Tamagawa-Caltech Joint Workshop

“Neural Mechanisms of The Social Mind”

Tamagawa University
Machida, Tokyo, Japan

December 6 – 8, 2007

Tamagawa-Caltech Joint Workshop "Neural Mechanisms of The Social Mind"

Tamagawa University
Machida,Tokyo,Japan
December 6 — 8, 2007

Organizers: **Ralph Adolphs** (Caltech)
John O'Doherty (Caltech)
Masamichi Sakagami (Tamagawa University)
Shinsuke Shimojo (Caltech)
Jun Tanji (Tamagawa University)

Thursday, December 6th

9:00-9:10 Opening Remark
Masamichi Sakagami (Tamagawa University)

——— Chairman: **Shinsuke Shimojo** (Caltech)

9:10-10:00 **Masami Sasaki, Tetsuhiko Sasaki** (Tamagawa University)
How does the learning capability develop in social system of the honeybee?

10:00-10:50 **Eiji Hoshi, Jun Tanji** (Tamagawa University)
Participation of the premotor cortex in controlling motor behavior

10:50-11:10 Coffee Break

——— Chairman: **John O'Doherty** (Caltech)

11:10-12:00 **Masamichi Sakagami, Xiaochuan Pan, Kensaku nomoto, Manami Yamamoto, Jiro Okuda, Kazuyuki Samejima** (Tamagawa University)
Multiple brain circuits for reward prediction

12:00-12:50 **Bernard Balleine** (UCLA)
Reward, reward prediction and choice in corticostriatal networks

12:50-14:10 Lunch

——— Chairman: **Masamichi Sakagami** (Tamagawa University)

14:10-14:50 **Shinsuke Shimojo** (Caltech)
Mind is social – from sensation to memory, and to decision

14:50-15:30 **Hackjin Kim** (Korean University)
Temporal isolation of neural processes underlying face preference decisions

15:30-16:10 **John O'Doherty** (Caltech)
Functional neuroimaging of decision making: From reward to social interactions

16:10-16:30 Coffee break

——————————————— Chairman: **Jiro Okuda** (Tamagawa University)

16:30-16:45 **Klaus Wunderlich** (Caltech):
Dissociating action and choice-values in the human brain

16:45-17:00 **Signe Bray** (Caltech):
How Pavlovian reward cues modulate instrumental choice: an fMRI study

17:00-17:35 **Naotsugu Tsuchiya, Ralph Adolphs** (Caltech)
Neuronal dynamics of face perception: decoding analysis of intracranial EEG recording in human epileptic patients

17:35-18:10 **Michael Campos**(Caltech)
MonkeyTV: Orbitofrontal responses during self-initiated video-watching

18:10-19:00 **Mitsuo Kawato, Masahiko Haruno** (ATR)
Internal models and heterarchical reinforcement learning for social interaction

19:00-20:30 Welcome party

Friday, December 7th

——————————— Chairman: **Kazuyuki Samejima** (Tamagawa University)

9:00-9:50 **Katsuhiko Miyazaki, Kayoko Miyazaki, Kenji Doya** (OIST)
Roles of dopamine and serotonin in actions for delayed rewards

9:50-10:40 **Tatsuyoshi Saijo** (Osaka University)
Are Japanese spiteful?

10:40-11:00 Coffee Break

——————————— Chairman: **Kenji Matsumoto** (Tamagawa University)

11:00-11:45 **Colin Camerer** (Caltech)
Bounded rationality and bounded deception in games

11:45-12:10 **Todd Hare** (Caltech)
Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors

12:10-12:25 **Dirk Neumann** (Caltech)
Connecting the brain (Diffusion based tractography and functional connectivity).

12:25-12:40 **Ian Krajbich** (Caltech)
Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex

12:40-14:10 Lunch

——————————— Chairman: **Naotsugu Tsuchiya** (Caltech)

14:10-15:00 **Jim Woodward** (Caltech)
Moral intuition: Neural substrates and normative significance

15:00-15:50 **Henrik Walter** (University of Bonn)
Intention, volition and the normative: From metaphysics to cognitive neuroscience

15:50-16:10 Coffee break

——————————— Chairman: **Ben Seymour** (UCL)

16:10-17:00 **Uta Noppeney** (Max Planck)
Audio-visual interactions within the cortical hierarchy

17:00-17:50 **Turhan Canli** (SUNY)
Neurogenetics of personality

17:50-18:40 **Lauren Stewart** (Goldsmiths)
When all the songs sound the same: Insights into the musical brain

19:00-20:30 Reception

Saturday, December 8th

Chairman: **Hackjin Kim** (Korea University)

9:00-9:50 **Ben Seymour** (UCL)
From aggression to altruism: Neurobiological insights into the nature of punishment

9:50-10:40 **Elizabeth Phelps** (NYU)
The social acquisition and alteration of learned fears

10:40-11:00 Coffee Break

Chairman: **Kenji Doya** (OIST)

11:00-11:50 **Mitsuru Kawamura** (Showa University)
Social cognitive impairments in Parkinson's Disease

11:50-12:40 **Nathaniel Daw** (NYU)
Semi-rational models of learned decision making

12:40-12:50 Closing Remark
Jun Tanji (Tamagawa University)

How does the learning capability develop in social system of the honeybee?

Masami Sasaki and Tetsuhiko Sasaki

Tamagawa University

The honeybee is a social insect and its colony consists of a queen bee and several thousands of workers. The workers are engaged in all the labor, except reproduction, that is necessary for maintenance of the colony. They possess a highly-developed brain and show a variety of complex social behaviors. For example, foraging bees can remember the location of the flowers and communicate the information about the foraging place to nest mates by means of a special dance, essentially an abstract language specific to the honeybees. The honeybee brain, composed of 950,000 well-organized neural cells, shows impressive cognitive abilities, including such human abilities as the ability to associate colors and shapes of flowers with food and the ability to acquire and understand concepts such as “same” and “different.”

We examined the influence of social experience on the honeybee’s ontogenic development and brain structure to increase our understanding of the honeybee’s learning and memory abilities. We used an associative learning paradigm and proboscis extension reflex, with menthone as a conditioned stimulus (CS) and 1.5 M sucrose as an unconditioned stimulus. The results showed that 90% of bees developed learning ability by the 9th day after emergence. However, when the bee was isolated from the hive by confining it in a small glass vial, learning was slowed down and smaller percentages of bees acquired the associations as the training progressed, suggesting that some social stimuli in colony is important for functional development of the honeybee brain (by Masami Sasaki).

In an attempt to clarify the function of a neurotransmitter, dopamine, in the social life of honeybee, we examined the expression level of the dopamine transporter (DAT), which is a protein on the presynaptic membrane of dopaminergic terminal. Because DAT is specifically expressed in dopaminergic neurons and recovers the released dopamine from the synaptic cleft, its expression in the brain is expected to reflect the activity of dopaminergic neurons. The levels of the dat transcript in workers, queens and males were determined by reverse transcription real-time quantitative polymerase chain reaction (RT-qPCR). The dat expression level was significantly higher in foraging workers than in workers working in the hive. In virgin queens and males that were ready for their mating flight, dat expression levels were as high as in the foraging workers. Mated queens, which stay in the hive and lay eggs, on the other hand, showed

lower dat expression than the virgin queens. Our data show that the dat transcript level is well correlated with the behavioral activity of honeybees (by Tetsuhiko Sasaki).

Participation of the premotor cortex in controlling motor behavior

Eiji Hoshi and Jun Tanji

Tamagawa University Brain Science Institute

The premotor cortex (PM) occupies the lateral surface of Brodman's area 6 in the frontal cortex. The PM has traditionally been viewed as participating in preparing and executing motor acts, or taking a part in sensorimotor transformation (or mapping). However, two studies that we conducted recently revealed that the PM is involved in much broader aspects of behavioral control. The first study examined the involvement of the PM in integrating different sets of information before the behavioral stage of motor preparation. We recorded neuronal activity from the PM while monkeys (*Macaca fuscata*) were performing a behavioral task. We found that neurons in the PM initially retrieved information on target location and arm use and, subsequently, integrated a pair of information. This result suggested that the PM plays an important role in retrieving and integrating crucial information that constitutes components of action, before motor planning. The second study examined the involvement of the PM in a process of transforming a concept for action into an actual motor act. Here, we define a concept for action as a motor behavior formulated at a conceptual level without specifying an actual movement. We recorded neuronal activity from the PM while monkeys were performing a behavioral task specifically designed to sort out the behavioral stage for concept specification and motor planning. We found that neuronal activity in the PM initially reflected an action concept, which was subsequently taken over by activity reflecting the direction of a planned movement. This result suggested that the PM is involved in the process of transforming a conceptual representation of action into a representation signifying actual movement being planned. Taken together, our recent findings indicate that the PM is not merely involved in motor planning or sensorimotor association. We propose that the PM is involved in behavioral stages that are required well in advance of motor planning. An intriguing possibility is its involvement in specifying a motor concept in the context of communication. It is possible that the PM could take part in transforming an action concept received with a verbal command into a motor plan that enables actual accomplishment of motor behavior.

Multiple brain circuits for reward prediction

Masamichi Sakagami, Xiaochuan Pan, Kensaku Nomoto,
Manami Yamamoto, Jiro Okuda, Kazuyuki Samejima
Tamagawa University Brain Science Institute

When we go to a grocery store to buy ingredients for dinner, we collect vegetables, meat, and spices necessary for a planned cuisine, say beef steak. During the behavior, we activate an internal model on the dish. However, we sometimes encounter unexpected excellent stuff, for example, plump mussels, which may activate our wonderful memory of delicious bouillabaisse in Marseille, formed through direct experience. Then we have to change our plan from beef steak to bouillabaisse with white wine, which leads to the activation of other internal model to follow.

What happens in our brain during the shopping? Two brain processes involved in reward prediction, at least, can guide our behavior. One is the process activating an internal model on the dish, which includes the systematic information on materials and methods (e.g. categorization and inference). The other is to code a specific reward value of a stimulus or event, dependent on direct experiences (e.g. classical conditioning and TD learning). According to the theoretical prediction by Daw, Niv & Dayan (2005), the former depends on the prefrontal cortex and the latter is the work by the striatum. However, we have not had direct evidence to support the prediction. We have executed three experiments by means of monkey single unit recording and human imaging with fMRI.

1) Simultaneous single unit recording from the monkey caudate and lateral prefrontal cortex in the reward inference task.
2) Human fMRI imaging during random-dot discrimination with asymmetric reward condition.
3) Single unit recording from the monkey dopamine neuron in the random-dot discrimination task with asymmetric reward schedule.

Results suggest that the prefrontal network contributes to the model-based reward prediction and the nigro-striatal network works for the model-free reward prediction. Additionally, the prefrontal network might have a compensatory function for decision making when circumstances are ambiguous.

Reward, reward prediction and choice in corticostriatal networks

Bernard Balleine
UCLA

Predictions about the occurrence of rewarding events can be based on environmental stimuli or the actions with which those events are associated. Although these two sources of predictive learning have long been thought to have much in common, recent evidence has emerged to counter these claims. In this talk I will illustrate these points by describing two theories of predictive learning, and current evidence for the involvement of an amygdala, midbrain and cortical network in establishing the relative validity of reward predictions. I will then contrast this form of learning with that involved in the acquisition of goal-directed action, as opposed to conditioned reflexes, and describe how predictions of goal value based on actions and based on stimuli differ. Recent research has significantly advanced our understanding of the behavioral and neural bases of goal-directed action. Although there is a literature linking the control of executive functions to the prefrontal cortex, more recent studies suggest that these functions depend on reward-related circuitry linking prefrontal, premotor and sensorimotor cortices with the striatum. Evidence from a range of species suggests that discrete cortico-striatal networks control functionally distinct decision processes involving (a) actions that are more flexible or goal-directed, sensitive to reward-related feedback and that involve regions of association cortices particularly medial, orbitomedial, premotor and anterior cingulate cortices together with their efferent targets in caudate/dorsomedial striatum; and (b) actions that are relatively automatic or habitual and involving sensorimotor cortices and dorsolateral striatum/putamen. These processes have been argued to depend on different learning rules and, correspondingly, different forms of plasticity. Furthermore, degeneration of these cortico-striatal circuits has been argued to result in the distinct forms of psychopathology such as that associated with Huntington's disease, obsessive compulsive disorder and Tourette's syndrome on the one hand and Parkinson's and multiple system atrophy on the other. Finally, distinct motivational processes appear to modulate these decision networks. We have evidence that opiate activity in the amgdala subserves the influence of reward on the performance of goal-directed actions whereas other studies suggests that midbrain dopaminergic control of striatal activity influences the execution of habits.

Mind is social – from sensation to memory, and to decision

Shinsuke Shimojo

Division of Biology /Computation and Neural Systems, California Institute of Technology

JST ERATO Shimojo Implicit Brain Function Project

I will describe two independent studies from our laboratory. A meta-analysis concerning evolution of the human trichromatic color perception, and a psychophysical study to investigate effects of memory on preference decision.

The first study aimed to answer an evolutionary question as to how the human color system evolved to be trichromatic with R-G and Y-B as the cardinal axes (Changizi et al., *Biol. Letters*, '06). The prevailing theory is that it is optimized to detect color changes of foliage or fruit, but it has several problems including that most of species eating leaves or fruits are actually not trichromats (at least not the regular human type). Instead, we argued that perhaps the human color system has been further modified to optimally detect subtle changes in skin color modulated by emotion (i.e. anger, shame, sadness, fear, etc.) . The peak of L and M cones turned out to be nearly optimal (within several nm offset) to detect such subtle changes caused by oxidization and density of hemoglobin. Also, evolution of facial bare skin showed a suspicious coincidence with this regard.

The second study was motivated by the paradox in the literature: although memory obviously affects attractiveness and preference, how exactly has been controversial with two seemingly-opposite principles proposed, Novelty and Familiarity. We examined three different categories of objects; faces, natural scenes, and geometric figures in a two-alternative forced-choice preference task between an old and a new stimuli (Shimojo et al., *VSS*, '07). The results indicate a surprising segregation. Familiarity preference became stronger over trials with faces, whereas novelty preference became stronger with natural scenes. No such strong tendency in geometric figures. The results partly resolve the paradox, while raising questions concerning what causes such a segregation.

Combining the two studies together, and inspired by the infant perception/cognition literature, I will conclude that the human mind is social, from sensation to memory, and to decision.

Temporal isolation of neural processes underlying face preference decisions

Hackjin Kim
Korea University

Decisions about whether we like someone are often made so rapidly from first impressions that it is difficult to examine the engagement of neural structures at specific points in time. Here, we used a temporally extended decision-making paradigm to examine brain activation with functional MRI (fMRI) at sequential stages of the decision-making process. Activity in reward-related brain structures—the nucleus accumbens (NAC) and orbitofrontal cortex (OFC)—was found to occur at temporally dissociable phases while subjects decided which of two unfamiliar faces they preferred. Increases in activation in the OFC occurred late in the trial, consistent with a role for this area in computing the decision of which face to choose. Signal increases in the NAC occurred early in the trial, consistent with a role for this area in initial preference formation. Moreover, early signal increases in the NAC also occurred while subjects performed a control task (judging face roundness) when these data were analyzed on the basis of which of those faces were subsequently chosen as preferred in a later task. The findings support a model in which rapid, automatic engagement of the NAC conveys a preference signal to the OFC, which in turn is used to guide choice.

Functional neuroimaging of decision making: From reward to social interactions.

John P O'Doherty
California Institute of Technology

In model-based functional magnetic resonance imaging (fMRI), signals derived from a computational model for a specific cognitive process are correlated against fMRI data from subjects performing a relevant task to determine brain regions showing a response profile consistent with that model. A key advantage of this technique over more conventional neuroimaging approaches is that model-based fMRI can provide insights into how a particular cognitive process is implemented in a specific brain area as opposed to merely identifying where a particular process is located. In this talk I will briefly summarize the approach of model-based fMRI, with reference to the field of reward learning and decision making, where computational models have been used to probe the neural mechanisms underlying learning of reward associations, modifying action choice to obtain reward, as well as in encoding expected value signals that reflect the abstract structure of a decision problem. Finally, I will show how model-based fMRI can be extended into the social domain in order to characterize the possible computations underlying mentalizing or theory of mind during strategic interactions.

Dissociating action and choice-values in the human brain

Klaus Wunderlich
Caltech

How Pavlovian reward cues modulate instrumental choice: an fMRI study

Signe Bray
Caltech

There is considerable evidence that a cue established as predictive of a specific outcome will bias choice towards actions associated with that outcome. Formally referred to as specific Pavlovian to instrumental transfer (PIT), this effect has been argued to model the influence of reward-related cues in advertising and drug addiction. To assess the neural bases of this influence we scanned 23 healthy subjects with functional magnetic resonance imaging (fMRI) as they underwent Pavlovian and instrumental training, followed by a transfer test. During the training phase, subjects learned to associate simple visual shape stimuli with one of four outcomes: orange juice, chocolate milk, cola and an affectively neutral tasteless control solution. The training session also included instrumental training trials in which the subjects chose from a pair of four possible button push actions that earned distinct outcomes: two of the button push actions led to reward outcomes and two led to the neutral outcome. Next, specific transfer was assessed, in extinction, by presenting the Pavlovian cues and assessing the choice between pairs of actions. We found a significant specific transfer effect: when subjects chose between actions, they favored the action corresponding to the outcome predicted by the concurrently presented Pavlovian cue. Neuroimaging results showed a significant difference in BOLD responses in ventolateral putamen on trials when subjects chose the action compatible with the Pavlovian cue compared to the incompatible action. These results provide important insight into the neural processes that mediate the influence of stimulus-outcome associations on decision making in humans.

Neuronal dynamics of face perception: decoding analysis of intracranial EEG recording in human epileptic patients

Naotsugu Tsuchiya and Ralph Adolphs
Caltech

How do regions of higher-order visual cortex represent information about emotions in facial expressions? This question has received considerable interest from fMRI, lesion, and electrophysiological studies. The most influential model of face processing argues that static aspects of a face, such as its identity, are encoded primarily in ventral temporal regions while dynamic information, such as emotional expression, depends on lateral and superior temporal sulcus and gyrus. However, supporting evidence comes mainly from clinical observation and fMRI, both of which lack temporal resolution for information flow. Recently, an alternative theory has been proposed which suggests that common initial processing for both aspects occurs in the ventral temporal cortex. To test these competing hypotheses, we studied electrophysiological responses in 9 awake human patients undergoing epilepsy monitoring, in whom over 120 sub-dural electrode contacts were implanted in ventral temporal (including fusiform face area, FFA) and lateral temporal (including superior temporal sulcus, STS) cortex. The patients viewed static and dynamic facial expressions of emotion while they performed either a gender discrimination or an emotion discrimination task.

We used a novel decoding method that quantified the information about the facial stimulus that is available from the time-varying neuronal oscillation in the field potential. We estimated the stimulus-induced oscillation from a time-frequency spectral analysis using a multi-taper method. This time-frequency representation of the response was then subjected to a multivariate decoding analysis.

Our analysis revealed that ventral temporal cortex rapidly categorizes faces from non-face objects within 100ms. We found that ventral temporal cortex represents emotion in dynamic morphing faces more quickly and accurately than lateral temporal cortex. Finally we found that the quality of represented information in ventral temporal cortex is substantially modulated by task-relevant attention.

MonkeyTV: Orbitofrontal responses during self-initiated video-watching

Michael Campos, Kari Koppitch, Richard A. Andersen and Shinsuke Shimojo

Biology/CNS, Caltech

Vision can be inherently rewarding. The reward circuitry in the brain supports an animal in identifying and obtaining rewards from its environment. The orbitofrontal cortex (OFC) is known to encode the subjective value of different juice reward options, and therefore supports decisions based on preferences in the context of appetitive rewards. It is unclear, however, whether the brain circuitry supporting the appetitive rewards is the same, distinct, or overlapping with that supporting non-appetitive rewards, which are important to modern human life. To investigate this issue we used a self-initiated free-choice paradigm in which a monkey pressed buttons to receive either the presentation of a 5 sec video clip in the video-watching period ("leisure"), or a drop of juice in a separate period ("work"). The leisure and work periods were run in separate blocks of 20 minutes each, while we simultaneously recorded 2-10 single OFC neurons. Neural activity was analyzed with respect to the button press. We first identified significant modulations in firing rate activity in any of five intervals defined with respect to the button press when compared to baseline. We found that two-thirds of the OFC neurons we encountered (394/585) were modulated in at least one interval in either the leisure or work period. Of these, approximately 40% were modulated in both periods, 40% were modulated in the work period exclusively, and 20% in the leisure period exclusively. The neurons that responded in only one period suggest that OFC contains at least two internal representations of distinct reward categories. The neurons that participated in both periods suggest that OFC also represents abstract commonalities between rewards of different kinds. These results are consistent with our intuition that perceptual experience itself is rewarding, and indicate that the neural correlates overlap with that for appetitive rewards.

Funding: JST.ERATO Shimojo Implicit Brain Functions Project, National Eye Institute to RAA.

Internal models and heterarchical reinforcement learning for social interaction

Mitsuo Kawato and Masahiko Haruno

ATR Computational Neuroscience Laboratories

Internal models are neural mechanisms that mimic input and output transfer characteristics of some dynamical systems, which reside in the external world, sensory and motor apparatus, or in some brain areas outside the neural systems that contain internal models. Dynamical systems that are to be modeled could be forward and inverse dynamics of motor apparatus, tools, and other people's brains in the context of communication. Neurophysiological and neuroimaging data suggest that at least some portions of internal models of motor apparatus (Shidara et al. 1993, Yamamoto et al. 2007), tools (Imamizu et al. 2000), speech related transformations (Callan et al. 2007) are located in the cerebellum with some topography, i.e. multiple models for different objects.

Wolpert, Doya and Kawato (2003) pointed out essential similarity in computational difficulty for sensory motor control and communication, and proposed a MOAIC framework with multiple internal models for communication. Internal models are also essential for efficient reinforcement learning algorithms with hierarchy (Kawato and Samejima, 2007). Plain reinforcement learning algorithms are too slow to be considered as realistic models of brain learning. Successful robotics demonstrations of reinforcement learning for any real world problems utilize both internal models and hierarchy. Haruno and Kawato (2006) extended the notion to heterarchical reinforcement learning with the emphasis on spiral neural connections between the striatum and substantia nigra and with some supporting neuroimaging data.

Finally, recent neuroimaging data on prisoner's dilemma (Haruno and Kawato, in preparation) suggest that reinforcement learning algorithm is an important theoretical framework also for social interactions, and the superior temporal sulcus could be related to modeling other agents' behaviors.

Roles of dopamine and serotonin in actions for delayed rewards

Katsuhiko Miyazaki, Kayoko Miyzaki and Kenji Doya

Neural Computation Unit, Okinawa Institute of Science and Technology

The deficits in the neurotransmitter serotonin is associated with disorders like depression and impulsivity. In order to understand the role of serotonin in goal-directed behaviors, we performed both chemical and electric recordings from the dorsal raphe nucleus, the major source of ascending serotonergic projection. Rats were trained to alternately visit food and water reward sites and delays were gradually introduced between the entry to the reward sites and the delivery of the rewards. In microdialysis experiments, we measured serotonin and dopamine efflux in the dorsal raphe nucleus at 5 minute time resolution. In the delayed reward condition, in which rats had to wait up to four seconds at the reward sites, serotonin efflux increased significantly compared to that in the immediate reward condition. In the intermitted reward condition, in which reward was given only once in three site visits, serotonin efflux did not change significantly while dopamine efflux decreased significantly. In single neuron recording experiments, we found that the serotonin neuron firing was significantly elevated during the waiting period before reward delivery. In the reward omission condition, the sustained serotonin neuron firing ceased before the rat gave up waiting and walked away from the reward site. These results suggest that the higher activity of serotonin enables the animal to wait for the delayed delivery of rewards.

Are Japanese spiteful?

Tatsuyoshi Saijo

Research Institute for Sustainable Science and Institute of Social and Economic Research, Osaka University

The talk is about voluntary public good provision in the laboratory, in a cross-cultural experiment conducted in the United States, China and Japan. The environments include both forced participation where subjects must participate in the voluntary contribution mechanism and voluntary participation where subjects first decide whether or not to participate in providing this non-excludable public good. In the forced participation case, Chinese subjects are relatively consistent with economic theory prediction, but we find that Japanese subjects are relatively spiteful. In the voluntary participation case, the participation decision is conveyed to the other subject prior to the subjects' contribution decisions. We find that only the American data are consistent with the evolutionary stable strategy Nash equilibrium predictions, and that behavior is significantly different across countries. Japanese subjects are more likely to act spitefully in the early periods of the experiment, even though our design changes subject pairings each period so that no two subjects ever interact twice. Surprisingly, this spiteful behavior eventually leads to more efficient public good contributions for Japanese subjects than for American subjects. Using Japanese subjects, f-MRI study is also conducted from the viewpoint of subjects who receive spiteful and altruistic actions.

Bounded rationality and bounded deception in games

Colin Camerer
Caltech

Equilibrium analysis in game theory assumes players correctly guess what others do and have the capacity to use "cheaptalk" to deceive others about private information. I discuss some theory and evidence of more general theories in which players only do a limited number of steps of strategic thinking, and find deception costly (which creates pupil dilation and distinctive neural activity). In the cognitive hierarchy theory, 0-step players randomize and higher k-step players choose best responses to players from steps 0 to k-1. This theory is illustrated with application to a Swedish lottery game (LUPI) and to reaction of moviegoers to a lack of information about movies. Cheaptalk is illustrated with a sender-receiver game and a "yard-sale bargaining" game in which players bargain over an object of unknown value.

Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors

Todd Hare, John O'Doherty, Colin Camerer, Wolfram Schultz and Antonio Rangel

Caltech

In order to make sound economic decisions, the brain needs to compute several different value related signals. These include goal values and decision values that are used to guide decisions to those actions with the largest net benefit, as well as prediction errors that are used to learn the value of actions. Goal values measure the predicted reward that results from the outcome generated by each of the actions under consideration. Decision values measure the net value of taking the different actions and prediction errors measure deviations from individuals' reward expectations. We use functional magnetic resonance imaging and a novel decision-making paradigm to dissociate the neural basis of these three computations. Our results show that they are supported by different neural substrates: goal values are correlated with activity in medial orbitofrontal cortex, decision values are correlated with activity in central orbitofrontal cortex, and prediction errors are correlated with activity in the ventral striatum. Furthermore, the results imply that dysfunctional processing in any one of these regions would lead to specific impairments that would disrupt some aspects of decision-making but not others.

Connecting the brain
(Diffusion based tractography and functional connectivity)

Dirk Neumann
Caltech

Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex

Ian Krajbich
Caltech

Damage to the ventromedial prefrontal cortex (VMPFC) has been reliably shown to impair concern for others. Reports of the real-life behavior of patients with such damage, together with impaired performances on tasks ranging from decision-making to moral judgment to economic games, support this conclusion. Yet there is to date no numerical quantification of the deficit in terms of a general mathematical model of choice, limiting our understanding of the underlying processes that might be impaired. Here we quantify these social deficits using a formal economic model of choice which incorporate terms for the disutility of unequal payoffs, manifested as guilt and envy. We gave seven patients with focal VMPFC lesions a battery of economic games that measure concern about their own payoffs and about the payoffs of others. Compared to brain-damaged (N=20) and healthy (N=16) comparison subjects, the VMPFC lesion patients donated significantly fewer points to others, and were both less trusting and less trustworthy. Pooling data across all the games, we show numerically that the VMPFC patients placed significantly less weight on allocating equal shares of payoffs. We propose that this general insensitivity to guilt may explain impaired performance on a number of social tasks, as well as impaired social behavior in real life.

Moral intuition:
Neural substrates and normative significance

Jim Woodward
Philosophy, Caltech

By "moral intuitions", contemporary philosophers mean quick, relatively automatic moral judgments that are typically reached with little awareness of the processing that leads to the judgment, as in the immediate judgment that many people have that incest is wrong. There is considerable disagreement within philosophy about the appropriate normative role of such intuitions, with some philosophers holding that acceptable moral theories should largely agree with our intuitive judgments, while others dismiss intuition as repository of superstition and prejudice, suggesting that it has no legitimate role to play in moral argument. Imaging and lesion studies associate moral intuition with aspects of social cognition involving processing in ventromedial prefrontal cortex, orbital frontal cortex, insula, anterior cingulate, and the amygdala. In addition, a number of researchers have associated processing in these areas with certain distinctive patterns of moral judgment; with reduced or impaired processing in these areas being associated with more "utilitarian" judgments, and increased activity with "anti-utilitarian" or "deontological' judgments. This talk, which is based on a forthcoming paper co-authored with John Allman ("Moral Intuition: Its Neural Substrates and Normative Significance"), will consider the implications of this understanding of the neural substrates of moral intuition for issues about the nature and status of moral intuition, the role of emotional processing in intuition, and whether appeals to intuition play some functional role in moral decision-making that would be lost if we abandon such appeals.

Intention, volition and the normative: From metaphysics to cognitive neuroscience

Henrik Walter
University of Bonn

Intentionality, Free Will and Moral Responsibility are eternal subjects in philosophy. Often, it is assumed that they are non-empirical by nature, i.e. that the concepts and problems associated with these terms are purely metaphysical and cannot, by their very nature, be subject to empirical investigations. Contrary to this assumption, I will argue that philosophical concepts themselves are shaped by our knowledge of the world and culturally based intuitions. Therefore, empirical investigations may serve as vivid examples and intuition pumps which will contribute to the assessment and possibly revision of some traditional concepts. After a very short introduction to the philosophical debate on free will and determinism I will present some studies from our lab investigating intentions, volition and normative decision making. Amongst others, I will present data on the role of social interaction in investigating the neural substrates of intentions. The main conclusion of my presentation will be that it is not the question if we are determined by our brains but how we are determined by them. This research question is empirically approachable and some results might modulate our abstract and normative concepts on philosophical issues considerably.

Audio-visual interactions within the cortical hierarchy

Uta Noppeney
Max Planck

To interact effectively with our environment, the human brain integrates information from multiple senses into a coherent and more reliable percept. Neurophysiological and functional imaging studies have revealed multi-sensory interactions in a widespread neural system encompassing subcortical structures, putative 'unisensory' and higher order association cortices. However, the types of information that are integrated in this multitude of integration sites remain unclear - spatial (where?), temporal (when?), stimulus content-related or semantic (what?), phonological and other types of information may be integrated at different levels of the cortical hierarchy. In a series of studies, we combined fMRI and psychophysics to investigate *where* and *how* different types of sensory features are combined within the cortical hierarchy. We presented subjects with object pictures and sounds while factorially manipulating the relative informativeness of the auditory and visual modalities. Collectively, our studies demonstrate that different types of information are combined at different levels of the cortical hierarchy with the anatomical locus determined by the type of information that is being integrated: While low level spatio-temporal interactions were found in Heschl's gyrus, higher order object features were integrated within the superior temporal sulci (STS) bilaterally. Furthermore, in primary sensory cortices, AV interactions were predominantly "automatic", in higher level association cortices (e.g. STS, ITG) performance dependent and in the fronto-parietal system driven by decisional processes. Consistent with the law of inverse effectiveness, the multi-sensory interactions were primarily subadditive and even suppressive for intact stimuli, but turned into (super)additive effects for degraded stimuli. Importantly, the (super)additive and suppressive modes at the neuronal level paralleled the multi-sensory benefit at the behavioural level across subjects suggesting that they may serve different aims: In the context of near-threshold unimodal inputs, bimodal stimulation decreases the thresholds of detection and identification leading to better categorization performance. In contrast, during categorization of intact stimuli that reach ceiling performance even when presented in one modality alone, suppressive interactions may reflect more efficient and faster processing through the dynamic weighting of the unimodal contributions according to their informativeness. In conclusion, the human brain integrates information that is abstracted from its sensory

inputs at multiple levels of the cortical hierarchy. The operational mode of audio-visual integration (in STS) is dictated by the informativeness of the auditory and visual modalities.

Neurogenetics of personality

Turhan Canli
Stony Brook University

Personality traits have a high degree of heritability, but are also influenced to a large extent by the non-shared environment. Thus, both unique life experiences and environmental exposure, alongside genetic factors, shape human personality. In this presentation, I will focus on the personality trait of neuroticism to summarize our current thinking on the role of genes and environment in personality research. Neuroticism, a risk factor for depression, is associated with a repeat length variation in the transcriptional control region of the serotonin transporter gene, which renders carriers of the short variant vulnerable for depression when exposed to life stress. The neural basis of this association is now being unraveled by a number of labs. We investigated the underlying neural mechanisms of these epigenetic processes in individuals with no history of psychopathology, using magnetic resonance-based imaging, genotyping, and self-reported life stress and rumination. Based on fMRI and perfusion data, we found support for a model by which life stress interacts with the effect of serotonin transporter genotype on amygdala and hippocampal resting activation, two regions involved in depression and stress. Life stress also differentially affected, as a function of serotonin transporter genotype, individuals level of rumination. We conclude that individual differences in vulnerability towards, or resilience against, mood disorders may be mediated by a gene x environment interaction. Neural correlates of these interactions are seen in brain regions previously associated with affective processing and brain response to stress, and may serve as biological vulnerability/resilience markers in future longitudinal studies.

When all the songs sounds the same: Insights into the musical brain

Lauren Stewart
Psychology Department, Goldsmiths, University of London.

The ability to make sense of musical sound has been observed in every culture since the beginning of recorded history. In early infancy, it allows us to respond to the sing-song interactions from a primary caregiver and to engage in musical play. In later life it shapes our social and cultural identities and modulates our affective and emotional states. But a few percent of the population fail to develop the ability to make sense of or engage with music. Individuals with congenital amusia (CA) cannot recognize familiar tunes, cannot tell one tune from another, frequently complain that music sounds like a "din" and avoid the many social situations in which music plays a role.

I will present evidence suggesting that congenital amusia can be explained by an insensitivity to pitch direction, the building block of musical contour. Amusic participants were found to have significantly higher thresholds than non-amusics when discriminating the direction of a pitch change. In many cases, thresholds for discrimination of pitch direction exceed one semitone (the building block and most common pitch change heard in Western music) and some had thresholds approaching an octave. However, while deficits in the perception of pitch direction may be sufficient to result in an amusic profile, some amusics have pitch direction thresholds in the same range as control participants, showing that elevated pitch thresholds cannot provide an exclusive explanatory account for the disorder. I will suggest that a deficit in pitch direction is just one possible endophenotype that underlies the behavioural manifestation of the disorder and outline several approaches that we are using in order to characterize other possible cognitive profiles. Considering amusia as a constellation of different cognitive endophenotypes not only reveals the complexity required for normal perceptual and appreciation of music but also helps to refine the behavioural phenotyping required for studies into the genetic and neurological basis of this developmental disorder.

From aggression to altruism: Neurobiological insights into the nature of punishment

Ben Seymour
Wellcome Trust Centre for Neuroimaging, UCL

The proclivity of humans to behave negatively towards one another incorporates behaviours that span aggression to strategic cooperative sanctioning, and plays a key part in the dynamics of social groups. One of the most interesting aspects of such behaviour is altruistic punishment, in which subjects punish others (typically free-riders who fail to cooperate in various forms of group interactions, but nevertheless take advantage of the group effort) at a pure cost to themselves (i.e., with negative immediate benefit), without any prospect of a direct return on this investment of effort or risk (i.e., with no long term pay-off at all). Economists have long mused about the proximate and ultimate basis of such behaviours, but critical insights may emerge from consideration (both psychological and neurobiological) of the underlying structure of motivation, learning and action systems. Viewing the evolution of altruism in the broader context of the evolution of learning systems, which need to balance the many other problems that arise when animals have to make decisions in uncertain environments, suggest new accounts of altruism that may not be quite as puzzling as previously thought.

In this talk, I will first review the basic structure of aversive learning and decision-making systems, distinguishing pavlovian and instrumental value and actions. The latter incorporates both habitual and goal-directed ('cognitive') systems. I will also mention observational learning, which is a key method for 'culturally' acquiring knowledge in social animals. I will then consider what happens when we embed these learning systems in social contexts, for example in game theoretic paradigms such as the iterated public goods game. This illustrates aspects of reciprocity that can be explained by reputation formation (a form of indirect reciprocity) and tit-for-tat (a form of direct reciprocity).

True altruism ('strong reciprocity') may be best thought of as a form of Pavlovian impulsivity, or as a generalisation error. First, innate responses to perceived unfairness may have evolved on the basis of punishment in non-altruistic circumstances, such as in groups or societies of small enough size such that individuals (and certainly their kin) would be likely to interact repeatedly with offenders, rendering the punishment 'selfish'. However, once established as an innate response, punishing non-cooperators could have

become blind to its proximal consequences for the individual (like other Pavlovian responses), thus appearing impulsive.

Second is the possibility that altruistic punishment arises from the structural inefficiency of instrumental control associated with habits, rather than the interference of Pavlovian imperatives over instrumental ones. Crudely, the idea is that choosing precisely whom to punish in a circumstance requires the detailed calculations of the consequences of punishment and likelihood of future interactions that only the goal-directed system could entertain. However, the habit system (and its observational-acquired counterpart) can engage in instrumental punishment in reciprocal cases and may therefore gain control over all similar such conditions, as discussed above. Its inability to calculate in detail the consequences of its output can then lead it to punish 'inappropriately' in altruistic situations.

Related reviews:

Ben Seymour, Tania Singer and Ray Dolan. The neurobiology of punishment. Nature Reviews: Neuroscience. April 2007 Vol 8; pp300-311

Peter Dayan and Ben Seymour. Values and actions in aversion. In Neuroeconomics: decision making in the brain. Edited by Glimcher, Fehr, Camerer and Poldrack. Elsevier 2008.

The social acquisition and alteration of learned fears

Elizabeth A. Phelps
NYU

Animal models of fear learning have relied on simple conditioning paradigms and emphasize a critical role for the amygdala. Research on the alteration and control of conditioned fears have highlighted the amygdala's interaction with the ventromedial prefrontal cortex and the hippocampus. In this talk, I will explore how these animal models extend to humans in a social context. Specifically, I will demonstrate how the neural circuitry of fear conditioning can be extended to fears learned through social communication. I will also discuss how the alteration of fears in humans through social and non-social means relies on overlapping neural mechanisms. Our studies suggest that social means of fear learning and their alteration in humans take advantage of phylogenetically shared systems of simple fear conditioning and extinction. Finally, I will present data suggesting that this flexibility of fear learning systems in humans may present unique challenges when trying to eliminate acquired fears through the disruption of reconsolidation.

Social cognitive impairments in Parkinson's Disease

Mitsuru Kawamura

Department of Neurology, Showa University School of Medicine

Although Parkinson's disease (PD) is primarily thought to be a dysfunction of the motor system, our recent studies show that PD patients have difficulties with social cognition such as facial-expression recognition and decision-making. In our study, recognition of fearful and disgusted faces was impaired in PD patients. Recognition of disgust was impaired in very early stage of PD. On the other hand, the recognition of the facial expression was normal in patients with Autosomal-Recessive Juveline Parkinson's disease (ARJP), in which the nigrostriatal dopaminergic system was selectively impaired whereas the mesocorticolimbic dopaminergic system was relatively intact. Decision-making may also be impaired in PD: although PD patients achieved the equivalent gain to controls in the Iowa Gambling Task (IGT), deck selection patterns of PD patients were different from those of normal controls. Compared with the normal controls, their decision-making was biased toward risky choices. This tendency toward risky choices was not correlated with age, education, intellectual ability, or severity of the disease. PD patients might have a potential preference for taking risky but high-reward behaviors. Skin conductance responses (SCRs) in PD during IGT were similar to those of patients with the amygdale damage. Response bias toward risky choices in PD may be explained by the dysfunction of the amygdala, which is known to evaluate risks. Both facial-expression recognition and decision-making were impaired in an early stage of PD, so those tasks may be useful for the early diagnosis of PD. The mesocorticolimbic dopaminergic system may play an important role in facial-expression recognition and decision-making, and the system may be impaired at an early stage of PD.

Semi-rational models of learned decision making

Nathaniel Daw
NYU

Decisions in the real world or the laboratory often involve substantial uncertainty about the task contingencies, which is only resolved via trial-and-error learning. Examples include "bandit" problems requiring choice between slot machines with unknown payoff probabilities, and morestructured tasks such as traversing an unfamiliar maze. While one can define optimal trial-by-trial choices during learning of such tasks, it is typically intractable actually to compute them. For this reason, the study of these problems in computer science ("reinforcement learning"; RL) concerns the development of tractable shortcuts and the analysis of under what conditions they approximate optimality. RL algorithms therefore offer a set of detailed hypotheses for how subjects (and their brains) might plausibly approach difficult decision problems. Moreover, their formal properties license novel inquiries such as whether subjects employ shortcuts well suited to the circumstances they face.

I discuss theory and data concerning what shortcuts organisms use to address some key problems in decision making: reconciliation of evidence received over time, the "credit assignment problem" for delayed rewards, and the exploration-exploitation dilemma.