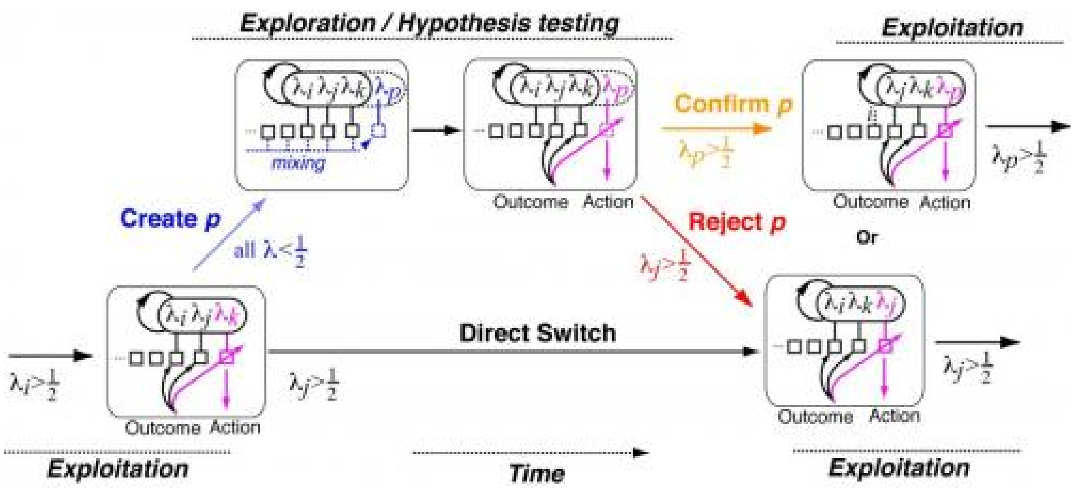
Upfront and personal: Scientists model human reasoning in the brain's prefrontal cortex

June 27 2014, by Stuart Mason Dambrot



A model of human reasoning. Solid squares, behavioral strategies stored in long-term memory. λi, λj, λk, λp denote absolute reliabilities of monitored strategies inferred from action outcomes (here, the inferential capacity is three). Purple, actor strategy learning external contingencies and selecting action maximizing rewards. In exploitation periods, the actor is reliable (i.e., λactor > 1 – λactor or λactor > 1/2), the others being necessary unreliable (because Σλ. ≤ 1). Otherwise, the system switches into exploration (all λ

(Medical Xpress)—Located at the forward end of the brain's frontal lobe, the mammalian prefrontal cortex (PFC) is the seat of many of our most unique cognitive abilities – collectively referred to as *executive function* – including planning, decision-making, and coordinating thoughts and actions with internal goals. That said, perhaps its most important attribute – one that is apparently unique to *H. sapiens* – is reasoning which, based on Bayesian, or probabilistic, inference, mitigates uncertainty by informing adaptive behavior. While the structural details of this remarkable process have historically remained elusive, scientists at Institut National de la Santé et de la Recherche Médicale, Paris, and Ecole Normale Supérieure, Paris and Université Pierre et Marie Curie, Paris have recently employed computational modeling and neuroimaging to show that the human prefrontal cortex involves two interactive reasoning pathways that embody hypothesis testing for evaluating, accepting and rejecting behavioral strategies. More specifically, their model describes behavior guided by reason in the form of an online algorithm combining Bayesian inference applied to multiple stored strategies with hypothesis testing that can update these strategies. In addition – as proposed in a previous work[1] – the scientists conclude that since the frontopolar cortex (FPC), located in the anterior-most portion of the frontal lobes, is human-specific and is a key component in executive function decision-making, the ability to make inferences on concurrent strategies and decide to switch directly to one of these alternative strategies is unique to humans as well.

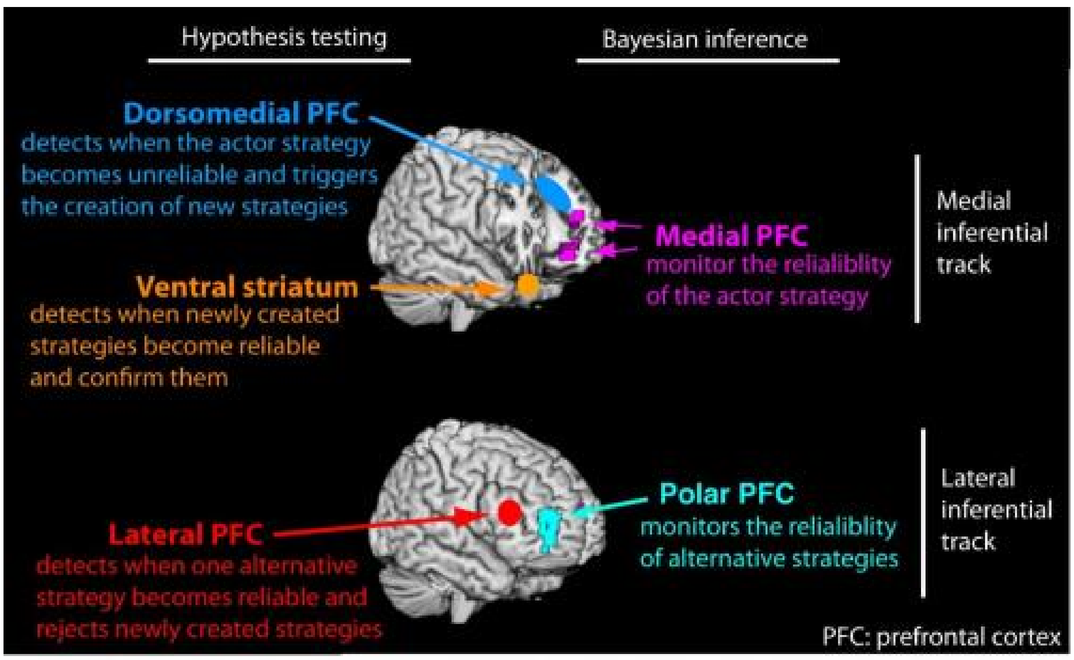
Prof. Etienne Koechlin discussed the paper that he, Dr. Maël Donoso and Dr. Anne G. E. Collins published in *Science*. "The main challenge using computational modeling and neuroimaging to show that the human PFC comprises two concurrent inferential tracks was to identify not only the strategies that guide subjects' behavior, but especially the alternative strategies which subjects may have in mind and might switch to," Koechlin tells *Medical Xpress*. "These alternative strategies, in essence, cannot be directly derived from observing subjects' behavior, but require a computational model describing how subjects create, adjust, monitor, replace, store and retrieve behavioral strategies." The problem, he points out, is that these issues are interdependent and cannot be solved separately – so the scientists needed a model that solves all these issues simultaneously. "Empirically," he says, "the challenge was to find a protocol rich enough to induce subjects to reason and develop multiple behavioral strategies, but simple enough to be tractable both for human subjects and for computational modeling."

Moreover, Koechlin continues, these two concurrent inferential tracks interact and – along with the striatum, a subcortical part of the forebrain involved in coordinating motivation with somatic movement – implement hypothesis testing for accepting or rejecting newly-created strategies. "If subjects use hypothesis testing to assess the significance of creating new strategies rather than simply adjusting previously learned ones, then there are specific time points when they decide to create a new strategy and – based on subsequent external evidence – to confirm or reject this new strategy," he explains. "These events occur online in the subjects' minds; are more dependent on their beliefs than protocol parameters; are not directly observable; and, most problematically, are rare." As a result, Koechlin notes, the challenge was to have an accurate model that could predict *precisely* when these events should occur in subjects' minds. In their experimental protocol, he illustrates, these events occurred about 30 times in a series of 1,500 trials that every subject performed – a rate at the limit of neuroimaging statistical power. The scientists therefore decided to test 40 subjects in two sessions, which Koechlin points out is about twice the standard number of subjects and sessions tested in neuroimaging experiments.

In devising a model that describes the process of human reasoning guiding behavior as a computationally tractable, online algorithm approximating Dirichlet process mixtures, Koechlin says that, in essence, the challenge was to imagine how human reasoning actually works, knowing that we are quite efficient in real-life environments. (A *Dirichlet process* is a probability distribution whose domain is itself a set of probability distributions, while *Dirichlet process mixtures* are optimal inferential processes that use Dirichlet processes to adapt to uncertain, variable and open-ended environments – as, Koechlin points out, we face in real life – by solving data point distribution problems that arise when it is not possible to determine *a priori* the number of clusters that generated the data.)

"These environments are very challenging," Koechlin notes, "because at any time, new never-experienced situations may occur. At those times, we need the ability to create, explore and learn new behavioral strategies, to exploit whenever appropriate the strategies you learned in previous situations, and – most importantly – to have an ability to understand when to explore and when to exploit." While this a difficult problem that Dirichlet processes theoretically solve, he adds, these processes are computationally intractable because the computing and memory requirements grow exponentially with time – meaning that these adaptive processes are certainly biologically implausible and cannot model human reasoning.

The scientists posited that reasoning has evolved under biologically-strong computing constraints, as well as by probably capturing some key features of Dirichlet processes. In so doing, they assumed two key biological constraints – the inability to (1) anachronistically revise past decisions about new strategy creation, this being a Dirichlet process component, and (2) make inferences concurrently on an unlimited number of behavioral strategies,



Reasoning processes in the prefrontal cortex. Credit: Etienne Koechlin

"Regarding first constraint, we say that mental inferences are online and forward – meaning that we infer from the past what should be done next, but we do not change our past decisions from the present," Koechlin explains. "The second constraint represents our inferential capacity." At the same time, he continues, human reasoning should capture the key feature of Dirichlet processes – namely, the ability to revise online the decision to create a new strategy. "This flexibility is crucial for preserving our limited inferential capacity and for dealing with the all-and-none nature of strategy creation. Hypothesis testing enables this flexibility, "Creating a new strategy is like setting a new behavioral hypothesis, which can be subsequently confirmed or rejected on the basis of new information."

As a result, the scientists ended up with the idea that human reasoning should combine forward Bayesian inferences on a limited number of concurrent behavioral strategies with hypothesis testing for possibly updating this inferential buffer with new strategies created from long-term memory. "However," Koechlin acknowledges, "this idea raised a new problem we had to address – that is, Bayesian inferences and hypothesis testing are somewhat incompatible processes, and in fact correspond to two radically different approaches in inferential statistics. Specifically, the former is usually used by theoreticians to compare models, whereas latter is commonly used in empirical sciences through T-tests, F-tests and so on."

The key insight the scientists used to address these challenges, Koechlin tells *Medical Xpress*, was to solve how to combine online Bayesian inference and hypothesis testing. "We propose a computational solution based on the idea that humans make absolute rather than relative judgments: this so-called *prefrontal algorithm* infers the absolute reliability of every monitored strategy – that is, to what extent the strategy is applicable or relevant to the current situation, given that possibly none could be applicable." This is equivalent, he says, to monitoring how likely the current external events and contingencies match those the strategy has already learned: If matching is more likely, then the strategy is reliable (or applicable or relevant) – but if non-matching is more likely, then the strategy is unreliable (or irrelevant).

"The algorithm then becomes simple," he explains. "As long as one monitored strategy is reliable – the other being necessarily unreliable – this strategy drives behavior, adjusting to learn external contingencies for maximizing rewards. On the other hand, if none are reliable, then hypothesis testing starts – and a new behavioral strategy is created from long-term memory for driving behavior." Initially unreliable, this new strategy learns and may become reliable, at which point it is confirmed and consolidated in long term memory, provided that the other monitored strategies remain unreliable. Conversely, the new strategy may remain unreliable, while one monitored strategy becomes reliable, so that, the latter is retrieved to drive behavior and the new strategy is rejected or disbanded as an unnecessary creation. "The algorithm thus predicts the occurrence of specific transitional, highly nonlinear events associated with hypothesis testing. Our results provide evidence that the prefrontal cortex implements this solution."

Their other important (and related) insight was to understand how new strategies are created from long-term memory, which Koechlin describes as "basically a weighted mixture of previous learned strategies stored in long-term memory, weighted by internal representations storing strategies reliability according to contextual cues." He repeats that the process is simple, adding that the overall algorithm has the important property of building a potentially unlimited repertoire of behavioral strategies online – a repertoire with sampling properties close to optimal Dirichlet processes models.

"Thus," he continues, "an important new insight of our study is to show how inferential and creative processes are tightly linked in the human prefrontal cortex. Another important insight is that the proposed algorithm provides a unified view of how the prefrontal executive function works – or equivalently, how the network of prefrontal regions including the ventromedial, dorsomedial, lateral and polar prefrontal regions forms an unified executive system that develops tractable inferences for guiding adaptive behavior and efficiently drive action in uncertain, variable and open-ended environments."

An interesting aspect of the paper is the explanation of how human reasoning involving Bayesian inference accounts for human responses deviating from formal logic. "Here's an example," Koechlin illustrates. "Assume you state that all birds are green, then that an unknown animal is green, and finally ask whether the unknown animal is a bird. The majority of human subjects probably respond yes, even though this response is of course wrong from the viewpoint of formal logic – obviously, animals other than birds can be green. However, the response is appropriate from a probabilistic inference viewpoint, since the green animal is more likely a bird than another animal. In other words, the response conforms to probabilistic inference principles."

The paper also explores the effect of inferential computational complexity problems on the evolution of higher cognitive functions. "Our findings show that the human prefrontal cortex has found a simple, approximate solution to make inferences in uncertain, variable and open-ended environments and consequently, arbitrates efficiently between staying with the same strategy by possibly adjusting it, switching to another strategy or creating new strategies for driving behavior," Koechlin tells *Medical Xpress*. "Anterior prefrontal regions make probabilistic inferences on the reliability or relevance of multiple – but not more than three or four at the same time – concurrent strategies, while posterior prefrontal regions implements hypothesis testing for possibly updating this anterior prefrontal inferential buffer with new strategies. These posterior prefrontal regions then make exclusive make true/false judgments of monitored strategies for deciding to create, confirm or reject new strategies."

In essence, then, the scientists' prefrontal algorithm arbitrates between staying with the ongoing behavioral strategy and possibly learning external contingencies for maximizing rewards, switching to other learned strategies, and forming new behavioral strategies. To achieve this, Koechlin points out, the ventromedial prefrontal cortex computes the reliability of the strategy driving ongoing behavior, while the dorsomedial prefrontal cortex detects when this strategy becomes unreliable for triggering the creation of a new behavioral strategy. This contrasts with the frontopolar prefrontal cortex, which concurrently computes the reliability of two or three alternative strategies, while the lateral prefrontal cortex detects when one among these alternatives become reliable for retrieving it to guide behavior. In that case, the newly created strategy is disbanded. In the converse case when the newly created strategy becomes reliable, while the alternatives remain unreliable, the ventral striatum reinforces and consolidates the new strategy in long-term memory.

Finally, in terms of the ability of humans alone to perform this inference-and-testing process, Koechlin points out that non-human primates do not have a prefrontal cortex region unique to humans known as the frontopolar cortex (FPC) – meaning that they can only infer whether to stay with the current strategy or to explore a new one created from long-term memory. "Alternatively, rodents have no lateral prefrontal cortex, so that they can only infer and decide to stay or to switch reactively after acting and experiencing action outcomes. Primates *with* the lateral prefrontal cortex can do that proactively before acting – but thanks to the frontopolar cortex, humans also make inferences on concurrent strategies and can infer and decide to switch directly to one of these alternative strategies."

Moving forward, Koechlin tells *Medical Xpress*, the scientists are conducting an experiment to investigate how inferential and creative processes are modulated by the context in which the person is acting. "We're also conducting some intracranial electrophysiological recordings in humans," he adds, "to better understand the neuronal dynamics of events associated with hypothesis-testing, exploration and exploitation behavior." Koechlin also says that their results have implications primarily in neuroscience and psychology, but also in robotics and artificial intelligence, and that they can help to understand dysexecutive syndrome (DES) – a group of usually simultaneous cognitive, behavioral and emotional symptoms, usually resulting from brain damage – in neuropsychiatric patients.

More information: Foundations of human reasoning in the prefrontal cortex, *Science*, Published Online May 29 2014, doi:10.1126/science.1252254

Related:
[1]Anterior Prefrontal Function and the Limits of Human Decision-Making, *Science* 26 October 2007: Vol. 318 no. 5850 pp. 594-598, doi:10.1126/science.1142995