarding the possibility of computer creativity is due to Lady Ada Lovelace who had this to say while remarking on “the Analytical Engine” designed by her friend Charles Babbage:

> It is desirable to guard against the possibility of exaggerated ideas that might arise as to the powers of the Analytical Engine. The Analytical Engine has no pretensions whatever to *originate* anything. (Lovelace 1842, italics added)

Though Lovelace does not frame her comments in terms of “creativity” as such, she explicitly denied that a computer could satisfy at least one condition that is plausibly required for creativity, namely *originality.* A computer cannot be the originator, the author, or the creator of anything new, she contends; it can only do what it is programmed to do. We cannot get anything out of a computer that has not already been programmed into it. Further, Lovelace may also be interpreted as expressing or implying doubt about whether a computer could satisfy the three other proposed requirements for genuine creativity. Insofar as a computer’s outputs cannot be original, one might also suspect that they cannot be *surprising.* The image of a machine strictly following rules invokes precisely the kind of mechanical procedure that is the antithesis of *spontaneity.* And it may seem that such a machine could not be a genuine *agent* either. The problem isn’t just that a computer can’t produce anything original; it’s that it deserves no credit for whatever it does produce. Any praise or blame for the outputs of a computer rightly go to the
engineers and programmers who made the machine, not to the machine itself. While these points may be intuitive, at least some of them are being challenged by modern technologies, which have come a long way since Babbage’s invention.

Consider AlphaGo again. This is a “deep learning” system, which involves two neural networks: a Policy network and a Value network. Very briefly: The system is trained using a vast number of legitimate moves made in actual games of Go played by professional human players (28.4 million moves from 160,000 games, to be precise; see Silver et al. 2016 and Halina 2021). The network is further trained, again using learning algorithms, by playing many games (some 100 million) against previous versions of itself (in the sense of a differently weighted neural network). The weights of nodes in the network are then adjusted by a learning algorithm that favors moves made in winning games. The value network is trained over a subset of these many games, with node weighting adjustments resulting in reliable probability assignments to moves vis-à-vis their potential to contribute to a win. Finally, the system employs a Monte Carlo search tree (MCT). Generally, this kind of algorithm is designed to simulate a decision process to optimize success given chosen parameters. In this case, the search algorithm selects a given path of moves, then adds some valid moves to this path, and then if this process does not terminate (end in win/loss), the system performs a “rollout”. A rollout essentially plays the game out for both players (using samples of possible moves) to its conclusion. The information that results from the MCT and processing by the value network are then fed back (back propagated) into the system. This entire process (once the system is trained) is rapid and determines how AlphaGo “decides” to move in any given game.

Here are some things to note. AlphaGo’s style of play is surprising. As commentators have noted, it is starkly unconventional relative to standards of human play (Halina cites Baker and Hui 2017). Indeed, Lee Sodol, the world champion Go player defeated by AlphaGo in 2016, remarked that AlphaGo’s play revealed that much of human play is, contrary to prior common opinion, not creative after all – intimating that at least some of the play of AlphaGo is. Note further that this system is flexible. While there are learning algorithms and rules that adjust network weights, the system is not mechanical or predictable in the same fashion as earlier, classical systems (including Deep Blue, for example). In a recent paper, Marta Halina has made
this argument (Halina 2021). She explicitly invokes Boden’s characterization, which requires novelty, value, and surprise of creativity. Again, the novelty and value should be plausibly attributed in this case. Regarding surprise, Halina suggests that it is AlphaGo’s employment of MCT that enables a kind of “insight”, flexibility, and unpredictable results. She writes, “It is the exploration parameter that allows AlphaGo to go beyond its training, encouraging it to simulate moves outside of those recommended by the policy network. As the search tree is constructed, the system starts choosing moves with the highest “action value” to simulate, where the action value indicates how good a move is based on the outcome of rollouts and value-network evaluations” (Halina 2021: 324). Halina grants that given its domain-specificity, as we have already noted, this system’s particular abilities do not generalize in a way that may be required to properly attribute genuine intelligence. But she suggests that the complex use of the MCT search may amount to “mental scenario building” or, we might say, a kind of imagination. And insofar as this search algorithm technology can be applied to other systems in other domains, and imagination is a general component of intelligence, perhaps here lies space for generalizability. AlphaGo also affords at least some reply to the traditional Lovelace worry. “Artificial systems do not act only according to preprogrammed rules hand-coded by engineers. Moreover, current deep-learning methods are capable of producing systems that are superhuman in their abilities to discover novel and valuable solutions to problems within specific domains” (Halina 2021: 327). If this is right, then AlphaGo exhibits originality. Finally, the flexibility with which this system operates may also satisfy Kronfeldner’s spontaneity requirement.

Some of these same features are found in a related approach in AI, namely research in evolutionary robotics. These systems also involve various forms of machine learning but in this case the learning is distributed, as it were, across a population of individuals rather than one individual. This approach can be understood, albeit imperfectly, as analogous to natural evolution. One begins, typically in computer simulation, with a population of agents. These agents are typically identified with individual neural networks, the connections and weightings of which are random to start. Relative to some task—for instance, avoiding obstacles, collecting objects, performing photo or phonotaxis—a genetic algorithm assigns a fitness value to each individual agent after a certain period of time or number of trials. Fitter agents are typically favored and used to generate the next population of agents. Also included in this generation are
random mutation and genetic crossover (digital breeding!). Although it can take hundreds of generations, this is a discovery approach to engineering or constructing a system that successfully performs a task; it is “gradient descent learning” (Clark 1997). In this bottom-up approach, no single individual, nor even an entire population, are in any strict sense programmed. Rather, successful agents have “learned” as a result of generations of randomness, crossover, and small fitness improvements (and lots and lots of failures). Early success cases evolved robots that can follow trails (Koza 1992), locomote in insect-fashion (Beer and Gallagher 1992), guide themselves visually (Cliff, Husbands, and Harvey 1993), and collect garbage (Nolfi and Floreana 2000). See Bird and Stokes (2006, 2007) and Stokes and Bird (2008) for analysis and study of creativity in the context of evolutionary robotics.

These systems most certainly produce novelty. Later, fit individuals achieve novelty at their aimed task relative to whole generations and populations of previous agents. And this novelty is often surprising to the engineers and programmers that build them, indeed sometimes even unpredictably independent of any relevant task for individuals in the population. There are many examples in the literature. Indeed Lehman et al. 2020 catalog a large range of cases where digital evolution surprises its creators, categorizing them in four representative groups: “mis-specified fitness functions”, “unintended debugging”, “exceeded experimenter expectations”, and “convergence with biology”. Here is one now relatively famous example of the first type of case. In early research in artificial life (A-Life), Karl Sims (1994) designed virtual creatures that were supposed to learn to walk (as well as swim and jump) in a simulated environment. The fitness function assessed individual agents on their average ground velocity across 10 seconds. Some of the fittest individuals to evolve were surprising: they grew tall and rigid and when they would fall over they would achieve high ground velocity, thus maximizing fitness given the (mis)specified parameters in unpredicted ways.

This is but one example of how systems like these can evolve is unpredictable or surprising ways. This unpredictability has occurred not just in simulated robotics, but in embodied robotics as well. In using a genetic algorithm to attempt to evolve oscillating sensors, researchers unintentionally evolved a radio antenna (Bird and Layzell 2002). This unexpected result arose
from a combination of the particular algorithm used (which was intended) and various physical features of the space such as proximity to a PC monitor (which the researchers had presumably deemed irrelevant but which the evolved system, in a sense, did not). And one might be further inclined to describe some of these achievements as creative (and not just in the trivial sense that they are original instances of robotic success), since they also produce value, at least insofar as they are useful at performing a task, whether it is locomoting or locating a source of light or sensing radio waves.

Some theorists in this domain might argue that these systems achieve spontaneity as well. Given the substantial inclusion of randomness in the system’s development—both at the outset when the individual’s neural networks are randomized and more importantly with random mutation across populations—it is intuitive to describe the system’s as not following a mechanical procedure. Indeed, the way in which systems exploit fitness functions and data patterns further underscores this point. (Again, see the rich catalog of cases offered by Lehman et al. 2020).

On the face of it, then, recent technologies in AI, evolutionary robotics, and artificial life, seem to fulfill many of the conditions proposed for genuine creativity. These systems produce things that are novel and valuable, and do so through computational processes that are plausibly surprising, original, and spontaneous. The one requirement we have yet to address, however, is agency. Recall the suggestion, implicit in Lovelace’s remarks, that whatever a computer produces is to the credit of the programmer, not the computer. Notice that as sophisticated as current technologies in artificial creativity may be, presumably they are still not subject to praise or blame for what they do. If any beings are responsible for the work of these programs, it still seems to be the programmers and engineers who make them, not the programs themselves. The programs themselves do not seem to “give a damn”. So, if the creative process requires agency, arguably we have not yet created, programmed, or evolved a computational system that is really creative, however much they might appear to be. In the pursuit of strong AC, agency might be the final frontier (Stokes and Paul 2021).
7 Conclusion

It should be clear from the above discussions that there are rich and lively research programs, across a range of scientific disciplines, studying human creativity. These approaches substantiate the view that, contrary to the romantic tradition, creativity can be explained. Psychological functions and neural correlates have been identified, and remarkable advances are being made with computational and robotics technologies. What may be less clear is that, despite these advances, the distinct research programs in question are largely disjoint or siloed.

In a recent paper, Geraint Wiggins and Joydeep Bhattacharya (2014) highlight this “gap” between scientific studies of creativity. Their particular emphasis is on the gaps between research in neuroscience and research in computer science, and they advocate a bridge in the form of a neurocomputational approach. This kind of bridging may be called for even beyond what these authors prescribe, since there are gaps not just between these disciplines, but also between these and behavioral psychology, AI and A-Life research, and philosophical analysis. Creativity is a deeply complex and deeply important phenomenon. Fully understanding it will require us to integrate a variety of theoretical perspectives, and, as this survey reveals, philosophy has a vital role to play in that endeavor.
Bibliography


https://doi.org/10.1016/j.cobeha.2018.08.013.


De Quincey, Thomas. 1827. “On Murder Considered as One of the Fine Arts.” *Blackwood’s Magazine*, 1827.

64


Nolfi, Stefano, and Dario Floreano. 2000. *Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines*. MIT Press.


https://doi.org/10.1207/s15326985ep3902_1.


