Neuron
Perspective

Neuroimaging of Cognition: Past, Present, and Future

R.J. Dolan[1,*]
[1]Wellcome Trust Centre for Neuroimaging, University College London, London WC1N 3BG, UK
*Correspondence: r.dolan@fil.ion.ucl.ac.uk
DOI 10.1016/j.neuron.2008.10.038

Neuroimaging, particularly that based upon functional magnetic resonance (fMRI), has become a dominant tool in cognitive neuroscience. This review provides a personal and selective perspective on its past, present, and future. Two trends currently characterize the field that broadly reflect a pursuit of "where"- and "how"-type questions. The latter addresses basic mechanisms related to the expression of task-induced neural activity and is likely to be an increasingly important theme in the future. This trend entails an enhanced symbiosis among investigators pursuing similar questions in fields such as computational and theoretical neuroscience as well as through the detailed analysis of microcircuitry.

A Beginning in Neuroimaging

In the late 1970s, I commenced training as a psychiatric resident at a large mental institution in North London's Friern Barnet Hospital, uncertain how my future would unfold. This crumbling Victorian institution, which gave rise to the cockney colloquialism for a madhouse, "Colney Hatch" was famously, and sympathetically, depicted in Richard Hunter's book "Psychiatry for the Poor" (Hunter and MacAlpine, 1973). The very same author, a maverick psychiatrist, used to conduct a weekly postmortem brain dissection demonstration on his deceased patients. At the time, within mainstream psychiatry, the mere idea of studying the human brain as a means to unravel the mysterious nature of psychiatric illness was viewed as arcane and treated with derision. Psychiatry was then a discipline paralyzed by a pervasive intellectual agnosia when challenged as to the likely causes of severe mental disorder.

I found Hunter's weekly brain dissection demonstrations fascinating and frustrating. Although the patients coming to postmortem had suffered life-long severe mental illness, it was rare for us to detect any macroscopic pathology. Indeed, I cannot recall a single instance, despite the proselytizing zeal of our demonstrator, where a convincing clinicopathological correlation could be established. As part of the weekly ritual, and no doubt to reinforce our attendance, we were served tea and chocolate biscuits, which provided the context for idle conversation among my fellow psychiatric residents. We often, subversively given the context, speculated that the only conceivable avenue for progress was not through brain dissection, however fine grained, but by dint of some future technological innovation. What we often imagined was a dissecting device for the living, one that would enable the physiological function of the brain to be revealed in its entire splendor. Less than 30 years later, I need only to remind myself of this time period to fully appreciate the extraordinary advances that have ensued in the interim, developments that have delivered the sophisticated technology of functional brain imaging.

By 1984, I had completed my first stint as a researcher, thanks to a generous training fellowship from the Wellcome Trust. At this time, there was growing excitement about a technology that seemed to presage a conversion of our postmortem dissection table fantasies into reality. A number of centers around the world were beginning to use positron emission tomography (PET) to measure regional cerebral metabolism using radiolabeled glucose or regional cerebral blood flow (rCBF) using radiolabeled oxygen. The scientific outputs from these centers were primarily resting-state investigations, where the emphasis was on physiological quantification of blood flow and metabolism. Yet, for those of us who took an interest, we could immediately intuit its future possibilities. Within the space of a few years, a number of these centers were extending resting-state applications of PET, capitalizing on the potential of perfusion techniques to measure task-induced brain activity. In hindsight, these techniques appear crude and unsophisticated, but at the time they were of enormous importance in realizing a dream within neuroscience, namely the prospect of a physiological window into the human mind.

I am certain that the excitement and expectation generated by these developments was the principal motivating factor for the many young scientists who committed to the field and were happy to be labeled by their colleagues, often mockingly, as "imagers." The intellectual promise offered by what I perceived as a revolutionary means to study the brain was certainly sufficient motivation for me to return to research once I completed my psychiatric training. In 1989, I arrived at the MRC Cyclotron Unit at the Hammersmith Hospital in London where I and my close colleagues over the next 20 years, Karl Friston and Chris Frith, joined a team led by Richard Frackowiak. We committed, as did a number of other pioneering groups worldwide, most notably the group of Marcus Raichle in St Louis, to an ambitious program to use noninvasive imaging to map the functional architecture of the living human brain. What follows in this review is a highly selective perspective on how this field has evolved, its current and future directions. As a consequence, and because of a remit that includes a word restriction, the review makes no claim to comprehensiveness.

A Revolution in Studying the Living Human Brain

To measure brain activity associated with discrete states of mind is the holy grail of cognitive neuroscience. In the late 1980s and

early 1990s, there were two significant developments, resting on nascent progress in perfusion imaging, that proved pivotal in realizing this dream. The first was the development of sophisticated data analysis tools, subsequently becoming a standard in the field, which provided whole brain imaging analysis in the form of statistical parametric maps (Fox et al., 1988; Friston et al., 1991). Second, the early 1990s saw the surprising and rapid emergence of a new approach to brain mapping using MRI based upon the blood-oxygen-level-dependent (BOLD) technique (Ogawa et al., 1990). This latter approach had significant benefits over PET, including greater spatial resolution and the absence of the potential hazards associated with radioactive tracers. The significance of this development derives from the fact that neuroscience now had a technology that was noninvasive and that afforded multiple repeated measures of brain activity across various task manipulations. The end result was fMRI, a technique that afforded much greater experimental flexibility.

Apart from obvious methodological advantages associated with fMRI, it had, in my view, an equally important cultural consequence. Access to PET as a technique, entailing by necessity the use of ionizing radiation, was vested in the hands of an elite. Their privileged position was predicated on either having nuclear medicine expertise, a medical license, or both. The fact that fMRI involved no inherent biohazard meant that it could now be deployed within less restricted environments, for example, within psychology departments of universities. In essence, fMRI democratized access to a powerful technology for investigating the living human brain, allowing a broad cross-section of academic disciplines to pursue new agendas. This, in my view, accounts for a virtual exponential rise in published neuroimaging output that, over the past 15 years, has been largely led by questions related to human cognition. Most people would agree that democracy is a good thing; equally, most people would agree that all things taken in excess should carry a health warning.

A Selective History of Imaging Human Cognition

The first applications of activation-based neuroimaging in the late 1980s involved the use of PET and utilized simple subtraction techniques based upon presenting alternating blocks of stimuli. In essence, this approach involved measuring brain activity in a condition of interest in one block and then subtracting activity associated with a carefully specified control condition acquired in a second block. This procedure was predicated on the idea of pure insertion of sequential cognitive processes and allowed, in principle, localization of discrete cognitive functions (Friston et al., 1996). Early examples of this methodology included identifying brain regions that process single words (Petersen et al., 1988; Posner et al., 1988) and the human homolog of V4 (Lueck et al., 1989) and V5 (Zeki et al., 1991), areas specialized for visual color and motion processing, respectively. These so-called ''block designs'' also characterize early fMRI studies. Where PET measurements necessarily required long data collection sequences, such temporal inflexibility was overcome using fMRI.

In the first instance, more flexible thinking on the part of experimental design provided new, more rigorous approaches that involved the implementation of parametric (Grafton et al., 1992) and factorial experimental designs (Friston et al., 1992). These provided a higher degree of experimental control and inference. Such developments coincided with a new class of questions, such as ''how does the presence or absence of selective attention influence brain activity associated with processing some other factor?'' (O'Craven et al., 1997) In addition, this increasing sophistication in factorial methodologies provided the platform for in vivo pharmacological challenges in conjunction with cognitive activation paradigms, enabling a characterization of how neuromodulatory influences impact discrete cognitive processes in health and disease (Dolan et al., 1995).

At the end of the 1990s, experimenters began to harness fMRI to its full potential. This was initiated by the implementation of event-related fMRI, a procedure akin to measuring evoked responses in electrophysiology, and obviated the shortcomings of the less flexible blocked designs. Researchers were now able to capitalize on known physiological properties of neurons, such as their susceptibility to adaptation with repetition of an input, and, using this knowledge, to enhance both spatial and functional sensitivity. This latter approach, akin to repetition suppression, allowed detection of representations that might be instantiated within topographically overlapping neuronal networks. This new approach was applied to both low-level and high-level stimulus attributes (Tootell et al., 1998: Grill-Spector and Malach, 2001; Winston, et al., 2004).

The 1990s was a decade that witnessed an explosion in applications of neuroimaging that generated data on virtually every conceivable aspect of human cognition, including sensori-motor learning (Karni et al., 1998), attention (Kastner et al., 1998), short- (Courtney et al., 1998; Jonides et al., 1993; Owen et al., 1996b) and long-term memory (Fletcher et al., 1995; Schacter and Buckner, 1998; Shallice et al., 1994; Wagner et al., 1998), perception (Haxby et al., 1991, 1996), language (Howard et al., 1992; Petersen et al., 1990; Price et al., 1992, 1997), and executive functions (Baker et al., 1996; Grasby et al., 1993; Owen et al., 1996a). While many findings recapitulated what would be predicted from lesion based models, it also turned out there were many surprising findings. For example, in the case of episodic memory, it turned out that performance of relatively simple tasks engaged regions of the brain that would not have been predicted from lesion models (Petrides et al., 1995; Tulving et al., 1994a, 1994b). This reflected, in part, the fact that attribution of function, based upon a lesion-deficit model, rested on an assumption that the consequences of a lesion solely reflected disruption of function in the damaged region. However, sophisticated perspectives on the impact of a lesion account for the fact that cognition arises not just out of functional differentiation but also out of functional integration. Consequently, the impact of a localized lesion is likely to reflect both local and distributed effects, evidenced by more widespread activations seen with neuroimaging than was predicted by a lesion-deficit model.

One far-reaching impact of neuroimaging has been the extent to which investigators were now empowered to tackle issues and questions that were not reflected in classical psychological parsing of the mind, as might be found in standard textbooks of psychology. In this regard, the impact of neuroimaging has arguably been greatest in relation to topics that were either ignored in neuropsychology or were heretofore difficult to tackle experimentally. Two pertinent examples involve the growth of studies

that address the biological underpinnings of consciousness and emotion, respectively.

In studies of consciousness, neuroimaging provided the means to address disparate questions inconceivable without this new technology. These include how the exercise of "free will" engages discrete circuits (Frith and Dolan, 1996), the fate of unseen but psychologically effective stimuli (Dehaene et al., 2001; Dehaene et al., 1998), the correlates of subjective perception (Lumer et al., 1998; Lumer and Rees, 1999), imagined future states (Sharot et al., 2007), as well as the level of residual neuronal function seen in patients with varying impairments of consciousness (Owen, 2008), to name a few. In the field of emotion, functional neuroimaging allowed a characterization of how the brain responds to external emotional stimuli (Morris et al., 1996; Whalen et al., 1998), the expression of emotional learning and extinction (Buchel et al., 1998; LaBar et al., 1998), and how emotion impacts on perception (Noesselt et al., 2005; Vuilleumier et al., 2004) and memory (Kensinger and Schacter, 2006). Even more importantly, the subjective nature of emotional measures were overcome with direct investigation, where studies characterized the neural basis of distinct feeling states (Damasio et al., 2000) and enabled a description of how interoceptive states, for example, changes in peripheral autonomic status, are mapped in the brain (Critchley et al., 2004). Interestingly, the representation of interoceptive states in anterior insular cortex was subsequently shown to also provide a substrate that enables a person to represent the subjective feeling states of others as, for example, expressed in empathy for pain (Singer et al., 2004).

Less than 20 years since its inception, the field of neuroimaging, using fMRI, has reached a high level of maturity and methodological sophistication. Neuroimaging now attracts interest not only from cognitive neuroscientists but also from a wide array of fields that lay outside those that might be thought to be concerned about the brain. Consequently, we have witnessed the emergence of a strong symbiosis between functional neuroimaging and a range of other disciplines, in many instances constituting unlikely bedfellows, including genetics (Drabant et al., 2006; Hariri et al., 2002), economics (Camerer, 2003; de Quervain et al., 2004; King-Casas et al., 2005), ethics (Hsu et al., 2008), and aesthetics (Winston et al., 2007), to name a few.

Current Perspectives within Neuroimaging

A repeatedly asked question over the past 20 years is what brain imaging has brought to neuroscience that was not already known. First, functional imaging realized what was never previously possible: namely, a characterization of the functional anatomy of the intact brain without the confound of pathology and the likely consequential plastic reorganization in response to disease or developmental abnormalities. Second, functional neuroimaging highlighted that even simple tasks engaged more widespread areas of the brain than would have been assumed from the lesion-deficit approach. This has led to a richer conceptualization of how brain function underpins cognition not only in terms of functional differentiation (localization) but also in terms of functional integration (distributed function). This latter characterization has motivated a new class of questions and methodological approaches that address how distributed brain regions interact during performance of a psychological task, as characterized in terms of functional connectivity (Fletcher et al., 1996) (which assess a significant dependence or mutual information between regional activities in different parts of the brain) and effective connectivity (the causal influence one regions exerts over another) (Buchel and Friston, 2001; Friston, 1994; McIntosh and Gonzalez-Lima, 1994). Furthermore, these developments have now furnished investigators with the tools to ask questions related to how brain regions comprising a distributed network interact within that network after some experimental manipulation: for example, the presence or absence of attention or the evolution of some form of learning (Plailly et al., 2008).

Neuroimaging has now consolidated its position as the major platform for a systems level understanding of the human brain. Currently within the field, there appears to be two broad conceptual trends. On the one hand, there is a "Kuhnian" culture of normative and largely descriptive imaging neuroscience using standard and well-established approaches based upon classical experimental manipulations and data analysis. This constitutes the majority of the field and is a testament to its scientific maturity. This contrasts with emerging approaches that are more embedded in theoretical and computational neuroscience and largely concerned with a broader and more mechanistic set of questions that speak to fundamental questions regarding how the brain works. Here, neuroimaging is seeing a striking convergence with expertise from other fields, such as engineering, dynamic programming, and computational neuroscience.

One recent development that is more allied with the former descriptive approach has been the use of multivariate pattern classification procedures. An example is the use of support vector machines (SVMs) to establish statistical dependence between distributed responses in a circumscribed part of the brain and some experimental variable. These approaches have excited much media attention and have been in the vanguard of claims related to mindreading or predicting future intentions. In simple terms, these procedures involve training a classifier with brain activity obtained during an experimental manipulation: for example, being presented with a house or a face. Then, in a subsequent phase of the experiment, conducted in the same subjects, the trained classifier can provide a reliable prediction of the sensory input (was it a house or a face?) (Kamitani and Tong, 2005). From a scientific perspective, these approaches have been claimed to be more sensitive than equivalent univariate analyses (Haynes and Rees, 2006). What emerges in most multivariate classification applications is a probabilistic mapping between activity in a subset of brain voxels and some aspect of the sensorium or behavior. Criticisms of these approaches include a suggestion that they represent a return to neophrenology and an absence of clarity regarding what new neurobiological questions are specifically amenable to these classifiers. Since these classification techniques are extremely sensitive to subject and paradigm specifics, the precise nature of what these signals represent is at best difficult to pinpoint. Newer multivariate approaches go some way to addressing these concerns, for example, the use of decoding based upon quantitative receptive-field models that have allowed people to decode natural images from an observer's brain activity (Kay et al., 2008).

One exciting, and less contentious, development that reflects a convergence with theoretical neuroscience has been the

implementation of experimental paradigms motivated by computational rather than psychological models. This approach owes much to the impact of the late David Marr, who argued that a sine qua non for understanding the human brain, or any system within the brain, involves specifying a computational goal (Marr, 1982). Such a perspective has two immediate consequences in terms of empirical research. First, it speaks to defining the algorithmic nature of the transformations between input and output and, second, specifying how this is implemented in the hardware of the brain (Marr, 1982). The application of this approach has involved harvesting the outputs from a computational model that provides a formal description of a transformation between a sensory input and behavioral output. For neuroimaging, the key variables are those proposed to mediate this transformation, an example being a prediction error signal underpinning reward-based or punishment-based learning. These variables are then correlated against fMRI data to determine brain regions manifesting a response profile consistent with the model variables. The key advantage of these model-based approaches derives from the fact that they address how a particular cognitive process is implemented in a specific brain area as opposed to identifying the location of its instantiation. An early example has been the implementation of temporal difference (TD) learning models, where the striking finding was that a TD-like prediction error signal was expressed in the ventral striatum during reward learning (McClure et al., 2003; O'Doherty et al., 2003). Model-based approaches also enable specific inferences regarding underlying neuromodulatory influences. For example, the likely dopaminergic nature of a reward prediction error signal seen in reward-based neuroimaging experiments was demonstrated in an investigation that used a computational learning model, Q-learning, in conjunction with an in vivo pharmacological manipulation of dopamine function (Pessiglione et al., 2006). This general approach has opened up a very rich field of investigation, particularly in the context of human decision making, that has seen a characterization of neuronal responses associated with exploratory and exploitative choice behavior (Daw et al., 2006), reward predictability and uncertainty (Berns et al., 2001; Critchley et al., 2001), intertemporal choice (McClure et al., 2004), and subjective utility (Kable and Glimcher, 2007).

In line with these developments, neuroimaging is also witnessing a more mechanistic emphasis that unambiguously reflects a shift within the field from "where-" to "how-" type questions. This emphasis parallels the development of generative modeling of brain responses using, for example, Dynamic Causal Modeling (DCM) (David et al., 2006; Stephan et al., 2008; Friston et al., 2003). In DCM, the data (be it fMRI, EEG, or MEG) is viewed as being caused by a perturbation of hidden neuronal states by some experimental input. These perturbations produce neuronal dynamics through neuronal interactions that are passed through biophysically motivated forward models to form the observed responses (for example, the BOLD response in fMRI). Inversion of these models furnishes conditional densities (e.g., conditional mean and precision) on the parameters of the underlying neuronal and hemodynamic models and their marginal likelihood for model comparison. The identification or inversion of generative models, given some data, enables conditional inferences about the parameters of these models and, critically, comparison of different generative models (Penny et al., 2004). The term "model comparison" may be erroneously understood as an artificial delineation of parameter space but, in fact, simply embodies hypothesis testing where these competing hypotheses can be framed in terms of competing generative models of the same data.

This aforementioned approach contrasts with standard techniques that rely solely on establishing statistical dependencies or probabilistic mappings between behavioral and physiological data or between physiological data acquired from different parts of the brain, which provide no machinery for model comparison beyond the existence of that mapping. Recently, attempts have been made to marry the different modeling approaches described above and create models that jointly represent computational and physiological mechanisms of neuronal systems. For example, predictive coding models of perception can be formulated in terms of hierarchically structured dynamic causal models (Friston, 2005). Alternatively, computational models of learning and decision-making can be embedded into dynamic causal models; this allows one to investigate how computational processes are reflected by time-dependent changes in specific connection strengths or neuronal state variables. An example is a recent study that integrated a Rescorla-Wagner learning model into a dynamic causal model, demonstrating that prediction errors drive the consolidation of connection strengths during associative learning (den Ouden et al., 2008).

Functional neuroimaging has been heralded as a technology that will impact upon our understanding of nervous system diseases. A sanguine view is that its impact has been, at best, modest. In neurology, for example, it has shed light on mechanisms in recovery of function following stroke (Ward et al., 2004, 2006). Classificatory approaches, using support vector machines (SVMs), show considerable promise in providing a classification of likely neurodegenerative process, based upon structural brain appearance, that may surpass the accuracy of neuroradiological experts (Kloppel et al., 2008). Paradoxically, it is in the problematic field of psychiatry that a so-called "translation" potential holds the greatest prospect of realization. There is a growing theoretical and empirical sophistication within cognitive neuroscience related to understanding emotional learning and extinction, motivation, and reward. The likelihood that this sophistication can provide both a more principled classification of psychiatric disorders and a high level specification of aberrant cognitive processes should not be underestimated. In the case of schizophrenia, the ability to characterize neuronal interactions using approaches such as DCM, including how neuromodulatory influences alter these interactions (Corlett et al., 2007; Honey et al., 2008; Murray et al., 2008), means that theoretical models of this disease, based upon the idea that the disorder reflects a failure of integration, can now be tested empirically.

Macro- and Microscopic Mapping of the Brain

Current general theories of the brain invoke the idea that a core function is to infer the causes of sensory input. Almost universally, such theories call upon hierarchical network models of how sensory data are generated. The relevance of systems neuroscience arises out of the fact that hierarchical optimisation in the brain is distributed, entailing recurrent self-organized

interactions among different levels of sensory and motor hierarchies. At the systems level, brain imaging using fMRI (as well as EEG and MEG) provides for a precise delineation of these functional architectures. As described, one can invert biologically realistic models with a large number of state variables and biophysical parameters, using state-of-the-art variational Bayesian techniques, with the inherent ability to quantify the uncertainty about any experimentally induced changes (Friston et al., 2007). Inversion of these models allows one to make inferences about neuronal processes that cannot be observed directly with available noninvasive techniques, for example, sensitization of glutamatergic postsynaptic responses or changes in neuronal spike frequency adaptation.

Ongoing validation studies in rodents have demonstrated the potential of this approach (Moran et al., 2008) and highlight a more general imperative to connect micro- and macroneurobiological scales by marrying and cross-validating techniques in animal and human neuroimaging. The increasing availability of intracranial recordings from patients provides a further bridge to validate noninvasive models and test the resolution of our predictions. In other words, one can optimize the level of modeling and assess, in a quantitative fashion, whether a particular imaging modality is sufficient to answer questions framed at a particular level of neurobiological model. However, future progress here is almost certainly going to benefit from much closer interactions between imaging neuroscience and scientists pursuing a microscopic characterization of underlying circuits. This knowledge is necessary to provide empirical constraints on microcircuitry and relative connection strengths that can then be transcribed directly into current models of EEG, MEG, or fMRI data, in essence informing macroscopic characterizations of brain function with microscopic observations of brain structure.

Conclusions and a Personal Coda

The dominant position of functional neuroimaging in systems neuroscience is unlikely to change in the immediate future. This reflects not only the absence of any emerging competition to existing technologies but also the necessity for system level descriptions of cognitive processing, no matter how one conceives of how the brain works. However, an increasing discourse between imaging neuroscience and experts from disciplines such as computational and theoretical neuroscience, as well as those working at the level of microcircuitry, seems inevitable. The largely untapped potential of functional neuroimaging to provide a richer characterization of aberrant processes in psychiatric disorders is likely to be of increasing importance. The more far-fetched claims made in relation to neuorimaging, evoking an Orwellian nightmare, are unlikely to threaten our human attachment to privacy. The reality is that deterministic predictions on behavior need to contend with the fact that even simple neuronal systems, and their expression in behavior, embody indeterminacy. Indeed, this indeterminancy is introduced by the brain itself, where it is suggested that it provides for a variance that optimizes long-term survival (Carpenter, 2004).

It is now nearly 20 years since I agreed to a career in imaging neuroscience, a period that coincides with the publication history of Neuron, and I find myself caught in an ironic twist of fate. I am sitting in a waiting room reflecting on having committed, at extremely short notice, to writing this opinion piece while preparing myself for an MRI scan on an increasingly painful right shoulder. I keep a rumbling anxiety in the background by an old trick of distraction, in this case realized by aimlessly flicking through an assortment of well-fingered magazines. Almost inevitably, I stumble across a magazine article that is a feature on brain imaging, entitled "The brain hacker," with an equally attention grabbing subtitle, "they can read your mind." The opening paragraph is indeed ominous and enough to make any mortal shudder at the prospect of submitting to any procedure suffixed with MRI, even ostensibly on a relatively unsophisticated piece of shoulder anatomy. Who can relax when reading: "Machines that read your thoughts have long been the stuff of science fiction. But now brain-scanning techniques are being used to see if you prefer Coke to Pepsi. Are secrets a thing of the past?" Reading on, I learn that the application of brain imaging is all-embracing when it comes to the human condition, and it can now unravel political preferences (neuropolitics), the likely choice of goods to purchase (neuromarketing), unmask covert racism, and distinguish truth from lies and even true love from mere lust. I momentarily remind myself, "It's only your shoulder." As I ease into the engulfing claustrophobia that is the bore of an MRI magnet, I reassure myself, without undue effort, that my personal secrets are safe and that the reality of brain imaging is more prosaic than the ubiquitous hype of newspapers and magazines.

ACKNOWLEDGMENTS

R.J.D.'s research is supported by the Wellcome Trust.

REFERENCES

Baker, S.C., Rogers, R.D., Owen, A.M., Frith, C.D., Dolan, R.J., Frackowiak, R.S., and Robbins, T.W. (1996). Neural systems engaged by planning: a PET study of the Tower of London task. Neuropsychologia *34*, 515–526.

Berns, G.S., McClure, S.M., Pagnoni, G., and Montague, P.R. (2001). Predictability modulates human brain response to reward. J. Neurosci. *21*, 2793–2798.

Buchel, C., and Friston, K. (2001). Interactions among neuronal systems assesed with functional neuroimaging. Rev. Neurol. (Paris) *157*, 807–815.

Buchel, C., Morris, J., Dolan, R.J., and Friston, K.J. (1998). Brain systems mediating aversive conditioning: an event-related fMRI study. Neuron *20*, 947–957.

Camerer, C.F. (2003). Psychology and economics. Strategizing in the brain. Science *300*, 1673–1675.

Carpenter, R.H. (2004). Contrast, probability, and saccadic latency; evidence for independence of detection and decision. Curr. Biol. *14*, 1576–1580.

Corlett, P.R., Murray, G.K., Honey, G.D., Aitken, M.R., Shanks, D.R., Robbins, T.W., Bullmore, E.T., Dickinson, A., and Fletcher, P.C. (2007). Disrupted prediction-error signal in psychosis: evidence for an associative account of delusions. Brain *130*, 2387–2400.

Courtney, S.M., Petit, L., Maisog, J.M., Ungerleider, L.G., and Haxby, J.V. (1998). An area specialized for spatial working memory in human frontal cortex. Science *279*, 1347–1351.

Critchley, H.D., Mathias, C.J., and Dolan, R.J. (2001). Neural activity in the human brain relating to uncertainty and arousal during anticipation. Neuron *29*, 537–545.

Critchley, H.D., Wiens, S., Rotshtein, P., Ohman, A., and Dolan, R.J. (2004). Neural systems supporting interoceptive awareness. Nat. Neurosci. *7*, 189–195.

Damasio, A.R., Grabowski, T.J., Bechara, A., Damasio, H., Ponto, L.L., Parvizi, J., and Hichwa, R.D. (2000). Subcortical and cortical brain activity during the feeling of self-generated emotions. Nat. Neurosci. *3*, 1049–1056.

David, O., Kiebel, S.J., Harrison, L.M., Mattout, J., Kilner, J.M., and Friston, K.J. (2006). Dynamic causal modeling of evoked responses in EEG and MEG. Neuroimage *30*, 1255–1272.

Daw, N.D., O'Doherty, J.P., Dayan, P., Seymour, B., and Dolan, R.J. (2006). Cortical substrates for exploratory decisions in humans. Nature *441*, 876–879.

de Quervain, D.J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., and Fehr, E. (2004). The neural basis of altruistic punishment. Science *305*, 1254–1258.

Dehaene, S., Naccache, L., Cohen, L., Bihan, D.L., Mangin, J.F., Poline, J.B., and Riviere, D. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. Nat. Neurosci. *4*, 752–758.

Dehaene, S., Naccache, L., Le Clec, H.G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., van de Moortele, P.F., and Le Bihan, D. (1998). Imaging unconscious semantic priming. Nature *395*, 597–600.

den Ouden, H., Friston, K.J., Daw, N.D., McIntosh, A., and Stephan, K.E. (2008). A dual role for prediction error in associative learning. Cereb. Cortex, in press. Published online September 26, 2008. 10.1093/cercor/bhn161.

Dolan, R.J., Fletcher, P., Frith, C.D., Friston, K.J., Frackowiak, R.S., and Grasby, P.M. (1995). Dopaminergic modulation of impaired cognitive activation in the anterior cingulate cortex in schizophrenia. Nature *378*, 180–182.

Drabant, E.M., Hariri, A.R., Meyer-Lindenberg, A., Munoz, K.E., Mattay, V.S., Kolachana, B.S., Egan, M.F., and Weinberger, D.R. (2006). Catechol O-methyltransferase val158met genotype and neural mechanisms related to affective arousal and regulation. Arch. Gen. Psychiatry *63*, 1396–1406.

Fletcher, P.C., Frith, C.D., Grasby, P.M., Shallice, T., Frackowiak, R.S.J., and Dolan, R.J. (1995). Brain systems for encoding and retrieval of auditory-verbal memory. An in vivo study in humans. Brain *118*, 401–416.

Fletcher, P.C., Frith, C.D., Grasby, P.M., Friston, K.J., and Dolan, R.J. (1996). Local and distributed effects of apomorphine on fronto-temporal function in acute unmedicated schizophrenia. J. Neurosci. *16*, 7055–7062.

Fox, P.T., Mintun, M.A., Reiman, E.M., and Raichle, M.E. (1988). Enhanced detection of focal brain responses using intersubject averaging and change-distribution analysis of subtracted PET images. J. Cereb. Blood Flow Metab. *8*, 642–653.

Friston, K. (1994). Functional and effective connectivity in neuroimaging: a synthesis. Hum. Brain Mapp. *2*, 56–78.

Friston, K. (2005). A theory of cortical responses. Philos. Trans. R. Soc. Lond. B Biol. Sci. *360*, 815–836.

Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., and Penny, W. (2007). Variational free energy and the Laplace approximation. Neuroimage *34*, 220–234.

Friston, K.J., Frith, C.D., Liddle, P.F., and Frackowiak, R.S. (1991). Comparing functional (PET) images: the assessment of significant change. J. Cereb. Blood Flow Metab. *11*, 690–699.

Friston, K.J., Frith, C.D., Passingham, R.E., Liddle, P.F., and Frackowiak, R.S. (1992). Motor practice and neurophysiological adaptation in the cerebellum: a positron tomography study. Proc. Biol. Sci *248*, 223–228.

Friston, K.J., Price, C.J., Fletcher, P., Moore, C., Frackowiak, R.S., and Dolan, R.J. (1996). The trouble with cognitive subtraction. Neuroimage *4*, 97–104.

Friston, K.J., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. Neuroimage *19*, 1273–1302.

Frith, C., and Dolan, R. (1996). The role of the prefrontal cortex in higher cognitive functions. Brain Res. Cogn. Brain Res. *5*, 175–181.

Grafton, S.T., Mazziotta, J.C., Presty, S., Friston, K.J., Frackowiak, R.S., and Phelps, M.E. (1992). Functional anatomy of human procedural learning determined with regional cerebral blood flow and PET. J. Neurosci. *12*, 2542–2548.

Grasby, P.M., Frith, C.D., Friston, K.J., Bench, C., Frackowiak, R.S., and Dolan, R.J. (1993). Functional mapping of brain areas implicated in auditory–verbal memory function. Brain *116*, 1–20.

Grill-Spector, K., and Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. Acta Psychol. (Amst.) *107*, 293–321.

Hariri, A.R., Mattay, V.S., Tessitore, A., Kolachana, B., Fera, F., Goldman, D., Egan, M.F., and Weinberger, D.R. (2002). Serotonin transporter genetic variation and the response of the human amygdala. Science *297*, 400–403.

Haxby, J.V., Grady, C.L., Horowitz, B., Ungerleider, L.G., Mishkin, M., Carson, R.E., Herscowitch, P., Schapiro, M.B., and Rappaport, S. (1991). Dissociation of object and spatial visual processing pathways in the human extrastriate cortex. J. Neurosci. *88*, 1621–1625.

Haxby, J.V., Ungerleider, L.G., Horwitz, B., Maisog, J.M., Rapoport, S.I., and Grady, C.L. (1996). Face encoding and recognition in the human brain. Proc. Natl. Acad. Sci. USA *93*, 922–927.

Haynes, J.D., and Rees, G. (2006). Decoding mental states from brain activity in humans. Nat. Rev. Neurosci. *7*, 523–534.

Honey, G.D., Corlett, P.R., Absalom, A.R., Lee, M., Pomarol-Clotet, E., Murray, G.K., McKenna, P.J., Bullmore, E.T., Menon, D.K., and Fletcher, P.C. (2008). Individual differences in psychotic effects of ketamine are predicted by brain function measured under placebo. J. Neurosci. *28*, 6295–6303.

Howard, D., Patterson, K., Wise, R., Brown, W.D., Friston, K., Weiller, C., and Frackowiak, R. (1992). The cortical localization of the lexicons. Positron emission tomography evidence. Brain *115*, 1769–1782.

Hsu, M., Anen, C., and Quartz, S.R. (2008). The right and the good: distributive justice and neural encoding of equity and efficiency. Science *320*, 1092–1095.

Hunter, R.A., and MacAlpine, I. (1973). Psychiatry for the Poor: 1851 Colney Hatch Asylum—Friern Hospital 1973: A Medical and Social History (Folkestone, England: Dawsons of Pall Mall).

Jonides, J., Smith, E.E., Koeppe, R.A., Awh, E., Minoshima, S., and Mintun, M.A. (1993). Spatial working memory in humans as revealed by PET. Nature *363*, 623–625.

Kable, J.W., and Glimcher, P.W. (2007). The neural correlates of subjective value during intertemporal choice. Nat. Neurosci. *10*, 1625–1633.

Kamitani, Y., and Tong, F. (2005). Decoding the visual and subjective contents of the human brain. Nat. Neurosci. *8*, 679–685.

Karni, A., Meyer, G., Rey-Hipolito, C., Jezzard, P., Adams, M.M., Turner, R., and Ungerleider, L.G. (1998). The acquisition of skilled motor performance: fast and slow experience-driven changes in primary motor cortex. Proc. Natl. Acad. Sci. USA *95*, 861–868.

Kastner, S., De Weerd, P., Desimone, R., and Ungerleider, L.G. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. Science *282*, 108–111.

Kay, K.N., Naselaris, T., Prenger, R.J., and Gallant, J.L. (2008). Identifying natural images from human brain activity. Nature *452*, 352–355.

Kensinger, E.A., and Schacter, D.L. (2006). Amygdala activity is associated with the successful encoding of item, but not source, information for positive and negative stimuli. J. Neurosci. *26*, 2564–2570.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C.F., Quartz, S.R., and Montague, P.R. (2005). Getting to know you: reputation and trust in a two-person economic exchange. Science *308*, 78–83.

Kloppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., Fox, N.C., Jack, C.R., Jr., Ashburner, J., and Frackowiak, R.S. (2008). Automatic classification of MR scans in Alzheimer's disease. Brain *131*, 681–689.

LaBar, K.S., Gatenby, J.C., Gore, J.C., LeDoux, J.E., and Phelps, E.A. (1998). Human amygdala activation during conditioned fear acquisition and extinction: a mixed-trial fMRI study. Neuron *20*, 937–945.

Lueck, C.J., Zeki, S., Friston, K.J., Deiber, M.P., Cope, P., Cunningham, V.J., Lammertsma, A.A., Kennard, C., and Frackowiak, R.S. (1989). The colour centre in the cerebral cortex of man. Nature *340*, 386–389.

Lumer, E.D., and Rees, G. (1999). Covariation of activity in visual and prefrontal cortex associated with subjective visual perception. Proc. Natl. Acad. Sci. USA *96*, 1669–1673.

Lumer, E.D., Friston, K.J., and Rees, G. (1998). Neural correlates of perceptual rivalry in the human brain. Science *280*, 1930–1934.

Marr, D. (1982). Vision (San Francisco, CA: WH Freeman).

McClure, S.M., Berns, G.S., and Montague, P.R. (2003). Temporal prediction errors in a passive learning task activate human striatum. Neuron *38*, 339–346.

McClure, S.M., Laibson, D.I., Loewenstein, G., and Cohen, J.D. (2004). Separate neural systems value immediate and delayed monetary rewards. Science *306*, 503–507.

McIntosh, A.R., and Gonzalez-Lima, F. (1994). Structural equation modeling and its application to network analysis in functional brain imaging. Hum. Brain Mapp. *2*, 2–22.

Moran, R.J., Stephan, K.E., Kiebel, S.J., Rombach, N., O'Connor, W.T., Murphy, K.J., Reilly, R.B., and Friston, K.J. (2008). Bayesian estimation of synaptic physiology from the spectral responses of neural masses. Neuroimage *42*, 272–284.

Morris, J.S., Frith, C.D., Perrett, D.I., Rowland, D., Young, A.W., Calder, A.J., and Dolan, R.J. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. Nature *383*, 812–815.

Murray, G.K., Corlett, P.R., Clark, L., Pessiglione, M., Blackwell, A.D., Honey, G., Jones, P.B., Bullmore, E.T., Robbins, T.W., and Fletcher, P.C. (2008). Substantia nigra/ventral tegmental reward prediction error disruption in psychosis. Mol. Psychiatry *13*, 267–276.

Noesselt, T., Driver, J., Heinze, H.J., and Dolan, R. (2005). Asymmetrical activation in the human brain during processing of fearful faces. Curr. Biol. *15*, 424–429.

O'Craven, K.M., Rosen, B.R., Kwong, K.K., Treisman, A., and Savoy, R.L. (1997). Voluntary attention modulates fMRI activity in human MT-MST. Neuron *18*, 591–598.

O'Doherty, J.P., Dayan, P., Friston, K., Critchley, H., and Dolan, R.J. (2003). Temporal difference models and reward-related learning in the human brain. Neuron *38*, 329–337.

Ogawa, S., Lee, T.M., Kay, A.R., and Tank, D.W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc. Natl. Acad. Sci. USA *87*, 9868–9872.

Owen, A.M. (2008). Disorders of consciousness. Ann. N Y Acad. Sci. *1124*, 225–238.

Owen, A.M., Doyon, J., Petrides, M., and Evans, A.C. (1996a). Planning and spatial working memory: a positron emission tomography study in humans. Eur. J. Neurosci. *8*, 353–364.

Owen, A.M., Milner, B., Petrides, M., and Evans, A.C. (1996b). Memory for object features versus memory for object location: A positron emission tomography study of encoding and retrieval processes. Proc. Natl. Acad. Sci. USA *93*, 9212–9217.

Penny, W.D., Stephan, K.E., Mechelli, A., and Friston, K.J. (2004). Modelling functional integration: a comparison of structural equation and dynamic causal models. Neuroimage *23* (*Suppl 1*), S264–S274.

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R.J., and Frith, C.D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature *442*, 1042–1045.

Petersen, S.E., Fox, P.T., Snyder, A.Z., and Raichle, M.E. (1990). Activation of extrastriate and frontal cortical areas by visual words and word-like stimuli. Science *249*, 1041–1044.

Petersen, S.E., Fox, P.T., Posner, M.I., Mintun, M., and Raichle, M.E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. Nature *331*, 585–589.

Petrides, M., Alivisatos, B., and Evans, A.C. (1995). Functional activation of the human ventrolateral frontal cortex during mnemonic retrieval of verbal information. Proc. Natl. Acad. Sci. USA *92*, 5803–5807.

Plailly, J., Howard, J.D., Gitelman, D.R., and Gottfried, J.A. (2008). Attention to odor modulates thalamocortical connectivity in the human brain. J. Neurosci. *28*, 5257–5267.

Posner, M.I., Petersen, S.E., Fox, P.T., and Raichle, M.E. (1988). Localization of cognitive operations in the human brain. Science *240*, 1627–1631.

Price, C., Wise, R., Ramsay, S., Friston, K., Howard, D., Patterson, K., and Frackowiak, R. (1992). Regional response differences within the human auditory cortex when listening to words. Neurosci. Lett. *146*, 179–182.

Price, C.J., Moore, C.J., Humphreys, G.W., and Wise, R.J.S. (1997). Segregating semantic from phonological processes during reading. J. Cogn. Neurosci. *9*, 727–733.

Schacter, D.L., and Buckner, R.L. (1998). Priming and the brain. Neuron *20*, 185–195.

Shallice, T., Fletcher, P., Frith, C.D., Grasby, P., Frackowiak, R.S.J., and Dolan, R.J. (1994). Brain regions associated with acquisition and retrieval of verbal episodic memory. Nature *368*, 633–635.

Sharot, T., Riccardi, A.M., Raio, C.M., and Phelps, E.A. (2007). Neural mechanisms mediating optimism bias. Nature *450*, 102–105.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R.J., and Frith, C.D. (2004). Empathy for pain involves the affective but not sensory components of pain. Science *303*, 1157–1162.

Stephan, K.E., Kasper, L., Harrison, L.M., Daunizeau, J., den Ouden, H.E., Breakspear, M., and Friston, K.J. (2008). Nonlinear dynamic causal models for fMRI. Neuroimage *42*, 649–662.

Tootell, R.B., Hadjikhani, N.K., Vanduffel, W., Liu, A.K., Mendola, J.D., Sereno, M.I., and Dale, A.M. (1998). Functional analysis of primary visual cortex (V1) in humans. Proc. Natl. Acad. Sci. USA *95*, 811–817.

Tulving, E., Kapur, S., Craik, F.I., Moscovitch, M., and Houle, S. (1994a). Hemispheric encoding/retrieval asymmetry in episodic memory: positron emission tomography findings. Proc. Natl. Acad. Sci. USA *91*, 2016–2020.

Tulving, E., Kapur, S., Markovitsch, H.J., Craik, F.I.M., Habib, R., and Houle, S. (1994b). Neuroanatomical correlates of retrieval in episodic memory: auditory sentence recognition. Proc. Natl. Acad. Sci. USA *91*, 2012–2015.

Vuilleumier, P., Richardson, M.P., Armony, J.L., Driver, J., and Dolan, R.J. (2004). Distant influences of amygdala lesion on visual cortical activation during emotional face processing. Nat. Neurosci. *7*, 1271–1278.

Wagner, A.D., Schacter, D.L., Rotte, M., Koutstaal, W., Maril, A., Dale, A.M., Rosen, B.R., and Buckner, R.L. (1998). Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. Science *281*, 1188–1191.

Ward, N.S., Brown, M.M., Thompson, A.J., and Frackowiak, R.S. (2004). The influence of time after stroke on brain activations during a motor task. Ann. Neurol. *55*, 829–834.

Ward, N.S., Newton, J.M., Swayne, O.B., Lee, L., Thompson, A.J., Greenwood, R.J., Rothwell, J.C., and Frackowiak, R.S. (2006). Motor system activation after subcortical stroke depends on corticospinal system integrity. Brain *129*, 809–819.

Whalen, P.J., Rauch, S.L., Etcoff, N.L., McInerney, S.C., Lee, M.B., and Jenike, M.A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. J. Neurosci. *18*, 411–418.

Winston, J.S., Henson, R.N., Fine-Goulden, M.R., and Dolan, R.J. (2004). fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. J. Neurophysiol. *92*, 1830–1839.

Winston, J.S., O'Doherty, J., Kilner, J.M., Perrett, D.I., and Dolan, R.J. (2007). Brain systems for assessing facial attractiveness. Neuropsychologia *45*, 195–206.

Zeki, S., Watson, J., Lueck, J., Friston, K.J., Kennard, C., and Frackowiak, R.S.J. (1991). A Direct Demonstration of Functional Specialization in Human Visual Cortex. J. Neurosci. *11*, 641–649.