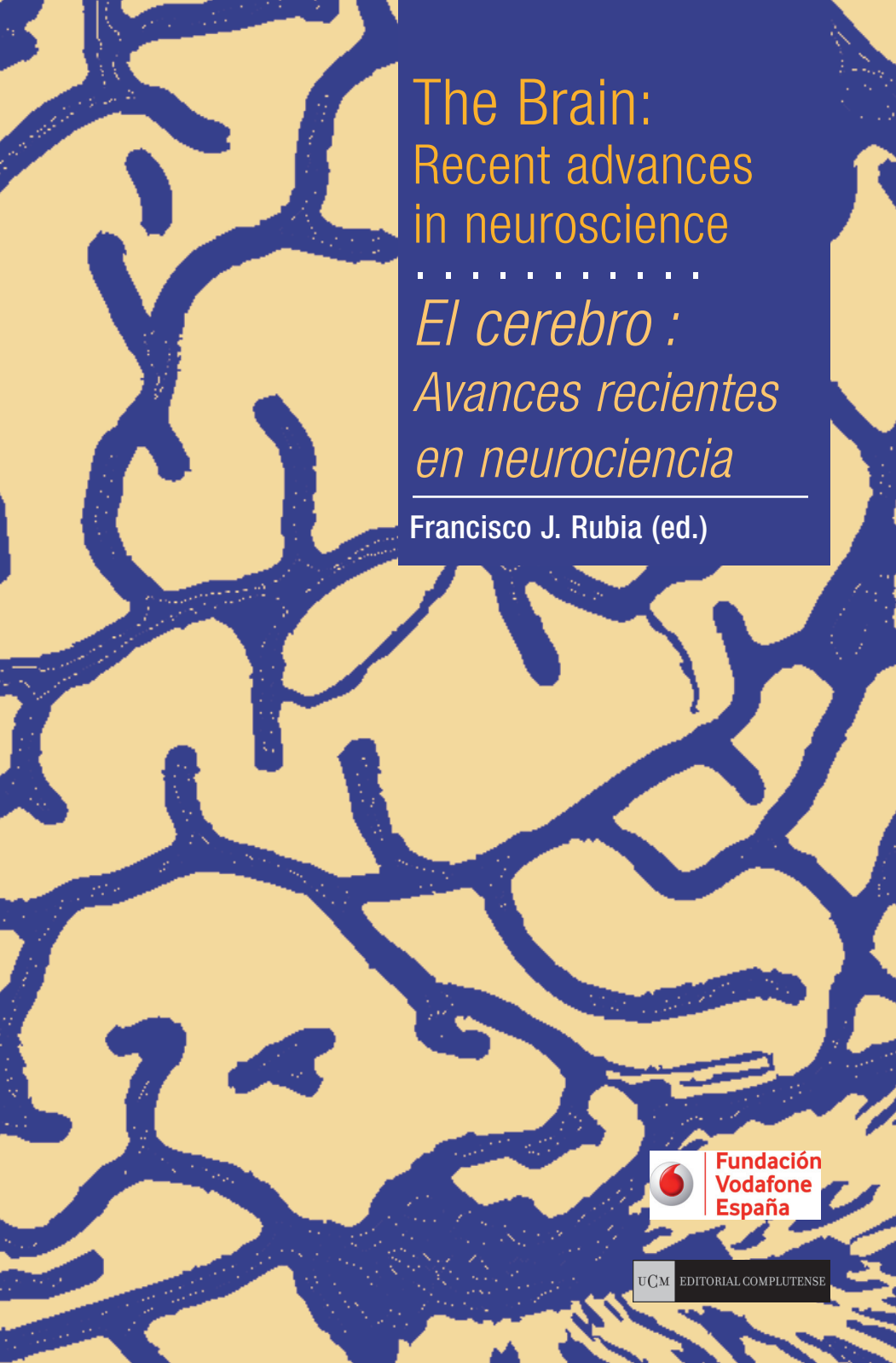


The Brain:
Recent advances in neuroscience

El cerebro :
Avances recientes en neurociencia

Francisco J. Rubia (ed.)

Fundación Vodafone España

UCM EDITORIAL COMPLUTENSE

THE BRAIN: RECENT ADVANCES IN NEUROSCIENCE

EL CEREBRO: AVANCES RECIENTES EN NEUROCIENCIA

EDITOR

FRANCISCO J. RUBIA

AUTORES

GERHARD ROTH

THOMAS F. MÜNTE

MICHAEL PAUEN

MARC JEANNEROD

DAVID PERRETT

UCM EDITORIAL COMPLUTENSE

Esta obra ha sido editada gracias a la colaboración de la Fundación Vodafone de España.

Queda rigurosamente prohibida sin la autorización escrita de los titulares del Copyright, bajo las sanciones establecidas en las leyes, la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la reprografía y el tratamiento informático, y la distribución de ejemplares de ella mediante alquiler o préstamo público.

Todos los libros publicados por Editorial Complutense a partir de enero de 2007 han superado el proceso de evaluación experta.

© 2009 *by* Francisco J. Rubia de la edición y los autores de sus textos

© 2009 *by* Editorial Complutense, S. A.
Donoso Cortés, 63 - 28015 Madrid
Tels.: 91 394 64 60/61. Fax: 91 394 64 58
ecsa@rect.ucm.es
www.editorialcomplutense.com

ISBN: 978-84-7491-950-9
Depósito legal: M-9727-2009

Primera edición: Marzo de 2009

Fotocomposición: MCF Textos, S. A.
Impresión: Top Printer Plus

Impreso en España - *Printed in Spain*

ÍNDICE

Prólogo 9

THE BRAIN: RECENT ADVANCES IN NEUROSCIENCE

Francisco J. RUBIA: Introductory Remarks 15

Gerhard ROTH: The Relationship between Reason and Emotion and its Impact for the Concept of Free Will 21

Thomas F. MÜNTE: A Neuroscience Look on Human Action Monitoring 37

Michael PAUEN: Human Self-Understanding, Neuroscience, and Free Will: A Revolution Ahead? 51

Marc JEANNEROD: Being Oneself: the Neural Basis of Self-Consciousness 67

David PERRETT: Using Visual Information to Understand and Predict What Comes Next 79

EL CEREBRO: AVANCES RECIENTES EN NEUROCIENCIA

Francisco J. RUBIA: Comentarios introductorios 97

Gerhard ROTH: La relación entre la razón y la emoción y su impacto sobre el concepto de libre albedrío 103

Thomas F. MÜNTE: Una aproximación desde la neurociencia a la monitorización de las funciones ejecutivas 119

Michael PAUEN: Autocomprensión humana, neurociencia, y libre albedrío: ¿se anticipa una revolución?..... 135

Marc JEANNEROD: Ser uno mismo: la base neuronal de la autoconciencia 153

David PERRETT: El uso de información visual para comprender y predecir lo que va a ocurrir..... 165

Sobre los autores 183

Quisiera dar las gracias a todos aquellos que han participado en la organización de este simposio, particularmente a Eva Moreno y Mar Santos, así como al personal de administración y servicios del Instituto Pluridisciplinar. También quiero dar las gracias a todos los que participaron en el simposio, y muy especialmente a la Fundación Vodafone España, representada por su director general, José Luis Ripoll, y sus colaboradores Javier del Arco y Mercedes Gómez.

Prólogo

José Luis Ripoll

Me es muy grato presentar un nuevo trabajo, apoyado por la Fundación Vodafone España, como es esta publicación, que difunde las ponencias que un panel internacional de expertos, de primer nivel, expusieron en el seminario «El cerebro: avances en neurociencia».

Bajo la coordinación del profesor Francisco Rubia y en colaboración con el Instituto Pluridisciplinar de la Universidad Complutense de Madrid, que tan sabiamente dirige, el simposio puso de manifiesto que existe un elevado potencial para crear vínculos importantes entre los dominios de las neurociencias y aquellos propios de las tecnologías de la información y de las comunicaciones (TIC).

Si algo han certificado las aportaciones de los distintos ponentes (que ahora amplían su eco gracias a esta edición) es la importancia de la comunicación coordinada y de la capacidad de interacción de la información para la generación eficiente de conocimiento o inteligencia.

Para conocer esta interacción, es fundamental encontrar las bases de la estructuración del conocimiento, capaz de generar economías de escala a nivel exponencial, como describen las neurociencias en el funcionamiento de la mente humana.

Los debates que se han suscitado en este simposio han puesto de manifiesto, entre otras cosas, que la combinación de los datos obtenidos en la investigación del cerebro van a sugerir un modelo computacional que defina la operatividad parcial del mismo y conjuntamente su funcionalidad como sistema único, lo que provoca-

ría un alto impacto en las tecnologías de computación, comunicación e información. El flujo bidireccional de ésta influirá en los productos y el funcionamiento tanto de la tecnología *hardware* como *software,* e impulsará, sin lugar a duda, los campos de la robótica y de la inteligencia artificial, entre otros.

Existe un alto potencial de creación de sistemas de retroalimentación positivos entre los dominios de la neuroinformática y las TIC, creando una única sinergia. La primera estimulará, según escuchamos, desarrollos importantes en los campos de la ingeniería neuromórfica o la biónica.

Hemos visto con evidente interés la presentación de distintas iniciativas que persiguen objetivos similares para el desarrollo del funcionamiento neuronal en las citadas TIC, como es el caso de la computación autónoma, que busca construir una nueva generación de tecnologías de la información autorreparable, autogestionable y autorregulable, análoga a la que presentan los organismos vivos.

Ciencias interdisciplinares como la neuroinformática o, en general, la bioinformática persiguen acelerar el progreso de comprensión del funcionamiento del cerebro, situándose en la intersección de la medicina, la biología, la psicología, la física, la computación, las matemáticas y la ingeniería, para generar aplicaciones que permitan el desarrollo de sistemas artificiales, que implementen los tipos de computación de procesamiento cerebral.

La convergencia de la nanotecnología, la biotecnología, las TIC y las neurociencias permite acelerar la mejora evolutiva en el aprendizaje, en la comunicación externa a la persona y en el interfaz hombre-máquina, así como posibilitar mejoras internas de la persona. ¿Qué aportará esta convergencia al dominio del aprendizaje? ¿Desarrollaremos nuevos métodos de aprendizaje virtuales? ¿Podremos obtener un mejor entendimiento de las capacidades y del funcionamiento del cerebro a través del análisis cartográfico de células?…

Estos y otros retos cruciales de nuestro tiempo se han tratado en este simposio y en esta publicación, ambos espléndidamente concebidos y dirigidos por el ilustre profesor Francisco J. Rubia Vila, catedrático emérito de la Facultad de Medicina de la Universidad

Complutense de Madrid y catedrático de la Universidad Ludwig Maximillian de Múnich. Su especialidad es la Fisiología del Sistema Nervioso, campo en el que tiene más de doscientas publicaciones como director de la Unidad de Cartografía Cerebral del Instituto Pluridisciplinar de la Universidad Complutense de Madrid. También es miembro numerario de la Real Academia Nacional de Medicina y vicepresidente europeo de la Delegación Española de la Academia Europea de Ciencias y Artes.

Enhorabuena, pues, a todos los conferenciantes y, especialmente, al profesor Rubia, por haber sabido dirigir el simposio y potenciar la realización de esta publicación, que, estoy seguro, despertará el interés de todos sus lectores.

Profesor José Luis Ripoll
Presidente
Fundación Vodafone España

THE BRAIN: RECENT ADVANCES IN NEUROSCIENCE

Introductory Remarks

Francisco J. Rubia

In a small piece, under the title of *A Difficulty of Psychoanalysis*, Sigmund Freud argued that human beings had suffered throughout history three important humiliations in their pride.

That of Nicolaus Copernicus, who did away with geo-centrism: that is, with the idea that the Earth was the centre of the Universe and of Creation. Earth was just a planet and not even one of the most important circling the Sun. Nowadays that idea has not only been confirmed but we also know that the Sun is merely one of the millions of suns that form one of the multiple galaxies of the universe, thus the importance of the Earth has been diminishing by leaps and bounds.

The second humiliation comes by way of biologist Charles Darwin with his Theory of Evolution, about which nowadays no one has any doubts, except a few Christian Creationist sects in the United States. And even after almost 150 years —because the theory, as you know, was published in 1859— there are still people around who have not digested the significance of this: that is, that our ancestry is from animals that have preceded us in the evolution ladder. This undoubtedly represented a hard blow for the idea that we were the jewel of Divine Creation and that we were created all of a sudden by a breath of divinity, as explained to us in Genesis. With this, the explanation given to us in the Bible became what it is: a myth or legend like many others. This does not mean that the mechanisms of evolution are not subject to study and debate, a theme that will continue to be controversial until it can be explained with total satisfaction.

For Freud the third humiliation was to come via the discovery, which was not really so, of the unconscious. The unconscious had already been described throughout the 19th century by various romantic German naturalist physicians, but Freud made it the focal point of his studies and gave it an importance that no one else had done. The result of these studies was the knowledge that consciousness was only the tip of an iceberg and that below the water was to be found an enormous majority of functions that, despite being unconscious, governed and directed human behaviour. The third humiliation was, therefore, that the human being was not even master of many of his actions. Nowadays it is estimated that of all the operations the brain carries out, only some 2% are conscious; the remainder are done without our being aware of them happening. In other words, Freud probably fell short in his theory.

Well now, awaiting us is a fourth humiliation that we only get a glimpse of at present: a neuroscientific revolution that is on the verge of placing on a tight-rope convictions as firm as the existence of the self, of external reality or free will. Subjects, that have not traditionally been the object of study by natural science, convinced as we were, all of us, that these were issues for philosophers, or at most psychologists, but which today are the object of the neurosciences that are leading us to believe that we were wrong up to now attributing to nature certain concepts that, quite possibly, were the fruits of our desires.

The human being does not have, for example, any reason at all to warrant thinking about the continuity of his person, of his self, which he considers is the same from the cradle to the tomb, knowing that nothing, whether in his body, or in his mind, or in his environment is permanent. And yet, who is going to convince us that this self, that is as subjectively present as the very external reality, does not exist? Our subjective impression is that we have inside us a kind of little being that ponders, acts and remains constant throughout our lives, but can we trust our subjective impressions?

Our sense organs have always fooled us and we know this, as did the ancient Greek philosophers of nature in Asia Minor. The

modern neurosciences tell us that neither colours, nor smells, nor sounds exist in nature, but that they are creations of the brain. Incidentally, in the 18th century, the Neapolitan philosopher, Giambattista Vico said exactly the same. However, have we taken on-board this reality? Are we not all convinced that these brain projections are not really so and that the qualities of our perceptions are a reflection of the reality that we are perceiving?

On the other hand, can we trust our own mind? Our cognitive capacities, are they not there to serve the cause of survival rather than for philosophical speculations or to reflect a reality that we are not sure to what extent it is a mental fabrication? Kant already warned us that our minds could be limited by *a priori* synthetic judgements that would make us see the world in a fashion that was maybe appropriate for our survival, but were not as freely made as we imagine. My opinion is that we have not totally taken on-board Kant's arguments of over two centuries ago in his *Critique of Pure Reason*. We know today that the idea that our mind is a clean slate, as the English empiricists proposed, is no more than yet another illusion. As William James said, if the animals that have preceded us with brains less developed than ours are born with a wide range of instincts sufficient for looking after themselves, human beings should not be endowed with less suitable instruments to cope with the environment, but on the contrary, many more.

We experience ourselves as beings that possess a sense of intention; we attribute to ourselves and to others responsibility for what we do; we think we can keep our states of mind in check and that we are in a position to say no to the factors determining our actions. We experience the sensation that we can control at all times what we think, do or want. But, is this so? Or are we subject to the determinism of the physical-chemical processes that take place in the brain?

We have data nowadays that implies that free will could be an illusion. Nobody can give assurance that this data is definitive. Definitive data in science does not exist, but it would seem logical to think that if we have accepted that the mind is a product of the brain and that this is pure matter, then logically the brain would have to be subject to the deterministic laws of nature, like

everything else. Nevertheless, who can convince us that we are not free to take decisions when the possibility of a choice exists between various options?

The majority of neuroscientists have abandoned the old idea that brain and mind, or body and soul, are of different natures, as Descartes proposed. But, does this mean we have ceased being or thinking in a dual manner? Are we not constantly analyzing what we think in antithetic terms, in antinomies? From the Orphic Mysteries of ancient history to the present day, duality has occupied our minds. Perhaps duality is the product of an innate tendency of the brain to see the world in opposing terms. This would explain why mythology, philosophy and ideologies are full of dualistic visions of reality. The adamant separation of body and soul in Descartes is practically ignored these days by neuroscientists, amongst other reasons because the interacting of both contradicts the laws of thermodynamics. Strange to say, this separation allowed at the time attention to be paid to the body, without entering into conflict with the Church, which signified the commencement of anatomy, physiology and —some say— modern science. But my opinion is that what then meant a considerable advancement over too long a time, has been an obstacle that has prevented the study of mental phenomena with scientific methods. We have no proof that operations exist in the brain that are not due to impulses that come from one's own nervous system. The hypothesis, therefore, of an extra-cerebral non-matter influence of this organ is not satisfactory nowadays for the majority of scientists, particularly because nobody can explain what this influence could be. An interaction with matter demands an exchange of energy. In order to mobilize energy, any non-matter body would have to have an energy source but in order to do so, it would have to cease being non-matter. But in addition, this dualistic explanation closes the door practically, as it has done in the past, to the natural scientific study of phenomena such as the conscience or the mind. The recent production of mystic, spiritual experiences by magnetic stimulation —electromagnetic stimulation of the temporal lobe— to my way of thinking, point to support of the Monist theory that the separation between

matter and spirit is no more than a product of the brain, but is not real outside our body, our mind.

As we can see, the neurosciences that before considered it inappropriate to deal with such subjects are now doing just that, but only to end up with the intuition that we have been completely mistaken with regard to what it meant. In other words, we are at the beginning of a systematic demolition, a deconstruction of concepts of which some are the very pillars upon which a large part of Western culture are built —no less—. Because, what are we going to do when they convince us that free will is a cerebral illusion? Does this not mean that our whole culture and civilization based on personal responsibility is being questioned?

These are the reasons that lead us to believe that a new humiliation for the human being is brewing up; a new revolution, the leading roles of which are the results obtained by neuroscience. Once again, science is on the point of opening our eyes to realities that have nothing to do with those we have been living with for centuries. Those were products of the mind and the realities that substitute them will be also. So that, as from this moment, to dream of any independent reality of the human brain will be possible, but not real. The world of the bat is as limited as is its brain. Ours, albeit of greater complexity, continues to be circumspect through the limitations imposed by this organ.

The discovery of neurons in the F5 pre-motor area of the brain that respond, not only when an animal makes a movement, but when it observes that same movement in its neighbours, has opened the doors to the study of the neurobiological basis of imitation, a highly important faculty in the acquisition of abilities and, as such, our culture and the ontogenic development of a child is unthinkable without it. Imitation is one of the most interesting examples of perceptive-motor coordination where the subject has to translate a very complex visual pattern in motor commands in such a way that the resulting movement coincides as much as possible with what was observed. It has been proposed that this system of mirror neurons is proof of the validity of a theory of simulation whereby humans would use their mental states to

predict, anticipate or explain the mental states of other persons. This theory, which has been named theory of the mind, is a very exciting theme that is related to the evolution of the brain and language.

Our capacity to understand others, which has always been called folk psychology, has given rise to a whole field of investigation of the cognitive sciences. Some philosophers expounded that this capability depended on our language ability. Studies carried out in non-human primates have also shown that they possess an intentional understanding that is useful for their lives in society, which led to the conclusion that the theory of the mind was a biological endowment independent of language. Non-human primates that have established a society with its rules and mutual relationships already possess what has been called a social intelligence, or Machiavellic intelligence, important for survival needs such as the capabilities for manipulation, deceit and cooperation. All this has also been considered as the development of independent language capabilities.

The fact that this is developed in a child at an age of approximately three to four years and that it is independent of other cognitive capabilities, makes one think that the theory of mind is a specific capability that can be lost, as an isolated factor, as is supposed to happen in autism. Autistic children fail in typical tasks that prove the capability of understanding the intentions of other people, independently of their intelligence or language.

Next chapters will complete and correct this introduction, and, overcoat, will clarify many of the erroneous concepts we have about ourselves and will serve to begin to adapt ourselves to the new knowledge that the neurosciences are bringing us.

The Relationship between Reason and Emotion and its Impact for the Concept of Free Will

Gerhard Roth

Whether reasons or emotions are guiding human behaviour, and whether or not we have free-will is a topic that has been discussed for more than two thousand years and, until recently, it was believed that science and neuroscience could say nothing about this problem and better leave it to theological, philosophical and, perhaps, psychological discussions. However, there have been lots of new empirical and experimental insights in neuroscience and experimental psychology that appear to reopen the entire discussion and, at least in German-speaking countries, during the past three years there has been enormous discussion about these topics, and once a week in the newspapers and journals there is some article about the debate between philosophers, psychologists, neuroscientists and politicians and, very recently, even people from law school, and there are now serious discussions whether or no the criminal law of Germany should be changed due to these new insights.

When we talk about the concept of free will, we need to at least briefly define what we understand under this concept. There is what I call the strong concept of free will, which is traditional, philosophical concept and also the basis of the criminal law —at least in Continental Europe—. In essence this concept means that my body actions are, at least partially, caused by an *immaterial* free will. This is called *mental causation*: something absolutely immaterial is

causing at least part of my material actions, therefore called "voluntary actions". This principle of mental causation by free will implies another principle, viz. "alternativism" which says that under identical physical-physiological and also psychological-motivational conditions, I could act otherwise if I only wanted it —even against all my motives—. This second principle is the basis of the concept of guilt in traditional criminal law: I am guilty only, if I am capable of deciding by pure will against any motive that would drive me towards the crime. This, in turn, implies that my actions are *not* completely determined by physical-chemical-physiological brain-process or by motives that are mediated by these processes. In this way, my actions *transcend* the realm of natural laws, where determinism holds and freedom is impossible. This is the concept of "indeterminism", which again is the basis of traditional, criminal law, at least in Continental Europe. The conclusion of this traditional concept in philosophy and criminal law is that determinism is incompatible with free will. Accordingly, it is also called *incompatibilism*.

However, other concepts of freedom of will and of action exist, and, according to Professor Pauen, there is a compatibilistic concept of free will. Briefly, this concept first means that my actions are free, if they are consequences or wishes, plans or volitions *grounded in my personality or character*. Second, my actions are free, if they are consequences of rational deliberations. And, essentially, free will then is self-determination or autonomy of action. And all this would then conclude that free will is compatible with determinism.

So, there are two major concepts: an incompatibilistic concept of free will and a compatibilistic one. However, from my point of view, it is important that in both concepts *rationality* plays a dominant role in decision-making and control of actions, because we usually believe that we are free in our decisions and actions to the degree in which we are guided by rationality and deliberation. But then, the question is: what is the specific role of rationality in decision-making and guidance of action in neurobiological and neuropsychological terms?

In a first step, we have to define reason (in German "Verstand") and intellect (in German "Vernunft"). Reason can defined by

"intelligent" problem solving by means of inductive and deductive thinking. And intellect (Vernunft) can be defined by medium and long-range planning by means of higher rational principles, especially regarding the social consequences of my actions. I am guided by my intellect if I am reflecting whether or not what I am doing is good or bad for my family, my colleagues and society.

Let us now have a look at the human brain. With 1.2-1.4 kg (or 1,200-1,400 cm^3) it is rather large, but by no means the largest brain of all animals (if we, as biologists, consider human beings as being animals): There are dolphins and whales that have up to 10 kg brain weight, and an African elephant has about 4 kg. It is not known what these animals are really doing with such large brains; by any reasonable measure of intelligence, they turn out to be less intelligent than humans.

Most of the human brain is covered by the cerebral cortex, which contains about 12 billion neurons and about 500 billion synapses. This six-layered cortex (called neo- or isocortex) is the seat of consciousness. Everything we are conscious of is bound to the activity of the cortex and everything that takes place outside the cortex is not accompanied by consciousness. So, whenever we have acts of free will in the traditional sense, which need to be conscious, they must reside in the cortex.

The cortex is usually divided into four lobes, viz. an occipital, temporal, parietal and frontal lobe. The occipital lobe and the adjacent parts of the temporal lobe are composed of areas that are involved in vision, the upper temporal lobe contains areas involved in audition and language (mostly in the left hemisphere), the posterior parietal lobe contains areas involved in spatial orientation, location of the body in space, control of movements (mostly eye, arm and hand), understanding of symbols including reading and writing, reading of maps, geometry, mathematics and music. The anterior portion of the parietal cortex contains the primary and secondary areas of somatosensation. In front of these somatosensory areas and situated in the posterior portion of the frontal lobe we find the cortical motor, premotor, supplementary and pre-supplementary motor areas responsible for the control, preparation and planning of fine motor actions. The pre-

supplementary motor area is also active, when we are imagining movement, for example, of the finger or the hand; also, whenever we are doing anything intentionally and consciously, this pre-supplementary motor area is active.

The upper anterior part of the frontal lobe is formed by the dorsolateral prefrontal cortex. This type of cortex is the seat of the so-called working memory; it is involved in the recognition of meaningful situations, of problems and of problem solving and of action planning. Generally speaking, it is the seat of intelligence. The lower and inner (medial) part of the frontal lobe is formed by the so-called orbital frontal cortex because it is located above the orbita of the eyes and by the anterior cingulate cortex. The orbitofrontal cortex is the site of considerations of the long-term consequences of one's own action, of moral and ethical considerations, the ethical and moral self; so, we are allowed to say this is seat of intellect, or "Vernunft" in the sense of Immanuel Kant. The anterior cingulate cortex is involved in cognitive attention, error recognition, reward expectations and emotional evaluation of stimuli.

In neuroscience, it is commonly thought that action planning starts with the activity in the dorsolateral prefrontal cortex. First we have plans and wishes and reflect, how to realize them in the most intelligent way. And then we are reflecting on the consequences —individual and especially social ones— and this is due to the activity of the orbitofrontal cortex. And now we decide what we are precisely going to do. This plan, as it is traditionally believed, is sent in parallel to the pre-supplementary and supplementary motor cortex, which correlates with the concrete sensation of will of action, and to the premotor cortex for the planning of gross motor actions, and eventually to the primary motor cortex for the control of fine muscle actions, which then activates, via spinal cord motor segments, the relevant muscles.

Many years ago, the neurophysiologist and philosopher, John Eccles, believed that the supplementary motor or as we know now, the pre-supplementary motor area, is the seat of free-will. Indeed, the activity of this part of the brain that connects most closely to the subjective awareness of free-will. Eccles thought that

the immaterial free will activated this cortical area (the "liason brain" as he called it) in order to materialize its intentions.

Above the primary motor, pre-motor and supplementary motor areas the so-called readiness potential can be recorded. This readiness potential (RP) is a slow negative wave of the cortex that is hidden within the electro-encephalogram (EEG) and can be visualized by averaging or filtering of EEG activity prior to a voluntary motor action. It starts 1-2 seconds before the onset of motor activity and is believed to represent the building-up of synchronous activity of neurons in the supplementary, premotor and motor cortex. The RP precedes any kind of voluntary action and must pass a certain amplitude threshold in order to elicit a movement. It starts symmetrically in the left and right hemisphere (hence called symmetrical RP) followed by an asymmetric part that occurs contralateral to the side of body and limb movement (asymmetrical RP).

With this in mind, we appear to be able to understand what goes on when we have wishes and plans to do things and we act accordingly. The prefrontal and orbitofrontal stimulate the pre-supplementary, supplementary, premotor and motor cortex to synchronize and build up the RP. Once the RP has passed a certain threshold, the pyramidal (cortico-spinal) tract carries motor activity to the spinal cord, which in turn activates the muscles such that our arm, our hand, our fingers, our head etc. move. This would be a neurobiological concept of how reason and intellect situated in the prefrontal and orbitofrontal cortex could control our actions and by activating the pre-supplementary cortex, induces the sensation of a freely willed action. However, if this were the end to the story, then there would be no patients with Parkinson's Disease.

Parkinson's Disease patients have difficulties with initiating willed actions, while they can exert externally triggered or highly automatised movements like marching. Importantly, PD patients have no primary cortical deficits; instead, they suffer from a deficit in dopamine (DA) production in the substantia nigra pars compacta, a small area in the midbrain tegmentum, where DA-producing neurons are located. In PD patients these neurons

have died to about 80%, when the disease becomes apparent. Consequently, as soon as Parkinson's Disease patients are given sufficient doses of L-dopa, which is then metabolised in the brain into dopamine, they can do at least for a while, what they want; so, ironically, L-dopa and dopamine induce the ability to exert free-willed actions.

How can we understand this? The substantia nigra is part of the basal ganglia, which reside in the deep parts of the endbrain (telencephalon), the diencephalon and midbrain tegmentum. When we make a cross-section through the endbrain and diencephalon, we see the enormously large compact nuclei on both sides of the third ventricle representing the nucleus caudatus and the Putamen, which together form the corpus striatum (often simply called "striatum") and the globus pallidus. Together, they contain about 600 million neurons, and they form the largest subcortical mass of the brain.

What are the basal ganglia doing? Initially it was believed that they are the seat of instinctive and reflex-like actions, but recently it was discovered that the basal ganglia are necessarily involved in all voluntary action, because they represent an *action memory*: "programs" of all actions that have been successfully executed are stored, and these programs improve with any repetition of a given action. At the beginning of the acquisition of a new movement pattern (playing piano, bike riding, skating, manipulation a device etc.) both cortex and basal ganglia have to interact; the more we execute that movement more elegantly, the less is the cortex involved and the more the job is done by the basal ganglia alone —and the less conciousness and attention are needed for the execution of the movement—.

If our cerebral cortex consciously decides, by means of the pre-frontal and orbitofrontal cortex, to do something, it cannot just go via the pyramidal tract to the motor centres in the spinal cord and to release a motor action, but the prefrontal, orbitofrontal and all related supplementary, premotor and motor areas in the cortex have to send neural activity to the striatum. Why?

The cortex does not possess all necessary information needed to execute a given motor program. It just contains some "plans", and

the basal ganglia have now to check how to realize them in a smooth and exact way. The reason for that is the following: Whenever we wish to execute a certain movement, our motor system has to release from our entire motor programme exactly those actions that are appropriate and at the same time suppress all other motor programmes. This is an enormously difficult task, because we have thousands and other thousands of possible motor programmes, and they all have to be suppressed and only one has to be released, otherwise our intentions to do something would end in spastic cramps. This suppression-activation is done in the basal ganglia in an absolutely fascinating way: the cortex stimulates the striatum, which activates in parallel the substantia nigra pars compacta, back and forth, then the globus pallidus and the substantia nigra pars reticulata, which then activate relay nuclei in the thalamus. But in between parts of the globus pallidus and the substantia nigra pars reticulata, there is the nucleus subthalamicus, which itself is inhibited, but exerts activation. In total, there is an enormously complex arrangement, mostly between an inhibitory effect and activity. In essence, there is a direct pathway that releases inhibition in the thalamus so it can activate the cortex, and there is an indirect pathway that suppresses the undesired motor plans.

Most importantly, inside the basal ganglia, there is a complicated interplay between the striatum, the globus pallidus and the substantia nigra pars compacta (the dopamine-producing part) and pars reticulata. Only when neurons in the substantia nigra pars compacta produce enough dopamine and send it to the striatum, the striatum can send its information, via the thalamus, back to the frontal, premotor and motor cortex. This feedback from the basal ganglia to the cortex is necessary for the RP to pass the threshold so that the premotor and motor cortex can activate the pyramidal tract, the spinal cord and the muscles.

In this way, the release of dopamine by neurons of the substantia nigra pars compacta into the striatum is something like a “go signal” for the entire motor brain. If —in the case of Parkinson’s Disease— there is no or an insufficient dopamine signal, the basal ganglia cannot activate the cortex, and no willed action can take place, only highly automatized movements that

can be started by the basal ganglia alone without the involvement of the cortex. So we have to accept that the conscious cortex cannot elicit voluntary acts without the contribution of the basal ganglia, which themselves act completely unconsciously.

But who or what controls the basal ganglia or more precisely: who or what controls the release of the dopamine signal as a "go signal"? This is where the emotions come in via the so-called limbic system. This system is exceedingly complex and includes centers in all parts of the brain. Here, I will comment only on the hippocampus, the amygdala and the mesolimbic system.

The hippocampus is situated in the lower part of the temporal cortex and is the organizer of so-called declarative memory, i.e., of all things that can be consciously recalled and (possibly) verbally reported. It also functions as a filter between the "preconscious" and the "conscious" controlling which memories, imaginations, thoughts and arguments come to my mind and which do not, for example in the process of reflection. The hippocampus is surrounded by the entorhinal cortex, which has also a control function in memory and learning. The amygdala is situated in the neighbourhood of the hippocampus and is closely connected to it. It is composed of functionally very different parts. One part, called *central amygdala*, has to do with the regulation of vegetative and visceral functions, with stress and affective states (together with two other limbic structures called hypothalamus and periaqueductal gray) and simple emotional conditioning; two other parts, the *cortical* and *medial amygdala*) deal with the processing of olfactory stimuli and pheromones (social-communicative olfactory signals). The largest part, called *basolateral amygdala*, deals with more complex emotional conditioning. Both the central and basolateral amygdala are the seat of inborn emotions. Here, our brain learns, already before birth, what is good and what is bad for us, in a completely unconscious way. Later, when the cortex and hippocampus develop, strong connections between amygdala and the prefrontal and orbitofrontal and other limbic cortices are formed, and in this way the amygdala is capable of strongly influencing these "higher" areas, and unconscious emotions may become conscious. In the same way, through these connections, the

prefrontal and obitofrontal cortex as well as the hippocampus try to control the activity of the amygdala, although this control is generally weaker than that in opposite direction. In this way, *impulse control* forms and is one of the major parts of what we call education and evolution of our moral and ethical selves. In some of us, this inhibitory influence is weak, then we have impulsive characters; if it is strong, as (hopefully) in most of us, then we are educated and controlled.

The amygdala is mostly concerned with more negative or surprising events in our life, and the antagonist of the amygdala is the so-called mesolimbic system, which contains the ventral tegmental area, the substantia nigra and the nucleus accumbens. The former two contain dopamine-producing neurons, which send DA either to the striatum (the substantia nigra pars compacta, see above) or in parallel to the nucleus accumbens, amygdala, entorhinal, prefrontal and orbitofrontal cortex (the ventral tegmental area). The nucleus accumbens (often also called ventral striatum) is part of the basal ganglia. It is strongly innervated by dopamine-producing cells and projects to the hippocampus, entorhinal and orbitofrontal cortex.

This mesolimbic system has two major functions: first, it represents the reward system acting by release of endogenous opiates (endorphins and enkephalins); everything that makes us happy is based eventually on the release of endogenous opiates. The second function is the reward-promising, -controlling and memory system, which is based on the release of dopamine system: whenever we experience something which makes us happy, dopamine is strongly increased, which also functions as a memory signal and forms reward expectations. Thus, whenever we are confronted with something that promises a reward, then the dopamine signal is very high. This is the basis of *motivation*. As a consequence, people who lack the endogenous opiate system have no pleasure in the world, and if they lack dopamine, they are not interested any more in doing anything because there is no stimulus to try to do something.

In this way, the limbic system is the origin of positive and negative emotional experience. It starts to develop very early, at the beginning, from the 7th week of gestation. Even long before birth,

it evaluates everything that the unborn child is doing according to the criteria "good" and "bad", "pleasant" and "unpleasant". The big time of this emotional limbic system are the first 3-5 years in our lives. During that period, the framework of character and personality is formed. All later experience is put into this framework and adjusted to it. Together with the amygdala and hippocampus/entorhinal cortex, the mesolimbic system represents the *unconscious* (and partly preconscious) *emotional experience memory*. This kind of memory guides our actions as child, young person and adult. Everything we are doing, wanting or wishing has to be compared with our past emotional experience. This guarantees, at least in principle that everything we are doing is being done in the light of past emotional experience. It is the most rational thing we can do.

The emotional experience influences the above described executive or "dorsal" loop between cortex, basal ganglia and thalamus via the "limbic" or "ventral" loop. This latter loop does not deal with the preparation and execution of actions, but with the origin of wishes and plans. It runs from orbitofrontal and adjacent cingulate cortex to the nucleus accumbens/ventral striatum, from there to the ventral pallidum, then to the thalamus and back to the orbitofrontal and cingulate cortex. Mediated by this loop, our wishes that originate unconsciously, can become conscious and are reflected consciously. Afterwards, they are sent back to unconscious evaluation by activity of the hippocampus, amygdala and mesolimbic limbic system. The ventral and the dorsal loop are interconnected at several points, the most important of which is the influence of the amygdala and nucleus accumbens onto the substantia nigra.

What does all this mean? The limbic system influences, with unconscious emotional experience, conscious cortical decisions in two ways. First, there is the influence of the amygdala and mesolimbic system on the orbitofrontal cortex regarding the ideas, wishes and plans, which means that all our wishes eventually come from the unconscious limbic system. Then these ideas, wishes and plans are consciously evaluated, sent back to the unconscious, then back to the conscious, and this may go on for days, months, even

years. Finally, after the conscious decision about what will actually be done, there is a *final control* of whether or not this decision is correct in the light of unconscious past emotional experience. This final control consists in the regulation of the release of the dopamine signal in the substantia nigra pars compacta, which acts as a "go" signal for the dorsal executive loop. If this final "go" signal does not occur, we may suddenly deviate from our momentary intentions and not do what we had planned to do —for example that we "forget" to place an important phone call or do not remember what we were looking for—. This mechanism also explains why Parkinson's Disease patients can have the wish to do something but are unable to realize that wish. Their ventral loop works correctly, but not the dorsal loop, because dopamine cannot be released by the substantia nigra pars compacta, and executive information cannot be sent from the striatum to the supplementary, premotor and motor cortex via thalamic relay nuclei. As a consequence, the readiness potential cannot build up sufficiently, and no threshold is being passed. This phenomenon is seen in RPs recorded from the cortex of PD patients.

If all this is the case, if we are controlled by emotional experience and memory, why do we need rationality and intellect at all? The limbic system could do everything by itself alone by using the content of the emotional memory!

There are three reasons, why at least in some cases rationality and intellect (Verstand and Vernunft) are necessary. First, the subcortical limbic centres are incapable of processing *details* of situations of problems. The amygdala can recognize objects such as fearful faces or threatening situations, but not their details. For the fine detail of an object or a situation including the context memory we need the cortex. Second, the sub-cortical unconscious limbic centres cannot deal with medium and long-term consequences of decisions including impulse inhibition. They can act only in the short run, as a young child cries: "I want to have it now, but not in a week… for Christmas… or next year!" We know little children cannot wait until Christmas because they don't know when Christmas is. To tell this exactly is a function of the cortex, especially of the frontal lobe. Third, these

centers do not understand language in a narrow sense, viz., as a grammatical-logical process, and, therefore, cannot utilize specific human communication. They can be influenced only by the emotional component of language.

Complex decisions need details of objects and situation, reflection of long-term consequences and —at least in the case of human beings— language, and this is why we have, among others, large cortexes —why we have rationality and intellect—. However, our conscious decisions are prepared and followed by emotional unconscious processes. They have the first and the last word. So, rationality is very important for decision making, but only plays the role of an *advisor*, presenting details of the situation and the possible consequences of decisions.

I will give an example: We have a good friend and he is in big trouble, for some problems with his family or for doing some crazy things, and he comes to us and asks for advice and we, as rational people, say: "Do you know what you are doing? Did you reflect the consequences of what you are planning to do? For example consequence A, B, C? And do you want these consequences?" And he says "No, I didn't reflect this". So just as the amygdala, he wants to do something "out of the gut", e.g. tell the boss his opinion or quit his job, and we tell him again "please reflect the consequences; do you have a new job offer, or what are you doing next week without a job? What will your family say?" "Oh, I didn't think about it". What we are advising is rationality.

Rationality presents choices, alternatives, gives advices on the basis of the medium and long-term consequences, but the final decision is always emotional. The advisor does not decide. Whether or not our friend will follow our advice is an *emotional decision*, not a rational one, and depends on his emotional experience, which is mostly unconscious. "I have to live with this decision!" This is what all industrial or political leaders are telling us as a basic rule. This means that there are no strictly rational decisions; rational decision-making always takes place in the framework of emotional limbic processing. Rationality is one thing that makes us superior —as we believe, at least— to all or most other beings, but rationality does not dominate our actions.

This model of neural guidance of voluntary actions is largely compatible with the weak, compatibilistic concept of free-will: Humans are free in the sense that they can act according to their conscious and unconscious will. This will, in turn, is determined by genetic and environmental neurobiological factors, as well as psychological and social and positive and negative experience, especially by those occurring early in life, which lead to structural and physiological changes in the brain. This together forms the emotional framework of our personality and character. In conclusion, this means that there is no free-will in the alternativistic and incompatibilistic sense, i.e., that under identical material and motivational conditions we could act otherwise by pure (immaterial) will, which is capable of transcending the laws of the material word. If we continue to use the term "free will", we can do it only in a compatibilistic sense.

Question. The specific question I would like you to discuss is whether the transition between an unconscious to a conscious state, where you say "all or nothing transition" or, by contrast, being quick is nevertheless a process as if —and this is my own bias in the interpretation— as if there would be a critical mass of neural activity required to reach this critical level in which you become, in a quick but nevertheless, not all or nothing transition, and this is asked specifically in the context of the role of the supramodal cortex area, which is thought to be critical to the definition of a conscious finger movement in this case.

Prof. Roth. First, there is of course a continuum between conscious and unconscious states. We can do things in a purely conscious way, and then we need to be highly attentive and concentrate on the situation or problem at hand. We also can have an *accompanying* consciousness: we do something while we don't know how we do it. In fact, most things in our daily life are done with only accompanying consciousness: we don't know the details, but we know what we are doing, for example when we are driving our car

through dense traffic. Finally, we do many things absolutely unconsciously, for example move our fingers while playing piano, moving our lips while speaking etc. The basal ganglia are the seat of things that we can do unconsciously or with only accompanying consciousness. Consciousness arises mostly when things and situations are new/unexpected, important and complicated. The unconscious brain then realizes that there are no already established executive programs at hand and accordingly induces consciousness by activating the cortex via the reticular formation, the basal forebrain, the hippocampus and other subcortical or allocortical centers in order to process complex details and to develop new executive programs. So, the most important thing is to understand that consciousness is, among others, always related to dealing with new, important and complicated things. This does not mean that that only the unconscious drives, in the Freudian sense, guide us, it is also everything we have experienced previously and consciously that that has sunk into the preconscious.

Q. Do you agree that consciousness is a process?

Prof. Roth. Yes. First there are many different kinds of consciousness —at least ten including phenomenal consciousness, attention, body awareness, authorship consciousness etc.— and consciousness is realised by very different mechanisms. Consciousness arises in the brain by processes which at least require 300 milliseconds, until the cortex is sufficiently activated. So consciousness is an instrument for the brain for solving new, important and complex problems.

Q. In the model, I missed —an error that I think is quite important— the cerebellum. The cerebellum has dopamine transporters and it is also recently has been attributed a role in motor attention, not only in cognitive transmittance, so I wonder how you think the cerebellum fits in?

Prof. Roth. In the diagram about the execute loops, for sake of simplicity I left the cerebellum out, but there is a "dorsal" basal ganglia loop and a cerebellum loop and they meet in the premotor and motor area. So even the extra input from the cerebellum is needed to bring the readiness potential above a certain level. However, what the cerebellum really does is still an open question

besides the fact that in brain imaging studies we almost always see it to be activated regardless of the nature of the tasks.

Q. There are many physical blocks and functional blocks also, but to work as a whole system, it is quite necessary but some synchronisation, coordination, interface, common function is, to my understanding, absolutely required. Where is this common functionality, assistant functionality located, according to your understanding?

Prof. Roth. This is a very difficult question! There is no single command system in the brain. But if we look at functional hierarchies inside the brain, we realize that there is a fundamental basis constituted by the hypothalamus, the periaqueductal gray and the autonomic-visceral centers in the hindbrain that secure our homeostasis, then on top of this lowest functional level are the amygdala and the mesolimbic system as the basis for emotional conditioning, and on top of this middle limbic level we have the limbic cortex (orbitofrontal, cingulate, insular, entorhinal), and on top of this highest limbic level —or better in parallel to it— we have the cognitive-linguistic cortex, mostly within the left hemisphere. The lowest limbic level develops first and the cognitive level latest. So there is no one single guiding centre, but there are various specific pathways that connect the various levels just described by excitation and inhibition. Thus, we are guided by a giant network comprising the different layers, and whenever we are doing something, these layers are promoting and inhibiting each other.

Q. There is one point that your discussion did not address and I think is critical for consciousness: this is the time. Many people think that consciousness is a question that takes time to appear, that requires time. So, one of the reasons why our actions are made unconsciously is because we do not wait for the consciousness to appear and, if we waited, then our actions would be impossible because we could not register the outgoing external stimuli and so we do most of our actions unconsciously because of this time factor and I think what is critical in our thinking about free will and consciousness and conscious actions is because we consider that we just need to go fast if we want to be accurate and if we want to be accurate we have no time to be conscious.

Prof. R. We have three ways of responding to external demands. The first is extremely fast and occurs by reflexes. This is executed by the spinal cord, and then we are doing it and only afterwards we become aware of it. In that case we usually say "it was my leg", "it was my arm", but not "it was me" —these reflexes are not attributed to my self—. Then we have fast, but learned and automatised responses, which need not be conscious but mostly are *accompanied* by consciousness. When we are doing something wrong, then we say "excuse me, it was me, but more or less it was my arm", so are doing something without explicit control, but we attribute it to our self. Only the third type of responses one is deliberate. We are confronted with new and complex problems and then we need time for conscious action. So there is always a battle inside the brain: either to do things unconsciously, reflex-like and maybe inappropriate, or appropriate but slowly, and also this may go wrong.

Q. In the end there are genetics and early social and emotional experiences, but, is there any hope from a scientific point of view to reconfigure a damaged brain via genetics, or emotional or social experiences? Is there any hope or will there be any hope in the future?

Prof. R. According to the last researches into the chances of psychotherapy, the earlier the damage takes place and the earlier therapy takes place, the more chances we have. Later in life —later meaning beyond ten years— thorough changes in our psyche, our personality, become increasingly more difficult. As a consequence, the efforts to change our personality need become stronger and stronger and more specific. If a person has substantial brain damage, there is almost nothing to do, if the brain did not already compensate for it, which however may occur in two thirds of all cases. In those cases where compensation does not occur, we don't know the reason. The reason for this increase in resistance of our personality against further changes is that the amygdala and the mesolimbic system very early on reduce their plasticity, while the cortex remains plastic for almost life-long. This apparently is one reason why psychotherapy is so difficult and often takes a long time —and often fails—.

A Neuroscience Look on Human Action Monitoring

Thomas F. Münte

To adapt their behavior to the environment, humans need to be able to monitor their performance and to detect and correct any errors. These abilities are part of the executive control system engaged in monitoring and detecting problems, planning, and adjusting the system's behavior. In the present chapter I will try to examine some of the basic neurophysiological characteristics of human action monitoring system. I will do this by presenting data mainly from my own laboratory but would like to stress at the outset that many different groups throughout the world have contributed to this research area over the past few years.

While action monitoring and error detection are important for virtually any task that we have to fulfil, e.g. driving a car, writing a letter, or preparing a meal, it is necessary to greatly simplify the problem in order to study this behavior in the laboratory.

Figure 1a illustrates the "working horse" of action monitoring research, the Eriksen flanker task, introduced in 1974 (Eriksen & Eriksen, 1974). It is quite simple: The subject views arrays comprising 5 letters with the task to respond to the middle letter. If this letter is an "H", the subject is supposed to press a button with the left hand; if it is an "S" a right hand response is required. The flanker task involves one class of stimuli, in which the relevant target letter is flanked by identical letters. These stimuli are called "congruent". Other stimuli feature flanking letters that are not identical to the target letter and would thus prime the

response with the opposite hand. These incongruent stimuli in conjunction with time pressure —imposed by a response dead line— lead to errors. In fact, the error rate is about 15% in normal subjects in this task. Importantly, in most cases subjects immediately realize that they have committed an error. If one records event-related brain potentials time-locked to the erroneous and correct button presses a prominent phasic negativity emerges for the error trials which has been called error-related negativity (ERN) or error negativity (NE) by its discoverers (Gehring & Fencsik, 2001; Falkenstein, Hoormann, Christ & Hohnsbein, 2000). The ERN is illustrated in figure 1c. The ERN has a latency of about 80 ms with regard to the button press. Moreover, the topographic map shows that its maximum is over medial frontal areas. This already suggests a generator within the medial frontal cortex. One can go a step further and try to determine where in the brain the generators of this ERP response reside. This, in principle, is a task that can not be solved, as Helmholtz has pointed out already in 1853. Nevertheless, modern source analysis techniques allow us to estimate the generators with a reasonable degree of confidence. Using multiple equivalent dipoles we arrived at a solution that features three dipoles, one located in medial frontal cortex and two others in lateral frontal locations. A medial frontal generator is also revealed using a different approach featuring distributed sources.

Still, one might argue that additional evidence is needed to imply medial frontal cortex in action monitoring. We (and many other groups) have therefore conducted an event-related functional MRI experiment. Functional MRI allows you to assess, which parts of the brain are active during a certain task by tracking the so called BOLD (blood oxygen level dependent) response. In the event-related variant of functional MRI, we can obtain separate activations for error and correct trials. If you do so, you observe an error-related activation in the medial frontal cortex involving the anterior cingulate gyrus (see figure 1b). This part of the brain is known to be involved in the executive aspects of many cognitive tasks. Interestingly, as in the brain potential source localization, two lateral frontal spots of activation are present in the MRI study

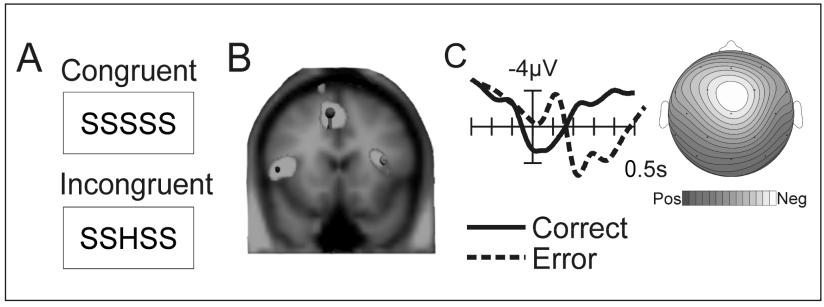


Fig. 1: A: Illustration of the typical flanker stimuli used in the Eriksen flanker task. The middle letter is relevant, while the surrounding 4 letters are distractors only. The subject is instructed to press a button with the left hand for the target letter S and to answer the letter H by a right hand button press. In the upper stimulus the flanking letters are identical to the target letters. Thus, the subject receives congruent information. In the lower stimulus, the flanker letters are non-identical to the target letters and prime the incorrect response. These stimuli are termed incongruent and lead to an increased rate of errors. B: Event-related brain potentials averaged time-locked to the subject's response show a phasic negativity for erroneous relative to correct responses, which has a peak latency of about 80 ms relative to the button press. The topographic map illustrates the distribution of this "error-related negativity" (after data of Rodríguez-Fornells *et al.*, 2002). C: Coronal slice of a flanker experiment conducted in conjunction with event-related functional magnetic resonance imaging (fMRI). Three main activations are observed: (i) anterior cingulate gyrus, (ii) left lateral frontal, (iii) insular cortex in the right hemisphere. In addition, dipoles obtained using brain electric source analysis (BESA) are mapped into the MR-Image (Rodríguez-Fornells and Münte, unpublished).

as well. It is worth pointing out that the temporal resolution of the MRI-method is much lower, with a peak of the BOLD-response at about 6 seconds after the error of the subject occurred. To sum up the combined electrophysiological and neuroimaging studies on error-detection, it appears that a network of brain areas comprising the anterior cingulate gyrus and lateral frontal areas is involved in action monitoring.

The next question that we asked was, whether or not the error-related negativity can be viewed as the neural correlate of the earliest error signal in the brain (Rodríguez-Fornells, Kurzbuch & Münte, 2002). As one of the consequences of the detection of an error is its correction, we hypothesized that the error-related negativity (if an indicator of the brain's earliest error signal) should precede any electrophysiological activity related to the correction. To answer this question we combined the recording of the error-

related negativity with that of an additional evoked brain response, the so-called lateralized readiness potential. This "LRP" can be recorded non-invasively from electrodes situated above the motor cortex and indexes the preparation of a movement with either the left or the right hand. It thus can serve as an index of the preparation of a corrective answer, which in our case is a second motor response (e.g., pressing the left hand button after the participant had erroneously pressed the right button). We thus recorded from young normal subjects in a situation in which they were encouraged to correct any errors. For the purpose of comparison we also included a condition in which subjects were not allowed to correct their errors. Figure 2 illustrates what we found.

The most important finding is that the LRP to the corrective response had an onset latency prior to the peak latency of the ERN. This in turn means that the ERN can not be considered as

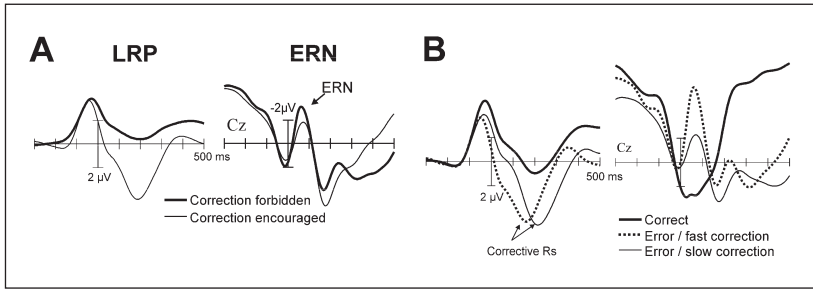


Fig. 2: Brain potentials from an experiment comparing conditions in which subjects are either encouraged to correct any performance errors or are instructed not to do so. A: on the left the lateralized readiness potential to error trials are displayed. An upward deflection indicates preparation of the erroneous response. The waveforms for the correction forbidden and correction encouraged conditions diverge even before the erroneous button press (which occurred at time 0). The peak of the error-related negativity (displayed on the right side) does not occur about 100 ms later. This was taken to indicate that correction was initiated prior to the process that is underlying the ERN. Hence, the ERN can not be viewed as a correlate of the earliest error signal in the brain. B: Comparison of the lateralized readiness potentials (left side) for fast and slow correction. The downward deflection that is indicating the preparation of the corrective response has a considerably earlier peak for the fast corrections. By contrast, the latency of the ERN (right side) does not differ between fast and slow corrections. Thus, there is no temporal coupling between corrective actions and the ERN-latency (Rodríguez-Fornells *et al.*, 2002).

a neural correlate of the earliest error signal in the brain, since this must precede, as said above, any corrective action. What is more, we also observed that the LRP for slow and fast corrections were of significantly different latencies but the latency of the ERN was indistinguishable for error trials that were corrected fast and slow. This, again, suggests that the ERN can not be viewed as the brain signal responsible for triggering the corrective action.

Thus, the take-home message of this study was that the ERN can not be the correlate of the earliest error signal in the brain and the question therefore arises, where and when this error signal is emitted. An important theory, inspired by recent work in animals, has linked the generation of the ERN to the actions of a subcortical brain system responsible for reward processing and reinforcement learning (Holroyd & Coles, 2002). This theory posits that our action monitoring system predicts the outcome of an action. If the outcome of an action (as is the case with a wrong button-press) turns out to be worse than expected, a phasic decrease in neural activity —mediated by the neurotransmitter dopamine— is observed in the basal ganglia and transmitted further to other brain structures such as the anterior cingulate cortex, where the ERN is issued. In the human, functional magnetic resonance imaging studies have shown activity of basal ganglia regions in response to error trials. As stated above, we are unable to determine, however, when exactly the neural events take place with the MR-technique. We therefore embarked on a project, together with neurosurgeon Volker Sturm and colleagues of the University of Cologne, Germany, in which we can directly record from critical structures of the brain's reward system during neurosurgical operations in awake patients. Sturm and colleagues are using deep brain stimulation devices (used since almost 20 years in Parkinson's disease patients to stimulate the subthalamic nucleus) to stimulate the nucleus accumbens in patients with obsessive compulsive disease. During such operations the patients are awake and are able to participate in psychological tasks. Our preliminary data suggest that neural activity recorded directly from the nucleus accumbens is different for error and correct trials. Moreover, this difference appears to precede the error

modulation observed previously in cortical areas. The next years will show, whether the signal in the nucleus accumbens might qualify as the earliest brain signal of error detections.

In real life we learn from our errors but we also learn from external feedback provided by our parents ("Leave that thing alone, its dangerous!" or "Very good, little boy!"), teachers, colleagues, computers or traffic lights. We thus asked, how the neurophysiological correlates of internally generated information about performance quality (i.e. the error-related negativity) would compare to external feedback information (Müller, Möller, Rodríguez-Fornells & Münte, 2005). In our study, a learning situation was created. During each experimental block participants were confronted with a new set of line drawings (of objects and animals) and had the task to learn for each object, whether a left or a right-hand button press was the correct answer for the object. As a learning signal feedback information (affirmative and negative) was provided to the participants 1100 ms after their button press. Thus, over the course of an experimental block, subjects were expected to learn the associations of stimuli and responses, which in fact they did. In this set-up we have the opportunity to observe brain signals indexing internally generated information on performance quality, i.e. the error-related negativity, by looking at brain responses in relation to the button-press, and brain signals indexing external information, by examining the brain potentials to the feedback stimulus. Here is what we observed: As expected, we obtained a reliable error-related negativity in response to erroneous button-presses. Moreover, as the association between a stimulus and the appropriate response became stronger over the course of an experimental block, we had predicted that the error-related negativity should also increase over the block. This was indeed the case. In addition, during the initial learning of the stimulus-response associations the feedback stimuli should provide more crucial information than during the later phases of an experimental block during which the stimulus-response associations were already firmly established. This was reflected by the amplitude of a phasic negativity, which we termed the feedback-related negativity, which was more pronounced

during initial learning than during later phases. Interestingly, a source analysis of this feedback related negativity revealed that, in addition to a prominent source in the anterior cingulate cortex (as in the error-related negativity) this brain response also received contributions from other more posterior regions (figure 3).

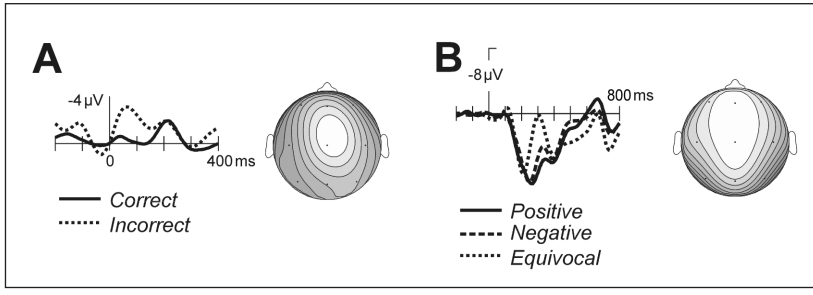


Fig. 3: Brain potentials from a learning task. Subjects had to associate a left or right hand button press with each of 16 line-drawings that occurred repeatedly during an experimental block. A: Brain responses time-locked to the button presses of the subject. Erroneous responses are associated with a clear negative component, i.e. an error-related negativity, that shows the typical medial frontal distribution. B: Brain potentials time-locked to the feedback stimuli indicating either a correct answer (positive), an incorrect answer (negative) or that the computer was unable to determine the correctness of the answer (equivocal, see text). Both, negative and equivocal feedback were associated with a phasic negativity which had a medial distribution that extended further back than that of the ERN. Source analysis (not shown) indicated a posterior midline source in addition to an anterior cingulate source (after data from Müller *et al.*, 2005).

Finally, we were interested in the brain responses to uninformative feedback information. To this end, we (falsely) informed the participants that the computer program used for stimulation was not quite finished yet and contained a "bug". This bug was said to cause the computer to be unable to determine whether the given response was correct or not every once in a while. This "equivocal" feedback, we hypothesized, should lead the subjects to an intense reexamination of their previous response, which in turn should be associated with an amplitude increase of the feedback-related negativity. This was in fact the case.

In keeping with the general theme of this book, I would like to turn to the question of error detection and awareness. If the error-related negativity (and the activations of the anterior cingulate

gyrus in fMRI studies) is not the first signal of error detection in the brain, it might be related to the error becoming aware. As Nieuwenhuis and colleagues (2001) put it: "If we want to examine the relation between error-related brain activity and the awareness of *slips* —the type of error that has been the focus of psychophysiological research on error processing— we need a paradigm in which the Ne system can easily derive a representation of the correct response, but participants are not (always) aware of errors in the execution of this response". Nieuwenhuis and colleagues opted for a so-called anti-saccade task (Nieuwenhuis, Ridderinkhof, Blom, Band & Kok, 2001). In such a task, a stimulus occurs either on the left or on the right side of a video-display. The subject has to overcome the natural tendency to look at this stimulus (i.e. to make a saccade towards it) but rather has to look to the other side (i.e. to make an anti-saccade). Nieuwenhuis and colleagues found that subjectively unperceived saccade errors were almost always immediately corrected, and were associated with faster correction times and smaller saccade sizes than perceived errors. With regard to electrophysiology they reported that, irrespective of whether the participant was aware of the error or not, erroneous saccades were followed by a sizable error-related negativity. Thus, these authors concluded that the ERN is not related to error-awareness. What they found, however, is that a positivity immediately following the ERN, often called the error positivity or Pe, showed a relationship to awareness of the error.

Before we conclude that the ERN indexes errors regardless of whether the subjects are consciously aware of them or not, one should consider another study by Scheffers and Coles (Scheffers & Coles, 2000), which is more closely related to the flanker paradigm that I have discussed so far. In their study, Scheffers and Coles found a close correlation between the amplitude of the ERN and the subjectively perceived accuracy of the response in a typical flanker task. Similar findings from our laboratory are illustrated in figure 4.

How can we resolve this apparent contradiction? Nieuwenhuis and colleagues have suggested that the ERN varies as a function of error awareness, whenever the degree of certainty about the accuracy of the response depends on data limitations, as in

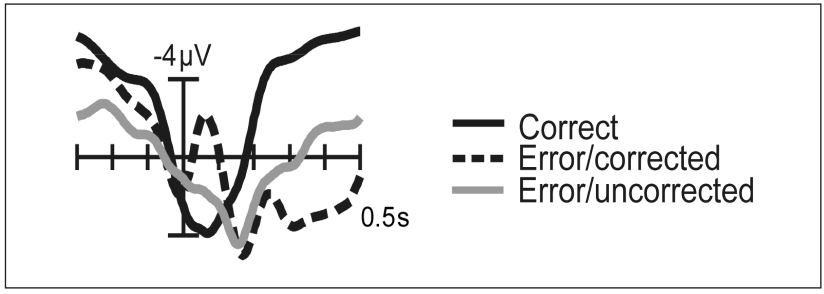


Fig 4: Brain potentials for correct responses, errors that were corrected and errors that were not corrected. The error correction rate in this experiment was about 90 %. Only the errors that were corrected were associated with a typical error-related negativity, while uncorrected errors were not (unpublished data).

the Scheffers and Coles study. Conversely, the ERN is unaffected by awareness, when there is uncertainty about the actual erroneous response, as in the Nieuwenhuis antisaccade experiment. With regard to fMRI experiments a similar flexible relationship between activations in medial frontal cortex and error awareness might be observed. In a recent brain imaging experiment Hester and colleagues (Hester, Foxe, Molholm, Shpaner & Garavan, 2005) showed that explicit awareness of an error (in this case subjects responded when they actually should have withheld a response) was associated with bilateral prefrontal and parietal brain activation. The anterior cingulate region showed similar activation for errors that the subjects were aware of and errors that they were not aware of. These results were taken to suggest that activity in the anterior cingulate gyrus is not sufficient for conscious awareness of errors but rather appears to feed information to other brain regions that might implement conscious measures of behavioral modification (such as post-error slowing). An important aspect that has to be considered when examining the relationship of brain events and awareness is the specific paradigm used. The correction of an error in the flanker task that has been used by us and many others, for example, requires to make a button press with the hand opposite to the hand used for the erroneous response. This will go hardly undetected. By contrast, other tasks, e.g. those requiring eye-movements, do not require a

change in the effector for correction. Thus, a correction of an error might be performed without becoming aware. To summarize, the relationship of electrophysiological and brain activations in MRI experiments on the one hand and awareness of performance errors on the other hand is far from clear and requires more work.

At the end this brief and personal summary of neurophysiological aspects of action monitoring I would like to address the question, whether the neurophysiological correlates of action monitoring, isolated using our almost ridiculously simple flanker task, might also be found in other, more natural tasks. The answer to this question is an unequivocal yes, as will hopefully become clear with the following example concerning speech errors or "slips of the tongue" (Möller, Jansma, Rodríguez-Fornells & Münte, 2006). Speaking is an amazing act. Within a several hundred milliseconds we go from the intention to speak about a certain topic to the vocalization of an utterance in a highly automatized fashion. During this complex process, which requires to look up the appropriate words in the mental lexicon, to derive their syntactic features, to identify their appropriate phonological makeup in the phonological output lexicon and to initiate the corresponding motor programs involving a few dozen different muscles, we make surprisingly few errors (about 1 in 1000 words). Most modern psycholinguistic models of speech production feature monitoring devices. For example, the popular model by Dutch psycholinguist Levelt incorporates several different monitoring loop, one on the conceptual level, another one monitoring "inner speech" and a third one examining one's own speech output. Recently, we have started to record brain potentials to speech errors induced by a classic laboratory task used for the induction of speech errors. In order to observe brain activity that is uncontaminated by vocalization artifacts subjects were required to utter the target word pair only after a speech prompt had been given. Prior to the error we could thus examine brain activity to those trials that were associated with a speech error and those trials that were not. Figure 5 illustrates some of the results. Clearly, speech errors were associated with a medio-frontal negativity, which had a topographical distribution very much like the error-related negativity. Also, our further investigations

regarding the neural generators of this speech-error effect revealed medial frontal structures as the likely source. From this pattern we concluded that a generic action monitoring system, as revealed by simple psychological experiments such as the flanker task, is also enlisted during the monitoring of speech errors. While these parallels are quite striking, we would like to stress at the same time that the particular type of speech error that we elicited in our task occurs rather late in the speech production system. It might well be that other, earlier types of speech errors might be detected and corrected within the linguistic system proper without recurrence to the generic action monitoring system that we have illustrated in this contribution.

To conclude, it seems to me that the investigation of action slips can be a rich source of information to psychologists, neuroscientists and maybe even philosophers. The combined use of neuroimaging techniques with high temporal (i.e. event-related brain potentials) and spatial (i.e., functional magnetic resonance imaging) resolution will greatly enhance our understanding of our action monitoring system and will allow us to determine which of these functions are carried out with and without awareness.

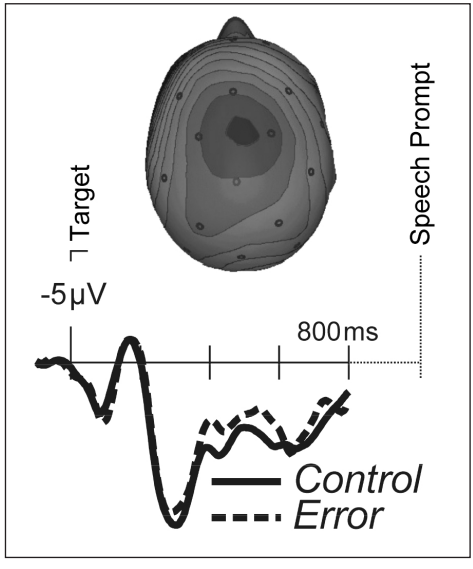


Fig. 5: Data from a task that was designed to elicit speech errors. The brain potentials shown were time-locked to the presentation of a word pair. After 1 second a speech prompt required subjects to vocalize the word pair. Those words that gave rise to a speech error were associated with an increased negativity. The scalp map shows a medial frontal maximum of this speech error effect that was similar to the distribution of the error-related negativity (after data of Möller *et al.*, in press).

Acknowledgement

I gratefully acknowledge the invaluable contribution of my friends and colleagues: Antoni Rodríguez-Fornells, Marcus Heldmann, Jürn Möller, Bernadette Jansma, Sandra Müller and Jordi Riba.
I am also indebted to the Deutsche Forschungsgemeinschaft who has supported our studies.

References

Eriksen, B. A., & Eriksen, C. W. (1974). Effect of noise letters on the identification of target letters in a non-search task. *Percept Psychophys,* 16, 143-149.

Falkenstein, M., Hoormann, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: a tutorial. *Biol Psychol,* 51, 87-107.

Gehring, W. J., & Fencsik, D. E. (2001). Functions of the medial frontal cortex in the processing of conflict and errors. *J Neurosci,* 21, 9430-9437.

Hester, R., Foxe, J. J., Molholm, S., Shpaner, M., & Garavan, H. (2005). Neural mechanisms involved in error processing: A comparison of errors made with and without awareness. *Neuroimage,* 27, 602-608.

Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol Rev,* 109, 679-709.

Möller, J., Jansma, B. M., Rodríguez-Fornells, A., & Münte, T. F. (2006). What the brain does before the tongue slips. *Cereb Cortex,* in press.

Müller, S. V., Möller, J., Rodríguez-Fornells, A., & Münte, T. F. (2005). Brain potentials related to self-generated and external information used for performance monitoring. *Clin Neurophysiol,* 116, 63-74.

Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: evidence from an antisaccade task. *Psychophysiology,* 38, 752-760.

Rodríguez-Fornells, A., Kurzbuch, A. R., & Münte, T. F. (2002). Time course of error detection and correction in humans: neurophysiological evidence. *J Neurosci,* 22, 9990-9996.

SCHEFFERS, M. K., & COLES, M. G. (2000). Performance monitoring in a confusing world: error-related brain activity, judgments of response accuracy, and types of errors. *J Exp Psychol Hum Percept Perform,* 26, 141-151.

Question. With regard to the error related negativity and activity of the anterior cingulate gyrus to errors that you have mentioned, we found no such activity in children.

Prof. Münte. Some researchers have tracked the developmental time-course of the ERN and it seems that it is rather a late addition to our repertoire of brain potentials. This might explain why you do not find the same pattern in children. Even normal children have a very small ERN.

Q. You mentioned that there is a conscious error correction or error awareness and an unconscious variant. Could you explain in greater detail when does it become conscious?

Prof. Münte. On the one hand it appears that whenever you correct an error or you signal an error you get an ERN. This is especially the case, if you need to change the effector, that is the response hand, for the correction. In this case, if you are able to correct an error you must have noticed it consciously before. This might suggest that the ERN is the correlate of error awareness, but the antisaccade experiment by Nieuwenhuis and co-workers has a somewhat different message. There are also other experiments that entail ballistic reaching movements, in which the target is displaced in such a way that the subjects make corrections to the ballistic movements without noticing them. In such situations, brain imaging studies have not revealed error-related activity in the anterior cingulate cortex. Thus, it appears to me that error awareness might be tight to the ERN and anterior cingulate activity whenever the experimental paradigm requires a change of effector.

Q. As was mentioned before, there are a lot of different sub-cortical structures involved in this control mechanism, right.

Why are you not able to see this sort of activity of this sub-cortical structure when you do your FMRI experiments?

Prof. Münte. On the contrary: It is quite possible to see activity in subcortical structures. For example, there is a recent paper from Magdeburg by Wittmann and co-workers (published in Neuron, 2005) which involves reinforcement learning. Here, a number of subcortical structures including the substantia nigra and the ventral tegmental area were shown to be active.

Q. Do you know if anyone found that thalamic nuclei involved in this sort of task?

Prof. Münte. A study by Ullsperger and von Cramon (published in Journal of Neuroscience, 2003) has actually found the habenula of the thalamus involved in an error monitoring task.

Human Self-Understanding, Neuroscience, and Free Will: A Revolution Ahead?

Michael Pauen

It seems obvious that some of the recent developments in neuroscience in Europe focussed on issues that are relevant for our human self-understanding and, among these issues are consciousness, self-consciousness, and free will.

Some think that this will lead us to a scientific solution of these problems. Among these are Benjamin Libet or Daniel Wegner. However, I think that this would be a misunderstanding of, let's say, the division of labor between philosophers and scientists.

On the one hand you have the philosophical problems —let's better say philosophical aspects of the problem—. If you want to say something about the free will problem, one of the questions you have to ask is: what exactly do we mean if we talk about freedom? So we have a conceptual question, or a question concerning the criteria. What would count as a free action? And that, obviously, is a philosophical issue. Philosophers try to make clear what the criteria are, and they try to come up with coherent concepts that really account for our pre-scientific intuitions.

On the other hand, you have empirical issues because, if you know what the criteria are: what we mean if we talk about consciousness of free will, then obviously we want to know whether we meet these criteria, and that is an empirical question. Do we have the relevant abilities?

The Traditional Concept of Free Will

Now the criteria of free will have been discussed for not only hundreds, but for thousands of years, and there is an old traditional philosophical and also pre-philosophical concept of free will, according to which freedom requires indetermination. That's what Gerhard Roth said is the strong concept of free will.

There are several well-known problems with this traditional concept of free will. One of the problems is that there would be no freedom in a deterministic world. So, imagine: yesterday evening I was standing at the train station in Heidelberg and I was thinking "Well, do I want to go to Madrid or will I just stay in Heidelberg?" If our world is deterministic, then it was determined that I went to Madrid, and, accordingly, I was not free when I went to Madrid and I was not responsible for doing so. That seems to be somewhat unattractive. Another point is that according to this traditional view, free acts of will would be beyond scientific explanation because only those acts would be free that cannot be explained, even if we know everything about the whole pre-history of these acts. If we could explain why I went to Madrid, provided that you knew my whole pre-history —my wishes and beliefs, etc.— then this act would not be free; it would be determined. And that would result in an idea of freedom that is just a mystery.

All these problems, of course, do not lead to a refutation of this traditional concept, but I think they give us reason enough for a revision of this concept, and that is what I would like to do.

What do philosophers do if they make such an evaluation? Imagine a biologist who tries to find out whether life needs vital force. What do they do? Well, they do something like this: they look at the relevant points and they look if a life force is necessary there. This is exactly what I do with the concept of free will. So I disassemble this concept and I look whether, somewhere at the relevant junctions, indetermination is needed in order to get freedom or, at least, in order to get a more demanding concept of free will. I would like to show you that this is not the case. So, put in indetermination wherever you want, you won't get a stronger idea of free will.

So my first hypothesis will be that freedom can be understood as a kind of self-determination. This kind of self-determination is compatible with determination in general, so self-determination may occur in a deterministic world.

Second, self-determination is compatible with the possibility that our cognitive acts including our acts of will may be physically realised. This concept does justice also to our pre-scientific intuitions. So can account for what we mean if we talk about freedom in a pre-scientific sense or in a pre-scientific environment. Third, it would follow that volitional acts are subject to scientific explanation. Finally, I will discuss *conceptual criteria*. I will not make any statement whether or not our world is deterministic; I will not say that our world is indeterministic and I will not make any effort to find a small, tiny loophole for indetermination in a deterministic world —just smuggling in a bit of indeterminism in order to save freedom—. My message will be: you don't need indetermination. Indetermination takes away control.

I will start with a minimal concept of freedom, then I will try to find out if this minimal concept of freedom may be just too weak and then I will discuss some empirical results that seem to show that even if freedom is compatible with determinism, it may not be compatible with some of our most recent scientific results.

Part I

A Minimal Concept of Freedom

I have already said that the most important reason why we need a philosophical clarification is that our pre-scientific intuitions or our pre-scientific understanding of free will is incoherent and unclear. The question, then, is: How do philosophers think they can come up with a coherent concept without diverging from our pre-scientific intuitions? The solution is that we have very stable and very coherent negative intuitions, that is, we have very clear ideas about what free actions are *not*, and that is my starting point.

There are basically two important points. The first point is that we distinguish between freedom and compulsion. Actions that are

performed under compulsion are not free. This gives us, what I would like to call the principle of autonomy. Free actions need to be autonomous, not externally determined and not performed under compulsion.

There is still a second, very important principle. We distinguish free, responsible actions from random events. Random events are not free, and the important difference between random events and free actions is that you can ascribe a free action to an agent or an originator. This is very important because we can not hold somebody responsible for a random event, because random events are beyond your control even if they happen in your brain.

Yes, imagine that someone is on trial because he has shot someone. Imagine, second, that his lawyer can prove that pulling the trigger resulted from a random event in the defendant's motor cortex. Then you would say that this was beyond the defendant's control. This illustrates that we distinguish freedom from chance and the important difference seems to be that free actions can be ascribed to an agent. So if there were a connection between the agent and the pulling of the trigger, then you would say he is responsible. However, if there is no connection because the pulling of the trigger was a random event, the you would say he is not responsible. That is the important thing.

Self-determination

It is quite easy to do justice to these two requirements, autonomy and authorship, if we translate freedom into self-determination. Self-determination implies a distinction from external determination, so it meets the requirement of autonomy. Second, self-determination implies also that the action is determined by the agent himself, so it meets the requirement of authorship or origination.

Imagine that I have a very strong belief that stealing is reprehensible. And imagine that I paid for the goods in my basket just because I have this belief. This would give us a very good reason to assume that what I did was self-determined.

If this is true then self-determination seems to give us an adequate analysis for freedom in the minimal sense discussed above. This minimal concept is based only on the assumptions of authorship and autonomy. Of course, it may be necessary to come up with additional criteria —I will discuss additional criteria in the second part—. However, at this point I would like to stress that, if we translate freedom into self-determination, then we have to say something about the self.

The Self

I can touch this point only shortly, but I would like to give you an idea of such a self. The first is that such a self should have some preferences, that is, some beliefs, some desires, and some dispositions that are really his own beliefs, desires, and dispositions. The important point is that if there are no such beliefs and desires and dispositions that are constitutive for the agent, then it does not make sense to talk about an agent, i.e. an originator of an action. And if it does not make sense to talk about an agent, then is does not make sense to talk about authorship. So if we know that we have to talk about authorship, then we know that we have to talk about such preferences, whatever these preferences may be. I have not said anything about that, but we do need that.

So, there is no authorship without those preferences. There is an obvious problem here because I may have some preferences; maybe I am an alcoholic or I may have some other compulsive behaviours. I would follow that I don't act freely, if I act according to these compulsive preferences.

So what we need is a distinction between those preferences that are really constitutive for a person and those preferences that are not, say because they are pathological, compulsive or whatever. This is a very difficult question and I can touch it only very shortly. I would like to give you a criterion and that is what we need if we want to have a clear concept of freedom. This criterion enables you to distinguish between those preferences that can be ascribed to an agent and those that cannot. And the criterion I want to present

you with is the agent's ability to decide effectively against the preference.

And now, go back to the examples I gave you earlier: my conviction that stealing is reprehensible; I might make a decision against this preference if you give me good reasons that stealing is a fine thing under certain conditions or that stealing is OK in certain conditions: maybe to save someone's life. So, this conviction that stealing is reprehensible seems to need this condition.

Second, let's think about alcoholism. If I am really an alcoholic, then you can give me the best reasons in the world, still I will not be able to get rid of my alcoholism just for those reasons. Of course, you may convince me to undergo a withdrawal treatment, but I will not be able to make an effective decision myself and so alcoholism and this preference do not meet the criterion. This seems to indicate that this criterion enables us to distinguish between those preferences that can be ascribed to an agent and those that cannot.

Preliminary Conclusions

This leads us to some interesting consequences with respect to the question we started with, namely whether freedom and determination and freedom and physical realization are compatible.

Imagine, first, that it turns out that my decision-making process was physically realized. Would that mean that it was not a self-determined decision? This is not the case. If we live in a purely physical world without an immaterial soul, then my beliefs, my motives, my preferences have to be physically realized, otherwise they cannot be effective with respect to my actions. So, physical realisation is nothing that interferes with my ability to make self-determined decisions; it enables me to make these decisions. That is the first important point.

The second point concerns the compatibility of freedom and determination. If what I said is true, then an action is self determined and therefore free, in this minimal sense, if it is

determined by those features that constitute the self. This follows from what I just said, and that means that the action can be explained with reference to the preferences of the agent. So if it can be explained with reference to my beliefs and my convictions that I went to Madrid rather than staying in Heidelberg, then it was a self-determined decision to come here and not something that was externally determined.

It would follow that determination does not interfere with my ability to perform self-determined actions either. That can be underlined, if we put in some indetermination. So imagine it is my firm belief that it is not fair to skip a talk that I have promised to give. And now imagine that, due to a moment of indetermination, I stay in Heidelberg. That would prevent me from acting according to my own beliefs and desires, namely to keep my promises. It would follow that indetermination, rather than leading us to a more demanding concept of freedom, may interfere with someone's ability to act on a self-determined manner.

So let's conclude that the decisive point seems not to be *whether* an action is determined. The decisive point seems to be *how* it is determined. Whether it is determined by myself, by my own beliefs and desires or whether it is determined by some external factors. Determination by myself is just by definition, self-determination.

Part II

Intuitions and Arguments

Given that what I have presented so far is only a "minimal concept of freedom", one might conclude that this concept is simply too weak. Something that is important for our intuitions concerning freedom might be missing. There are many candidates for what is missing, but I think the most important candidate is the ability to do otherwise.

Now of course, it seems really impossible to do otherwise in a deterministic world. If our world is deterministic, then it was determined a thousand years ago that I would go to Madrid rather

than staying in Heidelberg, and it would be just a queer idea that I was free under such conditions.

That is why many people who think that freedom and determination are compatible reject the idea that freedom implies the ability to do otherwise or, philosophically speaking, the "principle of alternative possibilities". These people think that our world is determined, and that determination leaves no room for the ability to do otherwise. If we want to keep the compatibility of freedom and determination, then we have to reject the "principle of alternative possibilities" in order to save freedom.

Proponents of this strategy have presented arguments and thought experiments in order to support this position. I do not think that these arguments are really conclusive but, rather than refuting these arguments in detail, I would like to show that we can dispense with the whole strategy, they are intended to support. This is so, because there are alternative possibilities in a determined world. In order to show this, I will give an analysis of what it really means to be able to do otherwise. In other words: Is it really necessary that we live in an indeterministic world in order to have the ability to do otherwise?

According to the standard interpretation, the ability to do otherwise requires that there are two different actions possible under identical internal and external circumstances. So just turn the movie back to the scene when I was standing at the station in Heidelberg. According to the standard interpretation, it must have been possible that I go either to Madrid or I stay in Heidelberg likewise, under exactly the same conditions, that is, with the same beliefs, the same desires, the same dispositions of myself.

It may well be, that such situations occur in our world. But should we say that they give us the ability to do otherwise? I think not. The reason is that it would not be possible, in principle, to explain that I went to Madrid rather than staying in Heidelberg with reference to my beliefs, desires, and dispositions, because the action can vary independently of my preferences. And if the action can vary completely independently from me, then it follows that I am not the author of this vary action because I cannot determine what happened. So if accept this interpretation and say that doing

otherwise means that two events can happen under the same conditions, then what happens does not depend on me and this means that I cannot be the real author of what happens. So, in this case, we could not say I could have done otherwise.

We could and would not say, that I could or would have done otherwise. We could say only that something else could have happened; and that this would have happened completely independently from my beliefs, desires, and dispositions. I conclude that this does not seem to be the right interpretation of what it means to be able to do otherwise.

But what *is* the right interpretation? Let's try to find out whether the idea of self-determination gives us a better interpretation of the ability to do otherwise.

If I have executed a self-determined action according to the minimal concept above, then it was up to me whether I went to Madrid or remained in Heidelberg. It was not determined by external factors, rather it can be explained with reference to my preferences that I did what I did. That is the criterion we have so far in this minimal concept.

Well, if it is up to me, then this implies, obviously, that I *can* go to Madrid and that I *can* stay in Heidelberg, likewise. Only if these two conditions are met, then it makes sense to say it is up to me. If I cannot stay in Heidelberg —say because I would have to pay a huge fine if I refused to go to Madrid— then we would not say it is up to me. But if it is justified to say that it is up to me —and we have already talked about the relevant conditions— then this implies that I *can* stay in Heidelberg and I *can* go to Madrid likewise. Consequently, we can say in advance that I *can* go to Madrid and I *can* stay in Heidelberg. And this means that it is justified afterwards to say I *could have* stayed in Heidelberg even if I actually went to Madrid. That is the important point.

So I conclude that, according to an adequate interpretation, someone's ability to do *x* rather than *y* means that it is up to the person whether she does *x* or *y*. This, however, is just an implication of self determination. It would follow that self determination gives you the ability to do otherwise, just for free.

And the idea that we need indeterminism in order to get the ability to do otherwise is just based on an inadequate understanding of this ability.

Still, it might be assumed that a more demanding concept of freedom should imply the requirement of indeterminism. I order to demonstrate that this is not the case, I would like disassemble the concept of freedom at some other intersections. It will turn out that you can introduce indetermination wherever you want, you won't get more freedom. The only thing you might get is chance and, consequently, a loss of control.

One of the widely shared intuitions of incompatibilists is that, in a deterministic world, all my actions are determined by my genetic dispositions and my upbringing. But then it is possible, in principle, to predict what someone will 10 or 100 years before her birth. However: How can someone be said to act freely if what she does is determined by factors that are clearly beyond her control because they were effective before the person's birth?

If determination by factors that happened before the person's birth is the issue here, then we should get more freedom if we waive determination just before the person's birth. So, let's interrupt the causal chain just before my birth, such that you cannot make predictions with respect to my future actions. What would be the result? It may well be that the interruption would result in additional room for freedom —I don't know. What I do know, however, is that I can't make use of it this freedom because, trivially, I am not living at this time, because it's before my birth. But if interrupting the causal chain at this point doesn't result in additional freedom then *not* interrupting it at this point cannot result in less freedom. It would follow that someone's ability to predict what I do 10 or 100 years before my birth cannot interfere with my ability to act freely.

Of course, it might be argued that the relevant point is a bit later in the causal chain. So maybe we have to disassemble the action at a different junction. So, let's interrupt the causal chain two days before the beginning of my process of decision-making. Before this interruption I believe that stealing is reprehensible and afterwards I accept it, just as a matter of fact.

Again: Would this enhance my ability to act freely? The important point is that this change of my preferences was not under my control. It occurred completely independently from all my other beliefs, desires, and dispositions. Of course, I may change or even give up some of my beliefs, if you give me good reasons for doing so, and this would be perfectly compatible with all our intuitions concerning freedom and control. But this is not what would happen in the above scenario. What would happen is that my preferences change or *could change*, just by accident. Given that such a change is clearly beyond my control, it would follow that this interruption, rather than enhancing my ability to act freely, actually interferes with this ability.

Again: If it changes anything, then it results in a loss of self-determination and control.

But maybe we have still chosen the wrong moment in time. So let's interrupt the causal chain just during the process of decision making. Imagine that I am wondering whether I should go to Madrid or stay in Heidelberg for quite some time. At the beginning of this process, I consider all the reasons in favour of the idea to go to Madrid, in the second half, I ponder over the reasons for remaining in Heidelberg. Then, just in the middle of the process, indetermination comes in and this means that all the previous considerations become ineffective with respect to the final decision. So it might turn out that although I have strong preferences for going to Madrid, I will decide to stay in Heidelberg. I take it as obvious that this would interfere any ability of making a rational decision.

To illustrate this even a bit further, imagine a parliament whose members are debating about some resolution, and imagine that all the arguments in favour of the resolution are brought forward at the beginning of the process while the arguments against the resolution are brought forward at the end of the process. If this process is interrupted in the middle, then all the arguments in favour of the resolution become ineffective with respect to the outcome. I take it as obvious that this does not result in additional freedom or self-determination. Again, what would happen would just be a loss of control, not an enhancement of freedom.

I conclude that waiving determination does not give us more freedom. If it gives us anything, then it will give us a loss of control. Again, the decisive point is not *whether* but *how* the action is determined. I should add that determination in the strong sense is not *required*. It may well be that there is only a strong probability relation between someone's intentions and what she does, such that we could give a non-deterministic explanation of the action with respect to the person's preferences. So there is no need for determination in a strict sense, the important point is that waiving determination does not give us more freedom but just leads to a loss of control.

Another important point concerns the physical realization of acts of will. If these acts are physically realized, then they might still be self-determined, provided that they would be determined by our own beliefs or by the physical realizers of these beliefs, respectively. If we live in a physical world, such a physical realization of my beliefs and desires would even be a condition of their becoming effective. It would follow, that not only freedom and determination, but also freedom and physical realization are compatible.

Part III

Empirical Results

So far I have talked about philosophical matters only, that is, about conceptual standards and criteria, and I have demonstrated that indetermination is not among these criteria. Of course, we want to know not only what the standards are but whether or not humans meet these standards: This is a question of empirical science.

This is a very critical point, given that some important experiments seem to refute our ability to act freely and in a self-determined manner. The crucial point is that self determination and free willed action requires intention. On a simplified picture, a self determined decision issues an intention that may be conscious or not and this intention, in turn, determines an action.

So if I make a self-determined decision to go to Madrid, this will result in the corresponding intention which, in turn, will bring about the necessary action, mediated by several loops that may or may not be conscious and that may, obviously involve emotion, apart from rational considerations.

The question, then, would be: How do we know that this is the case? One obvious answer would be that we have a sense of agency. We know that we have acted for certain reasons or for certain beliefs and this sense of agency discriminates between intentional and unintentional action. The problem now is that there is a whole body of evidence that raises doubts that our intentions are really responsible for our actions. Among them is the work done by J.M.R. Delgado. He was able make his research subjects execute certain actions by stimulating their brain: His subjects moved their bodies or they moved their heads. The important point is that, although the action was not caused by the subjects' intentions, the subjects reported that they performed an intentional action. They even gave reasons which mean that they confabulated justifications. They said, for instance, that they had turned their head because they had heard a noise, or because they were looking for their slippers. And that seems to show that our sense of agency is faulty, at least in come cases.

The same follows from evidence from the 1960s and 1970s concerning split-brain patients. Split-brain patients can execute commands that are addressed exclusively to their right, unconscious hemisphere. Imagine that the patient is asked to laugh and does so. However, when the patient was asked why she did so, she claimed that the experimenters were making jokes, which was not the case.

These and other experiments have led some to the idea that our sense of agency is deceptive in general, such that we do not act for the intentions we would declare, should someone ask us for our reasons and motives. Actually, we do what we do just because of unconscious causes and motives, and this means that the process of action generation involves a causal process that is beyond our conscious access and only after all this has happened we present a post hoc construction and which results in our —illusory— sense

of agency. The causation of action, in general, would turn out to be a subconscious process which is not determined by conscious intention.

If all this would be true, there would not only be no free will, there would even be no conscious will at all. This is the central hypothesis of Daniel Wegner:

> "This means, though, that conscious will is an illusion. It is an illusion in the sense that the experience of consciously willing an action is not a direct indication that the conscious thought has caused the action. Conscious will, viewed this way, may be an extraordinary illusion indeed —the equivalent of a magician's producing an elephant from the folds of his handkerchief—. How could it seem so much like our wills cause our actions if this isn't actually happening?" (Wegner, 2002, 2).

Is this true? I don't think it is. There is considerable evidence that intention is real and that it is effective, even if an adequate picture is much more complicated than the simple view that I presented at the beginning. Part of this evidence comes from experiments with so-called "alien hand" patients. Alien hand patients are not able to control one of their hands in certain situations. Consequently, these patients can discriminate between intentional and unintentional action, so their sense of agency is quite reliable, even if it is not infallible.

The same goes for the so-called "intentional binding phenomenon" that has been investigated by Patrick Haggard in recent years[1]. Haggard's subjects had to perform an action at a certain time. 250 milliseconds later they heard a sound and their task was to judge the interval between these events. If there were just an intentional action followed by a sound which was not produced by the action, then the subjects judged quite reliably. However, if the sound was an intentional effect of the subject's action, then the subjects *underestimated* the interval between the action and the sound. Conversely, if the subjects performed

[1] See Haggard, 2005; Haggard and Clark, 2003.

an unintentional action that was produced by magnetic brain stimulation, then they *overestimated* the interval by about 50 milliseconds, or some 20%.

This seems to show that our sense of agency is able to discriminate between intentional and unintentional action. So, even if our sense of agency may lead us astray, sometimes, it does not follow that our sense of agency is completely unreliable, nor can we conclude that intention does not play a causal role in the emergence of action.

Other studies show that the sense of agency is not produced immediately by our intentions, rather, it results from a post hoc cognitive mechanism that processes also data from the environment, and this does make sense, even if intentions are real and our sense of agency is reliable in general.

Imagine you are a soccer player and you have the intention of scoring a goal. However, even if a goal is scored, it does not follow that you were the one who performed the action. So just relying on your intentions would not be a good idea. There may have also been another soccer player, or you may have bribed the referee or the goalkeeper of the other team, so there may be many alternative causes for this goal to be scored. It seems obvious that the only way to account for these possibilities is a post hoc evaluation of the process. Otherwise, you cannot make sure that you ascribe those actions and only those actions to your self that you really performed. Of course such a process will be fallible and that is what the above experiments show; they do not show, however, that conscious will is an illusion or that intention is ineffective.

Conclusion

I tried to show that there is no revolution ahead and that current scientific results neither refute freedom nor responsibility. The important philosophical point is that freedom and determination are compatible. Consequently, if we demonstrate that our brain is a deterministic system, then it does not follow that we are unable to act freely. Still it may turn out to be necessary to revise our idea

of intentional action in some sense, given that the mechanism that underlies intentional action differs substantially from our pre-scientific ideas. The same may be true for self consciousness and for consciousness in general. This implies not only a revision of our beliefs but also a revision of the underlying concepts, or a revision of our ideas about these abilities.

Imagine, until the late 19th century, people thought that an immaterial soul-substance was the source of consciousness. Meanwhile, the concept of a soul-substance has been displaced by the concept of consciousness which does not require an immaterial substance. I think that we need similar revisions of other concepts, and that is what I have tried to show in the case of free will. We have to give up the traditional concept of free will in favour of a concept that really is able to evaluate the scientific findings we have. And if we do so, then this revision may enable us to make use of scientific results that tell us more about the foundations and the limits of our human abilities. So, this revision may help us to find out more about ourselves and of course, this will have consequences for education, for the legal system and, of course, also for everyday life. Still, I think, that these consequences need not be dramatic. Chances are that they just tell us much more about ourselves.

References

HAGGARD, P. (2005). Conscious Intention and Motor Cognition. *Trends in Cognitive Science,* 5 (6), 290-295.

HAGGARD, P., & SAM CLARK (2003). Intentional Action: Conscious Experience and Neural Prediction. *Consciousness and Cognition,* 12, 695-707.

WEGNER, D. M. (2002). *The Illusion of Conscious Will.* Cambridge MA: MIT Press.

Being Oneself: the Neural Basis of Self-Consciousness

Marc Jeannerod

Failure of action identification and attribution in schizophrenia

How do we recognize ourselves as an agent and how do we distinguish our actions from those of other agents? The identification of oneself as the origin of an action may seem a relatively simple task when movements are overtly executed. In this condition, sensory signals arise from the moving limbs and from the effects of the movement on the external world: these signals can be compared with those resulting from the action generation mechanism, and the outcome of this comparison can be used to label the action as self-generated. Which renders self-identification less easy to understand is the existence of a number of situations where the action generation mechanism is activated, but where the action remains covert, as in imagined actions (motor images), for example. Yet, imagined actions, as well as other mental states involving covert actions, such as intending or thinking about an action, are clearly self-attributed in spite of the lack of overt execution, and the lack of the corresponding sensory signals. In addition, actions performed by the people that we observe are also clearly distinguished from ours own (Jeannerod, 2003).

Subtle as it may be, the mechanism of action identification can be disrupted in pathological conditions. Pathological conditions offer many examples of misidentification of action: a typical case is that of schizophrenia. Misidentification, however, is not evenly

distributed among schizophrenic patients. This type of disturbance is preferentially observed in one particular class of patients who present the so-called "first-rank symptoms". According to Kurt Schneider, these symptoms refer to a state where patients interpret their own thoughts or actions as due to alien forces or to other people and feel being controlled or influenced by others (Schneider, 1955). First-rank symptoms (or *Schneiderian* symptoms, as they will be called here) reflect the disruption of a mechanism that normally generates consciousness of one's own actions and thoughts and makes possible their correct attribution to their author. Thus, a study of attribution behavior in schizophrenic patients would not only help us understand the factors responsible for misattribution in the patients, but it would also shed light on this critical function in normal life.

The pattern of action misidentification in these patients is twofold. According to the French psychiatrist Pierre Janet, patients may either attribute their own actions or thoughts to others rather than to themselves (under-attributions); or patients may attribute the actions or thoughts of others to themselves (over-attributions) (Janet, 1937). A typical example of under-attributions is hallucinations. Hallucinating schizophrenic patients may show a tendency to project their own experience onto external events. Accordingly, they may misattribute their own intentions or actions to external agents. During auditory hallucinations, the patient will hear voices that are typically experienced as coming from a powerful external entity, but which in fact correspond to sub-vocal speech produced by the patient. The voices are often comments where the patient is addressed in the third person and also include commands and directions for action (Chadwick & Birchwood, 1994). The patient may declare that he or she is being acted upon by an alien force, as if his or her thoughts or acts were controlled by an external agent. The so-called mimetic behavior observed at the acute stage of psychosis also relates to this category. Over-attributions, which Janet (1937) called "excess of appropriation", correspond for the patient to the illusion that actions of others are in fact initiated or performed by him/her and that he/she is influencing other people (the clinical picture of megalomania). In

this case, patients are convinced that their intentions or actions can affect external events, for example, that they can influence the thought and the actions of other people. Accordingly, they tend to misattribute the occurrence of external events to themselves. The consequence of this misinterpretation would be that external events are seen as resulting from their own intentions and actions.

1. Do Schneiderian schizophrenic patients fail to monitor their own actions?

The most prevalent theory for explaining attribution difficulties in schizophrenic patients is that proposed by Christopher Frith and his colleagues. This theory, which was first formulated in the early 1990s (e.g., Frith, 1992) has been recently reformulated within the framework of the action monitoring mechanism. The general idea is that the delusion of influence arises from a lack of awareness of the predicted limb position. The reasoning of the Frith school is the following: "Under normal circumstances the awareness of initiating a movement must depend on the predicted limb position because awareness of initiating a movement precedes the actual movement and any feedback about actual limb position. The patient with delusions of control is aware of his goal, of his intention to move and of his movement having occurred, but he is not aware of having initiated the movement. It is as if the movement, although intended, has been initiated by some external force" (Blakemore *et al.*, 2002: 240). Thus, the key problem for these patients would be that, due to their pathological condition, the endogenous signals related to the generation of their actions (the so-called efference copy) would either be absent or not properly used by the nervous system (Frith *et al.*, 2000).

These contentions are supported by experimental evidence showing that patients with Schneiderian symptoms fail to monitor the discrepancies between their intended and predicted limb positions. In these experiments, patients were shown moving visual targets that they had to catch by displacing a lever, in the absence of visual feedback from their movement. If the target moved in an unexpected direction, the patients tended not to correct their error

when they had started their movement in the direction where they expected the target to appear. By contrast, the same patients had no difficulty making the correction when they were provided visual feedback on the displacement of the lever (e.g., Frith and Done, 1989; Mlakar *et al.*, 1994). This type of experiment, however, mostly stresses the automatic part of the action monitoring system, but tends to overlook the conscious part of this mechanism, that which makes it possible to give an explicit judgement on the action. In other words, the link between the experimentally proven inability to make corrections and the clinically observed misattributions of the action to its agent is not firmly established.

In order to establish this link, one needs first to address the issue of conscious awareness of the action in schizophrenic patients. With this aim in view, Fourneret *et al.* (2001) used a paradigm similar to that of Frith and Done (1989) for testing the ability not only to produce automatic corrections but also to make conscious judgements on the motor task. They used an apparatus where the subjects could not see their hand. Instead they saw a spot moving on a screen, the position of which corresponded to the tip of a hand-held stylus. Their task consisted in drawing a line with the stylus to a target located in front of them on the screen in the direction of their sagittal body axis. In addition, a systematic bias was electronically introduced in the stylus output such that, when they drew the line in the proper direction, it appeared deviated by 15° to the right. Thus, the subjects had to deviate their hand movement in a direction opposite to the deviation in order to reach the target. Finally, in some blocks of trials, an opaque mask occluded the screen, such that the subjects could not see the line they drew except in the last one third of the trajectory to the target. Patients with and without Schneiderian symptoms, as well as normal control subjects were tested. It took approximately 10 trials for subjects from all three groups to learn to gradually compensate for the deviation and to maintain the same strategy of moving the arm to the left throughout the task, including when the mask occluded the view of the spot. This result indicates that the patients, like normal subjects, were able, in the absence of

visual feedback, to initiate a hand movement in the proper direction to reach the target. This implies that they monitored their central command signals, modified them so as to achieve the task and maintained this modification across the trials. In addition, most of the patients in the group with Schneiderian symptoms were found to be able to explicitly describe the strategy they had to use in order to reach the target. This experiment (Fourneret *et al.*, 2001) therefore demonstrates that consciously monitoring their actions is possible in these patients, provided they are tested with the appropriate experimental design.

2. A specific attribution deficit in schizophrenic patients

These findings suggest that the explanation for the misidentification and the mis-attribution of their own movements observed in Schneiderian patients should be looked for at a level of action consciousness directly related to their experience of being the agents of their actions (their sense of agency), rather than at the sensory-motor level. To address this point, we will examine the results obtained with groups of schizophrenic patients in experiments specifically designed for testing self-recognition and attribution. The basic methodology for these experiments is to record agency judgements about movements that are shown to the subjects and that may correspond, or not, to their own movements. The first experiment of this type was made by Daprati *et al.* (1997). Subjects were requested to make simple finger movements with their hidden right hand. At the same time, they were shown on a screen the movements of a right hand of an uncertain origin, that is, a hand that could equally likely belong to them or to someone else. Finally, the movements performed by that hand either could be the same as, or different from, those the subjects had performed. In the condition with the maximum uncertainty about the ownership of the hand, Schneiderian patients made attribution errors in 80% of trials, i.e., they massively mis-attributed the movements of the alien hand to themselves. The error rate in the same condition was only 30% in control subjects and around 50% in non-Schneiderian patients.

The results thus reveal the existence of impaired attribution of action in schizophrenia, especially in Schneiderian patients (see also Franck *et al.*, 2001). In addition, these patients tend to misattribute to themselves the moving hands they see more often than to misattribute their own hands to other agents.

The exaggerated tendency of schizophrenic patients to self-attribute alien actions testifies to a perturbed sense of agency. Note that this effect could still be possibly related to a difficulty in matching the action one produces with the sensory consequences of that action. For example, when such patients are asked to indicate the time at which they detect a sound occurring at a fixed delay after a self-produced movement, they tend to shrink the estimated time. In other words, they tend to over-bind the actions they produce and the sensory consequences of these actions (Haggard *et al.*, 2003).

3. The depth of misattribution in schizophrenic patients

According to Schneider, the first-rank symptoms observed in schizophrenic patients are the expression of false beliefs which lead to a feeling of "depersonalization" by impairing the distinction between the self and other people. If this assumption is correct, then misattribution should not only be observed for actions involving overt behavior; it should also extend deeper into the mechanisms of action generation, i.e., to covert action-related mental states. Among Schneiderian symptoms, verbal hallucinations represent a typical case of misattribution of covert behavior. As mentioned earlier, it is known that auditory verbal hallucinations in schizophrenic patients are in fact related to the production of inner speech by the patient. Some hallucinated patients even show weak muscular activity in their laryngeal muscles (e.g., Gould, 1949; David, 1994). Thus, the patients perceive their own inner speech as voices arising from an external source. McGuire *et al.* (1996) compared the brain activity of schizophrenic patients predisposed to hallucinations (whom they called "hallucinators"), as well as of patients with no history of hallucinations (non-hallucinators) and normal controls. subjects

had to either think simple sentences in their mind (inner speech) or imagine themselves hearing the same sentences spoken by an alien voice (verbal imagery). The authors verified that no vocal utterance was produced by the subjects. Brain activity was measured during these tasks, using PET. The main effect of the tasks was that, when compared to controls and non-hallucinators, patients predisposed to hallucinations showed a reduction of activity in their left middle temporal gyrus during the task of imagining sentences pronounced by another person. McGuire *et al*'s conclusion was that this reduced activity in an area specifically devoted to the processing of spoken language could interfere with the patients' recognition of their own language as self-generated. The reasoning of the authors was that regions concerned with the generation of language (e.g., in the frontal lobe) normally inform speech perception areas of imminent language output, so that the subsequently perceived speech is recognized as self-generated. This process would be dysfunctional in hallucinators, hence their vulnerability to hallucinations. Another complementary finding was obtained by Dierks *et al.* (1999) in Schneiderian patients. When these patients experienced hallucinations, their brain metabolism increased in the primary auditory cortex (Heschl gyrus) on the left side. Thus, during verbal hallucinations, the auditory temporal areas remain active, which suggests that the nervous system in these patients behaves as if it were actually processing the speech of an external speaker. On the contrary, in healthy subjects, self-generated inner speech is accompanied by a decrease in responsiveness of their primary auditory cortex.

These data do not speak in favor of a defective action monitoring mechanism. Indeed, because in verbal hallucinations there is no vocal utterance, the defective simulation mechanism should be looked for at the level of the representation of the actions, not only at the level of their execution. Accordingly, it can be proposed that the network which we normally use for attributing thoughts and intentions to their agent (be it ourselves or another person) is damaged in certain types of schizophrenic patients and therefore does not allow to properly "simulate" the covert operations needed for attribution. This hypothesis would be

consistent with the fact that, in clinical practice, the Schneiderian symptoms primarily affect the patients thoughts, rather than their executed actions.

4. The role of the frontal lobes

The nature of the dysfunction responsible for misattribution of actions in pathological conditions is still an open question. It is conceivable that changes in the pattern of cortical connectivity could alter the shape of the networks corresponding to different representations, or the relative intensity of activation in the areas composing these networks. This explanation fits with the classical observation of a defective function of prefrontal cortex in schizophrenia.

Prefrontal cortex is known to be hypoactive in many schizophrenic patients (the so-called "hypofrontality"; see Ingvar and Franzen, 1975, Weinberger and Berman, 1996), and its morphological aspect has been shown to be abnormal on post-mortem examination (Goldman-Rakic and Selemon, 1997). On the behavioral side, schizophrenic patients show signs of "behavioral hypofrontality" when tested with neuropsychological tests exploring prefrontal functions (e.g., tests for working memory or problem solving requiring the use of pre-instructed rules; see Posada *et al.*, 2001, Posada and Franck, 2002). Because prefrontal areas are known to normally exert an inhibitory control on other areas involved in various aspects of motor and sensorimotor processing, an alteration of this control in schizophrenic patients might result in aberrant behavior, due to the release of activity of normally inhibited parts of cortex, specially in the posterior parts of the hemispheres. Indeed, the above neuroimaging studies have revealed that patients presenting verbal hallucinations show abnormal activation of primary auditory areas in their left temporal lobe, as if they were processing an external auditory stimulus. Similarly, an increased activity in the right posterior parietal lobe has been observed in patients with delusion of influence during an action recognition task (Spence *et al.*, 1997). In a recent paper, Farrer and her colleagues examined the

brain activity of Schneiderian patients during a task where subjects saw a distorted feedback of their own movements. Normal subjects in this condition show an activation of their right inferior parietal lobule correlated with the degree of distortion: the larger the distortion, the higher the activation (Farrer *et al.*, 2003). Schizophrenic patients showed a different pattern. First, the activity of their right inferior parietal lobule in the baseline condition was already higher than in normal controls. Second, the increase in activity correlated poorly with the increased distortion (Farrer *et al.*, 2004). Thus, the patients had really no cues (as inferred from the change in activity in their parietal lobe) about whether they saw their own movements or the movements of an alien agent. It is not surprising that such patients had difficulties in correctly attributing their own movements to themselves, or in disentangling their actions from those of others. Finally, it is noteworthy that deluded schizophrenic patients show an increase in activity in their parietal lobes even at rest (Franck *et al.*, 2002), and that a permanent increase in activity of the same area is observed in non-psychotic subjects presenting factors of vulnerability for schizophrenia. These findings illustrate the importance of a trait by trait analysis of complex pathological psychotic syndromes (Jeannerod *et al.*, 2003).

References

BLAKEMORE, S.-J., FRITH, C. D., & WOLPERT, D. (2002). Abnormalities in the awareness of action. *Trends in Cognitive Sciences*, 6, 237-242.

CHADWICK, P., & BIRCHWOOD, M. (1994). The omnipotence of voices. A cognitive approach to auditory hallucinations. *British Journal of Psychiatry*, 164, 190-201.

DAPRATI, E., FRANCK, N., GEORGIEFF, N., PROUST, J., PACHERIE, E., DALERY, J., & JEANNEROD, M. (1997). Looking for the agent. An investigation into consciousness of action and self-consciousness in schizophrenic patients. *Cognition*, 65, 71-86.

DAVID, A. S. (1994). The neuropsychological origin of auditory hallucinations. In: David, A. S., Cutting, J. C. (eds.): *The neuropsychology of schizophrenia*. Hove: Lawrence Erlbaum Associates, pp. 269-313.

DIERKS, T., LINDEN, D. E. J., JANDL, M., FORMISANO, E., GOEBEL, R., LANFERMAN, H., & SINGER, W. (1999). Activation of the Heschl's gyrus during auditory hallucniations. *Neuron,* 22, 615-621.

FARRER, C.; FRANCK, N., GEORGIEFF, N., FRITH, C.D., DECETY, J., & JEANNEROD, M. (2003). Modulating the experience of agency: a PET study. *Neuroimage*, 18, 324-333.

—, FRANCK, N., GEORGIEFF, N., FRITH, C. D., DECETY, J., GEORGIEFF, N., D'AMATO, T., & JEANNEROD, M. (2004). Neural correlates of action attribution in schizophrenia. *Psychiatry Research: Neuroimaging*, 131, 31-44.

FOURNERET, P., FRANCK, N., SLACHEVSKY, A., & JEANNEROD, M. (2001). Self-monitoring in schizophrenia revisited. *Neuroreport*, 12, 1203-1208.

FRANCK, N., FARRER, C., GEORGIEFF, N., MARIE-CARDINE, M., DALÉRY, J., D'AMATO, T., & JEANNEROD, M. (2001). Defective recognition of one's own actions in schizophrenic patients. *American Journal of Psychiatry*, 158, 454-459.

—, O'LEARY, D. S., FLAUM, M., HICHWA, R. D., & ANDREASEN, N. C. (2002). Cerebral blood flow changes associated with schneiderian first-rank symptoms in schizophrenia. *Journal of Psychiatry and Clinical Neuroscience*, 14, 277-282.

FRITH, C. D. (1992). *The Cognitive Neuropsychology of Schizophrenia.* Hove: Lawrence Erlbaum Associates.

—, & DONE, D. J. (1989). Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. *Psychological Medicine*, 19, 359-363.

—, BLAKEMORE S. J., & WOLPERT D. M. (2000). Abnormalities in the awareness and control of action. *Phiosophical Transactions of the Royal Society London,* B355, 1771-1788.

GOLDMAN-RAKIC, P. S., & SELEMON, L. D. (1997). Functional and anatomical aspects of prefrontal pathology in schizophrenia. *Schizophrenia Bulletin*, 23, 437-458.

GOULD, L. N. (1949). Auditory hallucinations in subvocal speech: objective study in a case of schizophrenia. *Journal of Nervous and Mental Diseases,* 109, 418-427.

HAGGARD, P., MARTIN, F., TAYLOR-CLARKE, M., JEANNEROD, M., & FRANCK, N. (2003). Awareness of action in schizophrenia. *Neuroreport*, 14, 1081-1085.

INGVAR, D. H., & FRANZEN, G. (1975). *Lancet*, 21, 1484.

JANET, P. (1937). Les troubles de la personnalité sociale. *Annales Médico-Psychologiques*, tome II, 149-200.

JEANNEROD, M. (2003). The mechanisms of self-recognition in humans. *Behavioral Brain Research*, 142, 1-15.

—, FARRER, C., FRANCK, N., FOURNERET, P., DAPRATI, E., & GEORGIEFF, N. (2003). Recognition of action in normal and schizophrenic subjects. In: *The self in neuroscience and psychiatry*, Kircher, T., & David, A. (eds.). Cambridge: Cambridge University Press, pp. 380-406.

MCGUIRE, P. K., SILBERSWEIG, D. A., & FRITH, C. D. (1996). Functional neuroanatomy of verbal self-monitoring. *Brain*, 119, 907-917.

MLAKAR, J., JENSTERLE, J., & FRITH, C. D. (1994). Central monitoring deficiency and schizophrenic symptoms. *Psychological Medicine*, 24, 557-564.

POSADA, A., & FRANCK, N. (2002). Use of a rule in schizophrenia. *Psychiatry Research*, 109, 289-296.

—, FRANCK, N., GEORGIEFF, N., & JEANNEROD, M. (2001). Anticipating incoming events. An impaired cognitive process in schizophrenia. *Cognition*, 81, 209-225.

SCHNEIDER, K. (1955). *Klinische Psychopathologie*. Stuttgart: Thieme Verlag.

SPENCE, S. A., BROOKS, D. J., HIRSCH, S. R., LIDDLE, P. F., MEEHAN, J., & GRASBY, P. M. (1997). A PET study of voluntary movement in schizophrenic patients experiencing passivity phenomena (delusions of alien control). *Brain*, 120, 1997-2011.

WEINBERGER, D. R., & BERMAN, K. F. (1996). Prefrontal function in schizophrenia: confounds and controversies. *Philosophical Transactions of the Royal Society*, B351, 1495-1503.

Using Visual Information to Understand and Predict What Comes Next

David Perrett

I am going to talk about the brain responses during sequences of actions that are witnessed in the world. This is work in progress. We do not fully understand what is going on but it seems to be taking place for all sorts of visually recognized actions: actions of the hand, like grasping, and actions of the head and the body that are seen by a monkey.

I am going to concentrate on some of the mechanisms at cellular level that might analyse these actions. At first glance, the mechanisms look rather dumb because they interfere with perception. I want to shift the emphasis from mechanisms that seem to make us insensitive to the world; to suggest actually the same mechanisms help and they help in a fundamental way. The mechanisms help predict what is coming next: hence the title.

We were asked to consider awareness in the symposium. Something that is emerging our work is that our perceptual awareness is incorrect. Our perceptual experience is one in which we sense the future, the immediate future: something that is going to happen in about $1/10^{th}$ of a second. Our perceptual awareness is distorted; we are seeing not what is actually happening, instead we sense what is just about to happen. I will argue this may be so because of the way our brains process sensory information. The story of how we come to this position is long.

Physiologists know very well the way brain cells respond to stimuli. The response of a cell and the response can be "high" or "low". We say when the response is low that the stimulus is outside of the cell's tuning curve or receptive field. A stimulus that produces a high response is said to be in the center of the cell's receptive field. Virtually all sensory cells have a "bell shaped" tuning curve.

We can map out sensory coding as we approximate what the cell in the brain somewhere really likes. This is true for tactile position on the skin; it is true for visual position in the eye; it is also true from the work of Hubel and Wiesel, for orientation sensitivity of cells in the visual cortex. Indeed, the same is true in the temporal cortex for faces and hands as well, so the tuning curve is a ubiquitous function of cells in the sensory parts of the brain. The same may also apply to the motor system.

What I am realising with more recent work is that these tuning functions emerge when physiologists map out single stimuli, one at a time. For example, you take a stimulus, which is good, you try it and you get a big cellular response. You take a stimulus that is weak, and get a small response from the cell, but each stimulus is mapped in complete isolation of every other stimulus. In the real world, things change continuously. As my head turns round and you see a sequence of views. As my hand reaches out and you see a sequence of views. When we use such real life dynamic stimuli, we find that the cell tuning curves change and I am going to argue that the cells become "predictors".

The original work that drove our own studies was that of Charlie Gross[1] who described cells in the temporal cortex that would respond to the sight of hands or monkey paws. Some time later, we were working with the temporal cortex in the same region and found much the same; cells that would respond to the sight of hand actions. What Giacomo Rizzolatti has described in temporal, parietal and premotor brain areas are components of the mirror system. The "mirror system" is now famous because it links the

[1] Gross, C. G., Rocha-Miranda, C. E., Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology*, 35, 96-111.

sensory representation of a hand action (in the temporal cortex) with the motoric representation (in parietal and frontal cortex).

It seems from the work by the Parma group and ourselves[2], that the representations in temporal cortex is entirely sensory. Cells in this region are sensitive to what the monkey witnesses but are not sensitive to what the monkey does. So, for the sight of a hand manipulating some fur there is a discharge from specific cells in temporal cortex. This sight provokes a good response on multiple trials.

Films where the hand acts alone or the fur deforms alone do not produce any response. Even for a film which is like magic, where the hand moves but is not in spatial contact with the fur, you do not get any impression of causality and the cells remain silent. This is one of the characteristics of temporal lobe cells and it is a characteristic that Rizzolatti's group replicated in the "mirror neurones" of the frontal cortex (area F5)[3]. For cells to fire, there has to be something causing a relationship between the agent performing an action and the substrate or the goal of the action. Hence we refer to the cells as coding actions rather than just coding hand or hand movements.

Following from the Parma group's work, where it became obvious that it was not just with vision that one could activate mirror cells. The sound of actions could also activate the cells. We started to explore audio-visual integration in temporal cortex. This region gets input from sound and from sight, so we should not be surprised if cells in this region are sensitive to both sound and sight.

So, we set about testing with films of actions where paper could be ripped or people could walk through the world. What we do is test the cell's responses to the sight of an action, and find whatever action they "like". We then work with that action. We will present the action visually and silently, or the correct sound of the action

---

[2] Keysers, C., & Perrett, D. I. (2004). Demystifying social cognition: a Hebbian perspective. *Trends in Cognitive Science,* 8, 501-507.

[3] Di Pellegrino, G., Fadiga, L., Fogassi, V., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research,* 91, 176-180.

alone, or present both sight and sound together. Alternatively we do stupid things, showing the film of one action and playing the wrong action sound at the same time (for example showing someone walking while the soundtrack is played in reverse). So Rizzolatti described F5 cells that would process either the sight or the sound of actions and such cells seem to occur in the temporal cortex region[4]. When you have got both sight and sound, you get a big response from temporal cortex cells. Sound alone produced nothing, but sound modulated the response to the sight. So what we can say is that about one quarter of the cells in the sensory part of the "mirror system" are already coding actions in an abstract way. The cells are integrating information from different modalities: the sight and sound of actions.

Coding is a little bit more exact than that. If we take the sight of one action and play the wrong sound, then we get a reduced response compared with when we make the correct pairing. So we can say that the cells require some kind of match between the sight and the sound. It is an open topic as to how the cells do this integration.

So, this is just to introduce you to sensory processing in the temporal cortex for actions and, it may well be that there is a route from here, through parietal cortex, to the mirror neurons in F5. I will focus on processing of individual images, which can be presented in random sequence. With images are occurring rather rapidly (1/20th of a second per image) it is difficult to see any one. If I slow the presentation down (1/5th of a second per image), of course you can see each and every image and you can work out who the faces are as they come up. We have used this rapid presentation technique to map out the sensory coding in temporal cortex. It is efficient because if the cells respond when you are going very fast, you can test many images.

What happens when you present thousands of test images? One of the things we have realised is that the cells can be extraordinarily specific. With diverse sets of images presented, each one for only

[4] Barraclough, N. E., Xiao, D., Baker, C., Oram, M. W., & Perrett, D. I. (2005). Integration of visual and auditory information by STS neurons responsive to the sight of actions. *Journal of Cognitive Neuroscience*, 17, 377-391.

50 milliseconds (equivalent to an individual frame on a television program embedded in frames of different images), these very brief images can provoke a cellular response[5]. A given cell may respond reliably whenever a picture of monkey face or a skull or a frog face is presented. There are many cells tuned to something about faces that will not respond to other stimuli.

When you use rapid testing, you find that cells in the temporal cortex code particular meaningful items, like faces; other cells code the hands and other cells code the sight of bodies. So that is telling us that one image, when it comes up briefly, would be sufficient to provoke a brain cell response. Another property we have known for a long time is that cells are tuned. One may respond to the front view but not to the back view. Other cells are tuned to other views: for example one cell responding only to the right profile view.

The majority of the cells in the temporal cortex are tuned to specific views or specific objects and you must not think of just one view being coded, there is no one privileged view; there are multiple views being coded, each by different sets of cells. I am showing you this because this is the background. When we use rapid testing —50 milliseconds per image—, again we can find that for one face we get view tuning. For a given cell any one of 8 different identities tested we found similar view tuning. So, all the properties we see with slow presentations seem to be there when the images are presented exceedingly briefly; maybe only for 50 milliseconds or even more briefly —14 milliseconds—[6].

So, cells in the temporal cortex can process very minimal information when a single image comes up. Summarizing: rapid testing actually shows you the brain can process up to 70 unrelated images every second and the cells are tuned for meaningful objects.

The main part of the talk is to consider what happens when images are presented in sequence and I would argue that

---

[5] Földiák, P., Xiao, D., Keysers, C., Edwards, R., & Perrett, D. I. (2003). Rapid serial visual presentation for the determination of neural selectivity in area STSa. *Progress in Brain Research,* 144, 107-116.

[6] Keysers, C., Xiao, D., Földiák P., & Perrett D. I. (2001). The speed of sight. *Journal of Cognitive Neuroscience*, 13, 90-101.

actions inevitably involve sequences of postures. But rather than jumping in the deep end and showing you complex actions, I want to start very minimally with sequences of two: just a pair of images because that is a minimal sequence. What we do is we present one image, we wait and then we present another. Of course, we can wait a variable amount of time before we present the second image.

What happens? We get a response to the first image —a nice big response—. Shortly after a response to one image, when we present an image only 200 milliseconds later, the response to the 2nd image is depressed. With a long interval between stimulus pairs of 450 ms, the 2nd image response that was suppressed has largely recovered.

In psychophysics, this is referred to as "masking" and we did not realise how prevalent this masking was at the cellular level. We find masking exists in virtually every cell we test and it lasts for about half a second. What would be the consequences if you went to the cinema and you were watching a scene? Well, actually the consequences are a bit bizarre. You see one image and then your brain is going to go to sleep, because it is going to suppress responses to subsequent images. So a new film would start, you would see a little bit of the film and then it would seem that your brain gets very, very insensitive to the next few images. So, that is a bit of a paradox for us, because we do understand what goes on in videos, so it cannot be true that our brain is completely disabled by the masking process.

One thing I have not yet considered is the similarity between a pair of images. We can test a cell with one image and measure a response. When we find a second stimulus that also produces a response, then it is most likely that the 2 images will appear to us to look visually or conceptually similar. The extent to which cells respond similarly to visual stimuli maps closely to how perceptually similar the stimuli are to human observers. What we have tested is pairs of images that are similar and pairs of stimuli that are distinct. When we use dissimilar images, we get less response to the first image but we get more response to the second image. So masking seems to depend on how similar the images are. And that is true with all the intervals we test.

So, masking is severe if the images are very similar. That is just what happens when you watch a video, or go to the cinema. All the images are pretty much the same. There is only a gradual change between images; unless it is MTV most of the images are actually quite similar.

So now I am going to move to sequences and ask what happens when you are processing a video. Here I am going to test, say five images: 1-2-3-4-5, and they are all going to be related. For any cell with a tuning curve we can start at the centre of the tuning with an effective image, and then work gradually through less and less effective views stepping progressively from the center to the edge of the tuning curve. So we will start with a good image, then we will show a sequence going away from that good image. When we test each image in isolation 1-5 we get a gradual decline in response. When we test the images in rapid succession as a film sequence we get a good response to the first image and then nothing. We can expect this rapid cessation from masking. A cell shows response to one image and then things get shut down. Data from many cells tested with an image sequence produce results like that I have shown you with for a particular example cell.

OK, so that is starting with a good stimulus and going outside the receptive field. Now I would now like to look at the effect of tuning for stimuli moving through the entire receptive field. We test images in the sequence again and we start with the weak stimulus, go to the best stimulus (at the center of the field) and then go on to weak stimuli.

What happens? Consider first a single cell that responds, not a face movement but to images of a hand action. The first image tested was a hand moving into the field of view, it is probably unrecognisable (as a hand) but it produced a small response. The second image (clearly a hand) produced a bigger response. The third image can be recognized as a monkey's hand grasping an object and these stimuli were relatively effective. When we test the subsequent frames, they are less effective. So frame number 3 produced the biggest response and that one a slightly smaller response.

Again, we can test it as a video, where we run through the frames showing a movie sequence. By this stage in the talk you can

guess what happens. We are going to get a good response early in the film sequence, then the response fades away. This is what happens, the cell response is maximal before frame number 3 and fades away quickly.

Well, again, we want to combine data for sequence testing across multiple cells. We find out what the best stimulus for each cell: it is the face this cell and for this cell it is the half-profile etc. We collect together responses of lots of different cells and look at the tuning for each one, relative to the best stimulus when that stimulus is tested in isolation. For a population of cells we have tested, if we use measure responses to views separately, we can find the responses get bigger towards the center of their tuning curve and then get smaller again. This is a classic receptive field; small response at the field edges, big response in the centre. These results follow testing of images separately. When we put the images together as a moving sequence, we find the response profile has shifted again earlier than should occur from separate image testing.

I have shown you a lot of data —and now want to draw back to the consequences—. I have said over and over again that when we show dynamic movements, then the tuning of all of these cells is shifted. The response to dynamic sequences peak before the best static image is reached. For the majority of cells we have worked with, this seems to be true. The shift in the timing of the peak activity is quite dramatic; it is about a tenth of a second in time. That is about half of the tuning curve. So, what was a familiar tuning curve in classic terms now has been distorted just by playing the stimuli in a natural order.

What is the consequence of all this for perception? Well, if we consider tuning of a particular cell for a face. Let's say the cell "likes" the front view of the head and it is responding maximally when the full face view is shown. When we show a rotating stimulus, e.g. my head turning around, then the maximum activity occurs much earlier than you would expect. As the head is turning, this particular cell is responding most before you have got to the front view. And that is my point with these tuning curves, the whole profile of activity in the brain is distorted.

During any natural sequence of images where stimuli change gradually, then the maximal activity in the brain at any moment occurs amongst cells that are tuned to views that have yet to occur. The most responsive cells during a dynamic action are tuned to postures that will occur should the action continue. The whole sensory apparatus is acting in a predictive capacity heralding what is just about to happen. A consequence is that our perceptual world does not reflect what is occurring at a given moment in time, but actually reflects a prediction about the likely future state. If the world is static and unchanging then brain activity will provide a veridical assessment of what is arriving at the sensory systems. As soon as the world includes natural change then the systems I have described will come in to play and the brain will be predicting the future state of the world. Of course normally the future state will occur, and we will be ready for it. Indeed that is the advantage of a predictive system, it helps us anticipate and prepare motor responses.

Our notion of sensory systems derives from physiologists measuring steady state responses to isolated stimuli. What I have reviewed here are two basic mechanisms, the first is the concept of a tuning curve, the second is the phenomenon of masking, or adaptation. Shortly following one stimulus, response to related stimuli is reduced. In vision this is called masking. Adaptation also occurs in processing of sound, and of touch. Indeed short term adaptation is found in all sensory systems. Its effect will be the same in all modalities. When we couple tuning curves for isolated stimuli, with short term adaptation, then we create predictive coding for natural dynamic stimuli[7].

I must emphasise the word "natural". The world does not change randomly. Any change is progressive and usually gradual. Objects do not appear, disappear, and reappear at random locations. No they move through space gradually changing position. Of course acceleration and deceleration is possible for animals and for objects colliding but there is no possibility

[7] Berry, M. J., Brivanlou, I. H., Jordan, T. A., & Meister, M. (1999). Anticipation of moving stimuli by the retina. *Nature,* 398, 334.

for objects to jump around. This may become possible in the hands of movie makers, but such cinematic creations are not "natural". The sensory system we have is one that is evolved, or is built up through learning to be matched to changes that take place in our world.

What is needed is a new conception of the nature of sensory systems. We need to think of our sensory systems as predictors, all the time they operate they are providing information about the likely state of the world. We should not think of perception and sensation as reality but we should realise that our experience is a prediction.

Question. Of course, processing of information is a fundamental aspect: audio and visual information. Assuming as a right approach that the predicting approach you are presenting this afternoon, I am guessing that some kind of patterns are going to be needed, in order to make some comparisons or some predictions, in order to create more or less the feature, or the next step perception, either visual or audio information processing. In such a case, how are these patterns going to be restored and how really is this processing turning out?

A. Your question goes to the implementation stage and a recent paper on the implementation of predictive coding in the retina was pointing not cells but to synapses as the location of adaptation[8]. The paper was claiming adaptation was dynamic. If your question is one about implementation, I can speculate. Feedback from the neighbouring cells could do this, because the neighbouring cells are actually detecting the same stuff. They are detecting visually and conceptually similar stuff, so their signal would be good for shutting down the brain in that region. So when you have seen one image, the next image to occur will be processed by roughly the

[8] Hosoya, T., Baccus, S. A., & Meister, M. (2005). Dynamic predictive coding by the retina. *Nature*, 436, 71-77.

same brain region. Well, all the neighbours processing that stuff will have collateral feedback, they could clamp the system down.

If your question was really about implementation then the following may be relevant. In another set of experiments, what we did here was we waited for a cell to respond well or poorly to the same stimulus and then measured what happened to the following stimulus. So image 1 is always the same and you get a good response on some trials and a poor response on other trials. The magnitude of the response to the second stimulus did not depend on the magnitude of the response to the first stimulus. We get equivalent masking, so that suggests it is not the poor little cell getting tired that causes masking. It is the neighbours shutting the cell up. Even when the cell has no cause to be tired because it was not responding very well and yet, then when the next stimulus comes up, we get just as much masking.

I am trying to answer the question: how is this masking and predictive coding implemented? And I think it is collateral feedback from neighbouring cells. That would be one simple way of doing it. Take the organised output from the local surround and use it to clamp back on the system. That stops the brain going epileptic, but you could achieve masking in many ways. "I don't know" is the real answer, but I suspect that this feedback is one way of doing it. Perhaps that was not your question, though?

Q. No, but in any case it is very illustrative to consider the cell level, but my question is more orientated to the functional level.

A. Yes, we can take a system level interpretation. Many of the brain imaging experiments are now done using an fMRI adaptation paradigm. Show two stimuli and the second stimulus should evoke a smaller response if it is related to the first. It is a repetition effect. With evoked potentials the same happens for a stimulus pair. We can wipe out the second evoked potential if the stimulus is conceptually similar to the first. So these are system level responses.

For the organism, what is the function? Well, I have tried to argue that one function might be prediction, otherwise, it is a strange mechanism: just get tired, stop responding. That doesn't help anybody. I think we need to think of masking and adaptation

in a different context; a context of how the brain processes naturally changing stimuli. When we do this the functional consequence of masking and adaptation is the capacity to predict.

Professor Jeanerod

Q. When you show actions, the frequency of the images can change the precept.

A. Maggie Shiffrar and Jean Decety[9] showed an image of one leg behind the other, and then and then, 1/2 a second later showed an image with the legs swapped around. Everybody thought, "yes, the leg is just going round" the movement was anatomically possible. But if the images are separated by only 1/10th of a second, people just think, "Oh, this leg has moved through the other one!". The interpretation of an action sequence depends on the speed of the movement and the frequency of the images.

The timing of the images in a movie of a natural action does not invalidate what I have been saying. For different rates of presentation or speeds of movement, the prediction will be more or less stretched into the future. If a movement occurs at a higher speed then the predictive coding will work over the same time scale but the predicted position or posture anticipated will be progressed further along the movement trajectory.

Q. In this case, changing the frequency of the presentation means that the effect is at a response level the same as other receptive cells.

A. We have not really looked at the frequency but with the Shiffrar experiment, there is a change in space and time, so you compute an actual velocity. Hence the illegal movement occurs when it happens too quickly for the joints to work. There are two considerations: the degree of the limb displacement and the timing. The example I gave for the impossible movement is rather slow, 1/2 a second, is much slower than television frame rates.

---

[9] Stevens, J. A., Fonlupt, P., Shiffrar, M., & Decety, J. (2000). New aspects of motion perception: Selective neural encoding of apparent human movements. *Neuroreport*, 11, 109-115.

There is a whole cottage industry in perception, about representation momentum[10]. If you change the position of an object progressively in images 1, 2, 3, 4 and say, where is the object in the last image? People get it wrong. They overestimate the position along the motion trajectory. There is a time course for the prediction commonly found in representation momentum at about this speed, and even if the presentation rate is slightly slower. I think that whatever image frequencies we play with, so long as we are less than half a second between images, we are going to find anticipatory effects. If we go much slower, maybe over a second between images, I don't think we are going to see anticipation.

We have not done the experiments with slow image rates. We started off with an ethological approach. Let's take film of natural actions and let's take the film images apart. Let's take a film of walking and play it at natural frequencies and see what we get. You know, how we spend most of our time watching television and film but we don't know what is really happening to image processing at TV image presentation rates. So we have not played around much with timing. I think that within half a second, we will still get effects.

Q. You reported that when you present the same picture, the same image ten times, the amplitude of the response is weaker and weaker in the one cell.

A. If it is a pretty similar the image, the response is getting weaker, yes.

Q. It is weaker, yes, but psychophysically, if we present the same or similar pictures ten times, does our perception impression gets weaker or does it stay the same?

A. I would ask the visual psychophysicists here – Stuart Anstis described the slowing of movement, if you watch a continuous movement it gets slower phenomenally; if you adapt to the bright sun, it gets phenomenally less bright. That is the same image seen repeatedly across time. We tried masking psychophysically with pairs of different images. We do get an impairment of reaction

[10] Freyd J. J., & Finke R. A. (1984). Representational momentum. *J Exptl Psych: Learning, Memory and Cognition,* 10, 126-132.

time to detect a second stimulus, we have not shown this to be based on image similarity presenting the same image multiple times. Does anybody else know whether you get a diminution of the qualia with repeated exposure?

Q. The thing is, how close are these single cell responses to subjective experience? Sometimes, in optical illusions, the absolute similarity between cell responses and subjective experience is there. If it is not the case, only one other interpretation that our subjective experience here does not come from single cell directly but from memory. So, we often see things through the eyeglasses of memory and not from actual experience, so they would be crucial to how close it is?

The best experiments have been done by Nikos Logothetis[11] and in his hands you can get quite similar reports about perceptual experience from the monkey and the activity of single cells. Basically, in our studies if we measure perception in humans during rapid presentations, then performance is pretty close to the ability of cells to detect the same images in the same rapid presentations[12].

Q. It is astonishing how good sometimes these cells are —quite unexpectedly—.

A. Yes, and as Giacomo Rizzolatti was reporting: every time the monkey reaches for this, you hear a cell recorded responding "crrrrrr, crrrrrr, crrrrrrr!", there is no doubt, you don't have to do that fancy statistics. Single cells can be very high fidelity of coding. They are not always in the noise.

Q. If that is the case, from the beginning of the visual system to the end, you look at "Equity" for example. That is a brilliant example that you need more than one cell to see things better than one cell can do. And another thing is if you look at experiments from Bill Newsome in area MT, which is a brilliant example as well. The cell can predict perfectly what is the reaction of the

---

[11] Sheinberg, D. L., & Logothetis, N. K. (1997). The role of temporal cortical areas in perceptual organization. *Proc Natl Acad Sci USA*, 94, 3408-3413.

[12] Keysers, C., Xiao, D., Földiák, P., & Perrett, D. I. (2001). The speed of sight. *Journal of Cognitive Neuroscience,* 13, 90-101.

moment and you can manipulate that cell with electrical stimulation or whatever and you can change the judgement of the whole subject.

A. This is excellent physiology work where people made the task equivalent to what the cells were thought to be doing and then they get comparable performance from the whole observer and from individual brain cells. The work of Newsome was fundamental in that sense.

Q. I have a comment on your talk, Professor Roth. You were saying that the limbic system is pushing the basal ganglia as a kind of motivation but, which structure is pushing the limbic system? Or in other words, is it allowed to talk about initiation in the nervous system at all?

A. If you look at it ontogenetically, long before birth there is a most primitive evaluation system already in the amygdala, which is working as I said before, from the 7th week of gestation on. And there is a very primitive evaluation, which is not circular, it is spiral. So at the very beginning there is good and bad, so the unborn child makes some movement or does something and there is a very primitive evaluation: this consequence was good or it was bad. And this is made mostly by the so-called central amygdala and the BNST and, afterwards, the basal lateral amygdala, which is connected to the cerebral cortex after birth, is learning always more complicated things, so there is increased self-evaluation. This is a basal process in the limbic system, so it is gaining experience in a spiral way. This is as far as we know, the primary process that drives all our experience. There is sensory input, motor output and checking the consequences of the motor output against the sensory input. And this is extremely primitive before birth but after birth it is increasingly complex, with increasing cortex. So there is no other evaluative system in the brain.

Well, if you destroy the amygdala in adulthood, or have patients with the very rare Pick's disease that can specifically lesion the amygdala bilateraly, they are incapable of recognizing fearful faces or fearful events.

EL CEREBRO: AVANCES RECIENTES EN NEUROCIENCIA

Comentarios introductorios

Francisco J. Rubia

En el opúsculo *Una dificultad del psicoanálisis,* Sigmund Freud (1856-1939) argumentaba que el ser humano había sufrido a lo largo de la historia tres humillaciones importantes en su amor propio. La primera, consumada por Nicolás Copérnico (1473-1543), acabó con el geocentrismo, es decir, con la concepción de que la Tierra era el centro del universo y de la creación. La Tierra no era más que un planeta del Sol, y no de los más importantes. Hoy esta idea no sólo está confirmada, sino que sabemos que el Sol no es más que uno de los millones de soles que componen una de las muchas galaxias que existen, por lo que la importancia de la Tierra ha ido disminuyendo a pasos agigantados.

La segunda humillación provino del biólogo Charles Darwin (1809-1882) y de su teoría de la evolución, que en la actualidad nadie pone en duda, excepto algunos grupúsculos de cristianos creacionistas en Estados Unidos. Sin embargo, después de casi 150 años —esta teoría se publicó en 1859—, todavía hay personas que no han asumido lo que ella significa: nuestra procedencia de animales que nos han precedido en la evolución. Esto supuso sin duda un gran golpe para la idea de que éramos la perla de la creación divina que habíamos sido creados por un soplo de la divinidad, como se dice en el Génesis. Con ello, la explicación de la Biblia pasó a ser lo que es: un mito o leyenda como muchas otras. Esto no quiere decir que no se discutan ni se estudien los mecanismos de la evolución, tema que seguirá siendo controvertido hasta que pueda explicarse con satisfacción.

Por último, según Freud, la tercera humillación vendría dada por su descubrimiento, que no fue tal, del inconsciente. Éste ya ha-

bía sido descrito a lo largo del siglo XIX por varios médicos naturalistas románticos alemanes, pero Freud hizo de él el centro de sus estudios y le dio una importancia que otros no le habían dado. El resultado de esos estudios fue saber que la conciencia era sólo la punta de un iceberg y que debajo del agua se encontraba una inmensa mayoría de funciones que, a pesar de ser inconscientes, gobernaban y dirigían la conducta humana. La tercera humillación, pues, implicaba que el ser humano no era ni siquiera dueño de muchos de sus actos. Hoy se calcula que, de todas las operaciones que el cerebro realiza, sólo un 2% serían conscientes; el resto se lleva a cabo sin que sepamos que se están realizando. En otras palabras: probablemente Freud se quedó corto.

Pues bien, nos aguarda una cuarta humillación, de la que sólo vislumbramos hoy su comienzo: una revolución neurocientífica, que está a punto de poner en entredicho convicciones tan firmes como la existencia del yo, de la realidad externa o de la libre voluntad. Temas todos estos que tradicionalmente no han sido objeto de estudio por parte de las ciencias naturales, convencidos como estábamos de que eran objeto de la filosofía o, como mucho, de la psicología, pero que hoy sí que se cuentan entre los objetos de estudio de las neurociencias, descubriéndonos que hemos estado equivocados hasta ahora al dar carta de naturaleza a determinados conceptos que, muy posiblemente, eran fruto de nuestros deseos.

El ser humano no tiene, por ejemplo, ningún motivo para pensar en la continuidad de su persona, de su yo, que considera que es el mismo desde la cuna a la tumba, sabiendo que nada ni en su cuerpo, ni en su mente, ni en su alrededor tiene permanencia. Y, sin embargo, ¿quién nos va a convencer de que no existe ese yo que subjetivamente está tan presente como la propia realidad exterior? Nuestra impresión subjetiva nos dice que tenemos dentro una especie de homúnculo que es el que reflexiona, actúa y permanece constante durante toda la vida. Ahora bien, ¿podemos fiarnos de nuestras impresiones subjetivas?

Los órganos de los sentidos nos han engañado desde siempre y lo sabemos, como ya lo sabían los filósofos griegos de la naturaleza en Asia Menor. Las neurociencias modernas nos dicen que ni los colores ni los olores, ni los sonidos existen en la naturaleza, sino

que son creaciones del cerebro. Por cierto que ya en el siglo XVIII el filósofo napolitano Giambattista Vico decía exactamente lo mismo. Sin embargo, ¿hemos asumido esta realidad? ¿No estamos todos convencidos de que esas «proyecciones» del cerebro no son tales y que las cualidades de nuestras percepciones son un reflejo de la realidad que percibimos?

Por otro lado, ¿podemos fiarnos de nuestra propia mente? Nuestras capacidades cognitivas, ¿no están más al servicio de la supervivencia que para especulaciones filosóficas o para reflejar una realidad de la que no sabemos hasta qué punto es construcción cerebral? Ya Kant advirtió que nuestra mente podía estar limitada por juicios sintéticos *a priori* que nos harían ver el mundo de una manera apropiada quizá para nuestra supervivencia, pero no tan libremente como suponemos. Mi opinión es que todavía no se ha asumido totalmente lo que Kant planteó hace ya más de dos siglos en su *Crítica de la razón pura.* Hoy sabemos también que la idea de que nuestra mente es una tábula rasa, tal como proponían los empiristas, no es más que otra ilusión. Como decía William James, si los animales que nos han precedido, con un cerebro menos desarrollado que el nuestro, tienen al nacer toda una gama de «instintos» adecuados para desenvolverse en el mundo exterior, los seres humanos no deberían de tener menos instrumentos apropiados para ese entorno, sino todo lo contrario, muchos más.

Nos experimentamos a nosotros mismos como seres que poseen una intencionalidad, nos atribuimos a nosotros y a los demás la responsabilidad de lo que hacemos, pensamos que podemos dominar nuestros estados de ánimo y que estamos en condiciones de decir «no» a los determinantes de nuestras acciones. Tenemos la sensación de que podemos controlar en todo momento lo que pensamos, hacemos o deseamos. Pero ¿es eso cierto? ¿O estamos sometidos a los determinismos de los procesos fisicoquímicos que tienen lugar en el cerebro?

En la actualidad, tenemos datos que revelarían que la libre voluntad puede ser una ilusión. Nadie puede afirmar que estos datos sean definitivos. No existen datos definitivos en ciencia, pero parece lógico pensar que si hemos aceptado que la mente es producto del cerebro y éste es pura materia, lógicamente el cerebro tendrá

que estar sometido a las leyes deterministas de la naturaleza como todo lo demás. Y, sin embargo, ¿quién nos convence de que no somos libres de tomar decisiones cuando existe la posibilidad de elección entre varias opciones?

La mayoría de los neurocientíficos han renunciado ya a la vieja idea de que cerebro y mente, o cuerpo y alma, son de naturaleza distinta tal como proponía Descartes (1596-1650). Pero ¿significa eso que hemos dejado de ser o pensar de forma dualista? ¿No estamos constantemente analizando lo que experimentamos en términos antitéticos, en antinomias? Desde los misterios órficos en la Antigüedad hasta nuestros días, el dualismo ha ocupado nuestras mentes. Quizá el dualismo sea un producto de una tendencia innata del cerebro a ver el mundo en términos opuestos. Esto explicaría por qué la mitología, la filosofía y las ideologías están llenas de visiones dualistas de la realidad. La tajante separación de cuerpo y alma en Descartes apenas es considerada hoy por las neurociencias, entre otras cosas porque la interacción de ambos contradice las leyes de la termodinámica. Curiosamente, esta separación permitió en su momento ocuparse del cuerpo sin entrar en conflicto con la Iglesia, lo que supuso el comienzo de la anatomía, de la fisiología y —como dicen algunos— de la ciencia moderna. Sin embargo, en mi opinión, lo que entonces supuso un avance considerable, durante demasiado tiempo ha sido una rémora que ha impedido el estudio de los fenómenos mentales con métodos científicos. No tenemos ninguna prueba de que existan en el cerebro operaciones que no se deban a impulsos que provienen del propio sistema nervioso. La hipótesis, pues, de una influencia extracerebral, inmaterial, sobre este órgano no satisface hoy a la mayoría de los científicos, sobre todo porque nadie puede explicar cómo sería esa influencia. Una interacción con la materia exige un intercambio de energía; para movilizar energía, cualquier ente inmaterial tendría que disponer de ella, pero para ello tendría que dejar de ser inmaterial. Pero, además, esa explicación dualista cierra las puertas en la práctica, como ha hecho en el pasado, al estudio científico natural de fenómenos como la conciencia o la mente. La reciente producción de experiencias místicas, espirituales —con la estimulación electromagnética del lóbulo temporal— apunta, a mi entender, a

un apoyo a la teoría monista de que la separación entre materia y espíritu no es más que un producto cerebral, pero no real fuera de nuestro cerebro.

Como vemos, las neurociencias, que anteriormente consideraban inapropiado ocuparse de estos temas, ahora lo están haciendo, pero para terminar intuyendo que hemos estado completamente equivocados respecto a lo que significan. Es decir, estamos al comienzo de un derribo sistemático, de una deconstrucción de conceptos que, algunos de ellos, son pilares en los que se asienta —nada menos— gran parte de nuestra cultura occidental. Porque, ¿qué vamos a hacer cuando nos convenzamos de que el libre albedrío es una ilusión cerebral? ¿No significa esto poner en tela de juicio toda nuestra cultura y civilización basada en la responsabilidad personal?

Éstas son las razones para pensar que se avecina una nueva humillación del ser humano, una nueva revolución protagonizada por los resultados de las neurociencias. De nuevo, una ciencia está a punto de abrirnos los ojos a realidades que nada tienen que ver con las que hemos vivido durante siglos; éstas han sido producto de nuestro cerebro y las realidades que las sustituyan posiblemente también lo serán. Y es que a partir de este momento soñar con una realidad independiente del cerebro humano será posible pero no real. El mundo del murciélago es tan limitado como lo es su propio cerebro; el nuestro, aun siendo más complejo, sigue estando circunscrito por las limitaciones impuestas por este órgano.

El descubrimiento de neuronas en el área premotora F5 del cerebro, que responden no sólo cuando el animal realiza un movimiento, sino cuando observa el mismo movimiento en sus congéneres, ha abierto las puertas al estudio de las bases neurobiológicas de la imitación, una facultad importantísima en la adquisición de habilidades que ha hecho que nuestra cultura y el desarrollo ontogenético del niño sean impensables sin ella. La imitación es uno de los ejemplos más interesantes de la coordinación perceptivo-motora, en la que el sujeto tiene que traducir un patrón visual muy complejo en comandos motores de forma que resulten en un movimiento que coincida lo mejor posible con el observado. Se ha propuesto que este sistema de neuronas especulares es la prueba de la validez de una teoría de la simulación según la cual los seres hu-

manos utilizarían sus estados mentales para predecir, anticipar o explicar los estados mentales de otras personas. Es éste un tema apasionante, que se ha denominado teoría de la mente y que está relacionado con la evolución del cerebro y del lenguaje.

Nuestra capacidad para entender a los demás, lo que se ha llamado siempre psicología popular, ha dado lugar a todo un campo de investigación en las ciencias cognitivas. Algunos filósofos plantearon que esta capacidad dependía de nuestra habilidad para el lenguaje. Estudios realizados en primates no humanos han mostrado que también poseen un entendimiento intencional que es útil para su vida en sociedad, lo que llevó a la conclusión de que la teoría de la mente era una dotación biológica independiente del lenguaje. Los primates no humanos que han establecido una sociedad con sus reglas y relaciones mutuas ya poseen lo que se ha llamado una inteligencia social, o maquiavélica, importante para las necesidades de supervivencia, como son las capacidades de manipulación, engaño y cooperación. Todo esto también ha sido considerado como el desarrollo de capacidades independientes del lenguaje.

El hecho de que se desarrolle en el niño aproximadamente a los 3 y los 4 años y que sea independiente del desarrollo de otras capacidades cognitivas hace pensar que la teoría de la mente es una capacidad específica que puede perderse aisladamente como se supone sucede en el autismo. Los niños autistas fallan en tareas típicas que muestran la capacidad de entender las intenciones de los demás, independientemente de su inteligencia o de su lenguaje.

Los próximos capítulos completarán, matizarán o corregirán las hipótesis que he señalado en estas páginas, aclarando, gracias a los nuevos conocimientos que nos aportan las neurociencias, muchos de los conceptos erróneos que tenemos sobre nosotros mismos.

La relación entre la razón y la emoción y su impacto sobre el concepto de libre albedrío

Gerhard Roth

¿Qué es lo que guía la conducta humana? ¿La razón o las emociones? ¿Existe realmente el libre albedrío? Éstos son temas que el hombre ha discutido desde hace más de 2000 años y, hasta hace muy poco, se pensaba que la ciencia y las neurociencias no podían aportar nada a estos problemas, considerados teológicos, filosóficos o quizá psicológicos. Sin embargo, como probablemente se sabe, a raíz de un gran número de nuevos descubrimientos y métodos que han aparecido recientemente, se ha reabierto el debate. Al menos en los países de habla alemana, desde hace por lo menos 3 años existe un enorme debate sobre estos temas y todos los días aparece en la prensa algún artículo a favor o en contra de filósofos, psicólogos, neurocientíficos, políticos y, muy recientemente, incluso gente de la Facultad de Derecho. En este momento existe incluso una iniciativa para modificar el Código Penal en Alemania ante estos nuevos conocimientos.

Cuando hablamos del libre albedrío, necesitamos definir, aunque sea brevemente, lo que entendemos por este concepto. Hay algo que yo llamo el concepto firme del libre albedrío, que es la base del concepto tradicional y filosófico, así como del derecho penal, al menos en la Europa continental. Esto significa que las acciones de mi cuerpo se deben, al menos en parte, a un libre albedrío *no material*. Esto se conoce como *causalidad mental*: algo totalmente inmaterial que causa al menos parte de mis acciones, aunque no to-

das. Ése es el primer punto e implica que ante las mismas condiciones materiales y físicas, es decir, ante las mismas condiciones físico-fisiológicas, puedo actuar de distinta forma sólo mediante un acto de voluntad. Considerando todas las situaciones y todas mis motivaciones, yo podría, en un mundo puro e inmaterial, decidir hacer otra cosa. En eso se basa el concepto de culpa del derecho tradicional. Es lo que se conoce como «alternativismo».

En segundo lugar, mis acciones *no* están totalmente determinadas por procesos cerebrales. Ése es el concepto de «indeterminismo» y, una vez más, es la base del derecho penal tradicional, al menos en la Europa continental. Y eso significa o indica que el «determinismo» es incompatible con el libre albedrío. Esto se denomina *incompatibilismo*.

Por supuesto, se trata de conceptos distintos y, según el profesor Pauen, existe un resurgimiento del concepto débil o «compatibilista» del libre albedrío. Brevemente, esto significa que mis acciones son libres si hay consecuencias o deseos, planes, o actos de voluntad *asentados en mi personalidad o en mi carácter*. En segundo lugar, mis acciones son libres si son la consecuencia de una deliberación racional. Esto es muy importante. Básicamente, entonces, el libre albedrío es la autodeterminación o la autonomía de la acción. Y todo esto permitiría concluir que el libre albedrío es compatible con el determinismo.

Por tanto, hay dos conceptos del libre albedrío: el incompatibilista y el compatibilista. Sin embargo, desde mi punto de vista, es importante que en ambos conceptos la *racionalidad* tenga un papel dominante en la toma de decisiones y guíe las acciones, ya que por lo general, aunque no siempre, creemos que somos libres para actuar y tomar decisiones en la medida en que la racionalidad o la deliberación nos guían. Entonces, la pregunta es: ¿cuál es el papel específico de la racionalidad en la toma de decisiones y la guía de las acciones en términos neuropsicológicos?

Ambos conceptos, el libre albedrío y la relación entre la racionalidad y la irracionalidad están estrechamente relacionados. ¿Cómo podríamos definir la razón (*Verstand*, en alemán) o el entendimiento (*Vernunft*, en alemán)? En primer lugar, la razón se define habitualmente como la solución real de problemas mediante un pensa-

miento inductivo y deductivo. Eso es la inteligencia. Y el entendimiento puede definirse como la planificación a medio y largo plazo, mediante la aplicación de principios racionales superiores, en especial los referentes a las consecuencias sociales de las acciones. Es decir, ¿cómo interpretará mi familia, mis compañeros, y la sociedad, lo que estoy haciendo? ¿Como algo bueno o algo malo?

Veamos ahora el cerebro humano: la parte anterior, la posterior, la corteza, el cerebelo y la parte más frontal de esta gigantesca corteza, a pesar de que no es la mayor corteza que existe —hay animales que tienen 8 ó 9 kg de cerebro en lugar de 1,4 kg, y una corteza mucho mayor, pero afortunadamente, no sabemos lo que hacen con ella, por lo que es mejor no intimar mucho con elefantes y ballenas…—.

Esta corteza contiene aproximadamente 12.000 millones de neuronas, y casi medio billón de sinapsis. La corteza es la sede de la conciencia. Todo aquello de lo que somos conscientes está vinculado a la actividad de la corteza y todo lo que tiene lugar fuera de la corteza no va acompañado de conciencia. Así, siempre que tenemos actos de libre albedrío, en el sentido tradicional, deben estar situados en la corteza. No nos asombra, por tanto, saber qué partes de la corteza son la sede del entendimiento.

Hay también una parte inferior, la llamada corteza frontal orbital, debido a su posición sobre la órbita de los ojos, que es la sede del yo ético y moral, de las consideraciones éticas y morales, de la consideración de las consecuencias a largo plazo y, en consecuencia, podemos decir que también lo es del entendimiento o de la *Vernunft* de Kant.

Normalmente, podemos planificar una acción. Podemos imaginar que primero existen los planes y deseos y después, se diseña la forma más inteligente de llevarlos a cabo debido a las funciones situadas en esta corteza frontal dorsal. Después, reflexionamos sobre las consecuencias individuales y, en especial, sobre las sociales; esto se debe a la actividad de la corteza frontal orbital. Así decidimos lo que vamos a hacer. Esta actividad, según se cree tradicionalmente, se envía a las cortezas premotora y motora, que ordenan las acciones voluntarias. Desde hace mucho tiempo, se sabe que la llamada corteza motora primaria guía los movimientos simples

(por ejemplo, los de los dedos de una mano). Después, está la premotora, una corteza suplementaria que guía los movimientos más complejos, y un área específica, la motora presuplementaria, que también se activa cuando imaginamos el movimiento (por ejemplo, el movimiento de un dedo o de una mano). Siempre que hacemos algo de forma intencional y consciente, esta área motora presuplementaria debe estar activa.

Hace muchos años, el neurofisiólogo y filósofo John Eccles creía que esta área motora suplementaria o, como se conoce ahora, presuplementaria era la sede del libre albedrío. O más precisamente, la parte del cerebro conectada más estrechamente con el libre albedrío inmaterial. Podríamos decir que es el cerebro de «enlace».

Por encima de estas áreas, en la corteza motora primaria, en la premotora y en el área motora suplementaria, es posible registrar los llamados «potenciales preparatorios». Éstos son absolutamente necesarios para realizar cualquier tipo de acción voluntaria y tienen un papel muy importante en la discusión filosófico-neurológica actual. Se trata de potenciales negativos que reflejan, según creemos, la máxima actividad sincrónica de las neuronas del área motora suplementaria o el potencial de preparación somático y, posteriormente, el potencial de preparación lateral que precede inmediatamente a nuestras acciones. Y las acciones motoras sólo se producen cuando el potencial de preparación lateralizado ha alcanzado una determinada amplitud. Entonces, se positiviza la actividad premotora, lo cual induce directamente la acción motora.

Teniendo esto en cuenta, ahora podemos comprender mejor lo que ocurre cuando tenemos un deseo o un plan de hacer algo.

Las cortezas suplementaria, premotora y motora, así como la motora donde el potencial de preparación alcanza una determinada amplitud, estimulan el tracto piramidal y, por último, los músculos. Y se piensa que ésta es toda la explicación. Éste sería un concepto neurobiológico de cómo el libre albedrío, situado en la corteza prefrontal y frontal, podría controlar nuestras acciones. Sin embargo, parece que la racionalidad y el entendimiento, la corteza prefrontal y la frontal orbital, guían nuestras acciones activando a la corteza prefrontal, que a su vez activa a las áreas premotora y mo-

tora de la corteza. Sin embargo, si la explicación acabara en este punto, no habría pacientes con enfermedad de Parkinson.

Los pacientes con enfermedad de Parkinson tienen dificultades para iniciar las acciones que desean realizar, aunque pueden llevar a cabo movimientos iniciados externamente o sumamente automatizados, como andar. Esto se conoce bien a partir de los datos de los estudios clínicos. Y los pacientes con Parkinson no tienen deficiencias corticales primarias, sino una producción deficiente de dopamina en los ganglios basales y, por tanto, en cuanto se administra a estos pacientes una dosis suficiente de L-dopa, que se metaboliza en la dopamina, son capaces de hacer lo que desean, al menos durante un rato. Así que, extrañamente, la dopamina afecta de alguna manera al libre albedrío o, al menos, a su ejecución.

¿Cómo entender esto? Consideremos los ganglios basales, en la parte más profunda del cerebro. Cuando hacemos un corte transversal del hemisferio, nos encontramos con unos núcleos compactos de gran tamaño, el primer y el segundo ventrículos, y el principio del tercero. Tenemos también la sustancia gris de la corteza y la sustancia blanca de estos ganglios basales, así como los núcleos caudado, putamen y pálido, que contienen unos 600 millones de neuronas que constituyen la masa subcortical de mayor tamaño y la cortical, y el telencéfalo.

¿Qué hacen estas estructuras? Inicialmente, se pensaba que allí residían las acciones reflejas, instintivas o automatizadas, pero recientemente se ha descubierto que participan necesariamente en todas las acciones voluntarias de la siguiente forma.

Si la corteza cerebral decide hacer algo, a través de la corteza prefrontal, no puede hacerlo simplemente a través del tracto piramidal a los centros motores de la médula espinal, sino que debe enviar la actividad al estriado. La información tiene que ir a la sustancia negra y volver, y de ahí al pálido dorsal; luego al tálamo, la estación de relevo, y después volver a la corteza cerebral. Así, debemos aceptar que la corteza neuronal no puede solicitar acciones voluntarias sin estimular estas estructuras subcorticales que funcionan de forma totalmente inconsciente. De hecho, no se trata de un solo circuito, sino que hay al menos cinco circuitos distintos. Podemos distinguir dos circuitos motores, uno motor a estas áreas corticales y subcor-

ticales, y uno óculo-motor, ya que el sistema ocular es siempre un sistema neurobiológico especial, en el que la atención está ligada principalmente a las acciones motoras oculares. Después, tenemos el circuito ejecutor: éste no procede del movimiento, sino de la razón y del entendimiento; baja y luego vuelve; y después tenemos el circuito límbico, que es el más importante para el habla. De dos áreas límbicas, la corteza frontal orbital y la cingulada, éste es el circuito emocional, el ejecutor y el llamado circuito cognitivo.

¿Por qué esto es así? Para empezar, veamos cuál es el problema. El problema es que siempre que queremos hacer algo, la corteza y los ganglios basales, en su totalidad, deben decidir, primero, si es apropiado realizar esa acción planificada y, después, cómo llevarla a cabo, y la corteza realmente no lo sabe. Sólo hace algunas sugerencias y los ganglios basales deben determinar cómo hacerlo de forma sencilla y precisa. Debemos emitir el programa motor completo y exacto para las acciones que deseamos realizar y suprimir todos los demás programas motores. Esto es sumamente difícil. Tenemos miles y miles de posibles programas motores. Todos ellos, excepto uno, deberán ser inhibidos. Esto se hace en los ganglios basales de una forma totalmente fascinante, ya que la corteza estimula al estriado, la información va a la sustancia negra, vuelve, va al pálido y vuelve a la parte reticulada de la sustancia negra. De ahí va al núcleo subtalámico, después al tálamo y vuelve. Así, hay activación, inhibición, inhibición, activación; activación, inhibición, inhibición, inhibición, activación; e inhibición, activación. Es una organización sumamente compleja, en la que predomina un efecto inhibidor y algo de actividad. Y la razón es que hay una vía que libera sustancias inhibidoras al tálamo para poder activar la corteza y una vía indirecta que suprime los planes motores no deseados.

Así que esta organización sumamente compleja, primordialmente inhibidora y con cierto grado de activación, y que aún no comprendemos totalmente, es necesaria para ejecutar un programa e inhibir todos los demás.

Pero ¿cuál es la función global de los ganglios basales? En la actualidad, se cree que los ganglios representan la memoria de acción, en el sentido de que cualquier movimiento que se haya ejecutado con

éxito, incluso antes del nacimiento, se guarda aquí y se convierte en acciones más automatizadas. Así, una vez que hemos aprendido y automatizado algo, ya no necesitamos la conciencia, simplemente lo hacemos. Sin embargo, cuando debemos hacer algo voluntariamente, necesitamos la conciencia, los ganglios basales también deben estar de acuerdo. Esto significa que antes de que la corteza ejecute cualquier acción voluntaria, el deseo o el plan debe enviarse a los ganglios basales para compararlo con esta memoria de acción.

En otras palabras, todo lo que queremos hacer debe ser compatible con nuestra memoria de acción; es decir, no contradecir nuestra experiencia motora pasada. Se trata de una señal de la dopamina de la parte compacta de la sustancia negra. Y esta dopamina influye en dos tipos de neuronas, con un efecto excitador e inhibidor. De esta forma, aumenta esta vía inhibidora directa y con ello la inhibición, y se liberan sustancias inhibidoras de forma compleja; así que, en ausencia de dopamina, todo está bloqueado. Esto explica por qué los pacientes con Parkinson no pueden hacer ningún movimiento aunque lo deseen vehementemente: simplemente porque falta esta señal de la dopamina.

De este modo, la actividad inconsciente de los ganglios basales es absolutamente necesaria para la ejecución de las acciones voluntarias conscientes, lo que, en sí mismo, parece encerrar una contradicción. Para hacer algo consciente y voluntariamente, es preciso que los ganglios inconscientes lo acepten, aunque no nos demos cuenta.

Y ahora surge la pregunta: ¿qué o quién controla a los ganglios basales? Ya que en el cerebro no hay nada que no controle algo y que no sea, a su vez, controlado por otra cosa, la respuesta es el sistema límbico o, más específicamente, el hipocampo, la amígdala y el sistema límbico basal. Repasaremos muy brevemente la amígdala y el sistema límbico basal.

Por el momento sólo voy a mencionar a la amígdala, situada en la parte inferior de la corteza temporal y que está estrechamente asociada a los ganglios basales.

La amígdala cortical está ligada a las emociones relacionadas con todos los afectos, mientras que la amígdala medial está relacionada con las feromonas, esto es, sólo con los afectos sociales.

A continuación tenemos la amígdala central, estrechamente relacionada con la medial y la cortical. La amígdala central guía todas las funciones viscerales y autónomas y, junto con el hipotálamo, es el lugar principal de regulación del estrés.

Pero la parte más grande es la amígdala basolateral, donde reside el condicionamiento emocional y las emociones innatas. Aquí es donde aprendemos, desde antes de nacer, lo que es bueno y malo para nosotros, de forma totalmente inconsciente. Más tarde, cuando se forma la corteza prefrontal, se establecen las estrechas conexiones entre ésta y la amígdala y la corteza frontal orbital y el hipocampo, de forma que la amígdala ejerce una fuerte influencia sobre las cortezas motora y frontal y otras partes del sistema límbico, la corteza insular, la prefrontal media y el hipocampo; por lo que las emociones se producen de forma inconsciente y esta actividad se envía a la corteza, donde las emociones se vuelven conscientes. Al desarrollarse la corteza frontal orbital y el hipocampo, estas vías corticales intentan controlar la actividad de la amígdala en mayor o menor grado. De esta forma, el control de los impulsos es una de las partes principales de lo que podemos llamar la educación y la evolución de nosotros mismos: nuestros principios morales y éticos. En algunas personas, esta influencia inhibidora es débil, por lo que tendrán un carácter impulsivo; en cambio, si es fuerte, como ocurre en la mayoría de nosotros, la persona será educada y controlada.

La amígdala se ocupa principalmente de los acontecimientos más negativos o sorprendentes de nuestra vida, y el agonista y antagonista de la amígdala es el llamado sistema límbico basal, que contiene el área tegmental ventral, el núcleo accumbens y la sustancia negra. Allí se produce la señal de la dopamina que se envía a los ganglios basales, por ejemplo, al núcleo accumbens, al estriado dorsal y a la corteza prefrontal. Y este sistema mesolímbico tiene dos funciones principales: en primer lugar, representa el sistema de recompensa, mediante la liberación de opiáceos endógenos, en la que, en último término, se basa todo cuanto nos hace sentir felices. El segundo es el sistema dopaminérgico, que es también, en cierta medida, un sistema de recompensas, ya que cuando experimentamos algo que nos emociona o nos hace felices, aumentan no-

tablemente los niveles de dopamina. Sin embargo, se trata más bien de un sistema de espera de recompensas. Siempre que nos enfrentamos a algo, las promesas que queremos, la señal de la dopamina es muy elevada. Las personas que carecen de opiáceos endógenos (el sistema de recompensa) son antihedonistas; no obtienen ningún placer de la vida y, al no tener dopamina, ya no les interesa hacer nada porque no tienen ningún estímulo para actuar.

Muy someramente, el sistema límbico es el origen de la experiencia emocional positiva y negativa, y esto comienza a partir de la séptima semana de gestación. Así, incluso antes del nacimiento, el sistema límbico evalúa todo lo que hace el feto según criterios de lo bueno y lo malo. Por supuesto, la etapa más importante para el desarrollo de este sistema límbico emocional son los primeros 3 a 5 años de nuestras vidas. En este sentido, el trabajo del sistema límbico constituye el marco de la personalidad de cada uno; evoluciona durante los primeros meses y años de nuestras vidas y define nuestro carácter. Posteriormente, todas las experiencias se integran en este marco. Y ahora, antes de planificar cualquier acción, la intención se compara con la memoria límbica, de forma que todo lo que hacemos, queremos o deseamos debe compararse con nuestra experiencia emocional anterior. Por tanto, es de esperar que este sistema límbico domine nuestras acciones y, aún antes, nuestros deseos. Esto garantiza, tal como he escrito aquí, al menos en principio, que todo lo que hagamos se haga desde la perspectiva de la experiencia emocional previa. Eso es lo más racional.

Aquí tenemos los circuitos ventrales o límbicos. Éste no es un circuito de acción, sino de deseos y planes; puede ser consciente y va de la corteza frontal orbital y de la corteza al núcleo accumbens, al estriado ventral, al pálido ventral y de vuelta al tálamo. Así que éste es el círculo de nuestros deseos, que se pueden originar de forma consciente o bien, en su mayoría, de forma inconsciente. Cuando pasan a la parte consciente, se evalúan conscientemente; después, vuelven a la parte inconsciente, bajo la influencia del hipocampo, la amígdala y el sistema límbico basal. Por tanto, este circuito límbico emocional ventral está fuertemente controlado por nuestra experiencia emocional previa.

¿Qué significa todo esto? El sistema límbico influye de dos formas en la preparación cortical de las acciones. En primer lugar, hay proyecciones de la amígdala y el sistema límbico basal a la corteza prefrontal, frontal orbital y cingulada, así como al hipocampo, que están relacionados con el control inicial de las ideas, los deseos y los planes. Todos nuestros deseos proceden, en último término, del sistema límbico subconsciente. Después, se evalúan conscientemente, se envían de nuevo a la parte inconsciente y, a continuación, otra vez a la parte consciente… y esto puede continuar durante días, meses o incluso años. Y por último, el paso final, a través de la proyección, a los ganglios basales, al circuito dorsal, y el control final de la ejecución de movimientos específicos. Así que el sistema límbico tiene la primera y la última palabra. Y en medio, la voluntad consciente tiene un papel muy importante.

Podemos ilustrar esto de la siguiente forma: primero, tenemos deseos y planes que se envían de la parte consciente a la inconsciente; el sistema límbico aprueba estos deseos y les da «luz verde». Esto da lugar a la voluntad, que se envía al área motora, pero una vez más, antes de ejecutarla, se realiza una última prueba: ¿realmente quieres hacer esto ahora? ¿y exactamente de qué forma? «Exactamente de qué forma» significa que los ganglios basales dorsales deben aprobar de nuevo el plan y la forma de llevarlo a cabo.

En los pacientes con Parkinson, falta esta proyección de la sustancia negra al putamen, por lo que no es posible ejecutar este circuito dorsal. Esto explica por qué los pacientes con enfermedad de Parkinson pueden tener el deseo de hacer algo; el circuito ventral está bien, pero el dorsal no, así que no pueden ejecutar su deseo. No pueden dar un paso adelante. Esta liberación de dopamina está controlada por el sistema límbico, así que hay una señal de liberación controlada por todas las emociones subcorticales.

Si esto es así, si estamos controlados por la memoria y la experiencia emocional, ¿para qué necesitamos la racionalidad? El sistema límbico se encarga de todo.

En primer lugar, los centros límbicos subcorticales son incapaces de procesar los detalles de las situaciones o los problemas. La amígdala, como saben, puede reconocer fases y situaciones de peli-

gro, pero no los detalles; necesitamos a la corteza para los detalles finos y para la memoria contextual. Esto no es posible en la amígdala.

En segundo, lugar, los centros límbicos subcorticales inconscientes no pueden determinar las consecuencias a medio y largo plazo de nuestras decisiones. Sólo pueden actuar a corto plazo, como hace un niño pequeño: lo quiero ahora, no dentro de una semana, ni para Navidad ni para el año que viene. Sabemos que un niño pequeño no puede esperar hasta Navidad porque no sabe lo que es la Navidad. Ésa es una función cortical.

En tercer lugar, los centros límbicos no comprenden el idioma y, por tanto, no pueden utilizar una comunicación humana específica. La amígdala no comprende el idioma. Éste puede afectar a aquélla, pero sólo a través del componente emocional, no del lingüístico, lo que significa que sin racionalidad las decisiones complejas y la planificación a largo plazo no son posibles. Es por eso que tenemos, entre otras cosas, cortezas de gran tamaño y racionalidad. Sin embargo, nuestras decisiones *conscientes* se preparan y, en último término, se llevan a cabo mediante procesos emocionales *inconscientes*. La primera y la última palabra. Por tanto, la racionalidad es muy importante, pero su función es exclusivamente de consejero, que muestra los detalles de la situación y las posibles consecuencias de las decisiones. Imaginemos que tenemos un amigo que tiene un grave problema, sea con su familia o por algún error que ha cometido, y acude a nosotros para pedirnos consejo. Nosotros, como cualquier persona racional, le preguntaremos: «¿Sabes lo que estás haciendo o lo que quieres hacer?». «¿Has pensado en las consecuencias: A, B, C, D y E?» Y él nos responde: «No, aún no he pensado en eso». Así que, al igual que la amígdala, nuestro amigo quiere hacer algo, protestar o dejar su trabajo, y nosotros le decimos: «Piensa primero en las consecuencias. ¿Tienes alguna otra oferta de trabajo? o ¿qué vas a hacer los próximos días sin trabajo?». Quizá nos responda: «Tienes razón. No había pensado en eso». Ésa es la racionalidad: representa las consecuencias a medio y largo plazo, pero la decisión final siempre es emocional. El hecho de que nuestro amigo siga o no nuestro consejo es puramente emocional. Eso es lo que nos dicen todos los líderes políticos o empre-

sariales. Significa que no hay decisiones puramente racionales. La toma racional de las mismas siempre se lleva a cabo en el marco de este procesamiento límbico emocional. Y la racionalidad es lo que nos hace superiores —o al menos eso es lo que creemos— a todas las demás criaturas. Pero la racionalidad no domina nuestras acciones. Esto significa que la interacción entre lo consciente y lo inconsciente, entre los centros límbicos y los centros motores ejecutivos, garantiza que las acciones voluntarias se lleven a cabo dentro del marco de los sentimientos emocionales inconscientes y de la parte racional cognitiva consciente de cada persona.

Este modelo de guiado neural de las acciones voluntarias es compatible, en gran medida, con el concepto débil y compatibilista del libre albedrío. El hombre es libre en el sentido de que puede actuar en función de su voluntad consciente e inconsciente. Sin embargo, esta voluntad está completamente determinada por factores neurobiológicos, genéticos y del entorno, así como por las experiencias psicológicas y sociales positivas y negativas, en particular las que se producen en etapas tempranas de la vida, que dan lugar a cambios estructurales y fisiológicos en el cerebro. Esto significa que todas las influencias psicológicas y sociales deben producir cambios estructurales y funcionales; de lo contrario no podrían actuar sobre nuestro sistema motor. Por último, esto supone que no existe el libre albedrío en sentido firme, sino sólo en el sentido débil y compatibilista. Y también significa que nadie, ni los filósofos, ni los psicólogos, ni los neurobiólogos, pueden explicar cómo funciona el libre albedrío en el sentido fuerte.

Pregunta. ¿La transición entre el estado inconsciente y el estado consciente es «una transición todo o nada»? ¿o, por el contrario, y a pesar de ser rápida, es un proceso? Entiendo que habría una masa crítica de actividad neural necesaria para alcanzar este nivel crítico en el que se da esa transición rápida pero, en cualquier caso, no «todo o nada». Pregunto esto en el contexto de la función del área de

la corteza supramodal, que se considera crítica para definir el movimiento consciente de un dedo.

Prof. Roth. En primer lugar, existe un continuo absoluto entre los estados consciente e inconsciente. Podemos hacer cosas de forma especialmente consciente, cuando estamos atentos y tenemos que concentrarnos en un asunto. También podemos tener una conciencia *asociada*: cuando hacemos algo que no sabemos cómo hacer. La mayoría de las cosas las hacemos con la memoria asociada: no sabemos los detalles, pero sí lo que vamos a hacer. Esto es desde el punto de vista cualitativo, pero también podemos hacer cosas de forma totalmente inconsciente. En los ganglios basales reside lo que podemos hacer, en principio, de forma totalmente inconsciente. Cuando conducimos un coche por el centro de Madrid, la mayoría de las acciones las hacemos de forma inconsciente, y la conciencia entra en acción cuando aparece algo nuevo, importante y complicado. Estos estímulos subconscientes o inconscientes inducen la conciencia, y la corteza se activa cada vez más porque los centros límbicos inconscientes no pueden procesar los detalles ni la complejidad. Por tanto, lo más importante es comprender que la conciencia, entre otras cosas, siempre está relacionada con el procesamiento de información nueva, importante y complicada. Así, siempre que nos enfrentamos a decisiones nuevas e importantes, debemos hacerlo de forma consciente, pero en nuestra memoria inconsciente se guarda todo lo que hemos experimentado en nuestras vidas. La mayoría de las cosas conscientes residen aquí y en este momento, es totalmente imposible que el cerebro active y lea toda esa memoria previa, así que la deja a un lado, pero la consulta siempre. No se trata de que los impulsos inconscientes, en el sentido freudiano, nos guíen; es también todo lo que hemos experimentado antes, de forma consciente. Ésa es la diferencia. No guían nuestras necesidades, siempre en el sentido freudiano, sino que, para cada decisión, guían toda nuestra experiencia previa, incluidos los estados que anteriormente han sido decisiones conscientes.

P. ¿La conciencia es un proceso?

Prof. Roth. Sí. Para empezar, hay diferentes tipos de conciencia —al menos diez estados distintos— y la conciencia funciona a través de mecanismos muy distintos. Está la conciencia de nosotros

mismos, la conciencia motora, y muchas cosas distintas, y éste es un proceso que tarda al menos 300 milisegundos en llegar a la corteza. Es muy complicado, pero siempre está controlado, independientemente de si algo llega a ser consciente o no, por el estímulo o por el inconsciente. Por tanto, la conciencia es un instrumento para resolver problemas nuevos e importantes.

P. En este modelo falta el cerebelo, y creo que ésa es una omisión muy importante. El cerebelo tiene sistemas de transporte de dopamina y recientemente se le ha atribuido una función en la atención motora y no sólo en la transmisión cognitiva. ¿Cómo cree que encaja el cerebelo en su modelo?

Prof. Roth. En el diagrama no aparece el cerebelo, pero está el circuito de los ganglios basales y el del cerebelo, y ambos se unen en la acción voluntaria en las áreas premotora y motora. Así que el aporte adicional del cerebelo se necesita para llevar el potencial de preparación por encima de un cierto nivel. Sin embargo, introducirlo en mi descripción habría complicado excesivamente las cosas.

P. Existen muchos bloqueos físicos y también funcionales, pero para funcionar como un sistema completo, es necesaria cierta sincronización, coordinación e interfaz. El funcionamiento común es, desde mi punto de vista, absolutamente necesario. ¿Dónde reside esta funcionalidad común, auxiliar, en su modelo?

Prof. Roth. Ésta es una pregunta muy difícil de responder, ya que no hay un solo sistema encargado de dar órdenes en el cerebro. Sin embargo, si miramos las jerarquías funcionales, queda claro que hay una base fundamental en el cerebro, en el hipotálamo, que tiene un origen muy precoz. Después está la amígdala, la base de las emociones, luego, la corteza límbica y, por último, la corteza cognitiva. Estas estructuras aparecen progresivamente más tarde, durante el desarrollo del cerebro. Y siempre, en esas capas más básicas, hay centros que combinan el intelecto con las capas superiores. Podemos decir que no hay un solo centro que actúe como guía, sino varias vías específicas que conectan siempre la capa fundamental del hipotálamo y el tronco cerebral con las capas más evolucionadas. Así que la cuestión es siempre: ¿estás de acuerdo y puedo inhibirte? ¿Estás a favor o es necesario inhibir esta ley? Siempre hay una

batalla. Es una red enorme y siempre que hacemos algo, esta red y estas capas, interactúan, favorecen e inhiben. Esto se puede ver maravillosamente con las técnicas de adquisición de imágenes cerebrales actuales. Así que el cerebro es una red, aunque sea a gran escala y no sólo una red neurobiológica formada por neuronas, y siempre que se hace algo, hay una lucha entre dos partes, y todas las acciones que llevan a cabo esos centros están sujetas a una fuerte competencia.

P. Hay un punto que no se ha tratado y que creo que es crítico para la conciencia: el tiempo. Mucha gente piensa que la conciencia es algo que tarda en aparecer, que requiere tiempo. Por tanto, uno de los motivos por el que nuestras acciones se llevan a cabo a nivel inconsciente es porque de esta forma no tenemos que esperar a que aparezca la conciencia. Si tuviéramos que esperar, nuestras acciones no serían posibles, ya que no podríamos registrar los estímulos externos que están ocurriendo, por eso, debido a este factor de tiempo, realizamos la mayoría de nuestras acciones inconscientemente. Considero que lo que es crítico cuando pensamos en el libre albedrío, la conciencia y las acciones conscientes es que necesitamos un proceso rápido para ser precisos y que, si queremos serlo, no tenemos tiempo para ser conscientes.

Prof. Roth. Tenemos tres formas de responder. La primera es sumamente rápida, a través de reflejos. Esto lo hace la médula espinal y después de la acción nos preguntamos quién hizo eso. Y nos respondemos «fue mi pierna», «fue mi brazo»… pero no fui yo. Eso es lo más rápido y tarda unos 80 milisegundos. Pero es una acción que no puedo atribuir a mí mismo. Después, tenemos las respuestas automatizadas y relativamente rápidas, que no necesitan ser conscientes, aunque en su mayoría vayan *acompañadas* de conciencia, pero no controladas por ésta. En esos casos, hacemos algo y decimos «lo siento, fui yo, aunque más bien fue mi brazo», así que el punto básico es que estamos haciendo algo sin un control explícito, pero nos lo atribuimos a nosotros mismos, lo cual es muy interesante —es voluntario pero automatizado—. Sólo el tercer tipo de respuesta es deliberado. Nos enfrentamos a problemas nuevos y complejos, y entonces tenemos tiempo, pero el cerebro siempre quiere hacer las cosas rápido y sin interferencia de la conciencia, ya

que de esa forma es preciso y muy automático. Así que hay que obligarlo a hacer algo consciente, porque puede salir mal, pero en las situaciones complejas es necesario involucrar a la conciencia. Así que siempre hay una lucha en el interior del cerebro sobre si las cosas deben hacerse de forma inconsciente, tipo reflejo, y quizá inapropiada, o hacerse de forma apropiada, pero invirtiendo tiempo y asumiendo el riesgo de que puedan salir mal. Ése es el sesgo del conflicto en el cerebro.

P. Finalmente, tenemos la genética y las experiencias sociales y emocionales precoces, ¿hay alguna esperanza, desde el punto de vista científico, de reconfigurar un cerebro dañado a través de la genética o de experiencias sociales o emocionales? ¿O hay alguna posibilidad de que sea posible en el futuro?

Prof. Roth. De acuerdo con las últimas investigaciones acerca del proceso de las probabilidades y la función de la psicoterapia en las enfermedades psicológicas, parece que hay una regla: mientras más pronto se produzca el daño y se inicie el tratamiento, mayores serán nuestras probabilidades. Pero más tarde, a partir de los 10 años, se vuelve cada vez más difícil. Así que la psicoterapia o cualquier tipo de estímulo tienen que ser más fuertes y más específicos. Si hay daños cerebrales importantes, prácticamente no podemos hacer nada si el cerebro no ha conseguido compensarlos solo, lo que ocurre en el 75% de los casos. En el 60% de los casos, el cerebro compensa y se cura solo, pero en una tercera parte no lo consigue y no sabemos el motivo. Incluso en ese caso, si el daño no es demasiado importante, siempre existe la posibilidad del tratamiento, pero al aumentar la edad se vuelve cada vez más difícil, ya que la amígdala reduce muy pronto su plasticidad, mientras que la corteza la conserva casi durante toda la vida. Pero el sistema límbico reduce notablemente su plasticidad, y la psicoterapia y todos los demás métodos deben ser sumamente especiales para volver a abrirla, ya que reabrir las acciones límbicas es muy difícil. Éste es el motivo por el que la psicoterapia es tan difícil y lleva tanto tiempo, y por el que con frecuencia falla.

Una aproximación desde la neurociencia a la monitorización de las funciones ejecutivas

Thomas F. Münte

Para adaptar su conducta al medio, los seres humanos necesitan poder monitorizar sus acciones y detectar y corregir los errores. Estas capacidades forman parte del sistema de control ejecutivo que se encarga de la observación y detección de problemas, de la planificación, y del ajuste del funcionamiento del sistema. En el presente capítulo trataré de examinar parte de las características neurofisiológicas básicas del sistema de monitorización de la actividad humana. Para ello, presentaré datos provenientes, principalmente, de mi propio laboratorio, pero quisiera enfatizar desde el principio que muchos grupos diferentes de distintas partes del mundo han aportado contribuciones a esta área de investigación durante los últimos años.

Aunque la monitorización de acciones y la detección de errores son importantes a la hora de llevar a cabo prácticamente cualquier tarea, desde conducir un coche o escribir una carta, hasta preparar una comida, es necesario simplificar al máximo la presentación del problema para su estudio en laboratorio.

La figura 1a ilustra la base sobre la que se sustenta la investigación sobre monitorización de acciones, la llamada «tarea de los flancos» de Eriksen, presentada en 1974 (Eriksen y Eriksen, 1974). Es extremadamente sencilla: al sujeto se le muestran series de cinco letras y se le pide que responda a la letra que ocupa la posición central. Si la misma es una «H», el sujeto debe apretar un botón con la

mano izquierda; si es una «S», lo debe hacer con la mano derecha. La tarea de los flancos implica una clase de estímulos, en la cual la letra diana es flanqueada por letras idénticas. Estos estímulos son denominados «congruentes».

Otras pruebas presentan letras de flancos que no son idénticas a la letra diana, forzando el uso de la mano contraria. Estos estímulos incongruentes, sumados a la presión impuesta por la obligación de responder dentro de un límite de tiempo, conducen a errores.

De hecho, el coeficiente de error está cerca del 15% en sujetos normales al realizar esta tarea. Pero lo importante es que en la mayoría de los casos, los sujetos se dan cuenta inmediatamente de que han cometido un error.

Cuando se registran los potenciales cerebrales evocados relacionados temporalmente con las tareas de apretar un botón de manera errónea o correcta, se observa una negatividad fásica sobresaliente en referencia a las respuestas erróneas que ha sido denominada «negatividad asociada a errores» (ERN, del inglés *error-related negativity*) o «negatividad de errores» (NE) por sus descubridores (Gehring y Fencsik, 2001; Falkenstein, Hoormann, Christ y Hohnsbein, 2000).

La ERN, ilustrada en la figura 1c., tiene una latencia de cerca de 80 milisegundos con respecto al momento de presionar el botón. Además, el mapa topográfico muestra que su máximo está sobre las áreas frontales mediales. Este hecho, en sí mismo, ya sugiere la existencia de un generador dentro de la corteza frontal medial. Pero se puede ir un paso más adelante, y tratar de determinar en qué parte del cerebro residen los generadores de esta respuesta de ERN. Esto, en principio, es una tarea que no puede ser resuelta, como Helmholtz indicó ya en 1853. No obstante, técnicas modernas de análisis de fuente nos permiten estimar los generadores con un grado razonable de fiabilidad. Utilizando dipolos múltiples equivalentes llegamos a una solución que representa tres dipolos, uno situado en la corteza frontal medial, y otros dos en ubicaciones frontales laterales. También se puede detectar un generador frontal medial si se utiliza un enfoque diferente que implique fuentes distribuidas.

Con todo, no se puede negar la necesidad de evidencia adicional para implicar la corteza frontal medial en la monitorización de ac-

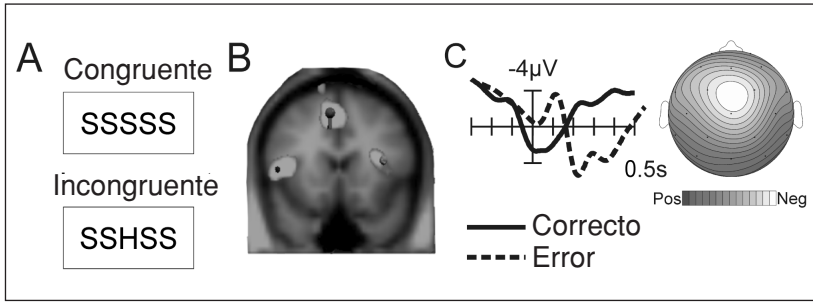


Fig. 1: A: Ilustración de los estímulos típicos de flancos que Eriksen utilizó en la tarea de flancos. La letra central es pertinente, mientras que las otras cuatro son solamente distractores. Se pide al sujeto que apriete un botón con la mano izquierda para la letra objetivo S y con la mano derecha para la letra H. En el estímulo superior, las letras que flanquean son idénticas a la letra diana, lo que supone que el sujeto recibe información congruente. En el estímulo inferior, las letras de flancos de la letra diana no son idénticas a las letras diana y priman una respuesta errónea. Estos estímulos son llamados incongruentes, y llevan a una tasa aumentada de errores. B: Los potenciales evocados promediados y relacionados temporalmente con la respuesta del sujeto, muestran una negatividad fásica en cuanto a respuestas erróneas en relación con las correctas, que tiene una latencia máxima de cerca de 80 milisegundos respecto al momento de presionar el botón. El mapa topográfico ilustra la distribución de esta «negatividad asociada al error» (según datos de Rodríguez-Fornells *et al.,* 2002). C: Sección coronal de un experimento de flancos realizado utilizando técnicas de resonancia magnética funcional (RMf). Se observan tres activaciones principales: (i) giro cingulado anterior, (ii) frontal lateral izquierdo, (iii) corteza insular en el hemisferio derecho. Además, se pueden observar los dipolos obtenidos utilizando el programa de análisis de fuente eléctrica cerebral (BESA) trazados en la imagen de la RM (Rodríguez-Fornells y Münte, no publicado).

ciones. Nosotros (y muchos otros grupos) hemos realizado, por tanto, un experimento con resonancia magnética funcional (RMf) relacionada con eventos. La RMf permite evaluar qué partes del cerebro están activas durante una determinada acción, rastreando la llamada respuesta BOLD (*blood oxygen level dependent,* 'dependiente del nivel de oxígeno en sangre'). En la variante de RMf relacionada con eventos podemos obtener activaciones separadas para ensayos de error y correctos, en cuyo caso se observa una activación asociada al error en la corteza frontal medial que implica el giro cingulado anterior (figura 1b). Es sabido que esta parte del cerebro interviene en aspectos ejecutivos de muchas acciones cognitivas. Resulta interesante constatar que, como en la localización de las fuentes de potenciales cerebrales, dos lugares frontales laterales de activación están también presentes en el estudio con RM. Convie-

ne indicar que la resolución temporal de la RM es mucho más baja, con un pico de la respuesta BOLD estimado en cerca de 6 segundos después del error del sujeto. Para resumir, la utilización conjunta de estudios electrofisiológicos y técnicas de RMf en relación a la detección de errores parece demostrar que una red de áreas cerebrales que comprenden el giro cingulado anterior y áreas frontales laterales participan en la monitorización de acciones.

La siguiente pregunta que nos planteamos fue si la ERN puede ser considerada como el correlato neural de la primera señal de error en el cerebro o no (Rodríguez-Fornells, Kurzbuch y Münte, 2002). Como una de las consecuencias de la detección de un error es su corrección, formulamos la hipótesis de que si la ERN (si, efectivamente, se trataba de un indicador de la primera señal de error del cerebro) debería preceder a cualquier actividad electrofisiológica relacionada con la corrección. Para contestar a esta pregunta, combinamos el registro de la negatividad asociada al error con la de una respuesta cerebral adicional evocada, el llamado «potencial preparatorio lateralizado» (LRP, del inglés *Lateral readiness potential*), que puede ser registrado de forma no-invasiva mediante electrodos situados sobre la corteza motora, y marca la preparación de un movimiento tanto con la mano izquierda como con la derecha, sirviendo, así, como índice de preparación de una respuesta correctiva, que en nuestro caso es una segunda respuesta motora (por ejemplo, apretando el botón izquierdo después de que el participante hubiera apretado erróneamente el derecho). Así, realizamos mediciones con sujetos jóvenes normales en una situación en la que se les pedía que corrigieran cualquier error. Con el fin de ofrecer una comparación, también incluimos una condición en la que a los sujetos no les estaba permitido corregir sus errores. La figura 2 ilustra lo que encontramos.

El hallazgo más importante es que el LRP a la respuesta correctiva tenía una latencia inicial anterior a la latencia pico de la ERN, lo que significa que la ERN no puede ser considerado como un factor neural correlacionado con la primera señal de error en el cerebro, puesto que ésta debe preceder, como dijimos antes, a cualquier acción correctiva. Además, también observamos que el LRP para correcciones lentas y rápidas mostraba latencias significativamente

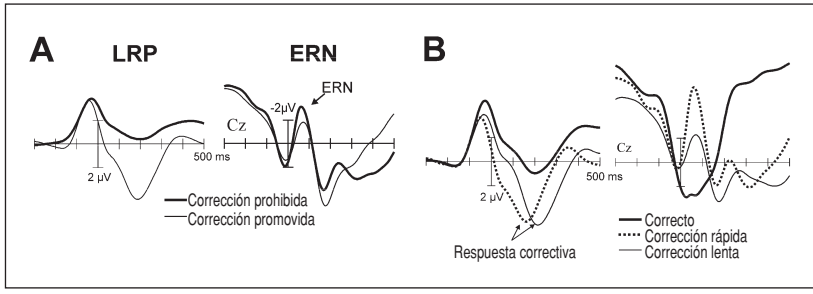


Fig. 2: Potenciales evocados cerebrales en un experimento que compara condiciones en las que al sujeto se le pide que corrija los errores, o que no lo haga. A: A la izquierda se muestra el potencial preparatorio lateralizado en pruebas de error. Un desvío ascendente indica preparación de la respuesta errónea. Las formas de ondas para las condiciones de la corrección prohibida y de la promovida se separan aún antes de presionar el botón incorrecto (que ocurría en tiempo 0). El pico de la negatividad asociada a error (mostrado en el lado derecho) no ocurre hasta cerca de 100 milisegundos más tarde, lo que parecía indicar que la corrección se iniciaba antes del proceso que subyace a ERN, por lo que no puede ser considerado como factor correlativo de la primera señal de error en el cerebro. B: Comparación de los potenciales preparatorios lateralizados (lado izquierdo) para las correcciones rápida y lenta. El desvío hacia abajo que indica la preparación de la respuesta correctiva presenta el pico considerablemente antes en el caso de las correcciones rápidas. En contraste, el estado latente de la ERN (lado derecho) no varía entre las correcciones rápidas y lentas. Es decir, no hay emparejamiento temporal entre acciones correctivas y estado de latencia de la ERN (Rodríguez-Fornells *et al.*, 2002).

diferentes, mientras que la latencia de la ERN era igual para pruebas de error tanto en casos de correcciones rápidas como lentas. Esto sugiere, una vez más, que la ERN no puede ser considerado como la primera señal cerebral responsable de provocar la acción correctiva.

Así pues, la conclusión de este estudio fue que la ERN no puede ser el elemento de correlación de la primera señal de error en el cerebro, lo que deja abierta, por tanto, la pregunta sobre dónde y cuándo se emite esa primera señal de error.

Una teoría importante, inspirada en trabajos recientes con animales, ha ligado la generación de la ERN a las acciones de un posible sistema cerebral subcortical responsable del procesamiento de recompensa y refuerzo de lo aprendido (Holroyd y Coles, 2002). Esta teoría postula que nuestro sistema de monitorización de acciones predice el resultado de una acción. Si dicho resultado (como en el caso de presionar el botón incorrecto) resulta ser peor de lo esperado, se observa una disminución fásica en la actividad neural

—mediada por el neurotransmisor dopamina— en los ganglios basales y transmitida posteriormente a otras estructuras del cerebro, como la corteza cingulada anterior, donde se produce la ERN. En el caso de los seres humanos, la técnica de RMf ha mostrado que existe actividad en las regiones de los ganglios basales en respuesta a pruebas de error. Sin embargo, como se ha indicado anteriormente, con la técnica de RM no podemos determinar cuándo, exactamente, se inicia la activación neural.

Con este fin, decidimos embarcarnos en un proyecto, junto al neurocirujano Volker Sturm y sus colaboradores de la Universidad de Colonia, Alemania, en el cual podíamos registrar directamente estructuras críticas del sistema de recompensa del cerebro durante operaciones neuroquirúrgicas en pacientes despiertos. Sturm y sus colaboradores utilizan dispositivos de estimulación cerebral profunda (usados desde hace casi 20 años en pacientes con enfermedad de Parkinson para estimular el núcleo subtalámico) con el fin de estimular el núcleo accumbens en pacientes con trastorno obsesivo-compulsivo. Durante tales operaciones, los pacientes están despiertos y pueden tomar parte en tareas psicológicas. Nuestros datos preliminares sugieren que la actividad neural registrada directamente del núcleo accumbens es diferente para pruebas de error y acierto. Además, esta diferencia parece preceder a la modulación del error observada anteriormente en áreas corticales. En los próximos años podremos confirmar si, efectivamente, la señal en el núcleo accumbens puede ser calificada como la primera señal cerebral en relación a la detección de errores.

En la vida diaria aprendemos de nuestros propios errores, pero también de la retroalimentación externa («Suelta eso, que es peligroso» o «Bien hecho, chaval»), proporcionada por nuestros padres, maestros, compañeros, por los ordenadores o los semáforos.

Consecuentemente, nos preguntamos qué tipo de relación existiría entre los elementos de correlación neurofisiológicos de información generada internamente sobre la calidad de nuestras acciones (es decir, la ERN) y la información de retroalimentación externa (Müller, Möller, Rodríguez-Fornells y Münte, 2005). Para ello, creamos una situación de aprendizaje. Durante cada bloque experimental, a los participantes se les presentaba un nuevo

conjunto de bosquejos (de objetos y animales) y tenían que aprender, para cada objeto, si la respuesta correcta para cada uno consistía en presionar el botón de la mano izquierda o el de la derecha. Como señal de aprendizaje, se proporcionó a los participantes información de retroalimentación (afirmativa y negativa) 1.100 milisegundos después de presionar el botón. La idea era que los sujetos aprendieran las asociaciones de estímulos y respuestas en el transcurso de un bloque experimental, cosa que, de hecho, hicieron.

En este contexto, se tiene la oportunidad de observar las señales cerebrales que clasifican la información generada internamente sobre la calidad de las acciones, es decir, la ERN, mediante la observación de las respuestas cerebrales en relación con la presión del botón; y las señales cerebrales que clasifican la información externa, mediante el examen de los potenciales cerebrales de reacción a la retroalimentación.

Éste es el resultado de nuestras observaciones: como esperábamos, obtuvimos una ERN fiable en respuesta a presionar el botón erróneo. Además, como la asociación entre estímulo y respuesta apropiada se hacía cada vez más fuerte en el transcurso de cada bloque experimental, habíamos predicho que la ERN también debía aumentar en el mismo transcurso. Y ése fue exactamente el caso. Es más, durante el aprendizaje inicial de las asociaciones estímulo-respuesta, los estímulos de retroalimentación debían proporcionar información más crucial que durante las fases posteriores de un bloque experimental, cuando las asociaciones estímulo-respuesta ya estaban firmemente establecidas. Esto quedó reflejado en la amplitud de la negatividad fásica, que denominamos «negatividad asociada a la retroalimentación», que fue más pronunciada durante las fases iniciales del aprendizaje que durante las posteriores. Curiosamente, un análisis de fuente de esta negatividad asociada a la retroalimentación reveló que, además de una fuente prominente en la corteza cingulada anterior (como en el caso de la ERN), esta respuesta del cerebro recibía también contribuciones de regiones más posteriores (figura 3).

Por último, decidimos examinar las respuestas cerebrales en casos de retroalimentación sin contenido. A este fin, informamos

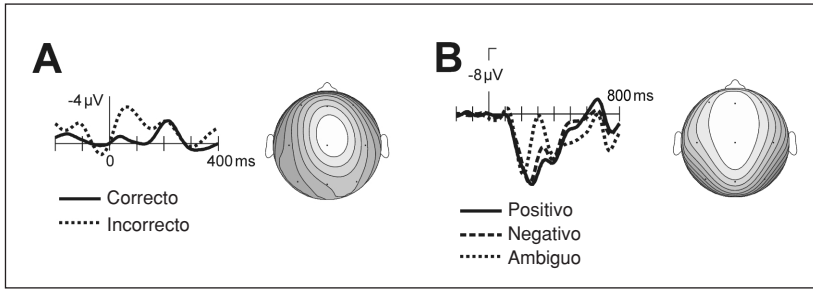


Fig. 3: Potenciales cerebrales en una tarea de aprendizaje. Los sujetos tenían que asociar el presionar un botón con la mano izquierda u otro con la mano derecha con cada uno de los 16 bosquejos que se sucedían, repetidamente, durante un bloque experimental. A: Respuestas cerebrales a presión del botón con tiempo de reacción por parte del sujeto. Las respuestas erróneas están asociadas con un componente negativo claro, es decir, una ERN, lo que muestra la distribución frontal, medial y típica. B: Potenciales cerebrales con tiempo de reacción a los estímulos de retroalimentación que indican: bien una respuesta correcta (positivos), bien una respuesta incorrecta (negativos), o que el ordenador no podía determinar con exactitud el acierto o no de la respuesta (ambiguo, ver texto). Tanto la retroalimentación negativa como la ambigua estaban asociadas a una negatividad fásica con una distribución medial que se extendía más atrás que la de la ERN. El análisis de la fuente (no mostrado) indicó la existencia de una fuente posterior de línea media además de una fuente cingulada anterior (según datos de Müller *et al.,* 2005).

(falsamente) a los participantes de que el programa utilizado para el estímulo no estaba todavía completamente terminado y que contenía un «virus». Les explicamos que ese virus era el causante de que el ordenador no pudiera determinar si una respuesta concreta era correcta o no, o por lo menos no podía hacerlo de forma continuada y fiable. Formulamos la hipótesis de que esa retroalimentación ambigua conduciría al sujeto a un intenso reexamen de su respuesta anterior, que debería estar asociada a un aumento considerable de la negatividad asociada a la retroalimentación. Y ése fue exactamente el caso.

De acuerdo con el contenido general de estas páginas, quisiera retomar el tema anterior: la detección de errores y la consciencia. Si la ERN no es la primera señal de detección de error en el cerebro, (ni tampoco lo son las activaciones del giro cingulado anterior mostradas en estudios con RMf), quizá aquélla esté relacionada con el hecho de que el error se haga consciente.

Como señalaron Nieuwenhuis *et al.* (2001): «Si queremos exa-

minar la relación entre actividad cerebral asociada a error y la consciencia en cuanto a lo que denominamos *deslices* —el tipo de error que ha sido el objetivo de investigación psicofisiológica respecto al procesamiento de errores—, necesitamos un paradigma en el que el sistema Ne pueda originar fácilmente una representación de la respuesta correcta, sin que los participantes sean (siempre) conscientes de los errores en la ejecución de esta respuesta».

Nieuwenhuis *et al.* optaron por la llamada «tarea antisacádica» (Nieuwenhuis *et al.*, 2001), en la que un estímulo tiene lugar bien a la izquierda, bien a la derecha de una presentación de video. El sujeto tiene que vencer la tendencia natural a mirar este estímulo (es decir, realizar un movimiento ocular sacádico en dirección al objeto) y mirar precisamente al lado contrario (es decir, realizar un movimiento ocular antisacádico). Nieuwenhuis *et al.* descubrieron que los errores sacádicos no percibidos subjetivamente eran corregidos casi de forma inmediata, prácticamente en todos los casos, y estaban asociados a tiempos más rápidos de corrección y tamaños sacádicos más reducidos que los errores percibidos.

Con relación a la electrofisiología, concluyeron que, independientemente de que el participante fuese o no consciente del error, los movimientos sacádicos erróneos iban seguidos de una considerable ERN. Por tanto, los autores concluyeron que la ERN no está relacionada con la consciencia de error. Lo que sí encontraron, sin embargo, es una positividad que sucede de forma inmediata a la ERN, a menudo llamada «positividad de error» o Pe, y que mostraba relación con la consciencia del error.

Antes de concluir que la ERN clasifica errores, tanto si los sujetos son plenamente conscientes de ello, como si no, se debe considerar otro estudio realizado por Scheffers y Coles (2000), que está más estrechamente relacionado con el paradigma de los flancos hasta ahora analizado.

En su estudio, Scheffers y Coles encontraron una estrecha correlación entre la amplitud de la ERN y la exactitud de la respuesta subjetivamente percibida en una tarea típica de flancos. La figura 4 ilustra conclusiones semejantes halladas en nuestro laboratorio.

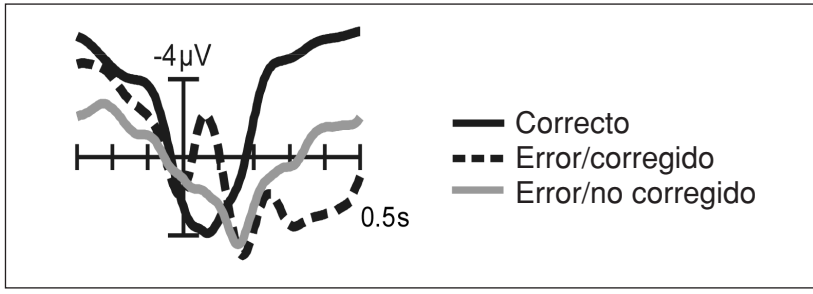


Fig. 4: Potenciales cerebrales para respuestas correctas, errores corregidos y errores no corregidos. La tasa de corrección de error en este experimento estuvo cercana al 90%. Sólo los errores corregidos fueron asociados a una ERN típica, mientras que los no corregidos no lo fueron (datos no publicados).

¿Cómo podemos resolver esta aparente contradicción? Nieuwenhuis *et al.* han sugerido que la ERN varía como una función de la consciencia de error, siempre que el grado de certeza acerca de la exactitud de la respuesta dependa de limitaciones de datos, como en el caso del estudio de Scheffers y Coles. Por el contrario, cuando existe incertidumbre acerca de la corrección o no de la respuesta, como en el experimento de movimientos antisacádicos de Nieuwenhuis, la ERN no es afectado por la consciencia.

Con respecto a los experimentos con RMf, se puede observar una relación flexible similar entre activaciones en la corteza frontal medial y la consciencia del error. En un experimento reciente realizado con técnicas de neuroimagen, Hester *et al.* (2005) mostraron que la consciencia explícita de un error (en este caso, los sujetos respondían cuando deberían no haberlo hecho) estaba asociada con la activación cerebral parietal y prefrontal bilateral. La región cingulada anterior mostraba una activación similar para errores de los que los sujetos eran conscientes y para errores de los que no lo eran.

Estos resultados permitieron sugerir que la actividad en el giro cingulado anterior no es suficiente para promover la consciencia activa de errores, sino que más bien parece enviar información a otras regiones del cerebro que quizá apliquen medidas conscientes de modificación de la conducta (como la ralentización post-error). Un aspecto importante que tiene que ser considerado al examinar la relación entre actividad cerebral y consciencia es el paradigma

específico utilizado. La corrección de errores en la tarea de flancos, que ha sido utilizada por nosotros y muchos otros, requiere presionar el botón con la mano contraria a la utilizada para la respuesta errónea, lo que pasa básicamente desapercibido. Sin embargo, otras tareas, por ejemplo las que implican movimientos de ojos, no requieren un cambio en el realizador para su corrección. Así, la corrección de un error puede ser llevada a cabo sin tener necesariamente consciencia de ella. Resumiendo, la relación entre activaciones electrofisiológicas y cerebrales en experimentos de RM, por una parte, y la consciencia de errores de acción, por otra, dista de estar clara y requiere mayor investigación.

Para terminar este breve y personal resumen sobre los aspectos neurofisiológicos implicados en la monitorización de acciones, quisiera responder a la pregunta de si los elementos neurofisiológicos asociados a la monitorización de acciones, aislados mediante el uso de nuestra sencilla tarea de flancos, pueden ser también encontrados en otras tareas más naturales.

La respuesta a esta pregunta es un sí categórico, como, sin duda, confirmará el ejemplo concerniente a los errores del discurso, los llamados «lapsus linguae» (Möller *et al.*, 2006). El habla es un acto asombroso. En unos pocos cientos de milisegundos pasamos de la intención de hablar sobre un tema determinado a la vocalización de una expresión sumamente automatizada. Durante este complejo proceso, que requiere la localización de las palabras apropiadas en el léxico mental, la conformación de sus características sintácticas, la identificación de la constitución fonológica apropiada y su posterior producción fonológica, y la puesta en marcha de los programas motores correspondientes que implican unas cuantas docenas de músculos diferentes, cometemos, sorprendentemente, muy pocos errores (aproximadamente 1 por cada 1.000 palabras).

La mayoría de los modernos modelos psicolinguísticos de producción del habla utilizan dispositivos de monitorización. Por ejemplo, el popular modelo del psicolingüista holandés Levelt integra varios bucles diferentes de monitorización, uno en el nivel conceptual, otro monitorizando el «habla interna» y un tercero examinando la vocalización propiamente dicha del lenguaje. Recientemente, hemos comenzado a registrar los potenciales cere-

brales con relación a errores del habla inducidos por una tarea clásica de laboratorio. Para poder observar la actividad cerebral no contaminada por artefactos de vocalización, se pidió a los sujetos que pronunciaran la pareja de palabras objetivo, sólo después de escuchar una señal para la iniciación del habla. De esta forma, con anterioridad al error, podíamos examinar la actividad cerebral implicada en las pruebas asociadas con un error del habla y en las pruebas en que no existía tal asociación.

La figura 5 ilustra algunos de los resultados. Claramente, los errores del habla fueron asociados con una negatividad mediofrontal, que mostraba una distribución topográfica muy similar a la ERN. Además, nuestras investigaciones posteriores respecto a los generadores neurales de este efecto de error del habla revelaron estructuras frontales mediales como posibles fuentes.

De ello concluimos que un sistema genérico de monitorización de acciones, como el revelado por experimentos psicológicos sencillos, tales como la tarea de flancos, también está implicado en la monitorización de los errores del habla.

Aunque estos paralelismos resultan asombrosos, querríamos,

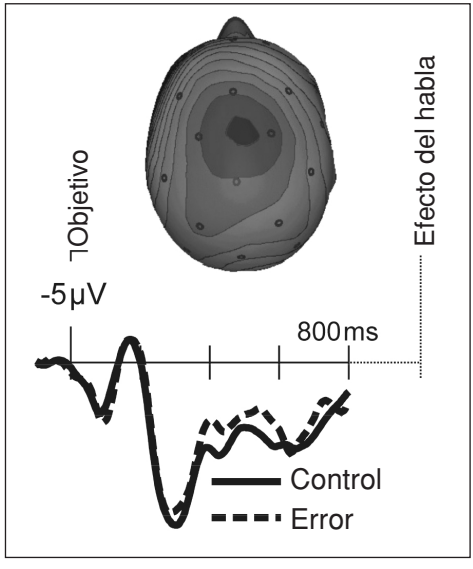


Fig. 5: Datos de una tarea diseñada para la detección de errores del habla. Los potenciales cerebrales mostrados corresponden en el tiempo a la presentación de una pareja de palabras. Transcurrido 1 segundo, se pedía a los sujetos que vocalizaran dicha pareja de palabras. Las palabras que dieron lugar a errores del habla estaban asociadas a una negatividad aumentada. El mapa recogido del cuero cabelludo muestra un máximo frontal medial de este efecto de error del habla similar a la distribución de la ERN (según datos de Müller *et al.,* en imprenta).

asimismo, enfatizar que el tipo concreto de error del habla que obtenemos en nuestra tarea ocurre más bien tarde en el sistema de producción del habla. Podría ser que otros tipos de errores del habla, que se producen con anterioridad, puedan ser detectados y corregidos dentro del sistema lingüístico en sí mismo, sin recurrir al sistema genérico de monitorización de acciones que hemos ilustrado en esta contribución.

Para concluir, considero que la investigación de las acciones que ocurren por desliz pueden ser una rica fuente de información para psicólogos, neurocientíficos y, quizá, filósofos. El uso conjunto de las técnicas de neuroimagen funcional de alta resolución temporal (como los potenciales evocados cerebrales) y espacial (como la RMf) nos ayudará enormemente a profundizar en nuestro conocimiento sobre el sistema de monitorización de acciones, y nos permitirá determinar cuáles de esas funciones se llevan a cabo de forma consciente, y cuáles no.

AGRADECIMIENTOS

Manifiesto mi más profundo agradecimiento y reconocimiento por su valiosísima contribución a mis amigos y colaboradores: Antoni Rodríguez-Fornells, Marcus Heldmann, Jürn Möller, Bernadette Jansma, Sandra Müller y Jordi Riba. Hago extensivo mi agradecimiento a la Deutsche Forschungsgemeinschaft (Fundación Alemana para la Investigación) por su apoyo a nuestras investigaciones.

Bibliografía

ERIKSEN, B. A., y ERIKSEN, C. W. (1974). Effect of noise letters on the identification of target letters in a non-search task. *Percept Psychophys,* 16, 143-149.

FALKENSTEIN, M., HOORMANN, J., CHRIST, S., y HOHNSBEIN, J. (2000). ERP components on reaction errors and their functional significance: a tutorial. *Biol Psychol,* 51, 87-107.

GEHRING, W. J., y FENCSIK, D. E. (2001). Functions of the medial frontal cortex in the processing of conflict and errors. *J Neurosci,* 21, 9430-9437.

HESTER, R., FOXE, J. J., MOLHOLM, S., SHPANER, M., y GARAVAN, H.

(2005). Neural mechanisms involved in error processing: A comparison of errors made with and without awareness. *Neuroimage,* 27, 602-608.

HOLROYD, C. B., y COLES, M. G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol Rev,* 109, 679-709.

MÖLLER, J., JANSMA, B. M., RODRÍGUEZ-FORNELLS, A., y MÜNTE, T. F. (2006). What the brain does before the tongue slips. *Cereb. Cortex, en prensa.*

MÜLLER, S. V., MÖLLER, J., RODRÍGUEZ-FORNELLS, A., y MÜNTE, T. F. (2005). Brain potentials related to self-generated and external information used for performance monitoring. *Clin Neurophysiol,* 116, 63-74.

NIEUWENHUIS, S., RIDDERINKHOF, K. R., BLOM, J., BAND, G. P., y KOK, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: evidence from an antisaccade task. *Psychophysiology,* 38, 752-760.

RODRÍGUEZ-FORNELLS, A., KURZBUCH, A. R., y MÜNTE, T. F. (2002). Time course of error detection and correction in humans: neurophysiological evidence. *J Neurosci,* 22, 9990-9996.

SCHEFFERS, M. K., y COLES, M. G. (2000). Performance monitoring in a confusing world: error-related brain activity, judgments of response accuracy, and types of errors. *J Exp Psychol Hum Percept Perform,* 26, 141-151.

Pregunta. Con respecto a la ERN y a la actividad de la circunvolución cingulada asociada a errores, que usted ha mencionado, nosotros encontramos que no existía tal actividad en los niños.

Prof. Münte. Algunos investigadores han rastreado el desarrollo temporal de la ERN y parece que se trata más bien de una adición tardía a nuestro repertorio de potenciales evocados, lo que quizá explique por qué no se encuentra el mismo patrón en los niños. Incluso los niños normales tienen una ERN muy reducido.

P. Mencionó usted que existe una variante consciente de corrección o de consciencia de errores y otra variante inconsciente. ¿Podría explicar con mayor detalle en qué momento se convierte en consciente?

Prof. Münte. Por una parte, parece que siempre que se corrige un error o se señala como tal, se desarrolla una ERN. Esto es así, especialmente, cuando se necesita cambiar el realizador, es decir, la mano encargada de responder para la corrección. En ese caso, si se es capaz de corregir el error, es evidente que se debe haber tomado consciencia de él con anterioridad. Esto podría sugerir que la ERN es el correlato de la consciencia del error, pero el experimento de antisacadas, realizado por Nieuwenhuis *et al.*, aporta un mensaje algo diferente. Existen también otros experimentos que conllevan movimientos de alcance balísticos, en los que el objetivo es desplazado de tal manera que los sujetos hacen correcciones a los movimientos balísticos sin advertir que las están realizando. En estas situaciones, las técnicas de neuroimagen no han revelado actividad asociada al error en la corteza cingulada anterior. Por tanto, creo poder deducir que la consciencia del error quizá debe estar muy próxima a la ERN y a la actividad cingulada anterior siempre que el paradigma experimental requiera un cambio de realizador.

P. En la tarea anterior, ya mencionada, existen muchas estructuras subcorticales diferentes implicadas en este mecanismo de control. Mi pregunta es: ¿por qué no pueden ustedes detectar este tipo de actividad de la estructura subcortical cuando llevan a cabo sus experimentos con RMf?

Prof. Münte. Justo al contrario: es totalmente posible observar actividad en las estructuras subcorticales. Por ejemplo, hay un trabajo reciente de Magdeburgo, realizado por Wittmann *et al.* (publicado en *Neuron,* 2005), en relación al aprendizaje del refuerzo. En él se observan una cantidad de estructuras subcorticales, incluidas la substancia negra y el área tegmental ventral, en plena actividad.

P. ¿Se ha descubierto si el núcleo talámico está implicado en este tipo de tarea?

Prof. Münte. Un estudio realizado por Ullsperger y von Cramon (publicado en el *Journal of Neuroscience,* 2003) demuestra que la habénula del tálamo está implicada en las tareas de monitorización de errores.

Autocomprensión humana, neurociencia y libre albedrío: ¿se anticipa una revolución?

Michael Pauen

¿El cerebro es un sistema determinista? En las próximas páginas examinaré las posibles consecuencias que deberíamos considerar si fuera así.

Es evidente que algunos de los avances recientes de la neurología en Europa han tenido que prestar atención a cuestiones relevantes para la comprensión de nuestra propia condición humana, entre las que figuran la conciencia, la conciencia de sí mismo y el libre albedrío.

Hay quienes piensan que esto nos conducirá a la resolución de estos problemas, hasta llegar a una solución científica. De esa opinión son Benjamin Libet o Daniel Wegner. Sin embargo, creo que eso sería no comprender la división del trabajo entre filósofos y científicos, por así decirlo. Por una parte, tenemos problemas filosóficos —o mejor dicho, aspectos filosóficos del problema—. Para abordar la cuestión del libre albedrío, una de las preguntas previas es: ¿a qué nos referimos exactamente cuando hablamos de libertad?

Así pues, estamos ante una pregunta conceptual, o una pregunta referente a los criterios. ¿Qué constituye un acto libre? Y esa es, desde luego, una cuestión filosófica. Los filósofos intentan esclarecer cuáles son los criterios, e intentan formular conceptos coherentes que expliquen realmente nuestras intuiciones precientíficas. Después, desde luego, surgen las preguntas empíricas, pues a partir del momento en que sabemos cuáles son los criterios, qué quere-

mos decir cuando hablamos de la conciencia del libre albedrío, obviamente deseamos saber dónde necesitamos esos criterios, lo cual constituye una pregunta empírica: ¿tenemos las capacidades necesarias?

El concepto tradicional del libre albedrío

Existe un antiguo concepto filosófico sobre esta cuestión y sobre los criterios del libre albedrío que se viene discutiendo no ya desde hace cientos de años, sino desde hace milenios, y existe un concepto filosófico tradicional y también prefilosófico del libre albedrío, según el cual la libertad requiere determinación. Ése es el concepto del libre albedrío que Gerhard Roth considera sólido.

Este concepto tradicional del libre albedrío plantea varios problemas bien conocidos. Una de sus consecuencias sería que no hay libertad en un mundo determinista. Así que imaginemos: ayer por la noche, me encontraba en la estación de Heidelberg y estaba pensando «¿De verdad quiero ir a Madrid o prefiero quedarme en Heidelberg?». Si nuestro mundo fuera determinista, en ese caso estaría determinado que yo viajara a Madrid y, obviamente, yo no hubiera sido libre cuando vine a Madrid, ni responsable de ello, una visión que resulta escasamente atractiva. Otra cuestión es que según esta visión tradicional, los actos de libre albedrío estarían más allá de toda explicación científica, pues sólo serían libres aquellos actos que no se pudiesen explicar aunque conociéramos todos los antecedentes. Si pudiéramos explicar por qué he venido a Madrid, si ustedes conocieran todos los antecedentes —mis deseos y creencias, etc.—, entonces ese acto no sería libre; estaría determinado, lo cual resultaría en una idea de la libertad que hace de ella un simple misterio.

Obviamente, todos estos problemas no conducen a una refutación de este concepto tradicional, aunque creo que sí nos dan motivo suficiente para su evaluación, y eso es lo que me gustaría hacer.

¿Cómo proceden los filósofos cuando llevan a cabo este tipo de evaluación? Imaginemos a un biólogo intentado averiguar si la vida necesita fuerza vital. ¿Qué hace? Más o menos lo siguiente: observa los puntos pertinentes y comprueba si la fuerza vital es necesaria en ellos. Es lo mismo que hago yo con el concepto del libre albedrío.

De modo que desmonto este concepto y compruebo si, en las secciones pertinentes, es necesario tener indeterminación. O si la indeterminación nos conduciría a una idea o un concepto del libre albedrío más fuerte. Quisiera demostrarles que no es así. En efecto, podemos poner indeterminación donde queramos, pero eso no nos proporcionará una idea más fuerte del libre albedrío.

Por tanto, mi hipótesis será que la libertad se puede entender como una especie de autodeterminación, la cual es compatible con la determinación en general, de manera que es posible la autodeterminación en un mundo determinista, y, si es compatible, con la posibilidad de que todo lo que ocurre en nuestro cerebro se realice físicamente, incluso los actos de voluntad. Este concepto también hace justicia a nuestra intuición precientífica, o sea a lo que queremos decir cuando hablamos de libertad en un sentido o en un entorno precientífico. También tendría como consecuencia que los actos volitivos se presten a explicación científica. Finalmente, no diré que nuestro mundo sea indeterminista, ni tampoco haré ningún esfuerzo por encontrar una pequeña, diminuta escapatoria para la indeterminación en un mundo determinista. Por el contrario, me limitaré a colar un poco de indeterminismo con objeto de salvar la libertad. Desde mi punto de vista, no necesitamos la indeterminación. Ésta nos priva del control.

Empezaré por un concepto mínimo de la libertad, y luego intentaré averiguar si dicho concepto puede resultar demasiado débil, tras lo cual comentaré algunos resultados empíricos que parecen demostrar que, aunque la libertad sea compatible con el determinismo, podría no ser compatible con algunos de nuestros resultados científicos más recientes.

Parte I

Un concepto mínimo de libertad

Ya he dicho que una de las razones básicas por las que necesitamos una aclaración filosófica es que nuestras intuiciones o nuestra comprensión precientíficas del libre albedrío son incoherentes y confusas. Si nos preguntáramos qué entendemos por libertad cuando hablamos de ella, obtendríamos preguntas totalmente distintas;

por tanto, la cuestión es que si esas intuiciones precientíficas son confusas e incoherentes, ¿cómo es posible que los filósofos piensen que pueden formular un concepto coherente sin apartarse de nuestras intuiciones precientíficas? En el caso de otros muchos conceptos, esto podría resultar muy difícil, o incluso imposible, pero en lo tocante al libre albedrío, tenemos una intuición negativa y estable y muy coherente, por lo que tenemos ideas muy claras sobre lo que no son actos libres. Ése será mi punto de partida.

Esta cuestión presenta dos aspectos básicos importantes. El primero es que diferenciamos la libertad de la compulsión. Cuando estamos muy seguros, decimos de los actos realizados bajo compulsión que no son libres, y eso nos proporciona un principio de autonomía muy débil, según el cual los actos libres tienen que ser autónomos, y no estar determinados externamente ni realizados bajo compulsión.

Aún hay un segundo principio muy importante. Distinguimos los actos libres, responsables, de los hechos fortuitos o aleatorios. Los hechos aleatorios no son libres, y la diferencia importante entre éstos y los actos libres es que un acto libre se puede atribuir a un autor, a un agente, lo cual es muy importante, pues, en el caso contrario, no podemos responsabilizar a nadie de un hecho aleatorio, porque escapa a nuestro control, aunque suceda en nuestro cerebro.

Veamos algunos ejemplos de hechos aleatorios. Imaginemos que una persona es llevada a juicio acusada de haber disparado a alguien, y que su abogado consigue demostrar que si apretó el gatillo fue debido a un hecho aleatorio sucedido en su corteza motora. Un simple hecho aleatorio. En ese caso, diríamos que el hecho escapaba al control de esa persona. Es como si se hubiera producido un hecho aleatorio en el arma, o en cualquier otro sitio. Digo esto para ilustrar que diferenciamos la libertad del azar, cuya diferencia más importante parece ser que los actos libres se pueden imputar a una acción. Por tanto, si existió una relación entre el actor y la activación del gatillo, entonces diremos que es el actor responsable, pero si no existe relación porque la activación del gatillo fue un hecho aleatorio, entonces diremos que el actor no es responsable. Ésa es la diferencia importante.

Autodeterminación

Antes de continuar, me gustaría señalar sucintamente qué supone la noción de autonomía en relación con el sistema nervioso central. *Autónomo* significa 'independiente de la compulsión ejercida por una fuerza externa', una idea muy estable acerca de la libertad. Y sólo para entender, para hacer justicia a esta intuición, es por lo que existe la diferencia entre fuerza y libertad. De modo que es un concepto de autonomía muy débil, ya que únicamente separa la fuerza de la determinación externa.

Ahora bien, resulta bastante fácil hacer justicia a esos dos requisitos, autonomía e iniciación. Si traducimos libertad por autodeterminación, ésta implica una distinción de la determinación externa, de modo que exige este principio de autonomía y, en segundo lugar, la autodeterminación implica, asimismo, que el acto viene determinado por el propio actor y, por ende, requiere este principio de autoría o iniciación.

Así pues, imaginemos que yo tengo una creencia muy fuerte según la cual robar es reprensible. Eso es realmente lo que creo. E imaginemos que he pagado por los artículos que contiene mi cesta sólo porque tengo esa creencia. Eso nos brindaría una razón excelente para asumir que lo que hice estaba autodeterminado. Poco importa cómo lo averigüemos, pero si nos limitamos a asumir que ésa era mi creencia y que actué movido por ella y que esa creencia era constitutiva para mí, entonces tenemos una buena razón para suponer que lo que hice fue autodeterminado.

Si eso es cierto, en ese caso la autodeterminación parece proporcionarnos un análisis adecuado para evaluar ese concepto mínimo de libertad, que se basa únicamente en esos dos supuestos de autoría y autonomía. Por supuesto, puede ser necesario introducir criterios adicionales —a lo que me referiré más adelante—, pero si traducimos la libertad por autodeterminación, querremos saber algo sobre ese yo que se autodetermina, en estos casos de acto libre.

El yo

¿Qué hace un agente que ejecuta o realiza un acto libre?, ¿cuáles son los criterios relevantes? El primero es que el actor que es capaz

de llevar a cabo el acto tendrá alguna creencia, ciertos deseos y ciertas disposiciones que son realmente sus propios deseos y disposiciones. Y aquí vemos que no se trata solamente de criterios racionales, sino también de criterios emocionales, de motivos emocionales que se traducen o se pueden traducir por actos autodeterminados —en el sentido de lo que plantea el Roth—.

La cuestión importante es que si no existieran tales creencias, deseos y disposiciones que son los actores o que son constitutivas de ellos, no tendría sentido hablar de un actor. Y si no tiene sentido hablar de un actor, entonces no tiene sentido hablar de autoría. Por tanto, si sabemos que tenemos que hablar de autoría, entonces sabemos que tenemos que hablar de esas preferencias, sean cuales sean.

Así, no hay autoría sin esas preferencias. Y esto nos plantea un problema evidente, dado que yo puedo tener algunas preferencias, quizá sea alcohólico o tenga algún otro comportamiento compulsivo, debido a las cuales se pueda afirmar que no actúo libremente, sino en función de esas preferencias compulsivas, dado que si alguien bebe una cerveza o un whisky siendo alcohólico entonces no actúa libremente.

Por tanto, lo que necesitamos es una distinción entre las preferencias que son realmente constitutivas de una persona, y las preferencias que no lo son, que son tal vez patológicas, compulsivas, etc. Ante la amplitud de esta problemática, señalaré sólo un criterio, necesario para tener un concepto claro de la libertad: se deben diferenciar las preferencias que se pueden imputar a un actor de las que no se le pueden imputar. Me refiero al criterio de la capacidad del agente de decidir efectivamente contra la preferencia.

Retomemos el ejemplo de mi convicción de que robar es reprensible. Podría tomar una decisión contraria a esta preferencia si se me dieran buenas razones de que robar es conveniente en ciertas condiciones o que robar es aceptable en ciertas circunstancias, por ejemplo para salvarle la vida a alguien. Por tanto, esa convicción de que robar es reprensible parece necesitar dicha condición.

En segundo lugar, consideremos mi alcoholismo. Si de verdad soy alcohólico, se me podrán dar las mejores razones del mundo y ninguna bastará para que me libre de mi adicción. Quiero decir

que tal vez logren convencerme para que me dirija a una clínica o busque ayuda, pero no seré capaz de tomar una decisión efectiva por mi cuenta, por lo que el alcoholismo y esta preferencia no requieren ese criterio. De modo que parece que ese criterio tiene la capacidad de diferenciar las creencias y los motivos que alguien se puede atribuir a sí mismo de las que no.

Conclusiones preliminares

Si esto es cierto, nos conduce a algunas consecuencias interesantes con respecto a la cuestión inicial, a saber, la cuestión de la determinación y la de una posible realización física de nuestro estado mental, incluidos los actos de voluntad.

Si estoy en lo cierto, entonces un acto es autodeterminado y por ende libre, en ese sentido mínimo, si está determinado por las características que constituyen el yo; es algo que se desprende simplemente de lo que acabo de decir, y significa que el acto se puede explicar haciendo referencia a las preferencias del actor. Por tanto, si por referencia a mis creencias y convicciones se puede explicar que viniera a Madrid en lugar de quedarme en Heidelberg, entonces venir aquí es una decisión autodeterminada, y no un acto determinado externamente.

Imaginemos que resulta que mi proceso de toma de decisión se realizó físicamente. Un especialista podría colocarme en el túnel de un tomógrafo y demostrar que todo lo que sucedió en mi cerebro cuando tomé esa decisión es explicable a nivel neurobiológico. ¿Significa eso que la mía no fue una decisión autodeterminada? No. Si estamos en un mundo puramente físico sin un alma inmaterial, entonces mis creencias, mis motivos, mis preferencias se tienen que realizar físicamente, pues de lo contrario no podrían ser efectivas.

Así pues, la realización física no interfiere en modo alguno con mi capacidad de tomar decisiones autodeterminadas, sino que me ayuda o me permite tomar esas decisiones. Ésta es la primera consideración importante.

En segundo lugar, supongamos que mi acto es determinado. Si viene determinado por mis propias preferencias, entonces es un acto autodeterminado, por mera definición, de modo que de ello se

desprendería que la determinación tampoco interfiere en mi capacidad de llevar a cabo actos autodeterminados. Esto se puede demostrar muy bien (y lo desarrollaré un poco más adelante) si se añade cierto grado de indeterminación. Imaginemos que tengo la firme creencia de que no es justo ni correcto saltarse un discurso que se me ha propuesto pronunciar.

Y conjeturemos que en un momento de indeterminación me quedo en Heidelberg. Eso me impediría actuar de acuerdo con mis propias creencias y motivos y con mis propios deseos, de modo que lo que se desprendería de ello —tal como lo demostraré más adelante— es que la indeterminación puede interferir con mis preferencias y con mi capacidad de actuar de una manera determinada.

Por tanto, podemos concluir que la cuestión decisiva no radica en si un acto es determinado, sino *cómo* es determinado, si viene determinado por mí mismo, por mis propias creencias y deseos, o si es determinado por factores externos. La determinación por el yo, por mí mismo, es simplemente, por definición, autodeterminación.

Parte II

Intuiciones y argumentos

Ahora bien, éste era el concepto mínimo. Se podría plantear que falta algo fundamental en mis intuiciones relativas a la libertad. Por ejemplo, nuestro pensamiento de que la libertad implica la capacidad de actuar de otro modo, algo que Gerhard Roth ha llamado el «alternativismo fuerte».

El problema es que en un mundo determinista no se puede hacer otra cosa y, si lo he pasado por alto, el concepto que he presentado es demasiado débil y podríamos olvidarnos de él.

Desde luego, en un mundo determinista parece realmente imposible hacer otra cosa. Si nuestro mundo fuera determinista, entonces haría miles de años que estaba determinado que yo viajaría a Madrid en lugar de quedarme en Heidelberg, y mi sensación de libertad no sería más que una idea caprichosa.

Por consiguiente, mucha gente que piensa que la libertad y la determinación son básicamente compatibles rechaza el principio se-

gún el cual la libertad implica la capacidad de actuar de otro modo. Esa gente piensa que el mundo está determinado; y la determinación no deja espacio a la capacidad de actuar de otra manera. Pero como pensamos que la libertad y la determinación son compatibles, tenemos que rechazar ese requisito de la capacidad de actuar de otro modo para salvaguardar la libertad. A mí no me parece que ésta sea una buena estrategia. No puedo entrar en detalles en estas páginas, pero diré que aunque se han expuesto argumentos y experimentos del pensamiento para apoyar esa posición, yo no creo que dichos argumentos sean realmente concluyentes. De todas formas, una vez más, quisiera seguir otra estrategia: quisiera preguntarme si es realmente necesario que vivamos en un mundo indeterminista para que tengamos la capacidad de actuar de otro modo.

Según la interpretación corriente, la capacidad de actuar de otro modo implica que en circunstancias internas y externas idénticas son posibles dos actos distintos. Por tanto, volvamos al momento en el que me encontraba en la estación de Heidelberg cuando, según esta última interpretación, en esas mismas condiciones, habría sido igualmente posible tanto que viniera a Madrid como que me quedara en Heidelberg. En las mismas condiciones, lo cual implica con las mismas creencias, los mismos deseos, las mismas disposiciones de mi yo.

Si ése fuera el caso —si quisiéramos explicar lo que ocurrió después, es decir que viniera a Madrid—, no podríamos explicar mi decisión haciendo referencia a mí, a mis creencias pues si, como hemos dicho antes, a la luz de mis creencias y mis deseos ambos actos son posibles, quizá con el mismo grado y la misma probabilidad, en ese caso mis creencias y deseos no nos proporcionan una explicación de lo que hice. Es más, el acto puede variar completamente con independencia de mi calidad de autor o actor del mismo. No está determinado por mí, de lo cual se desprende que el acto varía independientemente de mí, de modo que no puedo ser el autor. No pude determinar lo que sucedió. Si entendemos que actuar de otro modo significa que se pueden producir dos hechos en las mismas condiciones, lo que suceda no dependerá de mí, de forma que no puedo ser el verdadero autor de lo que sucede. Por tanto, en ese caso, no podemos decir que yo podría haber actuado

de otro modo, sólo que podría haber sucedido otra cosa, y que habría sucedido con total independencia de mis creencias y mis deseos y mis disposiciones. Así pues, ésta no parece ser la interpretación adecuada de qué significa tener la capacidad de actuar de otro modo.

Ahora volvamos atrás. Comprobemos si la idea de autodeterminación nos da una explicación o una interpretación mejor de qué significa tener la capacidad de actuar de otro modo. Si he llevado a cabo un acto autodeterminado (cuando vine a Madrid en lugar de quedarme en Heidelberg), entonces el hecho de que viniera a Madrid en vez de quedarme en Heidelberg se podrá explicar a la luz de mis preferencias, creencias y deseos. Mi acto no estuvo determinado por factores externos. Éste es el criterio que tenemos claro hasta ahora de este concepto mínimo.

Siendo así, es decir, si el acto depende de mí, antes de que yo actúe se puede decir que yo decido si voy a Madrid o me quedo en Heidelberg. Y, obviamente, esto implica que puedo ir a Madrid y puedo quedarme en Heidelberg. Sólo si se cumplen estas dos condiciones tiene sentido decir que soy yo quien decide. Si no me puedo quedar en Heidelberg —porque tendría que pagar una penalización por no cumplir con mis obligaciones en Madrid—, entonces no diríamos que es decisión mía, porque no sería así, y el hecho de que yo viniera a Madrid o me quedara en Heidelberg dependería de otras personas. Pero si está justificado decir que es decisión mía —y ya hemos hablado de las condiciones—, esto implica que me puedo quedar en Heidelberg y que puedo venir a Madrid. Si esto es cierto, entonces, por supuesto que antes de que yo viajara a Madrid, era posible decir: «Puedo ir a Madrid y puedo quedarme en Heidelberg», con lo cual, a toro pasado, estaría justificado decir que podría haberme quedado en Heidelberg, aunque en realidad haya venido a Madrid. Ésta es la cuestión importante.

Por tanto, la mejor interpretación de la capacidad de actuar de otro modo consiste simplemente en preguntarse si lo que ocurre depende del actor, lo cual es simplemente una implicación de esta idea de autodeterminación, de manera que la autodeterminación te brinda la capacidad de actuar de otro modo, gratuitamente. Y la

idea de que necesitamos el determinismo para tener la capacidad de actuar de otra manera no es más que una comprensión errónea de esta capacidad de actuar de otro modo.

Probablemente a muchos lectores aún no le convenzan mis argumentos, por eso quiero demostrarlo mejor. Es posible que no estén convencidos porque piensen que quizá no sea ésa la cuestión pertinente; tal vez no hayamos desmontado el organismo en la intersección adecuada. Por tanto, quisiera desmontar este organismo por otras intersecciones para demostrarles que introducir la indeterminación no acrecienta la libertad.

Desmenucemos, pues, la causa del cambio que me conduce a la decisión de viajar a Madrid, justo inmediatamente antes de mi nacimiento. Muchos de ustedes pensarán que si lo que yo hago viene determinado por mis disposiciones genéticas y por las leyes de la naturaleza, es posible predecir lo que haremos 10 ó 100 años antes de nuestro nacimiento. Si fuera así, ¿cómo podríamos actuar libremente cuando se puede predecir lo que vamos a hacer? Porque en ese caso ya está determinado, y no podemos actuar libremente. Bien, desenvolvamos la causa del cambio, y veamos si de lo que se trata aquí es de determinación, en cuyo caso tendríamos más libertad si introdujéramos nuestra determinación. Así que vamos a interrumpir la cadena causal justo antes de mi nacimiento, de manera que no se puedan hacer predicciones. ¿Qué pasa entonces? Quizá quede algún espacio adicional para la libertad, no lo sé. Lo que sí sé es que no puedo hacer uso del mismo, porque no lo estoy viviendo. Así que lo que podría conferirme una mayor libertad, es decir, interrumpir la cadena causal justo antes de mi nacimiento, suprimiendo con ello la capacidad de que alguien predijera hace miles de años lo que yo iba a hacer, no puede ser la cuestión importante. Aunque tal vez debamos interrumpir la cadena causal un poco más tarde, es decir, que quizá deberíamos desmontar el organismo en un empalme distinto.

En ese caso, interrumpamos la cadena causal, por ejemplo, 2 días antes de que se iniciara el proceso de toma de decisión. Antes de esa interrupción de la cadena causal, yo creía que robar era reprensible, pero después lo he aceptado, como si tal cosa. Ahora bien, ¿incrementa eso mi capacidad de actuar libremente? La cuestión impor-

tante es que yo no tuve nada que ver con ese cambio. Éste surgió de forma totalmente independiente del resto de mis creencias. Por supuesto, puedo revisar mi creencia, como he dicho antes, si se me dan buenas razones para hacerlo, pero no ha sido así; simplemente, mi creencia ha cambiado. Así que si se produce cualquier cambio sin que yo tenga control alguno sobre el mismo, y a consecuencia de ese cambio yo me pongo a robar, entonces tampoco tengo control sobre mis robos. Por tanto, parece claro que esta interrupción y esta instalación de indeterminación, de renuncia a la determinación, interfiere con mi capacidad de llevar a cabo un acto autodeterminado.

En ese caso tal vez hayamos elegido otra vez un momento inadecuado en el tiempo. Interrumpamos, pues, el proceso de toma de decisión, la cadena causal, precisamente durante el proceso de toma de decisión. Imaginemos que estoy pensando si debería viajar a Madrid o quedarme en Heidelberg, y justo al principio de ese proceso pienso en todas las razones que parecen apoyar la idea de que es bueno ir a Madrid. Bien, pues ahora entra en juego la indeterminación, y la diferencia consiste en que todas esas razones no fructifican en mi decisión, así que puede resultar que aunque crea que debo ir a Madrid, y mis firmes convicciones me lo digan, a la hora de la verdad no vaya, contrariamente a mis creencias. Lo cual parece interferir con cualquier capacidad de seguir un proceso de decisión racional.

Imaginemos un Parlamento donde alguien está reflexionando sobre una legislación en el marco de un debate donde se esgrimen argumentos a favor de la misma al principio de la discusión, y argumentos contrarios a la legislación al final de la sesión, pero ésta se interrumpe a la mitad, por lo que todo lo que se ha dicho antes queda sin efecto. Esto no sería mayor libertad, sino simplemente una pérdida de control, y ésa es la cuestión importante. La indeterminación nos hace perder control, pero no nos brinda libertad. Alcanzamos mayor libertad cuando tenemos una capacidad adicional de llevar a cabo actos autodeterminados, no por el hecho de perder determinación. Por tanto, la conclusión parece ser que introducir indeterminación no nos brinda mayor libertad, no nos proporciona un incremento de la libertad. Una vez más, la cuestión decisiva

no es *si*, sino *cómo* se determina el acto. Y añadiría que estrictamente hablando no es necesaria la determinación, y que también podría ser que existiera una relación de probabilidad muy fuerte entre las intenciones o las creencias de alguien y lo que hace. Eso no supone el menor problema.

Por tanto, no necesitamos determinación en un sentido estricto; la cuestión importante es que la renuncia a la determinación no nos proporciona mayor libertad, sino que únicamente desemboca en una pérdida de control.

El segundo aspecto importante es que, con independencia de que se realicen moral o físicamente, nuestros actos de voluntad siguen siendo autodeterminados, a condición de que estén determinados por nuestras propias creencias o por su realización neutral. Pero aunque estos procesos se produzcan en las zonas del cerebro donde se desarrollan nuestras creencias y deseos, el acto sigue siendo autodeterminado, de lo cual se desprende que libertad y determinación y libertad y realización física son compatibles.

Parte III

Resultados empíricos

Hasta ahora sólo he hablado de aspectos filosóficos, de manera que he hablado de patrones, es decir, de patrones conceptuales, y he intentado demostrar que la indeterminación no figura entre ellos. Por supuesto, no sólo es una forma de aprender cuáles son los patrones, sino también de determinar si los humanos los necesitamos o no. Ésa es una pregunta para la ciencia empírica, y seguirá estando abierta mientras no tengamos suficiente ciencia empírica; por otro lado, también es una cuestión de sentido común.

Así las cosas, parece existir un punto muy crítico, si tenemos en cuenta algunos resultados no tan recientes de la ciencia empírica. La cuestión importante es que la autodeterminación, el acto realizado por libre voluntad, requiere intenciones, aunque una forma muy simplificada de verlo consistiría en decir que la decisión autodeterminada expresa una intención que puede ser consciente o no —eso no es lo importante—, y que esta intención, a su vez, provoca un acto. Por tanto, si tomo una decisión autodeterminada, a

consecuencia de la misma albergo la intención de viajar a Madrid, de modo que mi viaje a Madrid viene causado por una intención consciente, mediada por diversos bucles que pueden ser o no conscientes, y que, por supuesto, pueden implicar emoción, además de consideraciones racionales.

La pregunta que se nos plantea es: ¿cómo sabemos que es así? Una respuesta obvia sería que tenemos una sensación de albedrío. Sabemos que hemos actuado por ciertas razones o en función de ciertas creencias, y esa sensación de albedrío discrimina el acto intencional del involuntario. El problema que se plantea a continuación es que existe todo un cuerpo de investigación que parece avalar la idea de que nuestras intenciones no son responsables de nuestros actos. Este cuerpo incluye la obra de un científico procedente de Madrid, J.M.R. Delgado, radicado en Estados Unidos. Basándose en un estudio similar realizado por Wilder Penfield en la década de los años cincuenta, fue capaz de conseguir que las personas ejecutaran actos mediante estimulación cerebral. Estimuló el cerebro de algunos de sus pacientes, que fueron capaces de realizar ciertos actos: movían sus cuerpos, o movían sus cabezas. Por tanto, en ese caso, el acto no estaba causado por las intenciones del agente, sino simplemente por la estimulación cerebral. La cuestión importante es que sus pacientes registraron lo que hicieron como un acto deliberado. Incluso dieron razones, de modo que confabularon justificaciones. Dijeron, por ejemplo, que habían girado la cabeza porque habían oído un ruido, o porque estaban buscando sus pantuflas, o porque estaban nerviosos. Esto parece demostrar que nuestra sensación de albedrío es engañosa.

Eso mismo se desprende de un conjunto de investigaciones que se remonta a las décadas de los años sesenta y setenta sobre pacientes con el cerebro dividido.

Respecto del hemisferio derecho, se pidió a la paciente que riera. Hizo lo que se le ordenaba. La orden era inconsciente para ella, no podía informar sobre la misma, pero cuando se le preguntó por qué se había reído, dio un motivo y dijo que los experimentadores estaban bromeando, lo cual no era cierto. Los experimentadores no sólo no estaban bromeando, sino que simplemente no había motivo alguno para que ella se echara a reír. De modo que esto parece

respaldar la idea de que nuestra sensación de albedrío es engañosa. Esos y otros experimentos nos han llevado a la idea de que, en general, nuestra sensación de intencionalidad es irreal, es decir, que no actuamos movidos por las intenciones que se nos atribuyen. Si a toro pasado nos preguntaran por qué hemos hecho algo, cuando hemos actuado debido únicamente a causas o procesos inconscientes, de modo que el proceso de generación del acto implica un proceso causal que escapa a nuestro acceso consciente, daremos una interpretación *post hoc,* lo cual resulta en nuestra sensación de albedrío.

La causalidad del acto es un proceso cortical subconsciente, y no está determinada por la intención consciente. Si esto fuera cierto, entonces no sólo no habría libre albedrío, sino que ni siquiera habría voluntad consciente en absoluto. Ésa es la hipótesis desarrollada por Daniel Wegner en *La ilusión de la voluntad consciente* (2002). En realidad, visto así, el mero hecho del registro consciente puede ser una ilusión extraordinaria, el equivalente a un mago que se saca un elefante del pliegue de su pañuelo. ¿Acaso no es verdad? Yo no creo que lo sea. Existe una evidencia considerable de que la intención es real y de que es efectiva, aunque el cuadro sea mucho más complicado que la descripción simplificada que expuse al principio, según la cual la intención causa el acto, y éste, una sensación de albedrío. Así pues parte de esta evidencia procede de experimentos con los llamados pacientes de «mano ajena o extraña». Éstos no son capaces de controlar una de sus manos en ciertas situaciones, de modo que pueden discriminar entre actos intencionales e involuntarios, por lo cual su sensación de albedrío es bastante fiable, aunque no sea infalible.

Lo mismo cabe decir del supuesto «fenómeno de vinculación intencional» investigado por Patrick Haggard durante los últimos años. En este caso, el experimento estaba diseñado de tal modo que los sujetos tenían que realizar un acto en un momento determinado, tras lo cual escuchaban un sonido, al cabo de 250 milisegundos. La tarea consistía en evaluar cuándo se habían producido esos dos sucesos. La cuestión importante era que si sólo había un acto y un sonido sin relación con el mismo, los sujetos podían eva-

luar esos 250 milisegundos de forma bastante fiable, pero si el sonido era una consecuencia o un efecto intencional de su acto, entonces subestimaban el lapso entre el acto y el sonido. Por tanto, si pretendían producir el sonido, subestimaban el lapso transcurrido entre el acto y su efecto, y, a la inversa, si el acto o el movimiento se había producido por estimulación magnética transcraneal, es decir, si era resultado de una inducción externa únicamente, los sujetos sobrestimaban el lapso en unos 50 milisegundos, esto es, en torno al 20%. Así pues, esto parece demostrar que nuestra sensación de albedrío tiene la capacidad de discriminar entre un acto intencional y un acto involuntario. Por tanto, la idea de que aunque a veces nos equivoquemos con respecto a nuestra sensación de albedrío no demuestra que dicha sensación sea absolutamente indigna de confianza, ni nos permite concluir que la intención no desempeña una función causal en la génesis de nuestro comportamiento.

Lo que también se desprende de otros estudios es que el proceso, o esa sensación de albedrío, no es producto inmediato de nuestras intenciones, sino que surge a partir de un proceso o mecanismo cognitivo *post hoc* que también toma en cuenta los datos procedentes del entorno para darle sentido. Imaginen que soy un jugador de fútbol, que tengo la intención de meter un gol y que, efectivamente, se pita un gol; de ello no se desprende que sea yo quien ha realizado el acto. Puede haberlo marcado otro jugador, o puedo haber sobornado al árbitro, entre otras posibilidades; es decir, puede haber muchos motivos por los que se pita ese gol. Por tanto, sólo si cuento con una evaluación *post hoc* del proceso podré estar seguro de que me he atribuido los actos, y solamente los actos que realmente he llevado a cabo yo mismo. Pero si existe tal proceso, por supuesto que será falible, como han demostrado los experimentos presentados antes.

Queda pues por demostrar que existe una conexión entre intención, acción y nuestra sensación de albedrío, si bien esta conexión parece ser mucho más complicada que en el caso de nuestra creencia precientífica. No obstante, los datos de que disponemos hasta ahora no refutan la idea de que tenemos la capacidad de llevar a cabo actos intencionales.

Conclusión

Quisiera terminar demostrando que no hay ninguna revolución a la vista, es decir, que no existe una refutación científica de la libertad y la responsabilidad. La libertad y la determinación no son incompatibles, de modo que aunque hayamos demostrado que nuestro cerebro es un sistema determinista, todavía no hemos demostrado que no seamos capaces de actuar libremente. Aparte de eso, de todas formas los resultados científicos parecen dejar espacio a la responsabilidad, y creo que existe una necesidad de revisión, así que he intentado demostrarles que, por lo que se refiere a la sensación de albedrío, el mecanismo puede diferir en cierta medida de nuestras expectativas precientíficas, lo cual no sólo es cierto para la sensación de albedrío, sino también en relación con nuestra conciencia de nosotros mismos y con la conciencia a secas, naturalmente. Esto implica, además de una revisión de nuestras creencias, una de nuestros conceptos, de nuestras ideas sobre esas capacidades.

Hasta finales del siglo XIX las personas hablaban del alma como fuente de su conciencia o de la conciencia de sí mismas, de nuestro libre albedrío y de nuestra capacidad de tener vida interior. Esa teoría fue desplazada por el concepto de la conciencia. Considero que necesitamos proceder a revisiones similares de nuestros conceptos a la luz de los nuevos descubrimientos científicos. Eso es precisamente lo que he intentado demostrar en lo tocante al libre albedrío. Por tanto, debemos prescindir de ese antiguo concepto tradicional del libre albedrío, a favor de un concepto que sea realmente capaz de evaluar los hallazgos científicos de los que disponemos. Y si lo hacemos así, esa revisión quizá nos permita después hacer un uso de los resultados científicos que nos ilustre mucho más sobre los cimientos y los límites de nuestras capacidades humanas. Por consiguiente, esta revisión nos puede ayudar a averiguar más sobre nosotros mismos, algo que, por supuesto, tendrá consecuencias en la educación, en el ordenamiento jurídico y, desde luego, también en la vida cotidiana.

Bibliografía

HAGGARD, P. (2005). Conscious Intention and Motor Cognition. *Trends in Cognitive Science,* 5 (6), 290-295.

HAGGARD, P., & SAM CLARK (2003). Intentional Action: Conscious Experience and Neural Prediction. *Consciousness and Cognition,* 12, 695-707.

WEGNER, D. M. (2002). *The Illusion of Conscious Will*. Cambridge MA: MIT Press.

Ser uno mismo: la base neuronal de la autoconciencia

Marc Jeannerod

Atribución e identificación falsa de la acción en la esquizofrenia

¿Cómo nos reconocemos a nosotros mismos como agentes y cómo diferenciamos los actos que realizamos de los realizados por otros agentes? La identificación de uno mismo como generador de una acción puede ser una tarea relativamente fácil cuando las acciones realizadas se ejecutan abiertamente. En esta situación, las señales sensoriales provienen de las extremidades en movimiento y de los efectos que tales movimientos provocan en el mundo exterior: estas señales se pueden comparar con las producidas por el mecanismo generador de acción, de modo que, como resultado de dicha comparación, podríamos clasificar tales acciones como autogeneradas, lo que dificulta el entendimiento de un gran número de situaciones en las que, a pesar de que el mecanismo generador de acción está activado, ésta permanece encubierta, por ejemplo en el caso de las acciones imaginadas (imágenes motoras). Sin embargo, las acciones imaginadas, así como otros estados mentales que involucran acciones encubiertas (pensar en hacer algo, por ejemplo), son claramente autoatribuibles, aunque no sean ejecutadas abiertamente y no existan señales sensoriales asociadas a las mismas. Por otro lado, las acciones realizadas por las personas de nuestro entorno que podemos observar están también claramente diferenciadas de las que nosotros mismos realizamos (Jeannerod, 2003).

Por muy sutil que esto parezca, el mecanismo de identificación de acción puede estar alterado en condiciones patológicas. Tales situaciones ofrecen muchos ejemplos en los cuales existe una identificación falsa de la acción: un caso típico es el de la esquizofrenia. Sin embargo, la identificación falsa no se manifiesta de forma uniforme entre los pacientes de esquizofrenia. Este tipo de perturbación se observa preferentemente en un caso particular de pacientes que presentan los denominados «síntomas de primer orden». De acuerdo con Kurt Schneider, estos síntomas se refieren a un estado en el cual los pacientes interpretan sus propios pensamientos o actos como resultantes de fuerzas o personas ajenas y se sienten controlados o influenciados por otros (Schneider, 1955). Los síntomas de primer orden (llamados aquí síntomas «schneiderianos») reflejan una alteración del mecanismo que, normalmente, genera la conciencia de nuestros actos y pensamientos y hace posible la atribución correcta de los mismos a su verdadero autor. Consecuentemente, un estudio de la atribución del comportamiento en pacientes esquizofrénicos nos ayudará no sólo a entender los factores responsables de la atribución incorrecta en los pacientes, sino también a esclarecer esta función crítica en la vida normal.

El patrón de falsa identificación de los actos en estos pacientes tiene un doble sentido: de acuerdo con el psiquiatra francés Pierre Janet, los pacientes o bien atribuyen sus propios pensamientos y actos a otras personas en vez de a ellos mismos (*under-attributions*); o bien los pacientes pueden atribuirse a ellos mismos los pensamientos y actos de otros (*over-attributions*) (Janet, 1937). Un ejemplo típico del primer tipo de falsa atribución (*under-attributions*) son las alucinaciones. Los pacientes esquizofrénicos que padecen alucinaciones pueden mostrar una tendencia a incorporar los acontecimientos externos a su propia experiencia y, por consiguiente, atribuyen de manera errónea sus propias intenciones o acciones a agentes externos. Durante las alucinaciones auditivas, el paciente oye voces que son típicamente experimentadas como procedentes de una entidad externa poderosa pero que, en realidad, provienen de una subvocalización o habla interna generada por el propio paciente. Estas voces son comentarios dirigidos al paciente en tercera persona e incluyen órdenes e instrucciones para realizar una acción

determinada (Chadwick y Birchwood, 1994). El paciente puede declarar que está siendo dirigido por una fuerza externa, como si sus pensamientos o actos fueran controlados por un agente externo. El denominado comportamiento mimético, que se observa en la fase aguda de la psicosis, también se refiere a esta categoría. El segundo tipo de falsas atribuciones (*over-attributions*), que Janet (1937) denominó «exceso de crédito», corresponde a una ilusión del paciente en la cual las acciones de los demás son, en realidad, iniciadas o realizadas por el propio paciente y él/ella está influenciando a los otros (el cuadro clínico de megalomanía). En este caso, los pacientes están convencidos de que sus intenciones o acciones pueden afectar a los acontecimientos externos: creen, por ejemplo, que pueden influir en el pensamiento y en las acciones de otras personas. De este modo, tienden a atribuirse de forma errónea la autoría de acontecimientos externos a ellos mismos. La consecuencia de esta mala interpretación sería que los acontecimientos externos son vistos como el resultado de sus propias intenciones y actos.

1. ¿Los pacientes esquizofrénicos schneiderianos son incapaces de controlar sus propias acciones?

La teoría más extendida para explicar las dificultades relacionadas con la atribución de acciones en pacientes esquizofrénicos es la propuesta por Christopher Frith y sus colaboradores. Esta teoría, que fue formulada por primera vez a principios de la década de los años noventa (Frith, 1992), ha sido reformulada recientemente en el contexto del mecanismo para el control de la acción. La idea general es que el delirio de influencia se origina por una falta de conciencia respecto a la posición predecible de un miembro. El razonamiento de la escuela de Frith es el siguiente: «En circunstancias normales, la conciencia de iniciar un movimiento debe depender de la posición predecible que alcanzará un miembro, ya que la conciencia de iniciar un movimiento precede al movimiento real y a cualquier retroalimentación relacionada con la posición real del miembro. Los pacientes que sufren delirio de control son conscientes de su objetivo, de su intención de moverse y de que su movimiento ha ocurrido, pero no de haber iniciado el movimiento. Es

como si el movimiento, aunque intencional, hubiese sido iniciado por una fuerza externa» (Blakemore *et al.*, 2002, p. 240). Por tanto, el problema clave de estos pacientes sería que, debido a su condición patológica, las señales endógenas relacionadas con la generación de acción (la denominada «copia eferente») estarían o bien ausentes, o bien usadas de forma inadecuada por el sistema nervioso (Frith *et al.*, 2000).

Estas opiniones controvertidas se sustentan en pruebas experimentales con pacientes que presentan síntomas schneiderianos, mostrando su incapacidad para controlar las discrepancias entre las posiciones que intentan alcanzar con sus miembros y las previstas. En estos experimentos, a los pacientes se les mostraban unos objetivos visuales movibles que debían intentar coger mediante el desplazamiento de una palanca, sin retroalimentación visual de sus movimientos. Si el objeto se movía en una dirección inesperada, los pacientes no tendían a corregir su error cuando ya habían iniciado el movimiento en la dirección que consideraban adecuada. Por el contrario, los mismos pacientes no tenían dificultad para hacer las correcciones oportunas cuando tenían retroalimentación visual del desplazamiento de la palanca (por ejemplo, Frith y Done, 1989; Mlakar *et al.*, 1994). Este tipo de experimentos, sin embargo, incide principalmente en la parte automática del sistema de control de la acción y tiende a desestimar la parte consciente de este mecanismo que es la que puede justificar explícitamente la acción. Dicho en otras palabras, la conexión entre la incapacidad para hacer correcciones, probada experimentalmente, y la falsa atribución de la acción a su agente, observada clínicamente, no está definida de una manera clara.

Para poder establecer esta conexión, primero necesitamos enfrentar el tema de la consciencia de la acción en pacientes esquizofrénicos. Teniendo en cuenta este punto de vista, Fourneret *et al.* (2001) usaron un paradigma similar al de Frith y Done (1989) para evaluar la capacidad no sólo de producir correcciones automáticas, sino también de hacer juicios conscientes de la actividad motora. Usaron un aparato en el cual los sujetos no podían verse las manos; en cambio, veían un punto moviéndose en una pantalla, cuya posición se correspondía con la punta de una especie de pun-

zón que sostenían con la mano. La tarea de los pacientes era trazar una línea con el punzón hasta un objetivo localizado frente a ellos en la pantalla, en la dirección del eje sagital del cuerpo. Además, se introdujo electrónicamente una desviación sistemática en el trazado del punzón, de modo que, cuando trazaban una línea en la dirección correcta, aparecía desviada unos 15° hacia la derecha. Por tanto, los sujetos tenían que desviar el movimiento de sus manos para poder alcanzar el objetivo. Finalmente, en algunos bloques de ensayos, una máscara opaca ocultaba la pantalla, de modo que los sujetos no podían ver la línea que dibujaban completa, sino sólo el último tercio de la trayectoria hasta el objetivo. En el experimento participaron pacientes con síntomas schneiderianos y sin ellos, así como también sujetos de control normales. Aproximadamente se necesitaron realizar diez ensayos con cada sujeto perteneciente a los tres grupos, hasta que aprendieron a compensar gradualmente la desviación y mantener la misma estrategia de movimiento del brazo hacia la izquierda a lo largo de toda la actividad, incluyendo el momento en que la máscara ocultaba la visión del punto. Este resultado indica que los pacientes, así como los sujetos normales, eran capaces, en ausencia de una retroalimentación visual, de iniciar el movimiento de la mano en la dirección adecuada para alcanzar el objetivo. Esto significa que controlaban las señales del comando central, las modificaban para realizar la tarea y las mantenían a lo largo de todo el ensayo. Por otro lado, la mayoría de los pacientes pertenecientes al grupo de los que manifestaban síntomas schneiderianos fueron capaces de describir explícitamente la estrategia que utilizaron para alcanzar el objetivo. Por tanto, este experimento (Fourneret *et al.*, 2001) demuestra que este tipo de pacientes son capaces de controlar conscientemente sus acciones, siempre y cuando sean evaluados con un diseño experimental apropiado.

2. Un déficit específico de atribución en pacientes esquizofrénicos

Estos hallazgos sugieren que la explicación dada para la falsa identificación y por la atribución incorrecta de los movimientos observados en pacientes schneiderianos debe ser analizada a un ni-

vel de consciencia de la acción directamente relacionada con los experimentos en que los pacientes son los agentes de sus acciones (su sentido de la agencialidad) más que a un nivel motor-sensorial. Para abordar este tema, examinaremos los resultados obtenidos con grupos de pacientes esquizofrénicos en experimentos especialmente diseñados para evaluar la autoidentificación y la atribución. La metodología básica usada en estos experimentos consiste en registrar juicios de agencialidad sobre movimientos que se muestran a los sujetos y que pueden, o no, corresponder con sus propios movimientos. El primer experimento de este tipo fue realizado por Daprati *et al.* (1997). Se les pidió a los sujetos que realizaran movimientos simples con los dedos de la mano derecha, la cual debían mantener escondida. Al mismo tiempo, en una pantalla se les mostraba movimientos realizados por una mano derecha de procedencia desconocida, es decir, que podía ser de ellos o de alguna otra persona. Finalmente, los movimientos realizados por la mano de la pantalla podían ser iguales o diferentes a los que los sujetos habían realizado. En los casos de máxima incertidumbre respecto a la procedencia de la mano de la pantalla, los pacientes schneiderianos cometieron errores de atribución en el 80% de las pruebas, esto es, atribuyeron los movimientos de la mano extraña a sus propios movimientos en la mayoría de los casos. En las mismas condiciones, la tasa de error de los sujetos control fue de sólo el 30%, y de, aproximadamente, el 50% en los pacientes no schneiderianos. Por tanto, los resultados revelan la existencia de una deficiencia de atribución de la acción en la esquizofrenia, especialmente en los pacientes schneiderianos (ver también Franck *et al.*, 2001). Por otra parte, estos pacientes tienden, en mayor medida, a atribuirse incorrectamente a ellos mismos los movimientos realizados por las manos que ven más que a la atribución de sus propios movimientos a otros agentes.

La tendencia exagerada de los pacientes esquizofrénicos a atribuirse acciones ajenas evidencia un sentido de la agencialidad alterado. Es importante señalar que este efecto podría estar asociado a una dificultad para relacionar las acciones realizadas por uno mismo con las consecuencias sensoriales de dicha acción. Por ejemplo, cuando a estos pacientes se les pide que indiquen el momento en el que detectan un sonido producido en un tiempo fijo determinado,

después de haber realizado un movimiento, tienden a acortar el tiempo transcurrido. Dicho de otro modo, tienden a enlazar estrechamente las acciones que realizan con las consecuencias sensoriales de las mismas (Haggard *et al.*, 2003).

3. La intensidad de la falsa atribución en los pacientes esquizofrénicos

Según Schneider, los síntomas de primer orden observados en pacientes esquizofrénicos son la expresión de falsas creencias que conducen a un sentimiento de «despersonalización», que deteriora la distinción entre uno mismo y los otros. Si esta suposición es correcta, entonces la falsa atribución debería observarse no sólo en las acciones que implican un comportamiento manifiesto, sino que también debería manifestarse más profundamente, hasta llegar a los mecanismos generadores de acciones, es decir, a los estados mentales relacionados con acciones encubiertas. Entre los síntomas schneiderianos, las alucinaciones verbales representan un caso típico de falsa atribución de comportamiento encubierto. Como se ha mencionado anteriormente, se sabe que las alucinaciones auditivas verbales en los pacientes esquizofrénicos están, en realidad, relacionadas con el habla interna producida por el propio paciente; incluso algunos pacientes con alucinaciones muestran una leve actividad muscular en los músculos laríngeos (por ejemplo, Gould, 1949; David, 1994). De este modo, los pacientes perciben su habla interna como voces procedentes de una fuente externa. McGuire *et al.* (1996) compararon la actividad cerebral de los pacientes esquizofrénicos con propensión a las alucinaciones (a los que denominó «alucinadores»), con pacientes sin propensión a las alucinaciones («no alucinadores») y con sujetos normales de control. Se les pidió que pensaran una oración simple (habla interna) o que se imaginaran a ellos mismos escuchando esas mismas oraciones dichas por una voz exterior (visualización verbal). Los autores del estudio verificaron que los sujetos no produjeran sonidos vocales. Durante el desarrollo del experimento se midió la actividad cerebral utilizando las técnicas de tomografía por emisión de positrones (PET). El efecto principal causado por las actividades propuestas fue que los

pacientes con propensión a las alucinaciones, en comparación con los sujetos del grupo control y los no alucinadores, mostraron una reducción de la actividad en el gyrus temporal medio izquierdo durante la tarea de imaginar oraciones dichas por otra persona. La conclusión de McGuire *et al.* fue que la reducida actividad en un área específicamente involucrada en la producción del lenguaje oral podría interferir con la capacidad de los pacientes para identificar su propia producción oral como una acción generada por ellos mismos. El razonamiento del experimento llevó a los autores a concluir que las regiones involucradas en la producción del lenguaje oral (por ejemplo, el lóbulo frontal) normalmente informan a las áreas de percepción del habla de una inminente producción oral, de tal modo que el habla percibida posteriormente se identifica como una acción generada por el propio sujeto. Este proceso sería disfuncional en los alucinadores, de ahí su vulnerabilidad a las alucinaciones. Otro hallazgo complementario fue obtenido por Dierks *et al.* (1999) en pacientes schneiderianos. Cuando los pacientes experimentaban alucinaciones, el metabolismo cerebral aumentaba en la corteza auditiva primaria (gyrus de Heschl) en el lado izquierdo. Por tanto, durante las alucinaciones verbales, las áreas temporales auditivas permanecen activas, lo que sugiere que el sistema nervioso de estos pacientes actúa como si, realmente, estuviera procesando el habla de un hablante externo. Por el contrario, en el caso de sujetos sanos, el habla interna autogenerada está caracterizada por una disminución de la respuesta de la corteza auditiva primaria.

Estos datos no hablan a favor de un mecanismo de control de la acción anormal. Sin embargo, debido a que en las alucinaciones verbales no hay expresión vocal, el mecanismo de simulación anormal debería buscarse en el ámbito de la representación de acciones y no sólo de la ejecución de las mismas. Por consiguiente, se podría pensar que la red de conexiones que normalmente usamos para atribuir los pensamientos e intenciones a sus agentes (ya sea nosotros mismos u otra persona) está dañada en ciertos pacientes esquizofrénicos y, como resultado, no les permite «simular» adecuadamente las operaciones encubiertas necesarias para la atribución. Esta hipótesis podría concordar con el hecho de que, en la práctica clínica, los síntomas schneiderianos afectan fundamental-

mente a los pensamientos de los pacientes más que a las acciones realizadas.

4. El rol de los lóbulos frontales

La naturaleza de la disfunción responsable de la falsa atribución de acciones en condiciones patológicas es aún una pregunta sin respuesta. Se considera que los cambios en el patrón de conectividad cortical podrían alterar la forma de las correspondientes conexiones con las distintas representaciones o la intensidad relativa de la activación de las áreas que componen estas redes de conexión. Esta explicación concuerda con la clásica observación de una disfunción de la corteza prefrontal en la esquizofrenia.

Se sabe que la corteza prefrontal presenta una hipoactividad en muchos pacientes esquizofrénicos (la denominada «hipofrontalidad», véanse Ingvar y Franzen, 1975; Weinberger y Berman, 1996), y se ha visto, en exámenes post mórtem, que su aspecto morfológico es anormal (Goldman-Rakic y Selemon, 1997). Desde el punto de vista del comportamiento, los pacientes esquizofrénicos muestran signos de «hipofrontalidad comportamental» al ser evaluados con pruebas neuropsicológicas para explorar las funciones prefrontales (por ejemplo, pruebas de memoria operativa o de resolución de problemas que requieren el uso de reglas preestablecidas; véanse Posada *et al*, 2001; Posada y Franck, 2002). Dado que las áreas prefrontales ejercen normalmente un control inhibitorio sobre otras áreas involucradas en varios aspectos de los procesos motores y sensomotores, una alteración en este control podría derivar en una conducta aberrante en los pacientes esquizofrénicos. Esto se explica por la activación de ciertas partes de la corteza que, por lo general, están inhibidas, especialmente en las partes posteriores de los hemisferios. Sin embargo, los estudios de imaginería neurológica, mencionados anteriormente, han revelado que los pacientes que padecen alucinaciones verbales muestran una activación anormal de las áreas auditivas primarias en el lóbulo temporal izquierdo, como si estuvieran procesando un estímulo auditivo externo. Del mismo modo, se ha observado un incremento de la actividad en el lóbulo parietal posterior derecho, en pacientes con delirio de in-

fluencia, durante una prueba de identificación de la acción (Spence *et al.*, 1997). En un estudio reciente, Farrer *et al.* examinaron la actividad cerebral en pacientes schneiderianos durante una prueba en la que los sujetos veían una retroalimentación distorsionada de sus movimientos. En las mismas condiciones, los sujetos normales muestran una activación del lóbulo parietal inferior derecho correlacionada con el grado de la distorsión: cuanto más amplia es la distorsión, mayor es la activación (Farrer *et al.*, 2003). Los pacientes esquizofrénicos mostraron un patrón diferente. Para empezar, la actividad del lóbulo parietal inferior derecho, en condiciones basales, era mayor que en los sujetos normales de control. En segundo lugar, el aumento de actividad no se correlacionaba adecuadamente con el que se producía en la distorsión (Farrer *et al.*, 2004). Por tanto, los pacientes no tenían, en realidad, ninguna señal (tal como se pudo inferir por el cambio de actividad en sus lóbulos parietales) para saber si lo que veían eran sus propios movimientos o los realizados por un agente externo. No es sorprendente que este tipo de pacientes tuvieran dificultades para atribuir correctamente los movimientos a ellos mismos o para desentrañar si tales acciones eran propias o ajenas. Finalmente, es importante mencionar que los pacientes esquizofrénicos con delirios muestran un aumento en la actividad de los lóbulos parietales, incluso en reposo (Franck *et al.*, 2002), y que se observa un permanente aumento en la actividad de esta misma área en sujetos no psicóticos que presentan factores de vulnerabilidad a la esquizofrenia. Estos hallazgos demuestran la importancia de llevar a cabo un análisis exhaustivo de los síndromes psicóticos patológicos complejos (Jeannerod *et al.*, 2003).

Bibliografía

BLAKEMORE, S.-J., FRITH, C. D., y WOLPERT, D. (2002). Abnormalities in the awareness of action. *Trends in Cognitive Sciences*, 6, 237-242.

CHADWICK, P., y BIRCHWOOD, M. (1994). The omnipotence of voices. A cognitive approach to auditory hallucinations. *British Journal of Psychiatry*, 164, 190-201.

DAPRATI, E., FRANCK, N., GEORGIEFF, N., PROUST, J., PACHERIE, E.,

DALERY, J., y JEANNEROD, M. (1997). Looking for the agent. An investigation into consciousness of action and self-consciousness in schizophrenic patients. *Cognition*, 65, 71-86.

DAVID, A. S. (1994). The neuropsychological origin of auditory hallucinations. En: David, A. S., y Cutting, J. C. (eds.): *The neuropsychology of schizophrenia*. Hove: Lawrence Erlbaum Associates, pp. 269-313.

DIERKS, T., LINDEN, D. E. J., JANDL, M., FORMISANO, E., GOEBEL, R., LANFERMAN, H., y SINGER, W. (1999). Activation of the Heschl's gyrus during auditory hallucniations. *Neuron,* 22, 615-621.

FARRER, C., FRANCK, N., GEORGIEFF, N., FRITH, C.D., DECETY, J., y JEANNEROD, M. (2003). Modulating the experience of agency: a PET study. *Neuroimage*, 18, 324-333.

—, FRANCK, N., GEORGIEFF, N., FRITH, C. D., DECETY, J., GEORGIEFF, N., D'AMATO, T., y JEANNEROD, M. (2004). Neural correlates of action attribution in schizophrenia. *Psychiatry Research: Neuroimaging*, 131, 31-44.

FOURNERET, P., FRANCK, N., SLACHEVSKY, A., y JEANNEROD, M. (2001). Self-monitoring in schizophrenia revisited. *Neuroreport*, 12, 1203-1208.

FRANCK, N., FARRER, C., GEORGIEFF, N., MARIE-CARDINE, M., DALÉRY, J., D'AMATO, T., y JEANNEROD, M. (2001). Defective recognition of one's own actions in schizophrenic patients. *American Journal of Psychiatry*, 158, 454-459.

—, O'LEARY, D. S., FLAUM, M., HICHWA, R. D., y ANDREASEN, N. C. (2002). Cerebral blood flow changes associated with schneiderian first-rank symptoms in schizophrenia. *Journal of Psychiatry and Clinical Neuroscience*, 14, 277-282.

FRITH, C. D. (1992). *The Cognitive Neuropsychology of Schizophrenia.* Hove: Lawrence Erlbaum Associates.

—, y DONE, D. J. (1989). Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. *Psychological Medicine*, 19, 359-363.

—, BLAKEMORE S. J., y WOLPERT D. M. (2000). Abnormalities in the awareness and control of action. *Phiosophical Transactions of the Royal Society London* B355, 1771-1788.

GOLDMAN-RAKIC, P. S., y SELEMON, L. D. (1997). Functional and anatomical aspects of prefrontal pathology in schizophrenia. *Schizophrenia Bulletin*, 23, 437-458.

GOULD, L. N. (1949). Auditory hallucinations in subvocal speech:

objective study in a case of schizophrenia. *Journal of Nervous and Mental Diseases,* 109, 418-427.

HAGGARD, P., MARTIN, F., TAYLOR-CLARKE, M., JEANNEROD, M., y FRANCK, N. (2003). Awareness of action in schizophrenia. *Neuroreport*, 14, 1081-1085.

INGVAR, D. H., y FRANZEN, G. (1975). *Lancet*, 21, 1484.

JANET, P. (1937). Les troubles de la personnalité sociale. *Annales Médico-Psychologiques*, tome II, 149-200.

JEANNEROD, M. (2003). The mechanisms of self-recognition in humans. *Behavioral Brain Research*, 142, 1-15.

—, FARRER, C., FRANCK, N., FOURNERET, P., DAPRATI, E., y GEORGIEFF, N. (2003). «Recognition of action in normal and schizophrenic subjects». En: *The self in neuroscience and psychiatry*, Kircher, T., y David, A. (eds.). Cambridge: Cambridge University Press, pp. 380-406.

MCGUIRE, P. K., SILBERSWEIG, D. A., y FRITH, C. D. (1996). Functional neuroanatomy of verbal self-monitoring. *Brain*, 119, 907-917.

MLAKAR, J., JENSTERLE, J., y FRITH, C. D. (1994). Central monitoring deficiency and schizophrenic symptoms. *Psychological Medicine*, 24, 557-564.

POSADA, A., y FRANCK, N. (2002). Use of a rule in schizophrenia. *Psychiatry Research*, 109, 289-296.

—, FRANCK, N., GEORGIEFF, N., y JEANNEROD, M. (2001). Anticipating incoming events. An impaired cognitive process in schizophrenia. *Cognition*, 81, 209-225.

SCHNEIDER, K. (1955). *Klinische Psychopathologie*. Stuttgart: Thieme Verlag.

SPENCE, S. A., BROOKS, D. J., HIRSCH, S. R., LIDDLE, P. F., MEEHAN, J., y GRASBY, P. M. (1997). A PET study of voluntary movement in schizophrenic patients experiencing passivity phenomena (delusions of alien control). *Brain*, 120, 1997-2011.

WEINBERGER, D. R., y BERMAN, K. F. (1996). Prefrontal function in schizophrenia: confounds and controversies. *Philosophical Transactions of the Royal Society*, B351, 1495-1503.

El uso de la información visual para comprender y predecir lo que va a ocurrir

David Perrett

Trataremos aquí de las respuestas del cerebro mientras transcurren secuencias de acciones que acontecen en el mundo. Esto es aún un proyecto, ya que no conocemos del todo qué ocurre en realidad; sin embargo, parece que estas respuestas se dan para todo tipo de acciones visualmente reconocidas: acciones manuales, como agarrar, y acciones relacionadas con la cabeza y el cuerpo, observadas bajo el punto de vista de un mono.

Me centraré en algunos de los mecanismos que, a nivel celular, pueden analizar estas acciones. A primera vista, estos mecanismos parecen ser bastante elementales, ya que afectan a la percepción. Me gustaría desviar la atención de los mecanismos que parecen hacernos insensibles al mundo y sugerir que, realmente, constituyen una ayuda fundamental. Dichos mecanismos permiten predecir lo que ocurrirá a continuación: de ahí el título.

En este simposio nos pidieron que consideráramos la consciencia. Lo que revela nuestro trabajo es que nuestra consciencia perceptiva no es precisa. Nuestra experiencia perceptiva es aquélla en la que sentimos el futuro inmediato: algo que va a ocurrir en una décima de segundo, aproximadamente. Nuestra consciencia perceptiva está distorsionada: no vemos lo que en realidad está ocurriendo, sino que sentimos lo que está a punto de suceder. Argumentaré que esto puede ser así por la forma en que nuestro cerebro

procesa la información sensorial. La historia de cómo hemos llegado a esta conclusión es larga.

Los fisiólogos conocen muy bien el modo en que las células cerebrales responden a los estímulos. La respuesta de una célula puede ser «alta» o «baja». Cuando la respuesta es baja, decimos que el estímulo está fuera de la curva de sintonización de la célula o campo receptivo. Se dice que un estímulo que produce una respuesta alta está en el centro del campo receptivo de la célula. Prácticamente todas las células sensoriales tienen una curva de sintonización campaniforme.

Podemos planificar la codificación sensorial a medida que aventuramos lo que a una célula, en algún lugar del cerebro, le gusta en realidad. Esto es correcto en lo que se refiere a las posiciones táctil en la piel y visual en el ojo, y también lo es, según lo expuesto por Hubel y Wiesel, con respecto a la sensibilidad de la orientación de las células en la corteza visual. De la misma manera puede aplicarse a la corteza temporal, en lo que se refiere a rostros y manos, de modo que la curva de sintonización es una función omnipresente de las células en las partes sensoriales del cerebro. Esto mismo también se puede aplicar al sistema motor.

La idea que se va consolidando con los trabajos más recientes es que estas funciones de sintonización aparecen cuando los fisiólogos cartografían los estímulos individuales, uno por uno. Por ejemplo, se coge un buen estímulo, se pone a prueba y se consigue una gran respuesta celular. Se toma un estímulo débil y se consigue una respuesta pequeña de la célula, pero cada estímulo se cartografía de forma completamente aislada de cualquier otro. En el mundo real, las cosas cambian continuamente. Al girar la cabeza, se puede ver una secuencia de visiones; al estirar la mano, se puede observar otra. Cuando empleamos tales estímulos dinámicos de la vida real, nos damos cuenta de que las curvas de sintonización de la célula cambian, y lo que, a continuación, voy a exponer es que las células se convierten en «predictoras».

El trabajo original que guió nuestros propios estudios fue el de Charlie Gross[1], quien describió las células que, en la corteza tem-

[1] Gross, C. G., Rocha-Miranda, C. E., y Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology,* 35, 96-111.

poral, respondían a la visión de unas manos de mono. Algún tiempo después, estuvimos trabajando con la corteza temporal en la misma zona y descubrimos prácticamente lo mismo: células que respondían al contemplar acciones manuales. Lo que Giacomo Rizzolatti ha descrito en las áreas temporal, parietal y premotora del cerebro son componentes del sistema espejo. El «sistema espejo» ya es famoso porque vincula la representación sensorial de una acción manual (en la corteza temporal) con una representación motora (en las cortezas parietal y frontal).

Los estudios llevados a cabo por el grupo de Parma y por nosotros mismos[2] indican que la representación en la corteza temporal es totalmente sensorial. Las células, en esta región, son sensibles a lo que un mono presencia pero no a lo que éste hace. Así, al observar una mano manipulando la piel, se produce una descarga proveniente de unas células específicas en la corteza temporal. Esta visión provoca una respuesta buena en múltiples ensayos.

En las películas donde la mano actúa por su cuenta o donde la piel se deforma sola, no se observa ningún tipo de respuesta. Incluso en una película donde, como si fuera magia, la mano se mueve, pero sin contacto espacial con la piel, no se puede obtener ninguna impresión de causalidad y las células permanecen en silencio. Ésta es una de las características de las células del lóbulo temporal y que el grupo de Rizzolatti repitió en las «neuronas espejo» de la corteza frontal (área F5)[3]. Para que las células descarguen, tiene que existir algo que origine la relación entre el agente que lleva a cabo una acción y el sustrato o la meta de la misma. Por tanto, nos referimos a las células que codifican acciones en lugar de meros movimientos manuales.

Volviendo al trabajo del grupo de Parma, se hizo evidente que no sólo se podían activar las células espejo con la visión. El sonido de las acciones también podía activar dichas células. Comenzamos a explorar la integración audiovisual en la corteza temporal. Esta re-

---

[2] Keysers, C., y Perrett, D. I. (2004). Demystifying social cognition: a Hebbian perspective. *Trends in Cognitive Science,* 8, 501-507.

[3] Di Pellegrino, G., Fadiga, L., Fogassi, V., Gallese, V., y Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research,* 91, 176-180.

gión recibe información proveniente del sonido y la visión, por lo que no deberíamos sorprendernos si las células de esta región son sensibles a ambos.

Así, comenzamos las pruebas con películas en las que se podía rasgar un papel o donde la gente podía caminar. Lo que hacemos es poner a prueba las respuestas de las células al contemplar una acción y descubrir qué acción «les gusta» a ellas para, después, trabajar nosotros. Presentamos la acción visualmente y sin sonido, o sólo el sonido correcto que acompaña a la acción, o bien presentamos tanto la imagen como el sonido de forma conjunta. También hacemos cosas absurdas, como mostrar, al mismo tiempo, la sucesión de imágenes de una acción junto con el sonido que no le corresponde (por ejemplo, se muestra a alguien andando mientras la banda sonora suena en orden inverso). Rizzolatti describió las células F5, que podían procesar tanto la visión como el sonido de acciones, y dichas células parecen encontrarse en la corteza temporal[4]. Al tener, simultáneamente, la visión y el sonido, se obtiene una gran respuesta de las células de la corteza temporal. El sonido, por sí solo, no produjo nada, pero moduló la respuesta a la visión. Por tanto, lo que podemos afirmar es que, aproximadamente, un cuarto de las células de la parte sensorial del «sistema espejo» son acciones que codifican de un modo abstracto. Las células integran información de diferentes modalidades: la visión y el sonido de las acciones.

La codificación es algo más exacta que eso. Si tomamos la visión de una acción y la acompañamos del sonido equivocado, entonces obtenemos una respuesta reducida en comparación a las acciones acompañadas del sonido correcto. Por tanto, podemos decir que las células requieren algún tipo de coincidencia entre la vista y el sonido. Éste es un tema abierto, igual que el modo en el que las células llevan a cabo esta integración.

Esto es sólo una introducción al procesamiento sensorial de las acciones en la corteza temporal y bien puede ser que haya una ruta que llegue desde aquí hasta las neuronas espejo en F5, a través de la

---

[4] Barraclough, N. E., Xiao, D., Baker, C., Oram, M. W., y Perrett, D. I. (2005). Integration of visual and auditory information by STS neurons responsive to the sight of actions. *Journal of Cognitive Neuroscience,* 17, 377-391.

corteza parietal. Me centraré en el procesamiento de imágenes individuales, las cuales pueden mostrarse en una secuencia aleatoria. Las imágenes que se suceden de forma muy rápida (1/20 de segundo por imagen) son difíciles de distinguir. Pero si hacemos la presentación más lenta (1/5 de segundo por imagen) es evidente que se pueden ver todas y cada una de ellas e, incluso, se pueden reconocer los rostros cuando aparecen. Hemos utilizado esta técnica de presentación rápida para cartografiar la codificación sensorial en la corteza temporal. Es un método eficiente, ya que, si las células responden cuando se va muy deprisa, se pueden evaluar muchas imágenes.

¿Qué ocurre cuando se exponen miles de imágenes de prueba? Una de las cosas que hemos descubierto es que las células pueden ser específicas de un modo extraordinario. Con la exposición a diversos grupos de imágenes, cada uno durante sólo 50 milisegundos (el equivalente a un fotograma individual de un programa de televisión, insertado en otros fotogramas de diferentes imágenes), estas imágenes muy breves pueden provocar una respuesta celular[5]. Una célula determinada puede responder de forma fiable cada vez que aparece la imagen de la cara de un mono o el cráneo de una rana. Existen otras muchas células sintonizadas a algo referente a rostros que no responderán a otros estímulos.

Al utilizar las pruebas rápidas, se descubre que las células de la corteza temporal codifican elementos significativos concretos, como los rostros; otras células codifican las manos, y otras, la visión de los cuerpos. Esto nos indica que una imagen, cuando aparece brevemente, podría ser suficiente para provocar una respuesta celular en el cerebro. Otra propiedad que conocemos desde hace tiempo es que las células están sintonizadas: una puede responder a la visión frontal pero no a la visión posterior. Otras células están sintonizadas con otras visiones: por ejemplo, una célula que responde solamente a la visión del perfil derecho.

La mayoría de las células de la corteza temporal están sintonizadas con la visión específica o con objetos concretos y no se debe

---

[5] Földiák, P., Xiao, D., Keysers, C., Edwards, R., y Perrett, D. I. (2003). Rapid serial visual presentation for the determination of neural selectivity in area STSa. *Progress in Brain Research,* 144, 107-116.

pensar que hay sólo una visión: no existe una visión privilegiada, sino muchas que están siendo codificadas, cada una de ellas por diferentes grupos de células. Hago esta aclaración porque éstos son los antecedentes. Cuando hacemos la prueba rápida —50 milisegundos por imagen—, de nuevo hallamos que se produce sintonización de la imagen para un rostro. Para cualquiera de las células de las ocho diferentes identidades analizadas, encontramos una imagen sintonizada similar. Por eso, todas las propiedades que vemos con las presentaciones ralentizadas parecen estar ahí cuando las imágenes se presentan durante un tiempo muy reducido, tal vez durante sólo 50 milisegundos, o incluso aún menos —14 milisegundos—[6].

Por consiguiente, las células, en la corteza temporal, pueden procesar una información mínima cuando aparece una imagen sencilla. En resumen, la prueba rápida realmente demuestra que el cerebro puede procesar hasta 70 imágenes, sin relación entre sí, cada segundo, y que las células están sintonizadas para objetos significativos.

Consideraremos qué ocurre cuando las imágenes se exponen en una secuencia, y yo afirmaría que las acciones implican inevitablemente secuencias de posturas. Pero en vez de empezar por lo complicado y mostrar las acciones complejas, quiero comenzar, muy brevemente, con las secuencias de dos: sólo un par de imágenes, porque ésa es una secuencia mínima. Lo que hacemos es exponer una imagen, esperar y después exponer otra. Desde luego, podemos esperar un período variable de tiempo antes de exponer la segunda imagen.

¿Qué ocurre? Obtenemos una buena respuesta a la primera imagen. Poco después de la respuesta a una imagen, cuando exponemos otra sólo 200 milisegundos después, la respuesta a la segunda imagen es menor. Con un intervalo largo entre pares de estímulos de 450 milisegundos, la respuesta a la segunda imagen, que era reducida, se ha recuperado en gran parte.

---

[6] Keysers, C., Xiao, D., Földiák P, y Perrett D. I. (2001). The speed of sight. *Journal of Cognitive Neuroscience,* 13, 90-101.

Esto se denomina, en psicofísica, «enmascaramiento» y no nos dimos cuenta de la frecuencia de dicho enmascaramiento a nivel celular. Descubrimos que se da, prácticamente, en todas las células que analizamos y que dura, aproximadamente, 1/2 segundo. ¿Cuáles serían las consecuencias si fuéramos al cine y viéramos una escena? Pues, en realidad, son un tanto singulares: veríamos una imagen y el cerebro se iría a dormir porque suprimiría las respuestas a las imágenes siguientes. De este modo, comenzaría una nueva película, veríamos una pequeña parte de la misma y luego parecería que el cerebro se vuelve muy insensible al resto de las imágenes. Esto nos resulta paradójico, ya que no comprendemos qué ocurre en los videos, así que no puede ser cierto que el proceso de enmascaramiento inutilice nuestro cerebro por completo.

Una cuestión que aún no he tenido en cuenta es la similitud entre un par de imágenes. Podemos analizar una célula con una imagen y medir una respuesta. Cuando encontramos un segundo estímulo que también produce una respuesta, entonces es muy probable que las dos imágenes nos parezcan visual y conceptualmente similares. El grado en el que las células responden de forma similar a estímulos visuales determina, de forma rigurosa, lo semejantes que son los estímulos, desde el punto de vista perceptivo, para los observadores humanos. Lo que hemos analizado son pares de imágenes similares y de estímulos diferentes. Cuando utilizamos imágenes distintas, obtenemos una respuesta menor a la primera imagen, pero una respuesta mayor a la segunda. Por tanto, el enmascaramiento parece depender de la similitud de las imágenes; esto ha ocurrido así en todos los intervalos que hemos analizado.

Así pues, el enmascaramiento es alto si las imágenes son muy parecidas. Esto es justo lo que sucede cuando vemos un video o vamos al cine. Todas las imágenes son prácticamente iguales. Sólo hay un cambio gradual entre ellas: salvo que se trate de la MTV, la mayoría de las imágenes son realmente bastante parecidas.

Ahora nos referiremos a las secuencias y veremos qué ocurre cuando se está procesando un video. Aquí voy a analizar, por ejemplo, cinco imágenes: 1-2-3-4-5, todas ellas relacionadas entre sí. Con cualquier célula, con una curva de sintonización, podemos co-

menzar con una imagen efectiva, desde el centro de la sintonización y, después, trabajar gradualmente, mediante visiones cada vez menos efectivas que surjan, progresivamente, desde el centro hasta el borde de la curva de sintonización. Comenzaremos con una buena imagen y luego mostraremos una secuencia que se aleja de la misma. Cuando analizamos cada imagen 1-5 por separado, obtenemos un descenso gradual en la respuesta. Cuando analizamos las imágenes en sucesión rápida, como una secuencia de una película, obtenemos una buena respuesta a la primera imagen y después nada. Podemos esperar este cese rápido por el enmascaramiento. Una célula muestra una respuesta a una imagen y luego todo se apaga. Los datos obtenidos, después de analizar muchas células con una secuencia de imagen, producen resultados como los que he mostrado para una célula tomada como ejemplo concreto.

Pues bien, esto es lo que ocurre cuando se comienza con un estímulo adecuado y se va fuera del campo receptivo. Ahora me gustaría ver el efecto de la sintonización en los estímulos que se mueven por todo el campo receptivo. Analizamos, de nuevo, imágenes en la secuencia y comenzamos con el estímulo débil, después seguimos con el mejor estímulo (en el centro del campo) y volvemos al estímulo débil.

¿Qué ocurre? Consideremos, en primer lugar, una célula simple que responde, no a un movimiento facial, sino a imágenes de una acción manual. La primera imagen analizada era una mano moviéndose hacia el campo de visión, que probablemente es irreconocible, pero que produjo una pequeña respuesta. La segunda imagen (una mano que aparece con claridad) produjo una mayor respuesta. La tercera imagen puede reconocerse como la mano de un mono agarrando un objeto y estos estímulos resultaron relativamente efectivos. Cuando analizamos los fotogramas siguientes, vemos que son menos efectivos. Así, el fotograma número 3 produjo la mayor respuesta, y el número 2, una respuesta algo menor.

Nuevamente podemos analizarlo como en un video, donde mostramos los fotogramas en una secuencia de película. Llegados a este punto, ya puede adivinarse lo que ocurre. Obtenemos una respuesta buena al comienzo de la secuencia de imágenes y luego la respuesta desaparece. Esto es lo que ocurre: la respuesta de la célu-

la es máxima antes del fotograma número 3 y después desaparece rápidamente.

De nuevo pretendemos combinar datos para el análisis de la secuencia en múltiples células. Averiguamos qué estímulo es el mejor para cada célula: para una, es el rostro; para otra, es el medio perfil, etc. Unimos las respuestas de muchas células diferentes y analizamos la sintonización de cada una en relación con el mejor estímulo, cuando éste se analiza por separado. Para una población de células analizadas, si empleamos respuestas de medida a las visiones de forma separada, encontramos que estas respuestas son mayores hacia el centro de su curva de sintonización y después disminuyen. Éste es un campo receptivo clásico: una respuesta baja en los bordes del campo y una respuesta alta en el centro. Estos resultados acompañan al análisis de imágenes de forma separada. Cuando unimos las imágenes y creamos una secuencia en movimiento, descubrimos que el perfil de respuesta ha cambiado otra vez antes de lo que ocurriría en el caso del análisis de imágenes por separado.

Tras presentar los datos, volvamos ahora a las consecuencias. He dicho repetidamente que cuando mostramos movimientos dinámicos, la sintonización de todas esas células varía. Se alcanza el pico de respuesta de las secuencias dinámicas anteriores a la mejor imagen estática. Esto parece cumplirse en la mayoría de las células con las que hemos trabajado. El cambio en el ritmo del pico de actividad es bastante drástico: alrededor de una décima de segundo, lo que es casi la mitad de la curva de sintonización. Así, lo que se conocía como una curva de sintonización en términos clásicos, ahora ha variado, simplemente, poniendo los estímulos en un orden natural.

¿Qué consecuencias tiene todo esto para la percepción? Pues bien, si tenemos en cuenta la sintonización de una célula para el rostro, digamos que a la célula «le gusta» la visión frontal de la cabeza y su respuesta es máxima cuando se muestra todo el rostro. Cuando mostramos un estímulo en rotación (por ejemplo, mi cabeza moviéndose de un lado para otro) la actividad máxima se da mucho antes de lo que se podría esperar. Mientras la cabeza gira, esta célula, en concreto, está respondiendo mucho antes de que hayamos llegado a la visión frontal. Y eso es lo que quiero puntualizar

sobre estas curvas de sintonización: todo el perfil de actividad en el cerebro está distorsionado.

Durante cualquier secuencia natural de imágenes donde los estímulos cambian de forma gradual, la máxima actividad cerebral se produce, en cualquier momento, entre células que están sintonizadas con visiones que aún no han sucedido. Durante una acción dinámica, las células más receptivas están sintonizadas con posturas que ocurrirán si la acción continúa. Todo el aparato sensitivo está actuando con capacidad predictiva, anunciando lo que está a punto de ocurrir. Por consiguiente, nuestro mundo perceptivo no refleja lo que está ocurriendo en un momento determinado en el tiempo, sino que, realmente, refleja una predicción acerca de un posible estado futuro. Si el mundo es estático e inmutable, entonces la actividad cerebral proporcionará una valoración veraz de lo que los sistemas sensoriales reciben. En cuanto el mundo incluya el cambio natural, los sistemas que he descrito se pondrán en marcha y el cerebro predecirá el futuro del mundo. Desde luego, el futuro llegará de modo natural y estaremos preparados para ello. Ésa es, precisamente, la ventaja de un sistema predictivo: nos ayuda a anticipar y preparar las respuestas motoras.

Nuestra idea acerca de los sistemas sensoriales proviene de las mediciones de las respuestas estáticas constantes a estímulos aislados llevadas a cabo por los fisiólogos. Lo que he repasado aquí son dos mecanismos básicos: el primero es el concepto de curva de sintonización; el segundo, el fenómeno del enmascaramiento o adaptación. Justo después de un estímulo, la respuesta relacionada con éste es reducida. En el campo de la visión, se denomina «enmascaramiento». La adaptación también se produce al procesar el sonido y el tacto. Efectivamente, la adaptación a corto plazo se da en todos los sistemas sensoriales. Su efecto será el mismo en todas las modalidades. Cuando unimos las curvas de sintonización para estímulos aislados con la adaptación a corto plazo, estamos creando la codificación predictiva para estímulos dinámicos naturales[7].

[7] Berry, M. J., Brivanlou, I. H., Jordan, T. A., y Meister, M. (1999). Anticipation of moving stimuli by the retina. *Nature*, 398, 334.

Es preciso remarcar la palabra «natural». El mundo no cambia al azar. Cualquier cambio es progresivo y, normalmente, gradual. Los objetos no aparecen, desaparecen y reaparecen en lugares aleatorios. Se mueven en el espacio, variando gradualmente su posición. Desde luego, se puede dar la aceleración y la desaceleración en animales y objetos que colisionan, pero no es posible que los objetos salten. Esto lo pueden hacer posible los cineastas, pero tales creaciones cinematográficas no son «naturales». El sistema sensorial que poseemos se ha desarrollado o formado a través del aprendizaje, para que coincida con los cambios que tienen lugar en nuestro mundo.

Lo que se necesita es una nueva concepción de la naturaleza de los sistemas sensoriales. Debemos pensar en ellos como predictores. Durante todo el tiempo en el que están en funcionamiento, nos están proporcionando información acerca del posible estado del mundo. No deberíamos considerar la percepción y la sensación como una realidad; deberíamos darnos cuenta de que nuestra experiencia es una predicción.

Pregunta. El procesamiento de la información es, por supuesto, un aspecto fundamental: la información auditiva y visual. Asumiendo como correcto el enfoque predictivo, imagino que será necesario algún tipo de pauta para hacer algunas comparaciones o predicciones, para crear, más o menos, las características o el siguiente nivel de percepción, en el procesamiento de la información tanto visual como auditiva. En tal caso, ¿cómo se van a restaurar dichas pautas y cuál va a ser realmente el resultado?

Respuesta. Su pregunta se refiere al nivel de implementación, y un reciente trabajo sobre implementación de la codificación predictiva en la retina señalaba como el lugar de adaptación[8] no a las células, sino a las sinapsis. Dicho trabajo reivindicaba que la adap-

[8] Hosoya, T., Baccus, S. A., y Meister, M. (2005). Dynamic predictive coding by the retina. *Nature*, 436, 71-77.

tación era dinámica. Si su pregunta es sobre implementación, puedo especular. La reacción de las células vecinas podría hacer esto porque, en realidad, dichas células están detectando lo mismo: material similar visual y conceptualmente, así que su señal sería buena para inutilizar el cerebro en esa región. Así, cuando se ha visto una imagen, la siguiente que aparezca será procesada, más o menos, por la misma región del cerebro. Ahora bien, el hecho de que todas las células vecinas procesen esa información tendrá una reacción colateral y podría limitar el sistema.

Si hablamos realmente sobre implementación, lo siguiente también puede ser relevante: en otra serie de experimentos, lo que hicimos fue esperar a que una célula diera una respuesta buena o pobre al mismo estímulo y, entonces, medimos lo que sucedió con el siguiente estímulo. Así, la imagen 1 es siempre la misma y se obtiene una respuesta buena en algunas pruebas y una respuesta pobre en otras. La magnitud de la respuesta al segundo estímulo no dependió de la magnitud de la respuesta al primero. Obtenemos un enmascaramiento equivalente, lo que sugiere que éste no está provocado por la célula pequeña pobre que se cansa: son las vecinas las que apagan dicha célula. Incluso cuando ésta no tiene ningún motivo para estar cansada, porque no estaba respondiendo muy bien, todavía obtenemos el mismo enmascaramiento al aparecer el siguiente estímulo.

Creo que la respuesta a la pregunta: «¿Cómo se implementan el enmascaramiento y la codificación predictiva?» es una reacción colateral de las células vecinas. Ésa sería una manera sencilla de hacerlo. Se toma el rendimiento organizado desde el marco local y se emplea para restringir el sistema. Esto impide que el cerebro caiga en la epilepsia, pero se puede obtener el enmascaramiento de muchas formas. La verdadera respuesta es: «No lo sé», pero sospecho que esta reacción es una forma de hacerlo. Aunque tal vez ésa no era su pregunta.

P. No, pero, en cualquier caso, es muy ilustrativo tener en cuenta el nivel celular, aunque mi pregunta está más orientada al nivel funcional.

R. Sí, podemos adoptar una interpretación a nivel del sistema. Muchos de los experimentos sobre formación de imágenes en el ce-

rebro se han realizado utilizando un paradigma de adaptación RMf. Se muestran dos estímulos y el segundo provocaría una respuesta menor si está relacionado con el primero. Es un efecto de repetición. En el caso de un par de estímulos, sucede lo mismo con potenciales evocados. Podemos eliminar el segundo potencial evocado si el estímulo es conceptualmente similar al primero. Así que estas respuestas se producen en el nivel del sistema.

Para el organismo, ¿cuál es la función? Pues bien, he intentado explicar que una función podría ser la predicción. Por lo demás, es un mecanismo extraño: justo cuando se cansa, deja de responder. Esto no es de gran ayuda. Yo creo que necesitamos considerar el enmascaramiento y la adaptación en un contexto diferente, un contexto que considere el modo en que el cerebro procesa los estímulos cambiantes de forma natural. Si hacemos esto, la consecuencia funcional del enmascaramiento y la adaptación es la capacidad para predecir.

Profesor Jeannerod

P. Cuando se muestran acciones, la frecuencia de las imágenes puede cambiar la regla.

R. Maggie Shiffar y Jean Decety[9] mostraron una imagen de una pierna detrás de otra e, inmediatamente, 1/2 segundo después, mostraron otra imagen con las piernas intercambiadas. Todos pensaron: «Sí, la pierna se está girando», es decir, que el movimiento era anatómicamente posible. Pero si las imágenes se separan solamente una décima de segundo, la gente piensa: «¡Ah, esta pierna ha atravesado la otra!». La interpretación de una secuencia de acciones depende de la velocidad del movimiento y de la frecuencia de las imágenes.

El ritmo de las imágenes en una película de una acción natural no invalida lo que he dicho hasta ahora. A diferentes ritmos de presentación o velocidades de movimiento, la predicción abarcará un

[9] Stevens, J. A., Fonlupt, P., Shiffrar, M., y Decety, J. (2000). New aspects of motion perception: Selective neural encoding of apparent human movements. *Neuroreport*, 11, 109-115.

período de tiempo futuro mayor o menor. Si se da un movimiento a mayor velocidad, entonces la codificación predictiva trabajará sobre la misma escala de tiempo, pero la posición pronosticada o la postura anticipada se desarrollarán un poco más allá de la trayectoria de movimiento.

P. En este caso, el cambio de frecuencia de la presentación significa que el efecto es el mismo que en otras células receptivas en cuanto a la respuesta.

R. Realmente no hemos observado la frecuencia, pero con el experimento de Shiffar hay un cambio en el espacio y en el tiempo, así que se calcula una velocidad real. Por tanto, el movimiento «ilegal» se da cuando éste ocurre demasiado rápido para que las articulaciones funcionen. Existen dos consideraciones: el grado de desplazamiento del miembro y el ritmo. El ejemplo que puse para el movimiento imposible es bastante lento, 1/2 segundo, bastante más lento que la velocidad de los fotogramas que emiten por televisión.

Existe una industria artesanal en la percepción acerca de la velocidad de representación[10]. Si se varía la posición de un objeto, progresivamente, en las imágenes 1, 2, 3, 4, y se pregunta dónde está el objeto en la última imagen, la gente se equivoca, sobrestima la posición a lo largo de la trayectoria de movimiento. Hay un recorrido temporal para la predicción en la velocidad de representación, aproximadamente a esta velocidad, incluso si ésta es algo más lenta. Creo que cualquier frecuencia de imagen con la que juguemos, mientras sea inferior a 1/2 segundo entre imágenes, nos proporcionará efectos anticipatorios. Si vamos mucho más despacio, tal vez por encima del segundo entre imágenes, no creo que vayamos a ver anticipación.

No hemos realizado los experimentos a una velocidad lenta de imágenes. Comenzamos con un enfoque etológico. Tomemos una película de acciones naturales y separemos sus imágenes, una película en la que aparezca la acción de caminar, emitámosla a frecuencias naturales y veamos qué obtenemos. Todos pasamos casi todo

[10] Freyd J. J., y Finke R. A. (1984). Representational momentum. *J Exptl Psych: Learning, Memory and Cognition,* 10, 126-132.

nuestro tiempo viendo películas o televisión, pero no sabemos qué está ocurriendo realmente al procesar imágenes a la velocidad de presentación de imágenes de la televisión. Así que no hemos jugado demasiado con la velocidad. Creo que, hasta medio segundo, aún podemos obtener algún efecto.

P. Usted dijo que al presentar la misma fotografía 10 veces, la amplitud de la respuesta es cada vez más débil en la misma célula.

R. Si es una imagen muy similar, la respuesta se debilita, sí.

P. Pero, psicofísicamente, si presentamos 10 veces la misma imagen o imágenes similares, ¿se debilita la impresión de la percepción o permanece igual?

R. Yo preguntaría a los psicofísicos. Stuart Anstis describió la ralentización del movimiento. Si se observa un movimiento de forma continua, se ralentiza enormemente; si nos adaptamos a la luz radiante del sol, ésta se vuelve menos brillante. Ésta es la misma imagen, vista, repetidamente, a través del tiempo. Intentamos el enmascaramiento con pares de imágenes distintas. Obtenemos una alteración del tiempo de reacción para detectar un segundo estímulo, pero no hemos demostrado que esto se base en la similitud de la imagen al presentar la misma múltiples veces. ¿Sabemos si se obtiene una disminución de los qualia con una exposición repetida?

P. La cuestión es: ¿qué proximidad existe entre las respuestas de las células simples y la experiencia subjetiva? A veces, en las ilusiones ópticas, la similitud absoluta entre las respuestas celulares y la experiencia subjetiva está ahí. Si éste no es el caso, sólo existe otra interpretación: que la experiencia subjetiva no proviene directamente de una célula simple, sino de la memoria. A menudo vemos cosas a través de las lentes de la memoria y no de la experiencia real, así que, ¿serían decisivos para esa proximidad?

R. Los mejores experimentos al respecto los ha llevado a cabo Nikos Logothetis[11] y de él se pueden obtener informes bastante parecidos acerca de la experiencia perceptiva del mono y de la actividad de las células simples. Básicamente, si en nuestros estudios medimos la percepción en humanos durante presentaciones rápidas,

[11] Sheinberg, D. L., y Logothetis, N. K. (1997). The role of temporal cortical areas in perceptual organization. *Proc Natl Acad Sci USA,* 94, 3408-3413.

entonces el rendimiento se aproxima mucho a la habilidad de las células para detectar las mismas imágenes en las mismas presentaciones rápidas[12].

P. Resulta bastante sorprendente cómo se comportan de bien, a veces, estas células.

R. Sí, y tal como Giacomo Rizzolati describió, cada vez que el mono alarga la mano, se escucha una célula grabada que responde: «prrrrrr, prrrrrr, prrrrrr». No hay ninguna duda, no es necesario hacer esa estadística tan elaborada. Las células simples pueden tener una gran fidelidad de codificación. No están siempre en el ruido.

P. Si éste es el caso, usted enfoca la «equidad» de principio a fin en el sistema visual. Es un gran ejemplo de que se necesita más de una célula para ver mejor las cosas. Y otra cuestión es si tiene en cuenta los experimentos de Bill Newsome en la zona Temporal Medial, que también son un ejemplo muy interesante. La célula puede predecir perfectamente cuál es la reacción del momento y se puede manipular dicha célula con estimulación eléctrica o algo parecido, y se puede cambiar el cálculo de todo este asunto.

R. Es un trabajo de fisiología excelente, donde recrearon la labor equivalente a la que, supuestamente, llevaban a cabo las células y obtuvieron un rendimiento comparable del observador y de las células cerebrales individuales. El trabajo de Newsome fue fundamental en ese sentido.

P. En otro orden de cosas, usted ha afirmado que el sistema límbico está detrás de los ganglios basales, como una especie de motivación, pero ¿qué estructura está detrás de este sistema? O, en otras palabras, ¿se puede hablar, o no, de iniciación en el sistema nervioso?

R. Si se mira desde el punto de vista ontogenético, mucho antes del nacimiento, ya existe un sistema de evaluación primitivo en la amígdala, que funciona, tal como he dicho antes, desde la séptima semana de gestación en adelante. Y hay una evaluación muy primitiva que no es circular sino en espiral. Así que, desde el principio, existe el concepto de bien y mal y el niño que aún no ha nacido

[12] Keysers, C., Xiao, D., Földiák, P., y Perrett, D. I. (2001). The speed of sight. *Journal of Cognitive Neuroscience,* 13, 90-101.

realiza algún movimiento o hace algo y se produce una evaluación muy primitiva: esta consecuencia era buena o mala. Y esto lo llevan a cabo, en su mayor parte, la amígdala central, el núcleo del lecho de la estría terminal y, después, la amígdala lateral basal, que se conecta a la corteza cerebral después de nacer y que está aprendiendo siempre cosas más complicadas, de manera que se realiza una autoevaluación creciente. Éste es un proceso basal en el sistema límbico, de modo que adquiere experiencia en forma de espiral. Hasta donde sabemos, éste es el proceso primario que guía toda nuestra experiencia. Tenemos una entrada sensorial, una salida motora y una verificación de las consecuencias de la salida motora frente a la entrada sensorial. Y el proceso es sumamente primitivo antes de nacer, pero después se va haciendo cada vez más complejo, con la corteza cerebral en aumento. Por tanto, no existe ningún otro sistema de evaluación en el cerebro.

Ahora bien, puede darse el caso de que se destruya la amígdala en la edad adulta o, que la amígdala se dañe bilateralmente en el caso de pacientes con la enfermedad poco común de Pick; éstos son incapaces de reconocer expresiones faciales que expresen miedo o de reconocer situaciones amenazantes.

SOBRE LOS AUTORES

FRANCISCO J. RUBIA - Instituto Pluridisciplinar, Universidad Complutense, Madrid, España.

GERHARD ROTH - Instituto de investigación del cerebro, Universidad de Bremen, Alemania.

THOMAS F. MÜNTE - Departamento de Neuropsicología, Universidad Otto-von-Guericke, Magdeburg, Alemania.

MICHAEL PAUEN - Departamento de Filosofía, Universidad Otto-von-Guericke de Magdeburg, Alemania.

MARC JEANNEROD - Institut des Sciences Cognitives, Lyon, Francia.

DAVID PERRETT - Facultad de Psicología, Universidad de St. Andrews, Reino Unido.