

Human creativity, evolutionary algorithms, and predictive representations: The mechanics of thought trials

Arne Dietrich & Hilde Haider

Psychonomic Bulletin & Review

ISSN 1069-9384

Volume 22

Number 4

Psychon Bull Rev (2015) 22:897-915

DOI 10.3758/s13423-014-0743-x

Psychonomic Bulletin & Review

VOLUME 22, NUMBER 4 ■ AUGUST 2015

PB&R

EDITOR

Gregory Hickok, *University of California, Irvine*

ASSOCIATE EDITORS

Jessica Cantlon, *University of Rochester*

Greg de Zubicaray, *University of Queensland*

Stephen D. Goldinger, *Arizona State University*

Antonia Hamilton, *University College London*

Andrew Heathcote, *University of Newcastle*

Marc Howard, *Boston University*

Liina Pykkänen, *New York University*

John Serences, *University of California, San Diego*

Sarah Shomstein, *George Washington University*

Mark Steyvers, *University of California, Irvine*

A PSYCHONOMIC SOCIETY PUBLICATION

www.psychonomic.org

ISSN 1069-9384

 Springer



 Springer

Your article is protected by copyright and all rights are held exclusively by Psychonomic Society, Inc.. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Human creativity, evolutionary algorithms, and predictive representations: The mechanics of thought trials

Arne Dietrich · Hilde Haider

Published online: 11 October 2014
© Psychonomic Society, Inc. 2014

Abstract Creative thinking is arguably the pinnacle of cerebral functionality. Like no other mental faculty, it has been omnipotent in transforming human civilizations. Probing the neural basis of this most extraordinary capacity, however, has been doggedly frustrated. Despite a flurry of activity in cognitive neuroscience, recent reviews have shown that there is no coherent picture emerging from the neuroimaging work. Based on this, we take a different route and apply two well established paradigms to the problem. First is the evolutionary framework that, despite being part and parcel of creativity research, has no informed experimental work in cognitive neuroscience. Second is the emerging prediction framework that recognizes predictive representations as an integrating principle of all cognition. We show here how the prediction imperative revealingly synthesizes a host of new insights into the way brains process variation-selection thought trials and present a new neural mechanism for the partial sightedness in human creativity. Our ability to run offline simulations of expected future environments and action outcomes can account for some of the characteristic properties of cultural evolutionary algorithms running in brains, such as degrees of sightedness, the formation of scaffolds to jump over unviable intermediate forms, or how fitness criteria are set for a selection process that is necessarily hypothetical. Prospective processing in the brain also sheds light on how human creating and designing – as opposed to biological creativity – can be accompanied by intentions and foresight. This paper raises questions about the nature

of creative thought that, as far as we know, have never been asked before.

Keywords Aha effect · Consciousness · Creativity · Emulation · Evolutionary algorithms · Fitness function · Heuristics · Idea · Insight · Prediction · Selection · Simulation

Creativity has a dubious distinction in the psychological sciences. It is hard to think of a mental phenomenon so central to the human condition for which we have so little understanding as to how the brain does it. One might be tempted to argue that this position is held by the still bigger problem of consciousness, but at least there we have some proposals – global working space (Baars, 1988), social brain hypothesis (Humphrey, 2002), competing neuronal coalitions (Crick & Koch, 2003), higher-order thought (Rosenthal, 2005), etc. – that provide solid theoretical grounding. By contrast, all theories and approaches guiding empirical research on creativity – divergent thinking, defocused attention, right brains, low arousal, prefrontal activation, alpha enhancement, etc. – have not been supported by the neuroimaging evidence. Recent reviews that take all data into account have shown that for each of these proposals there is more evidence against it than supports it (Arden, Chavez, Grazioplene, & Jung, 2010; Dietrich & Kanso, 2010; Sawyer, 2011).

A number of reasons have been suggested for this state of affairs. One is surely the difficulty of formulating an operational definition. The commonly mooted definition of creativity – something useful, novel, and surprising – is too vague to be useful for either laboratory work or psychometric assessment. This elusiveness of the construct of creativity leads to subsequent complications. It is generally acknowledged that testing instruments for creativity have little real-world validity (e.g., Abraham, 2013; Arden et al., 2010; Dietrich, 2007; Dietrich & Kanso, 2010; Ward, Smith, & Finke, 1999). In

A. Dietrich (✉)
Department of Psychology, American University of Beirut,
Beirut 1107-2020, Lebanon
e-mail: ad12@aub.edu.lb

H. Haider
Department of Psychology, University of Cologne, Cologne,
Germany

addition, all commonly used creativity tests are understood to tap into a multitude of cognitive functions, such as working memory, executive attention, or perceptual processes (Dietrich, 2007; Ward et al., 1999). Attempts to tighten the definition and associate creativity with more specific mental processes – divergent thinking, defocused attention, remote associations, for instance – have not yielded valid divisions of what is and what is not associated with creativity. This is because they either ended up becoming false categories, for the simple reason that their exact opposites also precipitate creative products (Dietrich, 2007), or they are still compound constructs and too broad, consisting of many separate mental processes (Ward et al., 1999). This, in turn, renders neuroimaging technology of little use, at least if the goal is to isolate brain processes or regions that are selective for the creative component of problem solving (Arden et al., 2010; Dietrich & Kanso, 2010; Sawyer, 2011). A more targeted critique of the methodological problems facing the neuroscientific study of creativity can be found elsewhere (e.g., Abraham, 2013; Dietrich, 2007).

Despite this critique, research using the same general paradigm continues nearly unchanged. There have been well over a dozen studies since the last review papers (e.g., Aziz-Zadeh & Liew, 2013; Gansler, Moore, Susmaras, & Jerram, 2011; Green, Kraemer, Fugelsang, Gray, & Dunbar, 2012; Jauk, Benedek, & Neubauer, 2012; Shen, Liu, & Chen, 2010; Takeuchi, Taki, Hashizume, Sassa, Nagase, Nouchi, & Kawashima, 2011, among others) that use neuroimaging technology in conjunction with creativity tests, despite the lack of a solid theoretical grounding for them in terms of the individual cognitive processes involved or the demonstration of ecological validity. As a result, we think that they have not changed the negative conclusions of the earlier reviews or have contributed to a better understanding of the neural mechanisms of creativity. We have argued (Dietrich, 2007, Dietrich & Kanso, 2010) that until we develop a solid understanding of the cognitive processes involved in creativity – as opposed to non-creative thinking – we consider the current paradigm as failed. We wish to outline in this paper the beginning of a different path that attempts to develop exactly these kinds of needed cognitive and computational processes.

For this, we want to bring to the fore another possible reason for the failure to reverse-engineer the brain mechanisms underlying creative thinking. This reason is the reluctance to come to terms with a reductive research program, especially one in which proximate (neuroscientific) explanations are informed by ultimate (evolutionary) mechanisms. Although unexpected in a scientific arena, creativity, like consciousness, is so close to home, deemed so unique to us humans, that it is laden with and protected by many deep-rooted intuitions that strongly interfere with understanding its neural roots (Dennett, 2004). Of course, few scientists openly argue for mystical muses or light-bulb explanations, but there

seems nevertheless to be a strong instinctive tug against the idea that creativity can be formalized to any significant extent using an evolutionary approach (see, for instance, Sternberg, 1999).

Evolutionary approaches have a long history in the field of creativity (e.g., Campbell, 1960; Popper, 1972; Simonton, 1999) and so has the opposition to them (see commentaries to Simonton's 1999 target article and Dasgupta, 2004; Perkins, 1994; Russ, 1999; Sternberg, 1998, 1999). Human creativity is part of a larger debate in the social sciences that discusses how far Darwinism extends upwards into culture and, by extension, human creative thought (see Mesoudi, Whiten, & Laland, 2006 and its attending commentaries as well as Ingold, 2000, 2007). This debate, however, seems to cloud the overall consensus, held by virtually all parties, that the basic grammar and logic of evolutionary thinking applies to human creativity (Kronfeldner, 2010). Perhaps the mere existence of the debate has created a perception of whether an evolutionary approach is useful at all in the search for the brain mechanisms of creativity.

Oddly, this point is underappreciated, so it is worth stressing. All sides concur with the basic notion that we generate and evaluate ideas on a trial and error basis, in a manner similar to a variation-selection process (Kronfeldner, 2010), yet the two-step evolutionary rationale has been nearly universally ignored in setting up empirical protocols. In consequence, we submit that yet another reason for the lack of progress is that neuroscience studies have failed to ground their investigations in the one mechanism that brains most likely use to engage in creative thinking.

The first part of the paper flows from this. It brings into full view the fact that, underneath the wedge issues, there exists a common denominator on vital points, namely that human creativity is a variational system that involves the partial coupling of variation to selection. Of course, this is not uncharted territory, as it has been argued many times before (e.g., Boyd & Richerson, 1985; Richerson & Boyd, 2005; Eysenck, 1993), but by emphasizing the common denominator in this debate, it becomes self-evident that we can no longer foreclose this most obvious of paradigms, albeit only in its basic form, in our search for sound neurocognitive explanations of creative thinking.

The point of the present paper is not to establish this common ground but to propose, based on it, a neural mechanism of how brains can couple variation to selection and hence generate ideational combinations with some sightedness. So, to pursue further the question of how brains produce creative ideas, we examine, in the second part, the newly emerged prediction framework. The primacy of prediction in organizing neural computation has explanatory power with respect to creativity that has not been realized. This also requires that we distinguish between two processes – emulation and simulation – which, however, we will use interchangeably until we

introduce that distinction. In the third and final part of this paper, we bring this explanatory power to bear on creative thinking by applying the prediction framework to the evolutionary paradigm. In particular, we develop in some detail the idea that internal representations of the emulated future provide the process that enables brains to partially couple variation to selection and, in so doing, provide foresight and purpose to human creativity.

To clarify our theoretical proposal with respect to Marr's three levels of analysis (Marr, 1982), the mechanistic explanation we suggest here belongs to the second, so-called algorithmic level. According to Marr, the first level of analysis is the computational level, which provides a generic characterization of the problem. The second level is the algorithmic level, which focuses on the system's processes and how the problem is solved. It is at this level our framework operates. The third level is the implementational level, which, according to the model, involves the physical substrate. We do not engage this level, as our framework does not specify hardware components of creativity, either in terms of brain areas or neurotransmitter systems.

The evolutionary framework

Culture involves inheritance. We pass on cultural units – however defined – which leave behind a trail of gradual and cumulative changes (Richerson & Boyd, 2005; Mesoudi et al., 2006). The sticking point has been whether there exists an important analogy, or disanalogy as the case may be, between cultural and biological evolution. This debate over cultural Darwinism is largely obstructive for the present context because it does not bear on pressing into service an evolutionary framework in the search for the neural mechanisms of creativity. The focus of this section is to isolate key features of cultural evolution to show that extensive agreement exists on critical points. For this common denominator to find its way to the neuroscience laboratory, we also draw attention to several important parameters that seem to, on the face of it, distinguish cultural evolutionary algorithms from their biological cousins. These are: (a) cognitive coupling, such as the use of heuristics, providing degrees of sightedness, (b) establishment of potential fitness values for hypothetical selection processes, and (c) cognitive scaffolding for multistep thought trials. Once they are outlined, we propose that the brain's prediction imperative is the driving force behind these adaptations to the evolutionary algorithm.

More than half a century ago, Campbell (1960) proposed that creative thought results from the twofold process of blind variation (BV) followed by selective retention (SR), or BVSR. The idea that human creativity is based on blind generation has been re-stated by a number of authors since, most notably Popper (1972, 1974, 1984), Campbell (1974) and, more

recently, Martindale (1990) and Simonton (1997, 1999, 2003, 2007). Most of the work flashing out the details has been done by Simonton who has insisted, as did Campbell and Popper, that the variation process is fundamentally blind. As might be expected, much criticism has descended on these proposals. The attack has been aimed at two levels. One broadside has been directed against the idea of any type of Darwinian analysis of human creativity (see commentaries to Simonton's 1999 target article as well as, for instance, Dasgupta, 2004; Perkins, 1994; Russ, 1999; Sternberg, 1998, 1999), a move that has been flanked by a more targeted critique aimed specifically at the issue of blindness (e.g., Boyd & Richerson, 1985; Richerson & Boyd, 2005; Kronfeldner, 2010; Martindale, 1999; Mumford, 1999; Read, 2006; Schooler & Dougal, 1999).

Given the evidence of directedness in human creativity (for reviews, see Kronfeldner, 2010; Richerson & Boyd, 2005), even Simonton, perhaps the most forceful proponent of blindness in recent decades, seems to have retreated from this position and has conceded “degrees of blindness” (e.g., Simonton, 1999, p. 62). As Kronfeldner (2010, p. 202) points out, however, “there cannot be more or less blindness ... since undirectedness is defined as the absence of *any* coupling. Either a process is undirected or it is more or less directed.” For our purposes, the result is the same, there is agreement that human idea formation is directed to some degree.

Before we show this near universal agreement in more detail, it is useful to briefly mention the few theoretical attempts that have been made to apply evolutionary thinking more directly to creative processes in brains (e.g., Calvin, 1987, 2007; Dietrich, 2004b, 2007; Jung, Mead, Carrasco, & Flores, 2013), as Campbell (1960, 1970), Popper (1984), and Simonton (1997, 1999, 2007) did not relate their models to neuroscience. While the proposals by Calvin and Dietrich have been of a general nature that were not specific enough to inform a research program on the neuroscience of creativity, the Jung et al. (2013) proposal suffers from grave theoretical and conceptual problems.

First, it resurrects for some reason the concept of blind variation, which runs counter to the evidence and the consensus in the literature that human creativity is partially sighted. Second, it simply equates divergent and convergent thinking with, respectively, variation and selection, despite all the well known theoretical problems with the concept of divergent thinking (e.g., Ward et al., 1999; Dietrich, 2007), and the inconclusive neuroimaging results that it produced (Arden et al., 2010; Dietrich & Kanso, 2010; Sawyer, 2011). There is no basis for this equivalence. To repeat the main problems, divergent and convergent thinking are compound constructs that have not been decomposed into individual cognitive processes – variation, selection, or any other. No one knows what they involve in terms of working memory, attention or spreading activation, for instance. Also, since both divergent

and convergent processes are involved in creative and ‘non-creative’ thinking, we cannot isolate the creative component involved in either. This turns the research paradigm from a category error into a fully circular one. The question arises that if both divergent and convergent thinking can lead to both creative and non-creative thinking what is there about either divergent or convergent thinking that is creative or non-creative? Finally, both variation and selection seem to occur at multiple levels prior to the linguistic output stage. Note that this is also a limitation of the one study that has used a variation-selection rationale in neuroscience (Ellamil, Dobson, Beeman, & Christoff, 2012). While divergent thinking, as measured by divergent thinking tests, clearly contains selection elements, convergent thinking, as measured by the Remote Association Test for instance, still requires variants to be generated before a match can be made (e.g., Sio & Ormerod, 2009). In other words, both divergent and convergent tests require both variation and selection.

As mentioned, unless we better understand the cognitive or computational components of creativity, we are pessimistic that the current neuroimaging paradigm will bring any light to the quest of a brain mechanism for creativity. We propose here a different path that stays clear, at least for the time being, of any neuroanatomical speculations. Most of the present paper tackles *how* questions; they describe cognitive or computational processes that might play a part in instantiating variation and selection in the brain. Although these issues can inform *where* questions, any proposal on the localization of creative thinking in the brain must be approached with extreme theoretical care given the phrenological past of the field – the right brain idea, for instance.

The common denominator

To keep in focus the point pursued in the present paper, it is useful to draw attention to how change comes about in a Darwinian variational system and contrast it with a Lamarckian transformational system. The task of disentangling this knotty issue is easier when we abstract it from its roots in biology. The Darwinian variational system is an evolutionary algorithm based on a variation-selection method. In such a system, the individual units of a population are unique. They vary. Evolution occurs because these different units have different survival rates which shift their relative distribution in the population over time. Adaptive units spread and come to outnumber the less adaptive ones. The Darwinian algorithm is a statistical change that shifts the proportions of the different variants over time (Lewontin, 1970). This is to say, the reason why the population changes is because the statistical distribution of their different units changes.

A Lamarckian transformational system is also an evolutionary algorithm, but one that works in a different way. This type of change is not based on naturally occurring variation

but on adaptation-guaranteeing instruction. Importantly, there is no place for individuality or variability in this evolutionary algorithm. Change here acts on every individual unit in the same way. Variation exists, but it is treated as noise and thus evolutionarily unimportant. The Lamarckian system, in other words, includes no waste and no competition. Without variation, a selection process is superfluous since nothing exists that it could work on. All units of the population change together, at the same time, at the same rate. So the reason why the population changes is because all of its members are yoked together and transformed jointly in the direction of adaptation. Evolution occurs because the entire population is transformed in lockstep, pulled into the direction of adaptation by the influence of the environment. Clearly, the transformational system is not an accurate description of cultural change. Culture does not affect each individual in the same way. Culture is, fundamentally, a variational system (Lewontin, 1970).

The matter of blindness hits the same spot but from a different angle. Blindness, attached to Darwinism, is the concept that variation is undirected with respect to subsequent selection. The twin sub-processes of variation and selection are discontinuous, with selection only imposing a direction on variation *ex post facto*, that is, after the variation process is complete. Blind in this undirected sense means that there is zero correlation between the factors causing variation and those that sort it; they are uncoupled. Lamarckian evolution, by contrast, is not based on variation, so blindness plays no role in it – at all. Instead, it is a transformational system in which adaptation is guaranteed, from the start. The probability that any change in it is adaptive is, therefore, certain. Lamarckian evolution is *totally* directed, or a coupling of 100 %.

From this abstracted perspective, Darwin’s theory of evolution by natural selection, or more specifically the New Synthesis, is an evolutionary algorithm in which the coupling parameter of the function is set to zero. Lamarckian evolution belongs to a different class of evolutionary algorithm in which this parameter is set to one. The weight of the evidence, especially from the expert-novice research paradigm in psychology, is that creative thinking appears to be partially directed and thus fits, strictly speaking, neither into the rigid category requirements of Neo-Darwinian (total) blindness nor Lamarckian (total) sightedness (for a review, see Richerson & Boyd, 2005). Taking cultural evolution to be ‘Darwinian’, therefore, depends mostly on whether you are a lumpner or a splitter. The categorical elimination of directedness is simply not necessary for a variational system. An evolutionary system can be a variational one yet involve some coupling. This seems to be the case for culture and, by extension, human creativity (Kronfeldner, 2010).

This raises the obvious question of how brains manage to accomplish the partial coupling of variation to selection. It is the aim of this paper to propose a neural mechanism for this effect. Conceptually, this question seems to have a deceptively

simple answer: We produce mental models of the world that can simulate the consequences of generate-and-test trials. But how does this work exactly? Which parameters of the algorithmic functions are changed, and to which values? The answers, we propose, lie in the prediction machinery of the brain. But before we focus on neural mechanisms, a brief survey of three such adaptations is useful because a lack of clarity here, at the cognitive level, is prone to lead to conceptual errors once we bring in the prediction perspective.

Properties of cultural evolutionary algorithms

Consider first the use of heuristics. Being blind, evolutionary algorithms in nature explore problem spaces by running brute, uneducated searches. Heuristic search algorithms, on the other hand, ignore parts of the solution space and, in that way, limit the search from the outset (Gigerenzer & Gaissmaier, 2011). The capacity to constrain the search *a priori* arises because brains extract good design principles – folk physics, folk psychology, linear logic, etc. – from the common stock of design elements and feed them into search functions exploring the infosphere (Dennett, 1995). This causes a coupling effect; the occurrence of novel ideas (variation) is informed by the kind of problem (selection). That is to say, adaptive variants crop up more readily *because* the environment (the nature of the question) helps guide the novelty-generating process. Since these common solutions to common problems yield, effectively, a bit of foresight with respect to adaptation, they have accumulated over time, becoming integrated into the hardware as a result (Dennett, 1995; Pinker, 2002). Moreover, this already outstanding array of hardwired information is expanded further by expertise acquired during the creator's lifetime.

In terms of implementation, all that is required for domain-specific knowledge to contribute to this kind of constraint speculation is that brains enlist memory to serve as input into their simulations. Note, again, that this does not change the nature of the creative process itself. Once a heuristic process zooms in on a probable region of the solution space, the quality crank works the same. We generate different ideas, test them for functionality or some other fitness criterion, and enter those with higher adaptivity into the next round of thought trials.

For the mind's second adaptation, we first need to describe a complication inherent in dry runs. Evolutionary algorithms require a fitness function. Darwin's phrase "evolution by natural selection" denotes the fact that selection occurs in the real world, on individuals that are made flesh. The differential survival rates are determined by causal factors present in the environment. But in simulated thought trials, concrete gives way to hypothetical. Here the selection process depends on merit criteria that must also be modeled. But on what basis is this done? Since the very essence of creative thinking is to go

into uncharted territory, how do we establish what *would be* adaptive in that unknown topography? To look for answers, we turn below to predictive processes, and to the computational solutions that first evolved for action control.

A third adaptation improving the basic evolutionary algorithm is scaffolding. In nature, every variation-selection cycle in a species' trajectory is actualized and must, in its own right, be a viable form. The basic move in Darwinian evolution, in other words, is to generate-and-field-test. It is widely accepted that the cognitive toolbox of the explicit system – executive attention, working memory, temporal integration, and so on – enables the mind to buffer and manipulate several units of information at the same time. This computational capacity can also be put to use in a consecutive manner. In strategic planning, for instance, when we evaluate the potential course of a whole series of actions. In the context of decision-making, this has been termed emulation chaining (Colder, 2011; Hesslow, 2002).

Extending this concept to creativity, we show in more detail in the next section how emulation chaining would require that simulation processes take their own outputs and re-enter those that have, conceptually speaking, the highest probability values as inputs. By using these feedforward models, thought trials can breed multiple generations in a hypothetical manner. Upgraded like that, the basic move in cultural evolution becomes to generate-and-hypothesis-test. Looking ahead in this way, albeit just a few steps, produces a striking effect. Because some designs require elements that cannot be realized without a temporary scaffold, a mechanism that includes an instant pay-off requirement, such as biological evolution, can also not build them. What scaffolding permits is that trajectories in cultural evolution can bypass impossible intermediates. The benefit is a plethora of higher-order, discontinuous design solutions.

We propose in detail below that the cognitive coupling of variation to selection is the upshot of the brain's prediction machinery. This likely boosts the effectiveness of the blind, *ex-post-facto* search algorithm of the biosphere by orders of magnitude. While heuristics render cultural evolutionary algorithms faster without a proportionate error trade-off; scaffolding proliferates design options. Although one might expect such improvements in a process that inexorably bootstraps, these adaptations are typically framed in cognitive psychology in terms of expert systems and often considered at odds with evolutionary models of creativity (Dasgupta, 2004; Mumford, 1999; Russ, 1999; Schooler & Dougal, 1999; Sternberg, 1998, 1999). This critique only holds, however, if we insist on total blindness. Take it away and human creativity is best placed under the umbrella of evolution.

We turn next to the prediction framework. One of the most exciting yet completely unmapped aspects of this emerging perspective is the prospect of a brain mechanism that drives

these adaptations and thus our understanding of the neural basis of creativity.

The prediction framework

Theorists have been converging from quite different quarters on the idea of prediction as a central purpose of brain function. When seen through an evolutionary lens, this conception of information processing provides a host of new insights that have sweeping implications for the neural basis of creativity. To establish this link, we describe in this section what is meant by the notion of a proactive brain, highlight the cognitive domains which have recognized the central importance of prediction, and address briefly how neural emulators find solutions in an event space. We then outline recent work extending the framework to the predictive machinery of the explicit system and differentiate the two modeling processes of simulation and emulation. In our writing, we pass over semantic differences that have been attached to words describing prediction, such as prospection, anticipation, or expectancy, primarily because they do not come into play at the resolution with which we treat the concept of prediction (for more precise definitions, see Bubic, Von Cramon, & Schubotz, 2010).

This section is to set up the prediction framework to meet the stated aim of this paper, to show that predictive processing can offer a mechanistic explanation for the complex properties of cultural evolutionary algorithms running in brains. By exploring the computational parallels between motor control and creative thinking, we also use the prediction framework to account for the very experiences – intentions and foresight – that are often considered incompatible with an evolutionary paradigm.

The scope of predictive representations

An emerging organizing theme in neuroscience is that the brain has evolved, fundamentally, to make predictions (e.g., Bar, 2007, 2009; Friston & Stephan, 2007; Grush, 2004; Llinas, 2001; Pezzulo, Butz, Castelfranchi, & Falcone, 2008). The core idea is as follows. We can interact with the world in a nearly infinite number of ways. For behavior to be purposeful and timely in such a high-dimensional setup, the set of possible options must be pruned. We accomplish this by continuously, automatically, and unconsciously generating expectations that meaningfully inform – constrain – perception and action at every turn (Llinas & Roy, 2009; Wolpert, Ghahramani, & Jordan, 1995). Even when not engaged in a specific task, during stimulus-independent thought, the brain does not idle but actively produces predictions that anticipate future events (Buckner & Carroll, 2007; Moulton & Kosslyn,

2009). As we will see later on, this view of the brain's search engines as perpetually switched into 'drive mode' has direct relevance for the effects of incubation on idea generation.

That predictive computation is not an optional add-on comes into clear focus when we look at sensorimotor time-scales. Here, we first need to describe the problem to be solved by prediction. The motor system is a distributed, hierarchical set of brain structures that needs an appreciable amount of time to formulate a motor plan and send motor commands down to effector muscles. Sensory systems take an equal amount of time for perception. If the interaction between the two is based on actual motor execution and sensory feedback, with one having to await the outcome of the other, further delays are introduced. But a sensorimotor cycle that relies solely on such direct engagement, and is limited by real-time neural processing, cannot keep pace with the rate of change between the actor and the environments (Downing, 2009; Grush, 1997). If this were so, life would happen much more slowly than it does (Kawato & Wolpert, 1998; Wolpert, Doya, & Kawato, 2003).

The motor system solves this timing problem by relying on emulators. As soon as emulation is involved in sensory and motor systems, the processing speed can increase beyond reality-based action and perception (Kawato & Wolpert, 1998; Rao & Ballard, 1999). This entails that the brain constructs, in addition to simply engaging with the body's inputs and outputs, internal models that simulate the body's inputs and outputs (Grush, 2004). This layer of coding can anticipate the sensory consequences of actions (forward models) and invoke control processes that guide movement based on end states (inverse models) (Bubic et al., 2010; Grush, 2004; Pezzulo et al., 2008). It can also speed perception by reducing information load. This is possible because internal models save us from having to decode all sensory information 'de novo' at every turn (Diedrichsen, Verstynen, Hon, Zhang, & Ivry, 2007; Rao & Ballard, 1999). It streamlines perception to the much simpler computational task of resolving differences between the predicted and sensed environment (Colder, 2011). Finally, emulators facilitate the interpretation of sensory signals by activating associative memory networks (Bubic et al., 2010; Gross, Heinze, Seiler, & Stephan, 1999; Hesslow, 2002).

The capacity to create expectations of the future is likely common, and perhaps even the ultimate, of all global brain functions (Llinas, 2001). This emerging paradigm is so central to our proposal that we must examine the role played by prospection in other non-motor domains. We consider below learning, memory, and attention. Still others include observation, imitation, social interactions, emotion, or theory of mind (see Colder, 2011; Wolpert, Doya, & Kawato, 2003).

Prediction has long been recognized as the critical element of learning. What determines conditioning in a Pavlovian paradigm, for instance, is the discrepancy between the occurrence and the predicted occurrence of a reward, the so-called

reward-prediction error (Rescorla & Wagner, 1972). Thus, learning depends on the predictability of the reward or, more precisely, the degree of unpredictability (surprise) governs the rate of learning (Schultz, 2000). If the reward-prediction error is positive (more than expected), learning occurs. Learning ceases when the error falls to zero. If the error is negative (less than expected), extinction takes place (Rescorla & Wagner, 1972). This applies also to reward conditioning, except that in operant learning the difference signal is between expected and actual reinforcement (Downing, 2009). This places prediction at the heart of associative learning because it stresses the fact that learning occurs as a response to prediction errors (Bar, 2007). Seen in this way, learning is aimed at updating emulation, that is, at minimizing error by way of optimizing prediction (Wolpert et al., 2003).

Some studies implicate specific event-related potential (ERP) components and the release of dopamine from neurons in the ventral tegmental area (VTA) in signaling emulation errors. For instance, the error-related negativity (ERN), which is largest at medial frontal sites, is an ERP signal generated about 50 to 100 ms before an error is made (Taylor, Stern, & Gehring, 2007). The fact that it shows at the same time the erroneous motor sequence begins, and peaks with response onset, suggests that it recognizes discrepancies in expectations. Similarly, increased dopamine release in the VTA coincides with reinforcement learning, indicating that it processes errors in the prediction of a reward (Schultz, 2000). It has been hypothesized that the dopamine signal in fact produces the ERN in the anterior cingulate cortex (Holroyd & Coles, 2002).

Memory, too, is for prediction. We think of memory as being about the past while prediction is about the future. The new thinking turns this on its head. Contrary to common sense, the point of memory is not to remember the past but to make information available for simulating the future (Schacter & Addis, 2007). In this view, it is an epistemic device for simulation (Fisher, 2006). Memory is reconstructive and associative (Collins & Loftus, 1975), making it essentially the same process as imagination (Bar, 2009; Hassabis & Maguire, 2009; Moulton & Kosslyn, 2009). Accordingly, imagination has also been termed episodic future thinking, or EFT (Atance & O'Neil, 2001). That it is not semantic future thinking is due to its personal relevance. Memory, in short, serves to enable prediction based on prior experience.

One final exemplary phenomenon is attention capture (Ruz & Lupiáñez Castillo, 2002). It is well known that we focus only on what we expect to be (Meyer, Reizenstein, & Schützwohl, 1997). Perception is not the result of simple, bottom-up processing, a fact forcefully exploited by inattention blindness and change blindness (Simon & Chabris, 1999). Rather, visual perception is driven by top-down processing that contains predictions (Bar, 2007; Grush, 2004; Mehta & Schaal, 2002; Rao & Ballard, 1999). We do

not experience blindspots, or any other kind of gaps in the visual field, because we construct internal models that temporarily label, rather than fully interpret, the outside world (Gross et al., 1999; O'Reagan & Noë, 2001; Ramachandran, 1992).

This brief and selective survey shows that prediction is a fundamental component of many cognitive processes. To explore the links of this framework to creative thinking, we return to the home base of the concept to describe some pivotal details that are important later on.

Motor emulation

Predictive mechanisms challenge the long – since the behaviorist era mostly implicitly – held view that information processing is hierarchical and serial. The assumption is, roughly, that a process starts with the output of a lower-level process and terminates as soon as its output is passed to the next higher stage. However, the pervasiveness of feedforward information flow in cognition shows that this view does not apply to neural processing. And that error carries over to the needless insistence on blindness in BVSr thought trials.

Historically, the motivating insight is the ideomotor principle, first introduced to psychology by William James (1890) and further elaborated as the common coding theory (Hommel, Musseler, Aschersleben, & Prinz, 2001; Hommel, 2009; Prinz, 1987, 1997). It states that actions are represented like any other event, in terms of perceptual consequences (Shin, Proctor, & Capaldi, 2010). The motor system represents motor instructions in terms of joint and muscle positions. Motion, however, must be understood in terms of perception, or changes in spatial coordinates. So, how are motor plans interpreted then? The solution involves translating a joint-angle story into a retina-based story, so to speak. To represent upcoming motion in terms of changes to sensory systems, not motor systems, cortical motor regions send an efference copy to sensory regions – the posterior cortex, for instance – that co-register current motor instructions with sensory maps representing the body (Frith, 1992). This process, which is the forward model of the emulator (a predictor), converts the motor plan into a representation predicting the sensory consequences of the planned action. Central is that the mapping occurs to regions processing sensations so that the nature of the predictive coding is not what we are about to do but what we are about to perceive.

An emulator also needs a second component, an inverse model that works backwards. Given the degrees of freedom in a dynamic integration of sensory and motor states, a motor plan cannot specify all possibilities *a priori*. Plus, as noted earlier, actual sensory feedback is too slow to inform fluent motor commands. This is actually a well known problem in engineering. The solution is an inverse model (a controller) that gets hold of the problem from the other end (Diedrichsen

et al., 2007; Mehta & Schaal, 2002; Wolpert et al., 2003). It asks what motor coordinates I would need to realize a distal outcome, say, to grasp the coffee mug on the table. The inverse model makes a first pass, generating an efference copy, which goes to the error-predicting forward model to generate a preview of the sensory consequences. By running the efference copy on the brain's sensory maps of the body, error data is generated. That error, estimated by the two components of the neural emulator, is fed once more into the inverse model, updating its motor command, and thus the efference copy, as a result (Wolpert et al., 1995, 2003). This optimizing proceeds as the motion unfolds letting both, the predictor and the controller, converge (for excellent illustrations of these concepts, see Wolpert & Ghahramani, 2000). Note that this process occurs online and is not conscious.

Explicit predictions

Creativity researchers should adopt the prediction framework because it provides compelling leverage to deepen our understanding of the mechanics of variation-selection thought trials. We now add to the mounting evidence that computational solutions initially evolved for the control of movement have been exapted to address challenges arising in higher cognitive functions. It is not surprising that this adaptation appeared first in the motor system. After all, motricity must occur in real time and it is the only external manifestation of the brain's activity (with the exception of "endocrine and exocrine secretion and the control of vascular tone"), making it also the only way we have of interacting with the world (Llinas & Roy, 2009; p. 1301).

The co-registration of action plans to predicted sensory effects is a form of implicit prediction. Emulators in the implicit system work 'online'; that is, they regulate ongoing behaviors typically occurring within the range of seconds (Bubic et al., 2010). Due to these short timescales, and the stepwise updating, this form of online emulation is quite accurate (Pezzulo et al., 2008). Our discussion has so far centered on this category of predictors for the simple reason that the computational principles we needed to extract are best understood in this context. But complex strategies of prediction can also be found in the explicit system (Downing, 2009; Moulton & Kosslyn, 2009). The key difference is that explicit predictors can operate 'offline'; that is, on problems that are hypothetical and that can be solved outside real time (Grush, 2004). Given the adaptive advantages of looking ahead, it is perhaps equally unsurprising that explicit processes exploit the same computational principles to represent the expected future. After surveying this body of work, we extend it. More specifically, we claim that the principles of prediction can cope with several complications that arise when navigating abstract and unknown solution spaces, as is the case in creative thinking and imagination.

Based on evidence from neuroscience, the brain operates two largely independent cognitive systems to acquire and represent information (e.g., Dienes & Perner, 1999; Reber, 1996; Schacter & Buckner, 1998). The explicit system is rule-based, its content can be expressed verbally, and it is tied to conscious awareness. In contrast, the implicit system is skill or experience-based, its content is not verbalizable and can only be conveyed through task performance, and it is inaccessible to conscious awareness (e.g., Dienes & Perner, 1999; Reber, 1996; Schacter & Buckner, 1998). The existence of two distinct systems for knowledge representation indicates that each must be specialized in some way. The explicit system is understood to be a sophisticated system that is capable, thanks mostly to the computational infrastructure in the prefrontal cortex, to represent knowledge in a higher-order format; that is, it represents additional information about the information, such as the fact that it contains the information it contains. This permits the information to be broadcast to other parts of the system. In contrast, the implicit system is taken to be a more primitive and evolutionarily ancient system that does not form higher-order representations. As a consequence, the explicit system, or any other functional system in the brain, does not know about knowledge imprinted in the implicit system. This encapsulation makes its information unavailable for working memory, and thus consciousness. This dual architecture is a fundamental component of neurocognitive (Dietrich, 2004a, b) and computational models of creativity (Helie & Sun, 2010).

The same duality holds for the brain's predictive machinery as well (Downing, 2009). Neural circuits realizing implicit feedforward models of the sensorimotor arc have been identified in the basal ganglia (Houk, Adams, & Barto, 1995) and the cerebellum (Wolpert et al., 2003). Online emulation in these brain regions proceeds by the predictor-controller pair linking one step to the next immediate one. Internal models in the implicit system cannot anticipate more than one step at a time because each element of the sequence must critically depend on the directly preceding one (Dienes & Perner, 1999).

Brain regions associated with the processing of explicit predictions are the hippocampus (Gluck & Myers, 1989), thalamus (Downing, 2009), and cortex (Hawkins, 2004). Predictive processing in the explicit system is not limited to the step-by-step approach. The cognitive flexibility inherent in the system carries over to its predictive competence. Executive processes, taking advantage of the omnidirectional access of higher-order representations, construct internal models that are run offline to produce expectations that can be long-term as well as hypothetical in nature (Downing, 2009; Grush, 2004). Predictive representations in the explicit system are essential to the higher cognitive functions typically ascribed to the prefrontal cortex, such as strategic planning, imagery, estimating the outcome of a series of actions, and, to return to creativity,

the conception of alternate futures that have never before seen the light of day.

We are now in a position to untangle some complexities that we have been keeping clamped. Specifically, we cannot continue to ride roughshod over the difference between emulation and simulation because the distinction might matter a great deal for creative thinking. Generally speaking, simulation is the use of one process to acquire knowledge about another. All simulations, mental imagery included, make predictions. By answering ‘what if’ questions, simulations generate knowledge that makes explicitly accessible the possible consequences of being in a specific situation (Moulton & Kosslyn, 2009). But not all types of simulation deliver the same kind of predictive information.

First-order simulations only mimic abstract models; they do not represent the intermediate processing steps. A computer model of a hurricane is an example. Although the sequence of the simulation is functional – each step critically constraining the next – and not epiphenomenal, the algorithms transforming successive states differ categorically between the simulator and the real system (Moulton & Kosslyn, 2009). Simulation of this type occurs when we imagine the changing scenery on a drive through the countryside without also modeling the corresponding steps causing the change in scenery – turning the steering wheel, for instance. Emulation, by contrast, is second-order simulation. It is a special type of simulation that imitates not only the content but also the processes transforming the content (Fisher, 2006; Grush, 1997, 2004; for an alternative definition and usage of the term emulation, see Tomasello, 1999). This is to say, both simulation and emulation operate by sequential analogy, but one is theory-driven, the other is process-driven (Goldman, 1995).

Why does this matter for creativity? Anticipating a later discussion, we advance in this article the speculative possibility that most types of creative insights, in order to become conscious, require emulation, not simulation. We also argue that this is consistent with the concept of embodied cognition (Barsalou, 2009; Niedenthal, Barsalou, Winkielman, Krauth-Gruber, & Ric, 2005).

As an example, take the deceptively simple question of why we cannot tickle ourselves. The standard response is that we know what is coming. The problem with that explanation is that when someone tickles us, also telling us what exactly to expect does not diminish the effect. That is because a mere simulation, imagining how it would feel, does not provide sufficiently detailed information about the predicted sensory consequences. Self-tickle, on the other hand, involves emulation. In this case, our motor system generates a motor command which, naturally, holds instructions for state-to-state transitions. The efference copy, holding the same information, is used to match predicted outcome to sensory feedback. So, unlike a first-order simulation, an internal model of the entire sensorimotor loop also models the transformational steps.

Subtracting the emulation from the actual sensory feedback lessons – anticipates – the experience properly (Blakemore & Decety, 2001).

Based on differences in computational architecture, it would appear that predictions in the implicit system are necessarily emulations. The concrete operational setup of implicit processing lacks the capacity to represent hypothetical scenarios. Any predictive processing, therefore, can only proceed on a step by step basis, and by using the same operator as the real system (Downing, 2009). The explicit system, on the other hand, can abstract information from context. In so doing, simulation increases generalizability and cognitive flexibility in the imagined future. Naturally, there is a tradeoff. Given the added transformative detail it contains, emulation might be more accurate and powerful (Pezzulo et al., 2008). Also, emulation always has self-relevance and this, we propose below, might be the key element in making an ingenious idea conscious. Anecdotal evidence would seem to corroborate this. We often ‘get it’ only when we know how to do it. Why else did Einstein need to ride the beam of light himself?

Units of creativity

Let us summarize the path taken so far. Debates on Darwinizing culture routinely generate more heat than light. The origins of our creative ideas, and how they arise, have been a prominent and long-standing showcase in this tug-of-war. In the first part of the paper, our goal was to distill from this controversy a common denominator that, though unobjectionable, still has not found its way to the neurocognition laboratory. We devoted the second part of the paper to the emerging prediction framework, which we claim provides propulsive help on exactly that front.

We assert that the frameworks of evolution and prediction, together, constitute the bedrock on which we should anchor a research program on the neural basis of creativity. While the evolutionary approach – despite some support – has not achieved this paradigmatic status, the prediction framework has yet to be applied to creativity at all. Indeed, it seems that either framework alone lacks the explanatory power to convincingly account for the phenomena of creative thinking. Apart from the long-standing insistence on blindness, it is perhaps this reason why the efforts of several evolutionary theorists (e.g., Campbell, 1960; Simonton, 1999) have had almost no influence on neuroscience research so far. This changes, we argue, when one perspective is considered through the lens of the other. In what follows, we flesh out in more detail how the evolutionary framework gains traction as a mechanistic explanation at the neurocognitive level through the application of prediction. The reverse also holds: the relevance of the prediction approach to creativity comes

into clear view only when it is framed in terms of a generate-and-test paradigm.

We use the prediction framework to ask several questions about the nature of creative thought. First, we examine how prediction causes partial sightedness in the variation process. Second, we work out how prediction impacts the creative process at the level of selection; specifically, we propose that predictive goal representations serve as merit criteria for a fitness function. Third, we explore how emulation chaining over many iterations might underlie the scaffolding effect that enables thought trials to leap over unrealizable forms. In the fourth question, we discuss the extent to which first-order simulations may be limited in generating insights. We suggest that creative ideas are likely to require emulation processes to reach consciousness. Finally, in the last query, we make the case that predictive processing can explain the phenomenological experiences of intention and foresight that are so commonly recruited to set human creativity apart from creativity in the biosphere and artificial intelligence.

Partial sightedness in the variation process

Prediction mechanisms in the brain can give direction to the production of ideational combinations in a number of ways. Recall the workings of the motor system in which the forward model (predictor) and the inverse model (controller) converge on an answer. The controller of the neural emulator starts by forming a representation of the movement endpoint, the final body configuration to be reached (Grush, 2004). In the terminology of evolutionary algorithms, this goal state can be considered equivalent to the fitness function of a selection process because the eventual behavior is based on it. Evidence from motor neuroscience suggests that there are multiple motor plans competing at any one time, so that the brain also runs simultaneously multiple forward models to predict the behavior of the motor system. Each forward model emulating various sensory changes can thus be regarded as a hypothesis tester for the context that it models (Wolpert et al., 2003). In consequence, we can regard the computational principle behind forward models as the neural analogue of a possible (partially sighted) variation process. For each predictor-controller pair trial, the efference copy with the smallest prediction error represents the best fit with respect to the desired goal state and is used to determine the next generation of action commands.

Although we can phrase the forward and inverse motor models in the same computational framework as variation-selection thought trials, this should not hide that creative behavior is different in many important ways. Based on differences in computational architecture alone, we can expect motor prediction in the implicit system to lack the sophistication of representing higher-order goals (Haider, Eichler, & Lange, 2011). For example, creativity is an adventure into

parts unknown, which, in motor control, only occurs if we were to execute a totally new movement, say, an original skateboard trick on the halfpipe. The motor system would not have learned an inverse model for this action (Schubotz, 2007). And without such an action memory, the motor system would not even know the algorithm that it needs to solve (Wolpert et al., 2003). Lacking this paired controller, and therefore prediction error data, a forward model in this case would be rather blind with respect to selection. A predictive goal state for the new skating stunt could still be generated, but it would have to be done by the explicit system (Anderson & Lebiere, 1998).

For implicit processing, we imagine that it works like this. The implicit system operates on the basis of a stochastic process (Perruchet & Vinter, 2002). Geared towards optimizing behavior, it simply tries out, by trial and error, solutions to environmental contingencies (Ashby & Casale, 2002). Learning is reflected by shifting weights in the knowledge network, which introduces a bias for subsequent information processing (Cleeremans, 2008; Shanks, 2005). Note that the increased efficiency of the system is realized without the *a priori* computation of a goal. Strictly speaking, in the absence of a goal representation, one cannot classify this process as a prediction. At best, the implicit system can be said to form expectations, heuristics that are embedded into the structure of the knowledge network (Perruchet, Cleeremans, & Destrebecqz, 2006).

To clarify, the motor system does use inverse models, which are, of course, representations of the future. But such inverse models only exist for actions that have been learned previously, that is, in known solution spaces. A completely new movement poses a different challenge. Because there is no controller that can be paired with the forward model, the variation process of implicit search algorithms would be blind. In addition, the implicit system cannot engage in offline emulation because the entire sensorimotor loop cannot be modeled internally. Put another way, the prediction machinery of the implicit system only possesses partial sightedness for known problem spaces, a situation that does not really qualify as creativity. For explorations in terra incognita, the implicit system can still be creative but this creativity is limited to that of biological evolution. That is, variation is an internal process while selection is determined, *ex post facto*, by ‘natural selection’, or actual feedback from the environment. Thus, we claim that the implicit system cannot, by itself, be creative above and beyond the creativity exhibited by evolutionary algorithms in nature – in a blind manner.

This changes for the explicit system, in rather dramatic fashion. Without repeating standard conceptions, a key difference with respect to explicit prediction is the computational platform of a working memory buffer in the prefrontal cortex. This additional layer of the explicit system codes higher-order goals that are abstracted from the context and maintained separate from – as opposed to embedded in – the knowledge

structure (Cowan, 2001; Miller & Cohen, 2001; Dienes & Perner, 1999). In that way, we can imagine – predict the effects of – events that we cannot reproduce with our bodies, such as ocean waves for instance. The projection of a metacognitive goal into an uncharted vision space is the missing piece to generating a complete internal model of external events. Again, translating this into an evolutionary account, what this type of goal representation does, in our thinking, is to also move the selection process inside, because the parameters of the prediction – the goal representation – double as a fitness function. In sum, we argue that, unlike the implicit system, the explicit system is able to run simulations offline, without an external sorting process. Add to this the ability to chain single iterations of thought trials, an upshot of goal maintenance in working memory, and the creative capacities of evolutionary algorithms in the explicit system far surpass those of their biological cousins.

Accordingly, prefrontal regions have been hypothesized to hold representations of alternative potential futures (Downing, 2009; Schubotz, 2007; Willingham, 1999) and engage in counterfactual thinking through episodic analogue reasoning (Kent & Lamberts, 2008). Moreover, the predictive account of motor control has been generalized to, for instance, strategic planning, deliberate reasoning, and decision-making (Cisek, 2006; Colder, 2011; Downing, 2009). These applications of predictive processing by the explicit system are closely related to mental model theory (Johnson-Laird & Byrne, 2009) or the simulation hypothesis (Hesslow, 2002). Note that these accounts of cognitive function are different to creativity in that in those cases, as opposed to creativity, the solution space is, in principle, known. That means they do not have to proceed via a variation-selection algorithm.

But the game-changing advantage of internalizing the selection process is not the ability to run offline simulation *per se*. It is that selection can now be coupled to the variation component, resulting in degrees of sightedness. Put differently, we contend in this paper that it is by virtue of having both processes operating in the same system that one, the goal representation, can be used to guide the other, the production of ideational variants.

Mainstream cognitive psychology typically does not use evolutionary language to describe its phenomena (but see e.g., Geary & Huffman, 2002). Cognitive psychologists tend to neglect variation-selection algorithms as an important cognitive mechanism. In keeping with our efforts to build terminological bridges, we can also link this notion of cognitive coupling to a specific mechanism in the literature on prediction. The core idea is that executive processes configure the neural simulator with context, memory, folk wisdom (Colder, 2011; Wolpert et al., 2003), and, of course, domain specific expertise. Set up this way, the simulator generates variants that have a higher fitness value from the start. Heuristics are one

way of gaining partial sightedness. Such educated guesses must necessarily be driven via predictive processes.

Our proposal of more specific neural correlates for variation and selection has the potential to open up new avenues of research in the neuroscience of creativity. We previously theorized that those processes required for generating creative ideas and those carrying out evaluations are likely to engage different cognitive processes and different brain regions (Dietrich, 2004b, 2007). For the variation process, we suggested that neural circuits that process specific information to yield noncreative combinations of that information must be the same circuits that generate creative combinations of that information. This is to say, the recombination of bits and pieces of information into novel configurations must originate from the very neural circuits that normally handle those bits and pieces of information. This possibility must be conceded as a consequence of taking the brain to be a nonlinear information processor that is functionally compartmentalized. If so, this would mean that there might not be a region, or a set of regions, that is especially engaged in generating variation. Regions that might produce variation in music improvisation are unlikely to be similar to those producing variation in theoretical physics. It also means that neuroimaging technology would only find activity in the places the purported creativity test taps into. Although this seems like common sense, it runs counter to the prevailing interpretation of the neuroimaging data.

The selection process, on the other hand, cannot be inherent in every neural circuit but must depend instead on higher-order structures, such as the prefrontal cortex. Only such integrative brain regions are capable of assessing a set of very complex and constantly changing rules, such as the values implicit in a person's culture. Consistent with this conception, we proposed that the role played by the prefrontal cortex can be seen as a threefold process. It is involved in (1) making ideational combinations occurring in the brain fully conscious, (2) assessing their likely value, and (3) ultimately implementing their creative expression (see Dietrich, 2004b).

This first proposal to identify one possible analogue for variation and selection in terms of brain areas and processes was an important step forward because it disentangled the two twin processes of the creative process at the neural level. However, it was still too general to inform a research program. On the basis of our present theoretical considerations, we can now specify this further by suggesting that, for some forms of human creative thinking, the variation process can be tied to the computational principles of prediction – in the form of forward models, for instance – while the fitness function of the selection process can be taken to correspond to a predictive goal representation. Owing to our lack of understanding the creativity mechanisms in the brain, we are cautious in formulating this because we cannot, and do not want to, make the very strong claim that this is true for all creativity. This initial

correspondence does more than just bringing the concept of creativity into contact with mainstream cognitive neuroscience. It is a proposal that is specific enough to present a new point of attack on the problem of the neural basis of creativity.

Merit criteria for the selection process

There are several complexities attached to an evolutionary variation-selection algorithm that has a coupling parameter greater than zero. None is as critical as the matter of how we impose a direction on the variation process at all. Where do we get the foresight to determine beforehand what counts as adaptive? Unless we are prepared to consider creators a special class of prophets, we need a sound mechanism to explain what looks for all the world like clairvoyance. This issue does not arise for the (blind) creative process in the biosphere of course. Nor, as we have claimed, does it apply to the (blind) creative process in the implicit system. It also plays no role in the proposals suggesting that the explicit system draws on the same predictive toolbox the motor system has evolved to construct internal models of action effects. That is because these domains deal with cognitions operating in problem spaces that are known – at least in principle. This is where creativity in the explicit system is different.

It turns out that the matter of foresight, so often used as a counterargument against an evolutionary analysis of human creativity, yields to a straightforward evolutionary explanation once we refract it through the prism of prediction. The short of this explanation is a goal representation, which, by virtue of being a prediction, can direct the variation process. A longer explanation would entail drawing closer computational parallels between predictor-controller pair updating occurring in motor control and offline emulation processes taking place during creative thinking. This would likely require a detailed application of Bayesian decision-making theory (e.g., Lau, 2008; Tenenbaum, Kemp, Griffiths, & Goodman, 2011), which cannot be done within the limited scope of this paper. In the present context, we must be content to state the basic theoretical idea. Suppose we take the generation of several forward models to correspond to the variation component. Suppose further we similarly equate the inverse model's predictive goal representation with the fitness function. With the boundaries of a hypothetical solution space set up like this, well established Bayesian inference techniques could tell us how evolutionary search algorithms converge on creative solutions in the explicit system.

While this is best taken up by computational scientists, our present efforts remain focused on the goal representation itself. We merely use the prediction framework to explore the kind of information that might flow into such goal representations. Remember that the solution space is unknown in creative endeavors and that the specs of this goal

representation would become, in our reasoning, the fitness function for the ratcheting evolutionary search algorithm.

Consider the case of creative behavior that can lean on a fairly well defined goal state, say, an architectural design for a building. The building must likely conform to a number of precise and predetermined specifications. The designer is further bound by a host of fixed parameters – gravity, behavior of building materials, and so on. Add to this former experiences, context, or special characteristics in the environment and the neural simulator can already be preconfigured with a great deal of information. While this still leaves plenty of room for creative designing, a goal representation can now readily be generated. And as soon as such a goal is in place, we would argue that the exploratory activity of the explicit system can then proceed in principle in a manner similar – computationally speaking – to that in the motor system.

Consider now the somewhat different case of creativity in a totally unknown region of design space. A famous example that comes to mind is the state of physics in the aftermath of the Michelson-Morley experiment, a study that provided conclusive experimental proof that there was no luminiferous ether (Michelson & Morley, 1887). The negative result effectively put an end to the Newtonian paradigm, leaving physicists temporarily with no viable alternative. It would appear that there are two ways to move forward in this situation. In the absence of any kind of predictive goal representation, an evolutionary algorithm could always ratchet upwards blindly and stepwise, as is the case in biological evolution. The other, more exciting possibility is to engage in predictive processing nonetheless. A projection based on little or no information is of course riskier but might nevertheless simulate a goal representation with some adaptive criteria. This is what Einstein did by, for instance, making the speed of light a universal constant. The key to this kind of creativity, it would seem, is to come up with the correct (or at least partially correct) goal state. Such a projection would also enable scaffolding.

Cognitive scaffolding for the simulated future

In biological evolution, each round in the creative process must be instantiated. This is because selection is *ex post facto* and, importantly, occurs in the real world. Accordingly, variants are first produced and only then, after they have come into existence, do they get sorted. In brains, selection occurs neither (necessarily) after the fact nor is it (necessarily) located in reality.

In motor control, decision-making, and other domains, the concept exists that a sequence of emulation steps can evaluate, or reason through, a series of choice points all the while the process is bypassing the realization of each option (Colder, 2011; Hesslow, 2002). The game of chess is a vivid illustration. Recall that emulation chaining of this type, minor exceptions aside, does not count as creativity if we apply the

putative definition of novel, useful and surprising. The predictive goal representation itself is situated within the boundaries of a recognized and mostly mapped out event space. The algorithm that needs to be solved is principally known and the task becomes one of methodical problem solving.

This stepwise and analytical way to approach problems is typically considered anathema to the inspiration of a creative insight (Bowers, Regehr, Balthazard, & Parker, 1990). Ingenious ideas are said to be discontinuous, surprising, and perhaps even wild. Consider, for instance, the requirement of the U.S. Patent and Trademark Office that inventions must contain a ‘non-obvious step.’ In thinking about a brain mechanism for this discontinuity, we need not look any further than a predictive goal representation. Mental projections into a distant future can readily be independent and disconnected from the present. The creative move in this case is to make a good prediction. The task thereafter would again consist mostly of joining the detached initial state with the distant goal state, or to keep with the motor terminology, bringing the forward and inverse model together. This kick-and-rush form of creativity can produce creative designs of a different kind. A goal representation in this case acts like a beacon of light. It sets a distant target a search process can aim at. In this way, cognitive scaffolding can temporarily support design elements that are not viable forms in their own right, allowing them to figure into subsequent iterations of the variation-selection algorithm. An arch is the canonical example of an interlocking design that must leap over non-adaptive, intermediate forms. Biological evolution cannot do that.

Suppose the entire process, goal projection and predictor-controller convergence included, runs unconsciously, as might be the case during incubation. If we indulge in this speculative possibility for a moment, one can see how such a mechanism might account for two prominent experiences associated with creative ideas. First, given that we have no access to our unconscious brain activity, we might presume that the feeling of discontinuity associated with insights stems from the cognitive scaffolding falling away prior to conscious representation. A creative insight, being several steps removed from the initial problem configuration, can thus no longer be consciously traced to it in a straight line. Second, and perhaps even more of a stretch, we might also presume that the sense of certainty we often get that a creative idea is obviously right arises from predictor-controller interactions to manage to connect every step between the initial state and the goal state. Perhaps this might even be the strengthening signal for the unconscious process to generate a conscious aha representation in the first place. We do not wish to carry this conjecture too far at this point, but the message we might take from this is that the inability to report the steps leading to the solution does not mean that there were not any steps. Nor does it mean that the steps did not follow a variation-selection protocol.

Emulation versus simulation

Of course, we do not want to make the very strong claim that all creative ideas depend on a reference to an efference copy. Given the lack of data on this issue, we naturally leave open the possibility that creativity can also arise by way of simulation, episodic analogue reasoning, or perhaps blind variation. This, again, is simply a matter of cautious theorizing, as we cannot claim that the prediction mechanism we are proposing here fits all the various ways in which humans display creative abilities. However, offline emulation seems to us a more powerful candidate for mediating the necessary prospective dimensions of creative insight. We think so for two reasons. The first is self-evident; emulation contains additional ‘how to’ information. When neural emulators can be pressed into service, we get, over and above the final solution, a roadmap on how to get there from here. The second is likely a consequence of the first. Modeling the transformational steps should make a solution more compelling. Insights are accompanied by the strong intuition that the answer just *is* right. It is reasonable to think that this feeling is rooted in the creative processing itself. Although anecdotal evidence suggest that this intuitive knowledge too, can be wrong, it is unlikely that the initial sense of certainty is totally unrelated to correctness.

We argue, on three separate grounds, that definitive knowledge of an unbroken thread joining the problem constellation with the far-flung goal representation is a key contributing factor, if not a decisive trigger, in bringing about an explicit representation of an unconscious incubation trial. First, competence in knowing how to connect a link in a chain to the next seems necessary to select adaptive variants at each iteration of the sequence. After all, the fitness function must operate without actual feedback for many iterations. Second, planning for the future engages a number of motor structures, such as the cerebellum, premotor cortex, and supplementary motor area (Baker, Rogers, Owen, Frith, Dolan, Frackowiak, & Robbins, 1996; Dagher, Owen, Boecker, & Brooks, 1999), which suggests that the motor system might be involved in imagining, and perhaps accessing the viability of, future scenarios. A third reason to favor emulation over simulation is based on the paradigm of embodied cognition and demands a bit more elaboration.

Embodied cognition is based on the idea that thoughts, emotions, and actions are grounded in bodily interaction with the environment (Barsalou, 2009; Niedenthal et al., 2005; Shapiro, 2007; Wilson & Golonka, 2013). This concept is now integrated in the theories of several disciplines, including robotics, linguistics, philosophy, social psychology, and cognitive psychology. It is by no means a new idea. Early findings of, for instance, Köhler (1951) already suggested that the adaptation to prism spectacles is enhanced when people act in a stable environment. Gibson (1966) also postulated early on a close interaction between perception and action,

stating that we have to perceive in order to move and move in order to perceive. The idea that perception and action are inextricably linked has now been widely adopted and is most comprehensively formulated in the common coding theory (Prinz, 1997; Hommel et al., 2001).

Some empirical findings exemplify the key idea of embodied cognition. For instance, Reed and Farah (1995) asked participants to compare two poses of a human model in an unusual body configuration. The pictures were presented in two versions. Either the visual angle had changed but the body position remained identical or the body position had changed as well. The task was simply to determine whether or not the body position of the model had changed. While viewing the pictures, participants also engaged in a secondary movement task. They made unconstrained, nonrepetitive movements of either the arms or the legs. If a body schema exists, it should be used in both, the primary, visual task to encode the model's body position and the secondary, movement task to ensure that they are nonrepetitive. The results showed that arm movements reduced error rates when the model's arm position changed and leg movements were better recognized when the model's leg position changed. This suggests that representations of one's own body contribute to the performance in a visual matching task. In mathematical cognition, Dehaene, Bossini, and Giraux (1993) discovered the so-called 'snarc' effect (spatial numerical associations of response codes), which refers to the finding that a left-hand reaction to small numbers is faster than a right-hand one and a right-hand reaction to large numbers is faster than a left-hand one. Word meanings also seem to be grounded in sensorimotor representations. Reading a word activates those sensorimotor brain areas that are also active when we actually hear the sounds produced by this object (e.g., Kiefer, Sim, Herrnberger, Grothe, & Hoenig, 2008). In yet another experiment, Glenberg and Kaschak (2002) asked participants to judge the sensibility of a sentence while making either a forward or a backward movement. The meaning of the sentence was either compatible with the movement or not. Again, the movement mattered. Responses were faster when movement and sentence meaning were congruent. This action-sentence compatibility effect occurred even for sentences referring to abstract actions. Finally, and most interestingly perhaps, is that higher-order cognitive processes are also affected by movements. For instance, Thomas and Lleras (2009) instructed participants to solve Maier's two-string problem, which is a classical insight task. Before doing so, they were first asked to move their arms either back and forth or sideways. Problem solving was significantly higher in the back and forth condition.

Overall, there is overwhelming evidence that actions (or movements) play an important role in trying to understand the world. Actions can change our perception and, more importantly, our thoughts. Similarly, thinking about an object

activates associated embodied states (Hommel & Eglau, 2002). Cognition, then, is not merely a process of abstract symbol manipulation but relies heavily on the brain's modality-specific systems as well as on current bodily states. According to Wilson and Golonka (2013), embodied cognition is based on the central thesis that performance emerges from the interplay of task specific resources that are distributed across the brain, the body, and the environment. The type of body human beings possess and the manner in which this type of body interacts with the environment determines how we think about the world (Lakoff & Johnson, 1999). If we had other bodies, our thinking would also be different.

Extending the rationale behind embodied cognition to creative thinking, we can make use of the perception-action coupling (Gibson, 1979) to argue that the computational toolbox that first evolved to solve movement tasks is recruited to help with creative thinking. Embodied cognition replaces the classical information-processing conception that only recognizes brain structures as sources of representations about the world. A more dynamic model of interaction that also considers the role played by the body and the environment supports our claim that these other sources of representations are involved in generating ideational combinations.

To illustrate this with a concrete example, imagine you need to cut a sheet of paper but do not have a pair of scissors at hand. Scanning the environment for suitable alternatives you are unlikely to simply consider objects on the basis of their characteristics. Rather, you try, for each object you encounter, to emulate the motion of cutting, and it is this emulation process that helps you decide if an object is suitable for the task. Such an assessment cannot be done solely by analyzing an object's characteristics (Carramazza & Mahon, 2006; Thomas & Lleras, 2009). Emulation helps to detect if an object can be used for the intended goal. Moreover, the intention to cut guides the perception of the object (Gibson, 1979) so that you might not even perceive the object's original function. Linking this to the generation of creative ideas, the fact that our cognition is embodied implies that movements, or the emulation of movements, are important in finding new ways to solve problems. Creativity, in this case, is to generate a predictive goal representation as well as to emulate our way to it; that is, creativity is to find the evolutionary algorithm that binds the goal state to the problem by way of a series of motor plans.

Intention and purpose in human creativity

Finally, we propose that the prediction framework can be recruited to do explanatory work for the phenomena of intention and purpose that often accompanies human creativity. By universal agreement, the creative process in the biosphere is not teleological. It serves no end, and its designs are neither premeditated nor deliberately initiated in response to a

perceived need. Human creators, by contrast, act on purpose; they create with intent and with an objective in mind. This argument, in one form or another, remains the central line of defense for those emphatically opposed to evolutionary approaches to human creativity (Dasgupta, 2004; Perkins, 1994; Russ, 1999; Sternberg, 1998, 1999). To shed light on this purported difference, we must briefly examine the role of prediction in neural explanations of the sense of agency.

Current thinking about the brain's agency system is based on the idea that conscious will is the product of the same unconscious processes that also give rise to motion (Wegner, 2003). To experience ourselves as agents we must have advance notice that our body is about to do something. Taking a bird's-eye view of our earlier discussions, this sneak preview is provided by the predictive processing of the motor system. Motor plans originate at the highest level of the motor hierarchy and take time to reach muscles. It is thought that during these preparatory computations, a carbon copy of the plan – the efference copy – is leaked to consciousness, effectively providing prior knowledge of the upcoming action (Blakemore, Smith, Steel, Johnstone, & Frith, 2000; Frith, 1992; Shin et al., 2010). We become conscious only of the two end-products of this activity, the wish to act and the motion. Since both originate from the same source, and one precedes the other, we establish a causal link between them and get the feeling that we cause our own motion (Wegner, 2003). The sense of agency, in other words, is the misinterpretation of neural activity as conscious intention when in fact it is a representation resulting from the simulation of possible action outcomes.

If current motor instructions are not co-registered, either because the motor plan is executed too fast by the implicit system or because it is mapped only weakly on to sensory regions, we cannot predict what we are going to see and feel. As a result, our own motion would come as a surprise to us and we get the feeling that we did not produce it intentionally (Frith, 1992). So, a normal sense of agency depends on the adequate mapping of predicted movement and a confirmation report that the expected experience actually occurred, that is, the prediction is validated by sensory feedback.

This rough, conceptual overview shows that intentions and the experience of purpose are themselves products of the brain's predictive machinery. This insight not only shows that intentions cannot be recruited to argue for a categorical distinction between cultural and biological creativity. It also demonstrates the explanatory power of predictive processing in explicating the neurocognitive mechanisms underlying creativity in general. Given the lack of sound explanations in this field of psychology, we propose that the prediction framework, especially when embedded into a larger evolutionary frame, offers a new and promising direction in our search for the creative process taking place in brains.

Conclusion

Despite a surge of papers in recent years, the neuroscience of creativity finds itself in a theoretical void that has perhaps no equal in psychology. A recent, comprehensive review concluded that “not a single currently circulating notion on the possible neural mechanisms underlying creative thinking survives close scrutiny” (Dietrich & Kanso, 2010; p. 845). Because the neuroimaging work since this review has followed the same rationale, this negative conclusion, which was also reached independently by two other reviews (Arden et al., 2010; Sawyer, 2011) remains intact. While there are undoubtedly a number of reasons for this malaise, we contended that perhaps the most significant one is that empirical studies have not been grounded in the evolutionary paradigm, the only paradigm we know for sure to generate creative products.

We first highlighted that we can no longer foreclose this most obvious of paradigms in our search for the brain mechanisms of creativity by establishing that underneath the wedge issues of the debate on cultural evolution there exists a common denominator on vital points, namely that human culture is a variational system of change that involves partial coupling.

To specify in more detail how the generate-and-test algorithm might operate in brains, we then introduced the prediction framework. After presenting the key ideas of the emulation theory of representation, we reviewed how internal models of motor control, driven via efference copies, estimate the sensory effects of different actions. We also reported on efforts to extend the brain's prediction competence from the sensorimotor loop to neural emulators that can be run offline by the explicit system to produce mental imagery, aid in decision-making, facilitate social interactions, and provide knowledge about the topography of potential futures.

Readers who accept the importance of the brain's predictive processing should have little trouble seeing its profound implications for the creative process. In the paper's final section, we spelled out these implications in some detail by bringing the prediction framework into contact with evolutionary thinking. In particular, we developed the idea that internal representations of the emulated future provide the neural mechanism that enables brains to partially couple variation to selection and, in so doing, provide degrees of sightedness.

The concept of prospection as a fundamental principle of neural computation can, in fact, address a whole host of objections that have traditionally been raised against an evolutionary analysis of human creativity, such as blindness, setting hypothetical merit criteria, scaffolding, foresight or intentions. We showed, for instance, how the prediction framework can explain the violation of blindness we see in human creativity. Predictive processing also gives substance to the problem of how brains set merit criteria for a selection process that is hypothetical and must therefore ratchet without actual environmental feedback. In creativity, after all, the

properties of the eventually triumphant solution are – by definitional fiat – unknown. We proposed that neural emulators generate goal representations that furnish the required merit criteria for a sorting process that is imaginary and therefore must ratchet through a hypothetical landscape without actual environmental feedback. In using such predicted quality evaluators, neural networks could attach probability values to the simulated variation in order to assess their fitness. Those items with the highest values or with the fastest speed of processing (Corallo, Sackur, Dehaene, & Sigman, 2008; Whittlesea, 2002; Wilbert & Haider, 2010) are kept for the next iteration.

The prediction framework provides, for the first time, a way to address at the neural level yet another objection against an evolutionary account. In the biosphere, each iteration of the variation-selection process is instantiated so each generation must yield units that are adaptive in their own right. Human creators, by contrast, are capable of scaffolding to reach higher-order designs; they can build a thing of beauty even if all the intermediate steps are ugly contraptions. This must necessarily be driven via predictive processes.

Most importantly perhaps, the prediction framework can provide clear leverage to rethink the most commonly raised objection, the matter of intention and foresight. While human creators act with purpose and foresight, this argument goes, biological evolution is, famously, devoid of any teleological notion. But here, too, can our deepened knowledge of the brain's prediction capacity significantly deflate, if not resolve, a false category formation. We showed that intention and foresight are phenomenological experiences that result from representing the expected future. They merely constitute a change to the blindness parameters of evolutionary algorithms and are thus entirely compatible with an evolutionary explanation of human creativity.

We see these new theoretical proposals on the neuroscience of creative thinking as merely a promising starting point. As such, they contain significant limitations. For instance, we do not address what constitutes variation or selection processes in general, how the brain copies information, or how individual differences play into the generation of goal representations. We hope, however, that such issues are taken up in subsequent work.

References

- Abraham, A. (2013). The promises and perils of the neuroscience of creativity. *Frontiers in Human Neuroscience*, 7, 1–9. doi:10.3389/fnhum.2013.00246
- Anderson, J. R., & Lebiere, C. J. (1998). *The Atomic Components of Thoughts*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Arden, R., Chavez, R. S., Grazioplene, R., & Jung, R. E. (2010). Neuroimaging creativity: A psychometric view. *Behavioural Brain Research*, 214, 143–156. doi:10.1016/j.bbr.2010.05.015
- Ashby, G. F., & Casale, M. B. (2002). The cognitive neuroscience of implicit category learning. In L. Jiménez (Ed.), *Attention and Implicit Learning* (pp. 109–141). Amsterdam: John Benjamins Publishing Company.
- Atance, C. M., & O'Neill, D. K. (2001). Episodic future thinking. *Trends in Cognitive Sciences*, 5, 533–539. doi:10.1016/S1364-6613(00)01804-0
- Aziz-Zadeh, L., & Liew, S. L. (2013). Exploring the neural correlates of visual creativity. *Social Cognitive and Affect Neuroscience*, 8, 475–480. doi:10.1093/scan/nss021
- Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.
- Baker, S. C., Rogers, R. D., Owen, A. M., Frith, C. D., Dolan, R. J., Frackowiak, R. S. J., & Robbins, T. W. (1996). Neural systems engaged in planning: A PET study of the Tower of London task. *Neuropsychologia*, 34, 515–526.
- Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Science*, 11, 280–289. doi:10.1016/j.tics.2007.05.005
- Bar, M. (2009). The proactive brain: Memory for prediction. *Philosophical Transactions of the Royal Society B*, 364, 1235–1243. doi:10.1098/rstb.2008.0310
- Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 364, 1281–1289. doi:10.1093/acprof:oso/9780195395518.003.0016
- Blakemore, S.-J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience*, 2, 561–567. doi:10.1177/0956797612444612
- Blakemore, S. J., Smith, J., Steel, R., Johnstone, E. C., & Frith, C. D. (2000). The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: Evidence for a breakdown in self-monitoring. *Psychological Medicine*, 30, 1131–1139. doi:10.1017/S0033291799002676
- Bowers, K. S., Regehr, G., Balthazard, C. G., & Parker, K. (1990). Intuition in the context of discovery. *Cognitive Psychology*, 22, 72–110.
- Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. Chicago: University of Chicago Press.
- Bubic, A., Von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience*, 4. doi:10.3389/fnhum.2010.00025
- Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences*, 11, 49–57. doi:10.1016/j.tics.2006.11.004
- Calvin, W. H. (1987). The brain as a darwinian machine. *Nature*, 330, 33–34.
- Calvin, W. H. (2007). Why a creative brain? Evolutionary setups for off-line planning of coherent stages. In H. Cohen & B. Stemmer (Eds.), *Consciousness and cognition: Fragments of mind and brain* (pp. 115–126). Amsterdam: Elsevier.
- Campbell, D. T. (1960). Blind variation and selective retention in creative thought as in other knowledge processes. *Psychological Review*, 67, 380–400.
- Campbell, D. T. (1974). Unjustified variation and selective retention in scientific discovery. In F. Ayala & T. Dobszhansky (Eds.), *Studies in the philosophy of biology: Reduction and related problems* (pp. 139–161). London: Macmillan.
- Caramazza, A., & Mahon, B. Z. (2006). The organisation of conceptual knowledge in the brain: The future's past and some future directions. *Cognitive Neuropsychology*, 23, 13–38.
- Cisek, P. (2006). Integrated neural processes for defining potential actions and deciding between them: A computational model. *The Journal of Neuroscience*, 26, 9761–9770. doi:10.1523/JNEUROSCI.5605-05.2006
- Cleeremans, A. (2008). The radical plasticity thesis. *Progress in Brain Research*, 168, 19–33. doi:10.3389/fpsyg.2011.00086

- Colder, B. (2011). Emulation as an integrating principle for cognition. *Frontiers in Human Neuroscience*, 5, 1–12. doi:10.3389/fnhum.2011.0005
- Collins, A. M., & Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychological Review*, 82, 407–428.
- Corallo, G., Sackur, J., Dehaene, S., & Sigman, M. (2008). Limits on introspection: Distorted subjective time during the dual-task bottleneck. *Psychological Science*, 19, 1110–1117. doi:10.1111/j.1467-9280.2008.02211.x
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87–185.
- Crick, F., & Koch, C. (2003). A framework for consciousness. *Nature Neuroscience*, 6, 119–126. doi:10.1038/nn0203-119
- Dagher, A., Owen, A. M., Boecker, H., & Brooks, D. J. (1999). Mapping the network for planning: A PET activation study with the Tower of London task. *Brain*, 122, 1973–1987. doi:10.1093/brain/122.10.1973
- Dasgupta, S. (2004). Is creativity a Darwinian process? *Creativity Research Journal*, 16, 403–413.
- Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, 122, 371–396.
- Dennett, D. C. (1995). *Darwin's dangerous idea*. New York: Simon & Schuster.
- Dennett, D. C. (2004). Could there be a Darwinian account of human creativity? In A. Moya & E. Font (Eds.), *Evolution, From Molecules to Ecosystems* (pp. 273–279). Oxford: Oxford University Press.
- Diedrichsen, J., Verstynen, T., Hon, A., Zhang, Y., & Ivry, R. B. (2007). Illusions of force perception: The role of sensori-motor predictions, visual information, and motor errors. *Journal of Neurophysiology*, 97, 3305–3313. doi:10.1152/jn.01076.2006
- Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences*, 5, 735–808.
- Dietrich, A. (2004a). Neurocognitive mechanisms underlying the experience of flow. *Consciousness and Cognition*, 13, 746–761. doi:10.1016/j.concog.2004.07.002
- Dietrich, A. (2004b). The cognitive neuroscience of creativity. *Psychonomic Bulletin & Review*, 11, 1011–1026. doi:10.3758/BF03196731
- Dietrich, A. (2007). Who is afraid of a cognitive neuroscience of creativity? *Methods*, 42, 22–27. doi:10.1016/j.ymeth.2006.12.00
- Dietrich, A., & Kanso, R. (2010). A review of EEG, ERP and neuroimaging studies of creativity and insight. *Psychological Bulletin*, 136, 822–848. doi:10.1037/a0019749
- Downing, K. L. (2009). Predictive models in the brain. *Connection Science*, 21, 39–74. doi:10.1080/09540090802610666
- Ellamil, M., Dobson, C., Beeman, M., & Christoff, K. (2012). Evaluative and generative modes of thought during the creative process. *NeuroImage*, 59, 1783–1794.
- Eysenck, H. J. (1993). Creativity and personality: Suggestions for a theory. *Psychological Inquiry*, 4, 147–178.
- Fisher, J. C. (2006). Does simulation theory really involve simulation? *Philosophical Psychology*, 19, 417–432. doi:10.1080/09515080600726377
- Friston, K. J., & Stephan, K. E. (2007). Free-energy and the brain. *Synthese*, 159, 417–458. doi:10.1007/s11229-007-9237-y
- Frith, C. D. (1992). *The cognitive neuropsychology of schizophrenia*. Hove: Lawrence Erlbaum.
- Gansler, D. A., Moore, D. W., Susmaras, T. M., & Jerram, M. W. (2011). Cortical morphology of visual creativity. *Neuropsychologia*, 49(9), 2527–2532. Retrieved from: <http://www.sciencedirect.com/science/article/pii/S002839321100234X>.
- Geary, D. C., & Huffman, K. J. (2002). Brain and cognitive evolution: Forms of modularity and function of mind. *Psychological Bulletin*, 128, 667–698. doi:10.1037//0033-2909.128.5.667
- Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
- Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, 62, 451–482. doi:10.1146/annurev-psych-120709-145346
- Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9, 558–565. doi:10.3758/BF03196313
- Gluck, M., & Myers, C. (1989). *Gateway to learning: An introduction to neural network modeling of the hippocampus and learning*. Norwell, MA: Kluwer Academic Publishers.
- Goldman, A. (1995). Interpretation psychologized. In M. Davis & T. Stone (Eds.), *Folk Psychology: The theory of mind debate* (pp. 74–99). Cambridge: Blackwell.
- Green, A. E., Kraemer, D. J. M., Fugelsang, J. A., Gray, J. R., & Dunbar, K. N. (2012). Neural correlates of creativity in analogical reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(2), 264–272. doi:10.1037/a0025764
- Gross, H., Heinze, A., Seiler, T., & Stephan, V. (1999). Generative character of perception: A neural architecture for sensorimotor anticipation. *Neural Networks*, 12, 1101–1129. doi:10.1016/S0893-6080(99)00047-7
- Grush, R. (1997). The architecture of representation. *Philosophical Psychology*, 10, 5–23.
- Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27, 377–396. doi:10.1017/S0140525X04000093
- Haider, H., Eichler, A., & Lange, T. (2011). An old problem: How can we distinguish between conscious and unconscious knowledge acquired in an implicit learning task. *Consciousness and Cognition*, 20, 658–672. doi:10.1016/j.concog.2010.10.021
- Hassabis, D., & Maguire, E. A. (2009). The construction system of the brain. *Philosophical Transactions of the Royal Society B*, 364, 1263–1271. doi:10.1098/rstb.2008.0296
- Hawkins, J. (2004). *On intelligence*. New York: Henry Holt and Company.
- Helie, S., & Sun, R. (2010). Incubation, insight, and creative problem solving: A unified theory and a connectionist model. *Psychological Review*, 117, 994–1024. doi:10.1037/a001953
- Hesslow, G. (2002). Conscious thought as simulation of behaviour and perception. *Trends in Cognitive Sciences*, 6, 242–247. doi:10.1016/S1364-6613(02)01913-7
- Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, 109, 679–708. doi:10.1037/0033-295X.109.4.679j
- Hommel, B. (2009). Action control according to TEC (theory of event coding). *Psychological Research*, 73, 512–526. doi:10.1007/s00426-009-0234-2
- Hommel, B., & Eglau, B. (2002). Control of stimulus-response translation in dual-task performance. *Psychological Research*, 66, 260–273. doi:10.1007/s00426-002-0100-y
- Hommel, B., Musseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, 24, 849–878.
- Houk, J., Adams, J., & Barto, A. (1995). A model of how the basal ganglia generate and use neural signals that predict reinforcement. In J. Houk, J. Davis, & D. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 249–270). Cambridge, MA: MIT Press.
- Humphrey, N. (2002). *The mind made flesh: Frontiers of psychology and evolution*. Oxford: Oxford University Press.
- Ingold, T. (2000). The poverty of selectionism. *Anthropology Today*, 16, 1–2. doi:10.1111/1467-8322.00022
- Ingold, T. (2007). The trouble with 'evolutionary biology'. *Anthropology Today*, 23, 13–17. doi:10.1111/j.1467-8322.2007.00497.x
- James, W. (1890). *Principles of Psychology*. New York: Holt.

- Jauk, E., Benedek, M., & Neubauer, A. C. (2012). Tackling creativity at its roots: Evidence for different patterns of EEG α activity related to convergent and divergent modes of task processing. *International Journal of Psychophysiology*, *84*(2), 219–225. doi:10.1016/j.ijpsycho.2012.02.012
- Johnson-Laird, P. N., & Byrne, R. M. J. (2009). “If” and the problem of conditional reasoning. *Trends in Cognitive Sciences*, *13*, 282–287. doi:10.1016/j.tics.209.04.003
- Jung, R. E., Mead, B. S., Carrasco, J., & Flores, R. A. (2013). The structure of creative cognition in the human brain. *Frontiers in Human Neuroscience*, *7*, 330.
- Kawato, M., & Wolpert, D. (1998). Internal models for motor control. *Sensory Guidance of Movement*, *218*, 291–307. doi:10.1016/S0959-4388(99)00028-8
- Kent, C., & Lamberts, K. (2008). The encoding–retrieval relationship: retrieval as mental simulation. *Trends in Cognitive Sciences*, *12*, 92–98. doi:10.1016/j.tics.2007.12.004
- Kiefer, M., Sim, E.-J., Herrnberger, B., Grothe, J., & Hoenig, K. (2008). The sound of concepts: Four markers for a link between auditory and conceptual brain systems. *The Journal of Neuroscience*, *28*, 12224–12230.
- Köhler, I. (1951). *Über Aufbau und Wandlungen der Wahrnehmungswelt. Insbesondere über bedingte Empfindungen*. Wien: Rohrer.
- Kronfeldner, M. E. (2010). Darwinian “blind” hypothesis formation revisited. *Synthese*, *175*, 193–218. doi:10.1007/s11229-009-9498
- Lakoff, G., & Johnson, M. (1999). *Philosophy in the Flesh*. New York: Basic Books.
- Lau, H. (2008). A higher order Bayesian decision theory of consciousness. In R. Banerjee & B.K. Chakrabarti (Eds.), *Progress in Brain Science*, *168*, 37–48.
- Lewontin, R. C. (1970). The units of selection. *Annual Review of Ecology and Systematics*, *1*, 1–18.
- Llinás, R. R. (2001). *I of the vortex: From neurons to self*. Boston: MIT Press.
- Llinás, R. R., & Roy, S. (2009). The ‘prediction imperative’ as the basis for self-awareness. *Philosophical Transactions of the Royal Society B*, *364*, 1301–1307. doi:10.1098/rstb.2008.0309
- Marr, D. (1982). *Vision: a computational approach*. San Francisco, CA: Freeman & Company.
- Martindale, C. (1990). *The clockwork muse: The predictability of artistic change*. New York: Basic Books.
- Martindale, C. (1999). Darwinian, Lamarckian, and rational creativity. *Psychological Inquiry*, *10*, 340–341.
- Mehta, B., & Schaal, S. (2002). Forward models in visuomotor control. *Journal of Neurophysiology*, *88*, 942–953.
- Mesoudi, A., Whiten, A., & Laland, K. N. (2006). Towards a unified science of cultural evolution. *Behavioural and Brain Sciences*, *29*, 329–383. doi:10.1017/S0140525X06009083
- Meyer, W. U., Reisenzein, R., & Schützwohl, A. (1997). Toward a process analysis of emotions: The case of surprise. *Motivation and Emotion*, *21*, 251–274. doi:10.1023/A:1024422330338
- Michelson, A. A., & Morley, E. W. (1887). On the relative motion of the earth and the luminiferous ether. *American Journal of Science*, *34*, 333–345.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202. doi:10.1146/annurev.neuro.24.1.167
- Moulton, S. T., & Kosslyn, S. M. (2009). Imagining predictions: Mental imagery as mental emulation. *Philosophical Transactions of the Royal Society B*, *364*, 1273–1280. doi:10.1098/rstb.2008.0314
- Mumford, M. D. (1999). Blind variation or selective variation? Evaluative elements in creative thought. *Psychological Inquiry*, *10*, 344–348.
- Niedenthal, P. M., Barsalou, L. W., Winkelman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review*, *9*, 184–211. doi:10.1207/s15327957pspr0903_1
- O’Reagan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioural and Brain Sciences*, *24*, 939–1031.
- Perkins, D. N. (1994). Creativity: Beyond the Darwinian paradigm. In M. A. Boden (Ed.), *Dimensions of creativity* (pp. 119–142). Cambridge, MA: MIT Press.
- Perruchet, P., Cleeremans, A., & Destrebecqz, A. (2006). Dissociating the effects of automatic activation and explicit expectancy on reaction times in a simple associative learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 955–965.
- Perruchet, P., & Vinter, A. (2002). The self-organizing consciousness. *Behavioural and Brain Sciences*, *25*, 297–388. doi:10.1037/0278-7393.25.2.955
- Pezzulo, G., Butz, M. V., Castelfranchi, C., & Falcone, R. (Eds.). (2008). Anticipation in natural and artificial cognition. In G. Pezzulo, M. Butz, C. Castelfranchi, *The challenge of anticipation: A unifying framework for the analysis and design of artificial cognitive systems*, (pp.3-22). New York: Springer.
- Pinker, S. (2002). *The blank slate*. New York: Penguin.
- Popper, K. R. (1972). *Objective knowledge: An evolutionary approach*. Oxford: Clarendon Press.
- Popper, K. R. (1974). Campbell on the evolutionary theory of knowledge. In P. A. Schilpp (Ed.), *The philosophy of Karl Popper* (pp. 1059–1065). LaSalle: Open Court.
- Popper, K. R. (1984). Evolutionary Epistemology. In J. W. Pollard (Ed.), *Evolutionary theory: Paths into the future* (pp. 239–255). New York: Wiley.
- Prinz, W. (1987). Ideo-motor action. In H. Heur & A. F. Sanders (Eds.), *Perspectives on perception and action* (pp. 47–76). Hillsdale, NJ: Erlbaum.
- Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, *9*, 129–154.
- Ramachandran, V. S. (1992). Blind spots. *Scientific American*, *266*, 86–91.
- Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*, 79–87. doi:10.1038/4580
- Read, D. W. (2006). Cultural evolution is not equivalent to Darwinian evolution. *Behavioural and Brain Sciences*, *29*, 361. doi:10.1017/S0140525X0638908X
- Reber, A. S. (1996). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. Oxford: Oxford University Press.
- Reed, C. L., & Farah, M. J. (1995). The psychological reality of the body schema: A test with normal participants. *Journal of Experimental Psychology: Human Perception & Performance*, *21*, 334–343. doi:10.1037/0096-1523.21.2.334
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical Conditioning II: Current research and theory* (pp. 64–99). New York: Appleton Century Crofts.
- Richerson, P. J., & Boyd, R. (2005). *Not by genes alone: How culture transformed human evolution*. Chicago: University of Chicago Press.
- Rosenthal, D. (2005). *Consciousness and Mind*. New York: Oxford University Press.
- Russ, S. W. (1999). An evolutionary model for creativity: Does it fit? *Psychological Inquiry*, *10*, 359–361.
- Ruz, M., & Lupiáñez Castillo, J. (2002). A review of attentional capture: On its automaticity and sensitivity to endogenous control. *Psicológica: Revista de Metodología y Psicología Experimental*, *23*, 283–310.
- Sawyer, K. (2011). The cognitive neuroscience of creativity: A critical review. *Creativity Research Journal*, *23*, 137–154.

- Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philosophical Transactions of the Royal Society B*, *362*, 773–786. doi:10.1098/rstb.2007.2087
- Schacter, D. L., & Buckner, R. L. (1998). On the relationship among priming, conscious recollection, and intentional retrieval: Evidence from neuroimaging research. *Neurobiology of Learning and Memory*, *70*, 284–303.
- Schooler, J. W., & Dougal, S. (1999). Why creativity is not like the proverbial typing monkey. *Psychological Inquiry*, *10*, 351–356.
- Schubotz, R. I. (2007). Prediction of external events with our motor system: Towards a new framework. *Trends in Cognitive Sciences*, *11*, 211–218. doi:10.1016/j.tics.2007.02.006
- Schultz, W. (2000). Multiple reward signals in the brain. *Neuroscience*, *1*, 199–207. doi:10.1038/35044563
- Shanks, D. R. (2005). Implicit Learning. In K. Lamberts & R. Goldstone (Eds.), *Handbook of Cognition* (pp. 202–220). London: Sage Publication.
- Shapiro, L. (2007). The embodied research program. *Philosophy Compass*, *2*(2), 338–346. doi:10.1111/j.1747-9991.2007.00064.x
- Shen, W. B., Liu, C., & Chen, J. J. (2010). Neural basis of creativity: Evidence from structural and functional imaging. *Advances in Psychological Science*. Retrieved from: http://en.cnki.com.cn/Article_en/CJFDTOTAL-XLXD201009009.htm
- Shin, Y. K., Proctor, R. W., & Capaldi, E. J. (2010). A review of contemporary ideomotor theory. *Psychological Bulletin*, *136*, 943–974. doi:10.1037/a0020541
- Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst – sustained inattention blindness for dynamic events. *Perception*, *28*, 1059–1074. doi:10.1068/p2952
- Simonton, D. K. (1997). Creative productivity: A predictive and explanatory model of career trajectories and landmarks. *Psychological Review*, *104*, 66–89.
- Simonton, D. K. (1999). Creativity as blind variation and selective retention: is the creative process Darwinian? *Psychological Inquiry*, *10*, 309–328.
- Simonton, D. K. (2003). Scientific creativity as constrained stochastic behavior: The integration of process and person perspectives. *Psychological Bulletin*, *129*, 475–494.
- Simonton, D. K. (2007). The creative process in Picasso's Guernica sketches: Monotonic improvements or nonmonotonic variants. *Creativity Research Journal*, *19*, 329–344. doi:10.1080/10400410701753291
- Sio, U. N., & Ormerod, T. C. (2009). Does incubation enhance problem solving: A meta-analytic review. *Psychological Bulletin*, *135*, 94–120. doi:10.1037/a0014212
- Sternberg, R. J. (1998). Cognitive mechanisms in human creativity: Is variation blind or sighted? *Journal of Creative Behavior*, *32*, 159–176. doi:10.1002/j.2162-6057.1998.tb00813.x
- Sternberg, R. J. (1999). Darwinian creativity as a conventional religious faith. *Psychological Inquiry*, *10*, 357–359.
- Takeuchi, H., Taki, Y., Hashizume, H., Sassa, Y., Nagase, T., Nouchi, R., & Kawashima, R. (2011). Failing to deactivate: The association between brain activity during a working memory task and creativity. *NeuroImage*, *55*(2), 681–687. doi:10.1016/j
- Taylor, S. F., Stern, E. R., & Gehring, W. J. (2007). Neural systems for error monitoring: Recent findings and theoretical perspectives. *The Neuroscientist*, *13*, 162–172. doi:10.1177/1073858406298184
- Tennebaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure and abstraction. *Science*, *331*, 1279–1285. doi:10.1126/science.1192788
- Thomas, L. E., & Lleras, A. (2009). Swinging into thought: Directed movement guides insight in problem solving. *Psychonomic Bulletin & Review*, *16*, 719–723.
- Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.
- Ward, T. B., Smith, S. M., & Finke, R. A. (1999). Creative cognition. In R. J. Sternberg (Ed.), *Handbook of Creativity*, (pp. 189–212). Cambridge University Press.
- Wegner, D. M. (2003). The mind's best trick: How we experience conscious will. *Trends in Cognitive Sciences*, *7*, 65–69. doi:10.1016/S1364-6613(03)00002-0
- Whittlesea, B. W. A. (2002). False memory and the discrepancy-attribution hypothesis: The prototype-familiarity illusion. *Journal of Experimental Psychology: General*, *131*, 96–115.
- Wilbert, J., & Haider, H. (2010). The subjective experience of committed errors and the discrepancy-attribution hypothesis. *Acta Psychologica*, *139*, 370–381. doi:10.1016/j.actpsy.2011.11.010
- Willingham, D. B. (1999). Implicit motor sequence learning is not purely perceptual. *Memory Cognition*, *27*, 561–572. doi:10.3758/BF03211549
- Wilson, A. D., & Golonka, S. (2013). Embodied cognition is not what you think it is. *Frontiers in Psychology: Cognitive Science*, *4*, 1–12.
- Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society Series B*, *358*, 593–602. doi:10.1098/rstb.2002.1238
- Wolpert, D. M., & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature Neuroscience*, *3*, 1212–1228.
- Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, *269*, 1880–1882. doi:10.1126/science.7569931