Det Kongelige Danske VIDENSKABERNES SELSKAB

International meeting on:

Self-awareness - An emerging field in neurobiology

17 - 19 September 2014

Royal Danish Academy for Sciences and Letters H.C. Andersens Boulevard 35 DK-1553 Copenhagen V





P. S. Krøyer: "Et Møde i Videnskabernes Selskab", 1897 ("A Meeting at the Royal Danish Academy for Sciences and Letters")

The biology of self-awareness

- Synopsis of the scientific programme

The last two decades have seen a remarkable effort to include self-awareness and conscious experience as topics for the natural sciences. Neuroscientists, philosophers and psychologists share the goal of trying to understand how the mind relates to the brain. The effort represents a gathering conviction that a fruitful convergence of disciplines is timely. It is a co-evolutionary strategy, typified by interaction among research domains, where research at one level provides constraints, corrections, and inspiration for research at other levels. The interdisciplinary integration has, however, brought us to a methodological challenge. Neuroscience and cognitive science traditionally believe that objective data represent the primary source of evidence, and that subjective reports are secondary. Subjective experiences, however, cannot be observed "from the outside" – objectively.



At the symposium, this challenge will be addressed by opening lectures that describe the conscious experience and function of self-awareness. Important empirical data have been obtained by neuroimaging methods that are based on hemodynamics and energy metabolism. Several groups have shown that paralimbic cortical regions, and the lateral parietal and subcortical regions to which they are closely connected, are strongly active under conditions supposedly related to self-awareness. This includes the "resting state" which may in fact be dominated by self-reflection. The system has been referred to as the "default mode system" due to the fact that its energy consumption decreases markedly during goal-directed activity in the external world. The symposium includes a discussion of the basic nature of this system with presentation of different views. Its importance in human life has interested researchers in its ontogenetic and phylogenetic development, and new data will be presented.

As another example, the introduction of magneto-encephalography in the study of self-awareness and disorders has recently resulted in new understanding of the oscillatory activity that links the regions of the paralimbic "default mode" system. We now know more about the contribution of GABA-ergic interneurons to the genesis of such oscillations at the cellular level, and a separate section will deal with this area, and its promises with respect to the development of transmitter-based therapies. Exciting data have elucidated aspects of the pathogenesis of disorders of self-awareness and self-control, such as schizophrenia and pathological gambling.

The symposium aims to address the fundamental limitations of interpreting experimental results: To which extent can experimental findings provide insights into causal mechanisms behind the "emergence" or "creation" of conscious experience, and to which extent do they represent correlations only? Similarly, the symposium will discuss how empirical data best inspire, improve, and be incorporated in theoretical models of mind-brain relations in this field.

Self-awareness – an emerging field in neurobiology

PROGRAM - Wednesday 17 September 2014

08.30 - 08.45 Welcome address Leif Østergaard, Aarhus (leif@cfin.au.dk) Opening remarks Morten Overgaard, Aarhus (morten.overgaard@ki.au.dk)

Session I. Discussion Leader: Andreas Roepstorff, Aarhus

08.45 – 11.00 Awareness and self-awareness Dan Zahavi, Copenhagen (dza@hum.ku.dk)

Finding the Mind in the Brain: The opposing domains hypothesis Anthony Jack, Cleveland (anthony.jack@case.edu)

Action and self-awareness Patrick Haggard, London (p.haggard@ucl.ac.uk)

11.00 - 11.30 **Coffee**

11.30 - 13.00Awakening of the newbornHugo Lagercrantz, Stockholm (hugo.lagercrantz@ki.se)

Development of metacognition Josef Perner, Salzburg (josef.perner@sbg.ac.at)

13.00 – 14.00 Lunch

Session II. Discussion Leader: Chris Frith, London, Aarhus

14.00 – 16.15 *Prediction-driven interactive loops between self-awareness and theory of mind* Axel Cleeremans, Bruxelles (axcleer@ulb.ac.be)

> The self and others Andreas Roepstorff, Aarhus (andreas@pet.auh.dk)

Social aspects of metacognition Chris Frith, London (c.frith@ucl.ac.uk)

16.15 – 16.45 **Coffee**

Session III. Discussion Leader: Morten Overgaard, Aalborg, Aarhus

16.45 – 18.15 *That's me in the spotlight: The role of self-generated thought in understanding who we are* Jonathan Smallwood, Santa Barbara (smallwood@psych.ucsb.edu)

What is the fundamental empirical basis for self-awareness research? Morten Overgaard, Aarhus (morten.overgaard@ki.au.dk)

19.00 **Dinner** (Restaurant Peder Oxe, Gråbrødretorv 11)

Self-awareness – an emerging field in neurobiology

PROGRAM - Thursday 18 September 2014

Session IV. Discussion Leader Uta Frith, London, Aarhus

08.30 – 10.45 *Seeing without an I: Disturbances in minimal self-awareness* Shaun Gallagher, Memphis (s.gallagher@memphis.edu)

> Paralimbic activity in minimal and extended self-awareness Troels W. Kjaer, Copenhagen (neurology@dadInet.dk)

Detecting and evaluating social information Kai Vogeley, Cologne (kai.vogeley@uk-koeln.de)

- 10.45 11.15 **Coffee**
- 11.15 12.45 Self-awareness, emotions, and reward Morten L. Kringelbach, Oxford, Aarhus (morten.kringelbach@queens.ox.ac.uk)

How does a molecule affect self-reflection? Neuroimaging studies of oxytocin effects on neural correlates of self-referential processing Shihui Han, Peking University, Beijing (shan@pku.edu.cn)

12.45 – 14.00 Lunch

Session V: Discussion Leader: Hans C. Lou, Aarhus

14.00 – 16.15 *The frontal lobe and metacognition* Steve Fleming, London (sf102@nyu.edu)

> Precuneus and self-awareness Andrea Eugenio Cavanna, Birmingham (a.e.cavanna@bham.ac.uk)

Neurovascular coupling and the decoding of brain states: Is less more? Leif Østergaard, Aarhus (leif@cfin.dk)

20.00 - **D i n n e r** (Restaurant Nimb Terrasse, Tivoli)

Self-awareness – an emerging field in neurobiology

PROGRAM - Friday 19 September 2014

Session VI. Discussion Leader: Gitte Moos Knudsen, Copenhagen

19.00 -	Farewell Dinner (optional)
15.45 - 16.15	Coffee
15.15 – 15.45	Leif Østergaard, Aarhus: Concluding remarks
	From correlation to causality – Implications for pathophysiology Hans C. Lou, Aarhus (hanslou1@gmail.com).
	The neurochemical substrate for the actions of hallucinogens Gitte Moos Knudsen. Copenhagen (gitte@nru.dk)
13.00 – 15.15	Reprogramming of the self - teachings from placebo Martin Ingvar. Stockholm (martin.ingvar@ki.se)
12.00 - 13.00	Lunch
	<i>The Evolution of Self-Awareness</i> Julian P. Keenan, Montclair, New Jersey (keenanj@mail.montclair.edu)
10.30 - 12.00	Interneurons, gamma oscillations, and cognition Vikaas S. Sohal, San Francisco (vikaas.sohal@ucsf.edu)
10.00 - 10.30	Coffee
	Neural oscillations and predictive mechanisms in schizophrenia: Relevance to self-disturbances? Peter Uhlhaas, Glasgow (peter.uhlhaas@glasgow.ac.uk)
08.30 - 10.00	Brain oscillations and human behaviour Joachim Gross, Glasgow UK (joachim.gross@glasgow.ac.uk)



Self-awareness

– an emerging field in neurobiology

ABSTRACTS

The neuroscientific investigation of self and self-awareness has been going on for more than 15 years (cf. Damasio 1999, Kircher & David 2003, Feinberg and Keenan 2004, Feinberg & Keenan 2005, Gillihan & Farah 2005, etc.). During the last decade in particular many articles on the topics have been published. To that extent, one might wonder whether it is still appropriate to speak of self-awareness as an emerging field in neurobiology. The field seems pretty well consolidated. On the other hand, if one compares a few decades of neuroscientific work on self-awareness with philosophical theorizing on the topic, which has been going on for millennia, the former can indeed still be said to be in its infancy. When confronted with the disparity between the philosophical and neuroscientific exploration of these topics, the neuroscientist might opt for two rather different strategies. Either she do as Francis Crick (1995) and argue that there has been no real progress in philosophy during the last two thousand years, which simply shows that that approach has failed, and that it is now up to the empirical sciences to finish the job. Or she can adopt a slightly more modest attitude and remain open to the possibility that there might be resources in the philosophical tradition that can of relevance and importance for her own work. Not surprisingly, I favor the latter option, and as I will show in my talk, if neuroscience in its investigation of self-awareness chose to ignore philosophy, it does so at its own peril.

In my talk, I will discuss the relation between awareness and self-awareness. I will distinguish different notions of awareness and argue that we also need to distinguish different kinds of self-awareness. More specifically, I will defend the view that phenomenal awareness entails and is constitutively dependent upon what has been called pre-reflective self-awareness (Zahavi 1999, 2004, 2005, 2014). Failing to make the required distinctions, i.e., assuming that the notion of self-awareness is unequivocal, is bound to lead to confusion, not only regarding the relation between awareness and self-awareness, but also when it comes to any attempts to locate the neural correlate of self-awareness.

When reading research on self and self-awareness written by neuroscientists and neuropsychologists it is obvious that much effort is spent on explaining the experimental setup and on discussing and interpreting the experimental results. Much less time is typically devoted to discussing and clarifying the very notion of self and self-awareness at work. Indeed, to quote Klein, "most investigators sidestep these difficulties, relying on their readers' familiarity with the term 'self', derived from years of knowledge by acquaintance, to confer a sense of confidence (in my opinion, false) that he or she knows to what the author refers" (Klein 2010: 173). But a lack of clarity in the concepts used will lead to a lack of clarity in the questions posed, and thus also to a lack of clarity in the design of the experiments supposed to provide an answer to the questions. Time permitting, I will exemplify this with reference to recent work on autism and facial self-recognition.

- Crick, F. (1995). The Astonishing Hypothesis. Touchstone.
- Damasio, A. (1999). The Feeling of What Happens. Harcourt.
- Feinberg T.E., Keenan, J.P. (2004). "Not What, but Where, is your 'Self'." Cerebrum: The Dana Forum on Brain Science, 6:49-62.
- Feinberg, T. E., & Keenan, J. P. (ed.) (2005). The Lost Self. Oxford University Press.
- Gillihan, S. J., & Farah, M. J. (2005). Is Self Special? A Critical Review of Evidence From Experimental Psychology and Cognitive Neuroscience. Psychological Bulletin, 131(1), 76–97.
- Kircher, T., & David, A. (2003). The Self in Neuroscience and Psychiatry. Cambridge University Press.
- Klein, S. B. (2010). The Self: As a Construct in Psychology and Neuropsychological Evidence for Its Multiplicity. Wiley Interdisciplinary Reviews: Cognitive Science, 1/2: 172–183
- Zahavi, D. (1999). Self-Awareness and Alterity: A Phenomenological Investigation. Evanston: Northwestern University Press.
- Zahavi, D. (2004). Back to Brentano? Journal of Consciousness Studies, 11(10-11), 66–87.
- Zahavi, D. (2005). Subjectivity and selfhood: Investigating the first-person perspective. MIT Press.
- Zahavi, D. (2014). Self and other: Exploring subjectivity, empathy, and shame. Oxford University Press.

How can we be sure that others truly have a mind? Whether or not they know it by its philosophical label, the problem of other minds is surely familiar to virtually everyone. The problem derives from an obvious asymmetry in knowledge. We have internal access to our own conscious experience - we are self-aware. However, we can only know other minds indirectly, through externally observable features such as behavior. The inferences we make about the internal states of others on this basis are, by all indications, often wrong. For those willing to embrace radical skepticism, the epistemological problem of other minds leads directly to metaphysical solipsism - the view that there is no mind but one's own. Yet academic philosophers have not found solipsism to be a particularly appealing, or worrisome, metaphysical proposition about the mind. They have traditionally been more drawn to, or troubled by, a different metaphysical proposition: that of dualism. While solipsism is motivated by the apparent asymmetry between selfawareness and other-awareness, dualism is motivated by a less personal epistemological concern - how can we relate our knowledge of the physical to our knowledge of the mental? Not everyone believes this is a problem. Some have argued the appearance of a problem is merely an illusion (Dennett, 1991). Nonetheless, the history of philosophy tells a clear story. The explanatory gap (Levine, 2000) between these two types of knowledge drives what has been and continues to be one of philosophy's central and defining problems. It may seem curious that it is not the more intuitively obvious problem of other minds, but the mind-body problem, which has dominated debate in philosophy of mind. Why have philosophers been more concerned by the distinction between minds and machines than by the distinction between self-awareness and external awareness? Philosophy has been accused of many flaws, yet it does have the distinction of being the only academic discipline that directly concerns itself with studying the nature of knowledge. Perhaps the philosophers are on to something.

Cognitive neuroscience is based on a simple methodological principle, known as cognitive subtraction (Price & Friston, 1997). Participants are presented with more and less complex tasks, and increases in brain activity are attributed to the presence of additional cognitive components in the more complex task. However, 15 years ago it was noticed that some brain regions break the rules of cognitive subtraction - they were more active the less complex the task (Shulman et al., 1997). These brain areas came to known as the Default Mode Network (DMN). It was later found that the DMN could usually be seen to be in tension with the Task Positive Network (TPN) (Fox et al., 2005). During tasks and at rest (in the absence of any task), the DMN tended to be less active when the TPN was more active. What could account for this pattern? The main hypothesis, which is now often stated as established fact, is that the activity in these networks reflects a tension between external and internal attention (Buckner, Andrews-Hanna, & Schacter, 2008). When people attend to their memories, their emotions or to their selves, then the DMN tends to be more active. On the other hand, when people attend to external stimuli, such as visual stimuli, then the DMN tends to be less active. Furthermore, we know that the TPN overlaps brain regions involved visual perception and attention, and that the DMN overlaps brain regions involved in awareness of our own bodily states. In other words, it has looked like the tension between these networks was a perfect candidate for explaining the problem of other minds - such that the DMN is concerned with self-awareness, and the TPN with external perception. There is no doubt that much of the data fits this interpretation, but is it the full story?

My laboratory has been working on testing the external vs. internal attention hypothesis against a second hypothesis called the opposing domains hypothesis. These hypotheses have some points of overlap. The opposing domains hypothesis holds that we understand the world using concepts which are grounded in different inputs. A similar hypothesis was originally put forward by the philosopher Thomas Nagel in his highly influential paper on consciousness "What is it like to be a bat?" (Nagel, 1974). In a footnote, Nagel explains how the mind-body problem might be understood as arising from our possessing two independent and disparate types of imagination "where the imagination of physical features is perceptual and the imagination of mental features is sympathetic." Our understanding of the internal mental states of others does seem to be grounded in self-awareness. Consider, for instance, the psychopathic profile. Psychopaths are relatively insensitive to punishment – a psychological feature thought to be due to reduced

amygdala function. They are not well attuned to their own distress, and they are also not well attuned to the distress of others. They miss out on one of the most important cues which triggers our sympathy for others, although recent research indicates that they can model others' internal mental states perfectly well when they deliberately choose to do so, which is often for instrumental purposes.

We have been comparing the two hypotheses using both experimental tests and meta-analytic methods. First, we have shown that different types of external attention-demanding tasks can push the networks in different directions, depending on whether people are asked to think about internal mental states or physical mechanisms (Jack et al., 2012). These effects cannot be explained by differences in visual attention. Second, we have shown that activity in these networks is influenced by whether we view other people sympathetically versus unsympathetically, as a result of dehumanizing (Jack, Dawson, & Norr, 2013). Third, we have conducted a meta-analysis which demonstrates that the prior evidence taken to support the internal vs. external attention hypothesis has not been examined carefully enough to distinguish that hypothesis from the opposing domains hypothesis. We show that the default network is not consistently recruited during memory recall, but only when the content of the memory inclines us to a sympathetic way of thinking, e.g. autobiographical memories. We also show that thinking about the actions and behaviors of others, the external symptoms of the mind, recruits the task positive network whereas thinking about the internal states of others, such as their emotions, recruits the default network. Finally, we have found that psychopathic personalities tend to view the mind as nothing but a physical object, whereas more sympathetic individuals are inclined to dualism (Jack, in press).

In conclusion, it is suggested that if we wish to promote self-awareness, then we shall first need to recognize that there are two very different paths to self-knowledge: one that involves a scientific understanding of ourselves as biological machines, and a second that involves a sympathetic understanding of our own and others' experience. Care must be taken, first, to avoid confusing these distinct types of knowledge, second, to achieve a synthesis that respects the unique insights offered by each path.

- Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The Brain's Default Network: Anatomy, Function, and Relevance to Disease. Annals of the New York Academy of Sciences, 1124(1), 1-38.
- Dennett, D. C. (1991). Consciousness explained (1st ed.). Boston: Little, Brown and Co.
- Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc Natl Acad Sci U S A, 102(27), 9673-9678.
- Jack, A. I. (in press). A scientific case for conceptual dualism: The problem of consciousness and the opposing domains hypothesis. In J. Knobe, T. Lombrozo & S. Nichols (Eds.), Oxford Studies in Experimental Philosophy (Vol. 1): Oxford University Press.
- Jack, A. I., Dawson, A. J., Begany, K. L., Leckie, R. L., Barry, K. P., Ciccia, A. H., & Snyder, A. Z. (2012). fMRI reveals reciprocal inhibition between social and physical cognitive domains. Neuroimage, 66C, 385-401.
- Jack, A. I., Dawson, A. J., & Norr, M. E. (2013). Seeing human: Distinct and overlapping neural signatures associated with two forms of dehumanization. Neuroimage, 79C, 313-328.
- Levine, J. (2000). Conceivability, identity, and the explanatory gap. Toward a Science of Consciousness lii, 3-12.
- Nagel, T. (1974). What Is It Like to Be a Bat. Philosophical Review, 83(4), 435-450.
- Price, C. J., & Friston, K. J. (1997). Cognitive conjunction: a new approach to brain activation experiments. Neuroimage, 5(4 Pt 1), 261-270.
- Shulman, G. L., Fiez, J. A., Corbetta, M., Buckner, R. L., Miezin, F. M., Raichle, M. E., & Petersen, S. E. (1997). Common blood flow changes across visual tasks: II. Decreases in cerebral cortex. Journal of Cognitive Neuroscience, 9(5), 648-663.

Many investigations of self-awareness used actions as an important marker of self. In the well-known "rouge test", an animal is aware that it looks at itself if and only if it makes an appropriate targeted action towards the patch of rouge. However, action is also an important cue to self-awareness, and could even be partly constitutive of self-awareness. Several research traditions have investigated how we come to experience ourselves as the authors of our own actions. One class of experimental design compares the experience of "self-generated actions" with the experience of a physically similar body movement that is passive, involuntary, or reflexive. In this talk, I will review the findings of this experimental programme. Most results point to the conclusion that people know what they are doing, and that the experience of one's own action is not exhausted by the experience of physical body movement.

Experiments in which self-generated action and involuntary movement co-occur offer a particularly interesting case: do involuntary movements that co-occur with self-generated actions come to feel self-generated, by a form of Hebbian associative learning and transfer? Within the limited learning opportunity of a typical behavioural experiment, the answer appears to be "no". This suggests that the experience of self-generated action, and thus an important aspect of self-awareness, depends on an experience triggered rather directly by some internal signal within the brain's motor systems. I will describe a number of behavioural, neuroscientific and psychopathological studies that aim to identify this putative signal.

The newborn infant wakes up at birth when it takes its first breaths of air and cries. The eyes become wide open with large pupils. It is aroused triggered by the stress of being born¹ and the transition from the womb to the cooler extrauterine environment. The evaporation of amniotic fluid has this cooling effect even in a tropical mileu. The first breaths of air have since antique time been regarded as the ignition of life as indicated by the word spiritus. The newborn becomes animated in this way – i.e. the emergence of consicousness². In modern time a bioethics committee³ has stated that when the newborn 'encompasses the capacity to breathe either independently, or with the support of a ventilator is the moral and legal point when human life must be preserved independent of gestational age.

The stress of being born is probably mainly due to the squeezing and squahing of the fetal head during vaginal delivery. This triggers an enormous catecholamine surge resulting in about a 20-fold higher levels in umbilical arterial blood. There is probably a parallell surge of the noradrenergic activity in the brain originating from the locus coeruleus⁴, which is responsible for the arousal. After elective cesarean section less catecholamines are released, which also has been showed to delay the transition at birth. However, the cooling and also the clamping of the umbilical cord and the removal of placental suppressors seem to be sufficient to awake the newborn also after caesarean section.

Newborn infants react differently when they are touched by another person than by themselves, they must have some awareness of their own bodies⁵. They have a proprioceptive sense of their own bodies. They feel not just sense the world.

Already newborns imitate gestures. Reciprocal imitation games support social bonding and affiliation. The "like me" concept gives rise to a life-long ability to connect with other persons. This is vital for our survival as species.

Imitation is mediated by mirror neurons, which seem to mature early in life. Neural mirroring mechanisms establish prelinguistic mapping between self and others⁶. This includes the anticipation of emotional reactions of other people. Thus infant imitation is not an automatic uncontrolled impulse but under intentional control.

A spontaneous resting state activity has been found in newborn infants with fMRI⁷. This activity may correspond to the idea of William James that there is a "stream of consciousness". It involves five hubs including the somatosensory system and the auditory and visual cortex in the infants⁸. This is in contrast to adults⁹ where ten hubs were idenfied including the insula, precuneus and ventromedial prefrontal cortex⁸. These findings make sense, the infant is probably only aware of what it feels, sees and hears in present time, while the adult relates the sensory input to memories, itself and also plans for the future.

- 1. Lagercrantz H, Slotkin TA. The "stress" of being born. Sci Am. Apr 1986;254(4):100-107.
- 2. Lagercrantz H, Changeux, J-P. The emergence of human consciousness: From fetal to neonatal life. Pediatric Res 2009(65):255-260.
- 3. http://www.nuffieldbioethics.org. 2007.
- 4. Lagercrantz H. Stress, arousal, and gene activation at birth. New Physiol Sc 1996(11):214-218.
- 5. Rochat P. The self as phenotype. Consciousness and Cognition. 2011;20:100-119.
- 6. Marshall PJ, Melzoff, A.M. Neural mirroring mechanisms and imitation in human infants. Philosophical transactions of the Royal Society B. 2014.
- 7. Fransson P, Skiold B, Horsch S, et al. Resting-state networks in the infant brain. Proc Natl Acad Sci U S A. Sep 25 2007;104(39):15531-15536.
- 8. Fransson P, Aden U, Blennow M, Lagercrantz H. The functional architecture of the infant brain as revealed by resting-state fMRI. Cereb Cortex. Jan 2011;21(1):145-154.
- 9. Fox MD, Raichle, M. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nature Review Neuroscience. 2007;8:700-711.

I will concentrate on the foundational issue of when children develop the basic ability to engage in metacognition. Metacognition is commonly defined as the ability to know that one knows, or think that one thinks. More generally, one can take it as the ability to have a mental state about a mental state. So feeling that one does not know, or being aware that one knows, would also qualify as acts of metacognition. I take this recursive element of a mental state about a mental state as defining of metacognition.

It is notoriously difficult to show metacognition in this recursive sense in non-linguistic animals and even in children who are not yet able to clearly talk about their own mental states. In particular there are two methodological problems to be overcome (Perner 2012). To show that children are aware of their inner states one has to minimally show a behavioural difference when they are aware of this state as compared to when they are not. Animal studies are able to demonstrate that monkeys Couchman et al, 2012) and apes (Call 2012) and birds (Fujita et al 2012) can be trained to give different secondary responses (or opt out from their primary response; Hampton 2001) depending on whether they are certain about the correct primary response or uncertain (e.g., Kornell et al 2007). This is typically interpreted as "confidence" ratings which are anthropomorphically taken as signs of gamblers' insight into their uncertainty.

One problem with this interpretation is, however, the "Being vs. Knowing" distinction (Perner 2012), i.e., it remains an open question as to whether the secondary responses are really responses based on metacognitive awareness of being uncertain (KNOWING that one is uncertain) about their primary response or whether it is based on the animals' state of BEING uncertain (i.e., not their awareness of being in that state). To my knowledge, there is no solution to this problem at hand, yet.

Another, related problem of interpretation arises from the fact that judgements about the world are based on internal, mental processes. This raises the "internal-external" problem in form of the question whether a specific test response is based on awareness of these internal processes (which would provide evidence of metacognition) or on awareness of the state of the world represented by these internal processes. For instance, when one is uncertain what to do one judges a task to be difficult. If one know immediately what to do the task is perceived as easy. The animals' "confidence" response may, therefore, not be evidence for animals metacognitively representing their uncertainty but only evidence for animals distinguishing easy from difficult tasks. That is, they represent something about the world and not their mental state about the world.

I will discuss how these problems apparent in research with non-linguistic animals also affects research with young children (or even adults) whose linguistic competence or understanding of mental states falls short of mature metacognitive discourse. For young infants there is little evidence of metacognition (Sodian et al 2012). Some evidence for children from 3 years is based on animal paradigms (Balcomb & Gerken 2008) and on implicit measures of uncertainty (Paulus et al 2013) which suffers from the same interpretation problems as the evidence for animals. Also some language based evidence points to 3-year olds as having metacognitive insights (Ghetti et al 2013, Rohwer et al 2012). This evidence leaves unclear whether children really mean to talk about their internal state of knowing or whether they mean with "know" their ability to act or answer competently. Their ability to distinguish their knowledge from just answering competently and to distinguish knowledge based responses from lucky guesses tends to emerge around 6 or 7 years (Miscione et al 1978: Rohwer et al 2012). There is evidence that children are sensitive to their state of ignorance from as early as 2½ years (they check when being uncertain where something is before responding; Call & Carpenter 2001). Nevertheless, once they have to commit themselves to a response in absence of any information they seem to mistake their assumption/guess for knowledge until they are 6 or 7 years old (Rohwer et al, unpublished manuscript).

- Balcomb, F. K., & Gerken, L. (2008). Three-year-old children can access their own memory to guide responses on a visual matching task. Developmental Science, 11(5), 750-760. doi:10.1111/j.1467-7687.2008.00725.x
- Call, J. (2012). Seeking information in non-human animals: Weaving a metacognitive web. In M. Beran, J. L. Brandl, J. Perner & J. Proust (Eds.), Foundations of metacognition (pp. 62-75). Oxford: Oxford University Press.
- Call, J., & Carpenter, M. (2001). Do apes and children know what they have seen? Animal Cognition, 4, 207-220.
- Couchman, J. J., Beran, M. J., Coutinho, M. V., Boomer, J., Zakrzewski, A., Church, B., et al. (2012). Do actions speak louder than words? A comparative perspective on implicit versus explicit meta-cognition and theory of mind. The British Journal of Developmental Psychology, 30(Pt 1), 210-221. doi:10.1111/j.2044-835X.2011.02065.x; 10.1111/j.2044-835X.2011.02065.x
- Fujita, K., Nakamura, N., Iwasaki, S., & Watanabe, S. (2012). Are birds metacognitive? In M. Beran, J. L. Brandl, J. Perner & J. Proust (Eds.), Foundations of metacognition (pp. 50-61). Oxford: Oxford University Press.
- Ghetti, S., Hembacher, E., & Coughlin, C. A. (2013). Feeling uncertain and acting on it during the preschool years: A metacognitive approach. Child Development Perspectives, 7(3), 160-165.
- Hampton, R. R. (2001). Rhesus monkeys know when they remember. Proceedings of the National Academy of Sciences USA, 98, 5359-5362.
- Kornell, N., Son, L. K., & Terrace, H. S. (2007). Transfer of metacognitive skills and hint seeking in monkeys. Psychological Science, 18, 64-71.
 Miscione, J. L., Marvin, R. S., O'Brien, R. G., & Greenberg, M. T. (1978). A developmental study of preschool children's understanding of the
- words"know" and "guess". Child Development, 49, 1107-1113.
 Paulus, M., Proust, J., & Sodian, B. (2013). Examining implicit metacognition in 3.5-year-old children: An eye-tracking and pupillometric
- study. Frontiers in Psychology, 4(Article 145), 1-7.
 Perner, J. (2012). MiniMeta: In search of minimal criteria for metacognition. In M. Beran, J. L. Brandl, J. Perner & J. Proust (Eds.),
- Foundations of metacognition (pp. 94-116). Oxford: Oxford University Press.
 Rohwer, M., Kloo, D, & Perner, J. (unpublished manuscript). Explicit and Implicit Admission of Ignorance by Children. Unpublished manuscript, Department of Psychology, Unviersity of Salzburg.
- Sodian, B., Thoermer, C., Kristen, S., & Perst, H. (2012). Metacognition in infants and young children. In M. Beran, J. L. Brandl, J. Perner & J. Proust (Eds.), Foundations of metacognition (pp. 119-133). Oxford: Oxford University Press.

Starting from the radical idea that consciousness is something that the brain learns to do rather than a static property associated with some patterns of neural activity and not with others, I explore the links between theory of mind, self-awareness, and perceptual awareness. Considering first the link between self-awareness and perceptual awareness, I suggest, congruently with the Higher-Order Thought (HOT) Theory of consciousness developed by Rosenthal, that first-order representations are conscious if and only if they are targeted by appropriate higher-order representations, that is, metarepresentations. The main functions of such metarepresentations are (1) to redescribe the target first-order representations in such a way as to explicitly indicate mental attitude, and (2) to subserve prediction-driven control mechanisms. Crucially, (1) such metarepresentations do not need to be conscious themselves (as in HOT), and (2) they emerge over training and development as a result of unconscious learning and plasticity mechanisms.

Metarepresentations thus form the basis for self-awareness because they enable agents to "know that they know", that is, to be acquainted with the geography of their own representational systems. I illustrate these arguments with implemented computational models (connectionist networks) applied to different experimental paradigms.

Next, I turn to the link between self-awareness and theory of mind. The main argument here is that developing infants continuously attempt to predict not only the consequences of their actions on the world, but also the consequences of their actions on other agents. But there is a crucial difference between interactions with the world and interactions with other agents: Understanding the reactions of the latter, unlike the former, requires assuming the existence of hidden, unobservables states. Thus, when one learns to interact with other agents, one also forms mental models of the internal states of those other agents. But this is the same prediction-driven process as involved in forming metarepresentations of one own's mental states. There is thus a direct link between theory of mind and self-awareness, a point that was forcefully argued by Carruthers. Hence we bridge the gap from theory of mind to perceptual awareness through the joint involvement of prediction-driven, learned interactive loops that make it possible for agents to better anticipate the consequences of their actions.

"Consciousness is at once the oldest problems of philosophy and one of the youngest problems of science".

The quote may sound quaintly contemporary ... However, it was announced in 1902 in Pittsburgh in the *Address of the President of the American Association for the Advancement of Sciences*, given by the physiologist Charles Sedgwick Minot. In interesting ways, Minot's solution to *The Problem of Consciousness in its Biological Aspect* was almost opposite to the current 'standard interpretation'. He argued, forcefully, against epiphenomenalism and for consciousness as a causal factor, both in the life of organisms and in evolution. Are we ever since Minot caught in a never-ending dialectic of epiphenomenalism vs. mental causation? Perhaps, but I will argue that at least two things have changed. The first is epistemic, a much better understanding of how the specifics of experimentation is intertwined with particular conceptual configurations. The second is ontological, there may be a new appreciation of how self and other, your consciousness and my consciousness, may at times be intertwined. I will discuss these points with reference to current experimental literature.

What is metacognition?

Metacognitive processes have been defined as 'thinking about thinking', 'cognitive processes that monitor and control other cognitive processes', or 'control processes that represent the properties of other control processes.' Under all such definitions, self-awareness requires metacognition. On the other hand, metacognition is not restricted to self-awareness, since metacognitive processes can be applied to the cognition of others and can also occur in the absence of awareness.

Subpersonal metacognition

In earlier accounts (e.g. 1), the cognitive control achieved by metacognition was associated with conscious awareness. More recent experiments suggest that metacognitive control can happen without awareness. For example, expert typists slow down after making an error (post error slowing), but do not slow down after a false error is inserted by the experimenters. However, they believe that they hae made the false error (2). This shows a clear distinction between subpersonal (implicit) metacognition (monitoring and control of behaviour without awareness) and explicit metacognition (self-awareness at the personal level).

Personal, explicit metacognition

Metacognitive processes at the subpersonal level can be highly sophisticated and accurate. For example, the integration of vision and touch requires information about the relative precision of these two sensory processes (metacognitive information), which is used to achieve a statistical optimal result (3). In contrast explicit metacognition is frequently inaccurate and erroneous. For example, people are quite good at catching balls, but their beliefs about how the ball will seem to move from the point of view of the catcher are systematically wrong (4). The phenomenon of choice blindness (5), in which people will accept and justify a choice they have not actually made, also reveals the fragility of self-awareness. This latter phenomenon shows that, even though we have rather little awareness of our decision-making processes, we are more than willing to talk about and justify our decisions.

A social function for explicit metacognition

In the context of decision-making, explicit metacognition appears to be more engaged when justifying decision to others, rather than making the decisions in the first place. This observation suggests that explicit metacognition might have a role in the control of interactions between people, rather than within people (6). To explore this idea we created a paradigm, social psychophysics, in which visual signals experienced by two people have to be integrated (7). This was by analogy with previous studies in which visual and touch signals have to be integrated within a single person. The results confirmed that, as long as the two people had roughly equal perceptual abilities, there was an advantage for working as a pair. Furthermore, the advantage was statistically optimal, give that the pair were communicating confidence to each other. Analysis of the discussion through which the pairs reached their joint decisions confirmed that communication of confidence was critical for the group advantage (8).

The malleability of explicit metacognition: accuracy & consensus

At first sight, this result may seem at odds with my claim that explicit metacognition is often inaccurate. Presumably optimal integration of perceptual experience should require accurate reporting of confidence. However, an accurate representation of confidence is not sufficient for a successful interaction. It is also necessary for the reporting of confidence to be aligned. The over-confident report of an incompetent person should not be given the same weight as the under-confident report of a competent person. In the social psychophysics we observed a group disadvantage when there was a marked discrepancy in the perceptual abilities of the two partners (7). The reporting and, perhaps, the experience of the metacognitive quantity of confidence need to be modified to suit the current interaction. Through discussion partners need to reach a consensus about how to report their confidence.

The consensus that results need not be an accurate reflection of reality. There is a considerable consistency among peoples' erroneous belief about how a ball will appear to move from the point of view of the catcher (4). In some cases consensus may be more important than accuracy. For example, most people believe that we act of our own free will (9), even though the proposal that this belief is an illusion has a long history (10). However, even if belief in free will is an illusion, this consensus has important consequences for ideas about responsibility and for social cohesion (11).

Mechanisms for the modulation of explicit metacognition

How is it possible for self-awareness to be modified by discussions with others? I favour a Bayesian framework in which the brain/mind is conceived as a hierarchy of interactions between prior beliefs and evidence. Comparison of expectations derived from prior beliefs and incoming evidence generates prediction errors that can be resolved by adjusting our beliefs about the world. In this framework prior beliefs and expectations create top-down influences on lower-level cognitive processes. These top-down influences do not stop within the person, but can also arise from discussions with others and from culture more generally (12). For example, telling people that free will is an illusion can alter behaviour (13) as well as subpersonal cognitive and neural processes (14). These effects seem to result from a weakening belief in the effectiveness of intentional control. The precise mechanisms by which culture exerts its effects on self-awareness and on subpersonal processes remains to be elucidated, but a plausible account might involve alterations of the estimates of precision associated with prior beliefs and with evidence. For example, a strong prior belief could result in a lowering the precision associated with evidence so that only strongest signal would have an effect. In this way erroneous cultural beliefs could be maintained.

- 1. Norman DA, Shallice T. 1986. In Consciousness and self regulation: Advances in research, ed. RJ Davidson, GE Schwartz, D Shapiro, pp. 1-18. New York: Plenum Press
- 2. Logan GD, Crump MJ. 2010. Science 330: 683-6
- 3. Ernst MO, Banks MS. 2002. Nature 415: 429-33
- 4. Reed N, McLeod P, Dienes Z. 2010. Conscious Cogn 19: 63-76
- 5. Johansson P, Hall L, Sikstrom S, et al. 2005. Science 310: 116-9
- 6. Shea NJ, Boldt A, Bang D, et al. 2014. Trends Cogn Sci 18: 186-93
- 7. Bahrami B, Olsen K, Latham PE, et al. 2010. Science 329: 1081-5
- 8. Fusaroli R, Bahrami B, Olsen K, et al. 2012. Psychol Sci 23: 931-9
- 9. Nahmias E, Morris S, Nadelhoffer T, et al. 2005. Philosophical Psychology 18: 561-84
- 10. Spinoza B. 1677. In Ethics (translation 1994). London: Penguin
- 11. Frith CD. 2013. Neuropsychologia 55: 137-42
- 12. Frith CD. 2014. Nova Acta Leopoldina
- 13. Baumeister RF, Masicampo EJ, DeWall CN. 2009. Personality and Social Psychology Bulletin 35: 260-8
- 14. Rigoni D, Kuhn S, Sartori G, et al. 2011. Psychological Science 22: 613-8

Knowing who we are and what we want to do next is one of the most important consequences of the human capacity for self-awareness because it allows a person that chance to pursue goals and objectives that extend beyond the often meagre opportunities afforded by the here and now. Studies examining the contents of thought indicate that much of our time is spent engaged in thoughts relating to our selves in the past or future and several features of these experiences suggest that they are important because of their independence from events in the surrounding environment. First, these experiences arise through processes that are only indirectly related to the events in the immediate environment. This process of self-generation is important because it provides for mental content that is independent of external input. Second, the experience of these self-generated thoughts is accompanied by a disengagement of attention from on-going events a process known as perceptual decoupling. This decoupling of attention from perception allows cognition to proceed in a manner that is insulated from the demands of the external environment.

Together the combination of self-generation and perceptual decoupling provides cognition the chance to follow a train of thought that is distinct from the external environment and hence allows a person to pursue goals that are unique to their needs, rather than following the dominant opportunities that a given moment offers. This talk will present evidence that indicates that these self-generated experiences are a core mechanism that is central to the originality and creativity of human cognition. It will discuss experimental work that shows that a primary function of our capacity to self-generate cognition is to make plans about the future, and that this capacity for prospection is beneficial in terms of its capacity to alleviate unhappiness, reduce stress and is linked to functional integrity within key regions of the default mode network, including the hippocampus. As imagining the future depends criticality on being able to think in original and creative ways, this prospective bias reflects the importance of self-generated thought in the human condition.

Although the capacity to engage in thoughts that do not refer to events in the here and now is important, the accompanying perceptual decoupling can jeopardise the performance of demanding tasks in the external environment. Indeed, when these thoughts occur in opposition to external tasks they can derail acts such as reading or sustained attention. Maximising the value of original and creative thinking, therefore, requires that these experiences are regulated so as to only occur when tasks are relatively undemanding and they are less likely to cause error. Consistent with this context regulation hypothesis we have found that several different measures of executive control all predict greater self-generated thinking when the external environment lacks rigorous external demands. For example, the capacity to delay gratification, as assessed by making patient economic decisions, is linked to greater self-generated thought when tasks lack compelling external input. This association between regulating economic decisions to maximize long term gain, and the capacity to regulate self-generated thinking so as to minimize the negative consequences of absent minded errors depends on individual differences in regions of the brain that play a role in minimising conflict, such as the anterior cingulate cortex.

Together this body of empirical work reveals that an important aspect of human thought is the capacity to generate thoughts and feelings that are independent of the environment. It seems that this freedom from immediacy that characterises our thoughts is both a primary motivation for people in todays society, as well as a cognitive capacity that is necessary to navigate the complex social environment in which we as a species exist. Rather than states of mind-wandering and daydreaming as being viewed as a lapse in concentration of an idle mind, their high frequency in daily life, and their links to creativity and originality all suggest that these states are likely to reflect a fundamental aspect of how the mind and brain maintains a coherent sense of self and so is able to perform goals that are extended over time.

Moving forward a complete understanding of the role of self-awareness to the human cognition will likely require a shift from the task driven methodology that is central to much of cognitive science, to more nuanced experimental paradigm that captures the dynamic interplay between intrinsic and extrinsic processes which are well suited to the pursuit of goals that are both present in the moment and those with a much longer time frame. Such a paradigm shift raises important questions for future research: (i) How does the mind/brain balance the experiential frame to current or distant goals depending on the internal and external demands of the person? (ii) What is the difference between external and internally motivated cognitive processes in terms of their neural signature?

- Smallwood, J. (2013). Distinguishing how from why the mind wanders: a process-occurrence framework for the wandering mind. Psychological Bulletin.
- Smallwood, J. & Andrews-Hanna, J, (2013). Not all minds that wander are lost: towards a balanced perspective on the mind-wandering state. Frontiers in Psychology.
- Andrews-Hanna, J, Smallwood, J. & Spreng, N.S (2014). The default mode network and self-generated thought. Annals of the New York Academy of Sciences.

Introduction

Over the last decades, neuroscience has had an increasing influence on human self-understanding. Even selfunderstanding itself has become an object of scientific investigations as a "neuroscience of self-awareness". Whereas such research is highly intriguing, it is not obvious how it should be carried out. As it is the case with any scientific domain, empirical research requires at minimum a definition of the exact object or phenomenon to be investigated as well as choices regarding measurement and methodology.

Today, a number of experiments have used neuroscientific techniques to reveal the neural substrates of self-awareness. There is however only good reason to keep scrutinizing foundational issues regarding concepts and methods. One could think of conceptual, behavioural and neuroscientific investigations as ideally working in tandem so that conceptual issues guide behavioural measures, which in turn would result in more informative and fine-grained neuroscientific data.

The fundamental empirical basis

So what is the fundamental empirical basis for research on self-awareness? The concept of self-awareness can be taken to relate to, at least, two different kinds of data: Reports (or comparable behaviour) about one self and one's properties (as a very simple computer programme could give) and reports about one self and one's properties based on conscious experience. Arguably, the most challenging and most fundamental conception is the last: It is by way of our experience that we ever get the idea that we have selves and we have knowledge about it. One serious challenge to this idea was put forward by philosopher Fred Dretske who argued that we never have experiences of selves or of consciousness, but only of the contents of consciousness, e.g. worldly objects ⁽¹⁾. We "see in by looking out", so to say. However, various other arguments from e.g. Sidney Shoemaker have shown that we know about ourselves whenever we are conscious as a matter of principle ⁽²⁾. A possible solution could be that the self is present in consciousness not as an object but as an aspect or a kind of relation to an object.

In that case, introspection becomes central to the study of self-awareness as the primary way in which we gain knowledge about ourselves. This is not a problem, in and of itself, as that is also the case in consciousness research and other, more classical areas of scientific study. However, it leads to particular demands and limitations regarding how we might measure self-awareness. In particular, we lack a measure to directly target the conception above.

Metacognitive measures

We have today very little knowledge of the relation between consciousness and introspection, and whether introspection changes the contents of consciousness. The few experiments that have attempted to target this relation directly indicate that introspection is separate from consciousness, yet may affect or change it "online" as it occurs ^(3, 4).

Introspection and self-awareness may be considered specific varieties of metacognition ⁽⁵⁾. Accordingly, one could apply established measures of metacognition hoping that they universally capture all varieties of it and thus also self-awareness. Such an approach would obviously not be very exclusive and thus possibly confounded. Experimental studies that directly have compared different measures of metacognition, however, indicate a high degree of specificity for the individual measures. Confidence ratings measure how confident one feels about one's own reports, the Perceptual Awareness Scale measures how clearly one experiences a particular content in consciousness, etc ⁽⁶⁾. Accordingly, such subjective measures are highly relevant in the study of our awareness of particular mental properties (e.g. our knowledge of our experience of confidence or perceptual clarity) but not necessarily in the study of our awareness of self per se.

Conclusions and consequences

The positive outcome of the discussion above is that it may seem possible to have a coherent scientific approach to at least two conceptions of self-awareness. It is however more uncertain if we have yet found a proper way to measure self-awareness in the most fundamental meaning of the concept, our conscious experience of being a self.

Another consequence of the conceptual distinction between self-awareness relating to consciousness and selfawareness not necessarily relating to consciousness concerns the possibilities for a neuroscientific explanation. It may seem possible, at least potentially, to explain conceptions not relating to consciousness in completely functional terms derived from neuroscience. Conceptions relating to consciousness inherit the explanatory challenges related to consciousness per se.

REFERENCES

- 1. Dretske, F. (1995): Naturalizing the Mind, MIT Press
- 2. Shoemaker, S. (1996): The First-Person Perspective and Other Essays, Cambridge University Press
- 3. Overgaard, M, & Sørensen T.A. (2004): Introspection distinct from first order experiences, Journal of Consciousness Studies, 11 (7-8), 77-95
- 4. Overgaard, M, Koivisto, M., Sørensen, T.A., Vangkilde, S. & Revonsuo, A. (2006): The electrophysiology of introspection, Consciousness and Cognition, 15, 662-672.
- 5. Overgaard, M. & Sandberg, K. (2012): Kinds of access: Different methods for report reveal different kinds of metacognitive access, Philosophical Transactions of the Royal Society of London Series B: Biological Sciences, 367, 1287-1296
- 6. Sandberg, K., Timmermans, B., Overgaard, M., & Cleeremans, A. (2010). Measuring consciousness: Is one measure better than the other? Consciousness and cognition, 19, 1069-1078

Acknowledgements

Morten Overgaard was supported by a Starting Grant, European Research Council

The minimal self, as distinguished from the narrative self, involves an immediate, embodied, pre-reflective (nonobservational) self-awareness anchored in the first-person perspective. Although this minimal self-awareness is said to have a "thin" phenomenology, it nonetheless includes a number of aspects, such as a sense of ownership (subjectivity, or the mineness of experience) and the sense of agency. Some philosophers maintain that in a strictly circumscribed sense, minimal self-awareness has an epistemological privilege in regard to the subject of experience. This idea is referred to as 'immunity to error through misidentification' (IEM) and derives from Wittgenstein (1958). Shoemaker (1968) explains it in terms of identification-free introspection; others, such as Evans (1982) and Cassam (1995; 2011), suggest that proprioception is IEM. According to this view, I can make a mistake about what I am conscious of (e.g., that my legs are crossed), and about how I am conscious of this (I may think it is by tactile sensation rather than by proprioception), but I cannot be wrong about whose legs they are, or at least, I cannot be wrong about who is experiencing his legs being crossed.

I distinguish between two views of IEM: a conservative view (which I'll call the C-theory) which holds that IEM depends on mode of access (introspection or proprioception), and which leads to the idea that IEM is contingent or de facto (e.g., Campbell 1999; Jeannerod and Pacherie 2004), and a liberal view (the L-theory), which holds that IEM depends on something more basic (the first-person perspective) and is more than a contingent or de facto principle. I argue in favor of the L-theory. The C-theory of IEM is subject to challenges by a number of theorists who cite various pathologies (schizophrenic delusions, somatoparaphrenia) or experiments (the rubber hand illusion, whole body displacement) to show that in some circumstances one can misidentify oneself. I'll briefly review these challenges here and show why they lead to problems with the C-theory but do not affect the L-theory of IEM.

My main focus will be on a more serious challenge associated with a recent unique case of pathology in visual perception reported by Zahn, Talazko and Ebert (2008). This is the case of DP, who is said to experience anonymous vision. Zahn et al. describe this disorder as a selective loss of the sense of self-ownership specifically for visual perception of objects. DP has a sense of ownership in the proprioceptive domain and no problem with the sense of agency for his actions. But DP describes a two-step process that occurs during visual perception. When he first attends to a visual object, he is able to see it, but he does not immediately recognize that he is the one who perceives it; he needs a second step to become aware that he himself is the one who perceives the object. Zahn et al. specifically understand this case as a challenge to IEM because DP's access to his first-person experience is not direct or non-observational. For visual perception DP is sensibly able to ask the Wittgensteinian non-sensible question: "Someone is seeing this object, is it I?" Note that DP's answer to this question is always correct. The problem is that he has to ask the question (or engage in a process of identifying himself) at all.

Zahn et al., following the C-theory, focus on the sense of ownership for the visual experience. But if IEM is tied to something more basic, namely, the first-person perspective, as the L-theory argues, then the question is whether DP's seeing is anchored in his first-person perspective? Lane (2012) suggests that this is just what is at stake in the case of DP, and that when DP sees an object the seeing is not anchored in the first-person perspective. If DP's visual perception truly lacks first-person perspective (or is anonymous) this would violate IEM even on the L-theory. I argue that this is not the case, and is indeed apriori impossible. I suggest that the problem is with DP's reflective mode of access, rather than with the perspectival aspect of perception, and that this explains why DP, in the end, never does misidentify himself in this regard.

- Campbell, J. 1999. Schizophrenia, the space of reasons and thinking as a motor process. The Monist, 82 (4): 609-625.
- Cassam, Q. 1995. Introspection and bodily self-ascription. In: J. L. Bermudez, A. Marcel and N. E. Eilan (eds), The Body and the Self, pp. 311–336. Cambridge, MA: The MIT Press.
- Cassam, Q. 2011. The embodied self. In S. Gallagher (ed.), The Oxford Handbook of the Self (139-156). Oxford: Oxford University Press.
- Evans G. 1982. Varieties of Reference. Oxford: Oxford University Press
- Jeannerod, M. and Pacherie, E. 2004. Agency, simulation, and self-identification. Mind and Language, 19(2): 113-46.
- Lane, T. 2012. Toward an explanatory framework for mental ownership. Phenomenology and the Cognitive Sciences 11(2): 251-286.
- Shoemaker, S. 1968. Self-reference and self-awareness. Journal of Philosophy 65: 555–567.
- Wittgenstein, L. 1958. The Blue and Brown Books, Blackwell
- Zahn, R., J. et al. 2008 Loss of the sense of self-ownership for perceptions of objects in a case of right inferior temporal, parieto-occipital and precentral hypometabolism. Psychopathology 41: 397-402.

We and others have established a paralimbic network of self-reference centered on the anterior cingulate/medial refrontal cortex (ACC) and posterior cingulate/medial parietal cortex (PCC) but also involving subcortical regions. This network of paralimbic regions seem to play a central role in maintaining conscious self-awareness. When the external load on this system is high it may serve a central processing and redistribution role of cortical information. When on the other hand the external load on this system is low it works in a task-negative low-input mode where internal representations provides the information being processed in the default mode network.

Our approach to examining the system was a set of investigations of regional cerebral blood and neurotransmitter nvolvement in changed conscious states of sleep and meditation(1-3). Aspects of these phenomena like the "mind-wandering" during Yoga Nidra meditation or the hypnagogic hallucinations of light sleep let us to try to characterize the dynamics of the underlying networks.

We demonstrated based on scanning data from the meditation study that the ACC and PCC regions, together with striatum, were active independent on mental content (fig 1). This led us to the speculate that these regions were instrumental in assuring a sense of unity of consciousness during widely different tasks, forming a "self" as a reference point to all experiences (4). Further studies of the connectivity revealed a high number of white matter tracts connecting the anterior and posterior cingulate.

The self-specific aspect of this network has been examined by comparing activation patterns of the narrative self to a neutral third person - the Danish Queen(5). Similar activation patterns has been identified with various imaging techniques. The narrative self awareness is associated with more efficient information processing than when third person information is processed and this advantage may be inhibited by transcranial magnetics stimulation inhibition of the prefrontal midline(6).

Recent work based on broadening of these findings finds that dopamine enhances confidence and accuracy in seeing rapidly presented words(7), we speculated that dopamine might contribute to the regulation of self-awareness and its paralimbic network.

The functional extent of the paralimbic network has not yet been fully described, but much evidence points towards a crucial role of the paralimbic network in various processes of conscious processing, especially processing related to extended self-awareness, as opposed to processing related to third person processing. Further research aims at fully describing the role of the first person perspective in the paralimbic network.



Figure 1

Regions involved in the midline network identified with principal component analysis of meditation and control regional cerebral blood flow data.

- 1. Lou HC, Kjaer TW, Friberg L, Wildschiodtz G, Holm S, Nowak M. A 15O-H2O PET study of meditation and the resting state of normal consciousness. Hum Brain Mapp. 1999;7(2):98-105. PubMed PMID: 9950067.
- 2. Kjaer TW, Bertelsen C, Piccini P, Brooks D, Alving J, Lou HC. Increased dopamine tone during meditation-induced change of consciousness. Brain research Cognitive brain research. 2002 Apr;13(2):255-9. PubMed PMID: 11958969.
- 3. Kjaer TW, Law I, Wiltschiotz G, Paulson OB, Madsen PL. Regional cerebral blood flow during light sleep--a H(2)(15)O-PET study. J Sleep Res. 2002 Sep;11(3):201-7. PubMed PMID: 12220315.
- 4. Kjaer TW, Lou HC, editors. Interaction between precuneus and dorsolateral prefrontal cortex may play a unitary role in consciousness: a principal component analysis of rCBF. Conscious Cogn; 2000.
- 5. Kjaer TW, Nowak M, Lou HC. Reflective self-awareness and conscious states: PET evidence for a common midline parietofrontal core. Neuroimage. 2002 Oct;17(2):1080-6. PubMed PMID: 12377180.
- 6. Lou HC, Gross J, Biermann-Ruben K, Kjaer TW, Schnitzler A. Coherence in consciousness: paralimbic gamma synchrony of self-reference links conscious experiences. Hum Brain Mapp. 2010 Feb;31(2):185-92. PubMed PMID: 19621368. Epub 2009/07/22. eng.
- 7. Lou HC, Skewes JC, Thomsen KR, Overgaard M, Lau HC, Mouridsen K, et al. Dopaminergic stimulation enhances confidence and accuracy in seeing rapidly presented words. Journal of vision. 2011;11(2). PubMed PMID: 21346001.

Over the last decade, cognitive neuroscience has started to systematically study the neural mechanisms of social cognition or social information processing. Essentially, two different neural systems have been established in this research domain that appear to constitute two different routes of processing underlying our social cognitive capacities in everyday social encounters, namely the so-called "mirror neuron system" (MNS) and the "social neural network" (SNN, also theory of mind network or mentalizing network). The functional roles of both systems appear to be complementary. The MNS serves comparatively "early" stages of social information processing that are more related to spatial or bodily signals expressed in the behaviour of others and supports the "detection" of potential social salience, including observation of other persons actions. Complementary to the functional role of the MNS, the SNN serves comparatively "late" stages of social information processing that are more related to the "evaluation" of emotional and psychological states of others that have to be inferred as inner mental experience from the behaviour of this person. Empirical studies on the neural mechanisms of ongoing social interactions with others show that essentially SNN components are recruited during the experience of social encounters together with the reward system of the brain.

Biological evolution is a complex process allowing species and organisms to try to resist the second law of thermodynamics (Darwin, 1859). Over time, this has led to the evolution of increasingly sophisticated brains allowing organisms to adapt and survive longer by learning from experience and predicting future events. Yet, this increase in adaptability also leads to an increasingly difficult resource allocation problem of how to best select and learn from competing wants and likes (Kringelbach and Berridge, 2009).

Pleasure can be thought of as evolution's boldest trick in that it serves to motivate an individual to pursue the rewards necessary for survival (Kringelbach, 2005). Evolutionary, the fundamental pleasures are food, sex and conspecifics, as these serve to ensure short and long-term survival both for the individual and the species. During development these fundamental pleasure shape our relationship with self and others and help make the necessary survival-related predictions and decisions when allocating limited brain resources (Berridge and Kringelbach, 2013).

The underlying brain networks are starting to be revealed through animal studies as well as neuroimaging and wholebrain computational studies of the spontaneous dynamics of the brain (Cabral et al., 2014). These studies have shown that the necessary switching between different brain networks for optimizing resource-allocation over a pleasure cycle of appetitive, consummatory and satiety states depends crucially on a set of interrelated reward-related structures (Georgiadis et al., 2012). The brain regions in this pleasure system include the nucleus accumbens, ventral pallidum, anterior cingulate and orbitofrontal cortices.

The choreography of the pleasure system helps to facilitate the state transition between different phases in the pleasure cycle to optimize survival. As shown by the pervasive influence of anhedonia, the lack of pleasure, on affective disorders, the balance of this system can all too easily be disturbed. But as our understanding grow of the underlying reward mechanisms and how they are integrated in whole-brain networks, this can lead to new, more effective interventions. Deep brain stimulation has shown significant potential for rebalancing brain networks in disease (Kringelbach et al., 2011) but the perhaps most promising path is to balance brain networks during development by supporting the crucial parent-infant relationship (Parsons et al., 2013).

- Berridge, K.C., and Kringelbach, M.L. (2013). Neuroscience of affect: brain mechanisms of pleasure and displeasure. Current opinion in neurobiology 23, 294-303.
- Cabral, J., Kringelbach, M.L., and Deco, G. (2014). Exploring the network dynamics underlying brain activity during rest. Prog Neurobiol 114, 102-131.
- Darwin, C. (1859). The origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life, 1st edn (London: John Murray).
- Georgiadis, J.R., Kringelbach, M.L., and Pfaus, J.G. (2012). Sex for fun: a synthesis of human and animal neurobiology. Nature reviews Urology 9, 486-498.
- Kringelbach, M.L. (2005). The orbitofrontal cortex: linking reward to hedonic experience. Nature Reviews Neuroscience 6, 691-702.
- Kringelbach, M.L., and Berridge, K.C. (2009). Towards a functional neuroanatomy of pleasure and happiness. Trends in Cognitive Sciences 13, 479-487.
- Kringelbach, M.L., Green, A.L., and Aziz, T.Z. (2011). Balancing the brain: resting state networks and deep brain stimulation. Frontiers Integrative Neuroscience 5, 8.
- Parsons, C.E., Stark, E.A., Young, K.S., Stein, A., and Kringelbach, M.L. (2013). Understanding the human parental brain: a critical role of the orbitofrontal cortex. Soc Neurosci 8, 525-543.

How does a molecule affect self-reflection? Neuroimaging studies of oxytocin effects on neural correlates of self-referential processing Shihui Han Beijing shan@pku.edu.cn

Social cognition consists of mental processes of both others and oneself. Recent behavioral and neuroimaging research has shown that oxytocin, a neuropeptide that functions as both a neurotransmitter and a hormone, promotes otheroriented cognitive/affective processing. However, whether and does oxytocin affect self-oriented cognitive and neural processes remains unresolved. I'll talk about our recent brain imaging studies that examined whether and how oxytocin modulates the other side of social cognition, i.e., the processing of the self. One of our early EEG study revealed evidence that oxytocin decreased neural activity involved in self-reflection on personality traits. However, the exact neural correlates of self-reflection that undergo oxytocin influences remain unclear. In addition, how oxytocin interacts with self-oriented and other-oriented mindset to shape the brain activity in the medial prefrontal cortex engaged in self-reflection of personality traits. In addition, independent self-construal priming that promotes self-focus interacts with oxytocin that facilitates other-focus to modulate neural activity in the anterior cingulate cortex that is engaged in conflict monitoring and cognitive control. Taken together, our brain imaging findings suggest that self-focus and other-focus mindset provides a platform at which oxytocin interacts with cultural orientations to shape the brain activity involved in self-reflection. Metacognition concerns the evaluation of one's own cognitive processes and underscores the ability to "know that we know". For example, a student who has high confidence that they have learnt enough for an exam may put away the books and stop studying. In contrast, if they believe there are weak spots in their knowledge, they might decide to turn their attention to learning this new material. This "cognitive" aspect of self-awareness is often thought to be a hallmark of the human mind, and, as this example illustrates, is crucial for high-level control of behaviour. But metacognition can also be devastatingly impaired, as in anosagnosia where patients fail to recognize impairments in cognitive function (consider a dementia patient who lacks metacognitive knowledge that her memory is failing).

Despite the centrality of metacognition in human mental life, the neural and computational mechanisms that underpin metacognition, as opposed to primary cognitive functions such as perception and memory, remain poorly understood. Convergent evidence from neuropsychology and functional brain imaging has begun to sketch a model of how metacognition may operate in the human brain. In my talk I will describe what has been learnt from these studies. I start by reviewing seminal neuropsychological work in the 1980s that identified a double dissociation between metacognition of memory and memory performance in subsets of amnesic patients (Janowsky, Shimamura, & Squire, 1989; Shimamura & Squire, 1986). Shimamura and Squire found that only some amnesic patients were aware of their problems with recall. Those who had poor metacognition suffered from Korsakoff's syndrome, a disorder often associated with alcoholism that results in damage to the frontal lobes. Subsequent lesion studies converged to identify a key role for the prefrontal cortex (PFC) in metacognition (Modirrousta & Fellows, 2008; Pannu, Kaszniak, & Rapcsak, 2005; Schnyer et al., 2004; Vilkki, Servo, & Surma-aho, 1998; Vilkki, Surma-aho, & Servo, 1999).

A key obstacle for a true science of metacognition to overcome is the development of quantitative measures of metacognitive ability. In general, for healthy subjects endowed with metacognitive sensitivity, when one is confident, one is more likely to be correct. Thus the degree of association between task performance and confidence can be taken as a quantitative measure of metacognition. However, most early studies use a statistical correlation coefficient or its variant to assess this degree of association, and such measures are susceptible to changes in the subject's performance and response bias. In contrast, recent computational approaches, grounded in signal detection theory (SDT), are able to circumvent these potential confounds and provide a robust, objective measure of metacognitive "efficiency" in the lab (Fleming & Lau, 2014; Galvin, Podd, Drga, & Whitmore, 2003; Maniscalco & Lau, 2012).

In our work we apply these measures in concert with neuroimaging in healthy volunteers to elucidate the neural correlates of metacognitive efficiency. In a typical task we ask subjects to make simple perceptual judgments, such as which of two patches contains a greater number of dots. Following each judgment they provide a confidence rating, allowing us to compute a measure of metacognitive efficiency – how well can each subject discriminate between his own correct or incorrect decisions? We find that individuals with greater grey matter volume and white-matter connectivity in anterior sectors of the prefrontal cortex (aPFC) have greater metacognitive efficiency, despite primary task performance being matched across individuals (Fleming, Weil, Nagy, Dolan, & Rees, 2010). Using functional MRI, we find that the aPFC comes online when subjects are asked to make a metacognitive judgment of their performance, and that its activity is strongest when subjects have low confidence in their decisions (Fleming, Huijgen, & Dolan, 2012). In addition, the relationship between aPFC activity and confidence predicts subjects' metacognitive ability.

This work suggests the aPFC is a key node in supporting metacognitive efficiency, at least for simple perceptual judgments. This is consistent with the role of the aPFC at the top of a cognitive hierarchy, receiving multimodal input from a number of sensory association areas and other prefrontal areas. Intriguingly, the aPFC is disproportionately enlarged in humans compared to other primates (Semendeferi, Armstrong, Schleicher, Zilles, & Van Hoesen, 2001), suggesting that the neural substrates of metacognition may be particularly well established in humans. However, several questions remain unanswered. We do not yet understand the computational contribution of aPFC to metacognition, or how it acts in concert with a network of brain regions that includes other prefrontal areas.

An important open question is whether metacognition is domain-general – do we use a single metacognitive system to reflect on our memories, emotions and decisions? Or is metacognition domain-specific, with specialized subsystems for metacognition in different domains? Early evidence supports a domain-specific model. McCurdy, Lau and colleagues found that grey matter volume in the aPFC of healthy participants predicted individual differences in metacognitive efficiency in a visual discrimination task, whereas grey matter volume in a neuroanatomically distinct region of medial parietal cotex predicted metacognitive accuracy in a recognition memory task (McCurdy et al., 2013). In an experiment building on these findings, we have recently reported evidence for a selective impairment of perceptual metacognition following lesions to the aPFC, despite patients' metacognitive accuracy on an equivalent memory task remaining unimpaired (Fleming, Ryu, Golfinos & Blackmon, under revision).

Drawing together these studies, I will end the talk by sketching a model in which the lateral-anterior PFC interacts with posterior sensory cortices to support metacognition of external, perceptual information, whereas midline structures centered on the precuneus promote metacognition of internal information arising from memory (Baird, Smallwood, Gorgolewski, & Margulies, 2013; Fleming & Dolan, 2012). I will discuss how this framework predicts interplay between self-related processing and metacognition in some domains but not others.

- Baird, B., Smallwood, J., Gorgolewski, K. J., & Margulies, D. S. (2013). Medial and Lateral Networks in Anterior Prefrontal Cortex Support Metacognitive Ability for Memory and Perception. The Journal of Neuroscience, 33(42), 16657–16665.
- Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 367(1594), 1338–1349.
- Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. Frontiers in Human Neuroscience, 8, 443.
- Fleming, S. M., Huijgen, J., & Dolan, R. J. (2012). Prefrontal Contributions to Metacognition in Perceptual Decision Making. The Journal of Neuroscience, 32(18), 6117–6125.
- Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. Science, 329(5998), 1541–1543.
- Galvin, S. J., Podd, J. V., Drga, V., & Whitmore, J. (2003). Type 2 tasks in the theory of signal detectability: discrimination between correct and incorrect decisions. Psychonomic Bulletin & Review, 10(4), 843–876.
- Janowsky, J. S., Shimamura, A. P., & Squire, L. R. (1989). Memory and metamemory: Comparisons between patients with frontal lobe lesions and amnesic patients. Psychobiology, 17(1), 3–11.
- Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. Consciousness and Cognition, 21(1), 422–430.
- McCurdy, L. Y., Maniscalco, B., Metcalfe, J., Liu, K. Y., de Lange, F. P., & Lau, H. (2013). Anatomical coupling between distinct metacognitive systems for memory and visual perception. The Journal of Neuroscience, 33(5), 1897–1906.
- Modirrousta, M., & Fellows, L. K. (2008). Medial prefrontal cortex plays a critical and selective role in "feeling of knowing" meta-memory judgments. Neuropsychologia, 46(12), 2958–2965.
- Pannu, J., Kaszniak, A., & Rapcsak, S. (2005). Metamemory for faces following frontal lobe damage. Journal of the International Neuropsychological Society, 11(6), 668–676.
- Schnyer, D. M., Verfaellie, M., Alexander, M. P., LaFleche, G., Nicholls, L., & Kaszniak, A. W. (2004). A role for right medial prefontal cortex in accurate feeling-of-knowing judgements: evidence from patients with lesions to frontal cortex. Neuropsychologia, 42(7), 957–966.
- Semendeferi, K., Armstrong, E., Schleicher, A., Zilles, K., & Van Hoesen, G. W. (2001). Prefrontal cortex in humans and apes: A comparative study of area 10. American Journal of Physical Anthropology, 114(3), 224–241.
- Shimamura, A. P., & Squire, L. R. (1986). Memory and metamemory: a study of the feeling-of-knowing phenomenon in amnesic patients. Journal of Experimental Psychology. Learning, Memory, and Cognition, 12(3), 452–460.
- Vilkki, J., Servo, A., & Surma-aho, O. (1998). Word list learning and prediction of recall after frontal lobe lesions. Neuropsychology, 12(2), 268. doi:10.1037/0894-4105.12.2.268
- Vilkki, J., Surma-aho, O., & Servo, A. (1999). Inaccurate prediction of retrieval in a face matrix learning task after right frontal lobe lesions. Neuropsychology, 13(2), 298. doi:10.1037/0894-4105.13.2.298

The posteromedial region of the parietal cortex, alternatively referred to as the precuneus or the mesial extent of Brodmann's area 7, is located in the depth of the interhemispheric fissure, an anatomical location which makes this territory particularly difficult to study. Because of its hidden location, as well as the relative rarity of focal lesions selectively affecting it, the precuneus has traditionally remained one of the less accurately mapped areas of the whole cortical surface. However its strategic location and widespread connections suggest that this area might play an important role in a variety of behavioural functions. The modern era of neuroimaging has recently made it possible to explore the anatomical and functional aspects of this long-neglected part of the brain. However, recent functional imaging findings in healthy subjects suggest a central role for the precuneus in a wide spectrum of highly integrated tasks, including visuo-spatial imagery, episodic memory retrieval and self-processing operations, namely first-person perspective taking and experience of agency. Furthermore, the precuneus and surrounding posteromedial areas are among the brain structures displaying the highest resting metabolic rates ('hot spots'). According to the 'default mode of brain function' model, these areas are characterized by transient decreases in their tonic level of activation during engagement in a wide range of goal-directed actions which share the only feature of not being selfreferential ('task-independent decreases'). Therefore, it has recently been proposed that precuneus is involved in the interwoven network of the neural correlates of self-consciousness, which are constantly engaged in self-related mental representations during rest. Specifically, the default mode network seems to subserve tasks requiring mental simulation of alternative perspectives or imagined scenes, which are commonly engaged during rest. These range from autobiographical memory, to envisioning the future, theory of mind, and moral decision making. This hypothesis on the functional role of the precuneus as a pivotal node within the default mode network is also consistent with the selective hypometabolism in the posteromedial cortex reported in a wide range of altered conscious states, such as non-REM sleep, drug-induced anaesthesia and vegetative state. This model of the role of the precuneus within the neural correlates of consciousness and self-awareness is in line with the current knowledge about the macroscopic and microscopic anatomy of this brain region, together with its widespread connectivity with both cortical and subcortical structures. The extensive network of precuneus connections, first suggested by connectivity and neurophysiological findings in non-human primates, links these notions with the multifaceted spectrum of its behavioural correlates. A critical analysis of the precuneus activation patterns in response to different mental tasks provides a useful conceptual framework for matching the functional imaging findings with the specific roles played by this structure in the higherorder cognitive functions in which it has been implicated. The implications of these functional and anatomical observations suggest possible adaptive roles of the default network for using past experiences to plan for the future, navigate social interactions, and maximize the utility of moments when we are not otherwise engaged by the external world. The relevance of the precuneus and interlinked regions of the default network for understanding a wide range of neuropsychiatric disorders involving alterations of consciousness, including autism, schizophrenia and Alzheimer's disease, cannot be underestimated. The study of the transient and reversible alterations of consciousness reported by patients with epilepsy provide useful insights into the brain mechanisms subserving consciousness, with deactivation of the precuneus and fronto-parietal association areas in both primary and secondarily generalized seizures characterized by complete loss of consciousness. Finally, studies conducted over the last few years show activation patterns which appear to converge with anatomical and connectivity data in providing preliminary evidence for functional subdivisions within the posteromedial parietal cortex.

- Buckner RL, Andrews-Hanna JR, Schacter DL. The brain's default network: anatomy, function, and relevance to disease. Annals of the New York Academy of Sciences 2008;1124:1-38.
- Buckwalter JA, Parvizi J, Morecraft RJ, van Hoesen GW. Thalamic projections to the posteromedial cortex in the macaque. Journal of Comparative Neurology 2008;507:1709-1733.
- Cauda F, Geminiani G, D'Agata F, Sacco K, Duca S, Bagshaw AP, Cavanna AE. Functional connectivity of the Posteromedial Cortex. PLoS One 2010;5:e13107.
- Cavanna AE, Bertero L, Cavanna S. The functional neuroimaging of the precuneus. Neuroscience Imaging 2008;2:161-175.
- Cavanna AE, Trimble MR. The precuneus: A review of its functional anatomy and behavioural correlates. Brain 2006;129:564-583.
- Cavanna AE. The precuneus and consciousness. CNS Spectrums 2007;12:545-552.
- Parvizi J, Van Hoesen GW, Buckwalter J, Damasio A. Neural connections of the posteromedial cortex in the macaque. Proceedings of the National Academy of Sciences USA 2006;103:1563-1568.
- Trimble MR, Cavanna AE. The role of the precuneus in episodic memory. In: Dere E, Easton A, Nadel L, Huston JP (Eds) Handbook of Behavioural Neuroscience: Episodic Memory Research. Amsterdam: Elsevier Science 2008, p. 363-377.
- Vogt BA, Laureys S. Posterior cingulate, precuneal and retrosplenial cortices: cytology and components of the neural network correlates of consciousness. Progress in Brain Research 2005;150:205-217.

Early observations of increased intracranial pressure during brain activation in humans¹ and animals² lead to the realization that brain work causes intracranial blood vessel to dilate, and thereby blood flow to increase, in order to support increased metabolic demands. Several neurovascular coupling mechanisms are now known to converge on cerebral arterioles to adjust cerebral blood flow (CBF) according to the metabolic needs of the tissue, and methods to measure regional CBF and brain oxygenation during rest and task activation have revolutionized our knowledge of brain function and its localization^{3,4}.

Neurovascular coupling mechanisms are likely to reflect the brain's vulnerability to even short periods of oxygen depletion - and one might expect that CBF is tightly regulated to meet the metabolic needs of brain tissue. Nevertheless, mounting evidence suggest that CBF changes seemingly fails to match the brain's metabolic demands during task activation in normal brain, and in particular in disease^{5,6}.

A recent re-analysis of the relation between blood flow and tissue oxygenation reveals that, while CBF determines the immediate availability of oxygen in the tissue, the capillary distribution of blood across the capillary bed determines the extent to which this oxygen can be taken up by the tissue, and thus be used to support brain work⁷. This biophysical property implies that brain regions may cover their metabolic demands by signaling to nearby contractile, capillary pericytes, in addition to arterioles upstream. Indeed, it was recently demonstrated that neuronal activity relaxes pericytes, and that pericyte dilation occur prior to upstream arteriolar dilation⁸.

This neuro-capillary coupling mechanism may have profound implications for our interpretation of functional brain imaging results. While the neurovascular coupling paradigm associates increased neuronal activity with increased CBF and blood oxygen level dependent (BOLD) signal amplitude, active control of oxygen extraction by pericyte dilation would be expected to cause a reduction in these signals. As a result, activation of neurons which cover their metabolic demands entirely by regulating capillary flow patterns is predicted to result in negative BOLD signal intensity changes, but no changes in CBF. For population of neurons which relax both arteriolar and pericyte tone, the amplitude of brain imaging signal is thus predicted to reflect not only neuronal activity, but also the coupling of neuronal activity to arteriolar and capillary tone.

The talk will describe the neurocapillary coupling mechanism and discuss its implications in our quest to infer brain activity from resting state and activation-related neuroimaging signals in health and disease.

- 1. Mosso A. Über den Kreislauf des Blutes im Menschlichen Gehirn. Leipzig: Verlag von Veit und Comp.; 1881.
- 2. Roy CS, Sherrington CS. On the Regulation of the Blood-supply of the Brain. J Physiol (London). 1880;11:85-108.
- 3. Raichle ME, Mintun MA. Brain work and brain imaging. Annu Rev Neurosci. 2006;29:449-476.
- 4. Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL. A default mode of brain function. Proc Natl Acad Sci U S A. 2001;98:676-682.
- 5. Blicher JU, Stagg CJ, O'Shea J, Ostergaard L, MacIntosh BJ, Johansen-Berg H, et al. Visualization of altered neurovascular coupling in chronic stroke patients using multimodal functional MRI. J Cereb Blood Flow Metab. 2012;32:2044-2054.
- Donahue MJ, Stevens RD, de Boorder M, Pekar JJ, Hendrikse J, van Zijl PC. Hemodynamic changes after visual stimulation and breath holding provide evidence for an uncoupling of cerebral blood flow and volume from oxygen metabolism. J Cereb Blood Flow Metab. 2009;29:176-185.
- 7. Jespersen SN, Østergaard L. The Roles of Cerebral Blood Flow, Capillary Transit Time Heterogeneity and Oxygen Tension in Brain Oxygenation and Metabolism. J Cereb Blood Flow Metab. 2012;32:264-277.
- 8. Hall CN, Reynell C, Gesslein B, Hamilton NB, Mishra A, Sutherland BA, et al. Capillary pericytes regulate cerebral blood flow in health and disease. Nature. 2014.

Ever since Hans Berger first observed rhythmic variations in the human electroencephalogram, around 1929, scientists have been fascinated by the question of whether brain oscillations play a causal role in human behaviour. Over the years, a multitude of studies, supported by significantly improved recording and analysis techniques, have accumulated a wealth of evidence that modulations in the amplitude of specific brain oscillations occur consistently in relation to specific cognitive tasks. Only very rarely, however, has this evidence gone beyond the level of a correlation to establish a clear causal link between brain oscillations and behaviour. Do these oscillations support specific processes, or are they mere epiphenomena? And if the former is true, precisely which computational processes do they implement? How can we make sense out of the myriad of tiny changes in neuronal activity that can be recorded in humans at the neuronal population level through scalp encephalography?

Here, I will review our recent progress in addressing these fundamental questions.

One important characteristic of oscillations is their periodicity — the organisation of their temporal dynamics into cycles. Each time point within a cycle is uniquely defined by its phase. Oscillations become potentially powerful computational tools when their phase (but also frequency and amplitude) changes due to dynamic modifications in the generating system or in the input. In the following, I present evidence for a possible mechanisms through which oscillatory phase can affect behaviour. This mechanism relies on the fact that the oscillatory cycle establishes a recurrent temporal reference frame that allows for the coding of temporal relations between groups of neural elements and between neural elements and the environment. Importantly, this reference frame is not fixed but is subject to dynamic changes (phase resetting).

I will present four studies that demonstrate the importance of phase resetting for establishing a temporal reference frame.

The first study shows how sensory stimulation in the auditory domain can lead to cross-modal phase resetting in the visual domain in a way that leads to rhythmic modulations of neural excitability. This is a possible mechanism for cross-modal integration.

The second study investigates the interaction between rhythmic components in continuous speech and brain oscillations.

Continuous speech is organized into a nested hierarchy of quasi-rhythmic components (prosody, syllables, phonemes) with different time scales. Interestingly, neural activity in the human auditory cortex shows rhythmic modulations with frequencies that match these speech rhythms. Here, we used magnetoencephalography and information theory to study brain oscillations in participants as they process continuous speech. We show that auditory brain oscillations at different frequencies align with the rhythmic structure of speech. This alignment is more precise when participants listen to intelligible rather than unintelligible speech. The onset of speech resets brain oscillations and improves their alignment to speech rhythms; it also improves the alignment between the different frequencies of nested brain oscillations in the auditory cortex. Since these brain oscillations reflect rhythmic changes in neural excitability, they are strong candidates for mediating the segmentation of continuous speech at different time scales corresponding to key speech components such as syllables and phonemes.

Third, I will present a study that demonstrates that detailed sensory information can be coded in the phase of neural oscillations after phase resetting. In a biologically relevant cognitive task, we instructed six human observers to categorize facial expressions of emotion while we measured the observers' EEG. We combined state-of-the-art stimulus control with statistical information theory analysis to quantify how the three parameters of oscillations (i.e., power,

phase, and frequency) code the visual information relevant for behavior in a cognitive task. We make three points: First, we demonstrate that phase codes considerably more information (2.4 times) relating to the cognitive task than power. Second, we show that the conjunction of power and phase coding reflects detailed visual features relevant for behavioral response—that is, features of facial expressions predicted by behavior. Third, we demonstrate, in analogy to communication technology, that oscillatory frequencies in the brain multiplex the coding of visual features, increasing coding capacity.

Finally, I will aim at relating the more general computational properties of neural oscillations with the specific topic of this conference - self-awareness. I will try to do that using recent data from an ongoing study of sensory attenuation. Sensory attenuation refers to the phenomenon that self-generated stimuli elicit a weaker sensory evoked brain response compared to externally generated stimuli. However, little is known about the neural mechanism underlying the effect. We hypothesised that alpha/beta oscillations mediate the sensory attenuation effect by modulating excitability in sensory areas. In this study, we compared the neural responses under both conditions for auditory stimuli with MEG in 14 participants. We replicated the reduction of M100 component in the auditory cortex for self-generated stimuli. Following source localisation we performed time-frequency analysis and demonstrated that reduction of M100 is due to a decrease in both power and phase locking in a broad frequency band (3-40Hz). Related to our hypothesis, we found both increased power and phase locking in alpha and beta band in a 400ms window prior to the onset of self-generated stimuli compared to the onset of externally generated stimuli. A decrease in power (beta) and phase locking in line with previous reports suggesting an inhibitory role of alpha/beta oscillations. They also suggest multifaceted mechanisms behind sensory attenuation. Our results indicate that oscillatory activity in sensory areas is controlled by brain mechanisms that predict sensory consequences of actions.

In summary, recent evidence points to an important role of brain oscillations in establishing a temporal reference frame and gating information flow in the human brain.

- Gross J, Hoogenboom N, Thut G, Schyns P, Panzeri S, et al (2013) Speech Rhythms and Multiplexed Oscillatory Sensory Coding in the Human Brain. PLoS Biol 11: e1001752.
- Romei V, Gross J, Thut G (2012) Sounds Reset Rhythms of Visual Cortex and Corresponding Human Visual Perception. Curr Biol .
- Thut G, Miniussi C, Gross J (2012) The functional importance of rhythmic activity in the brain. Curr Biol 22: R658-R663.
- Schyns, Thut, Gross (2011) Cracking the Code of Oscillatory Activity. PLoS Biol 9: e1001064.

From phenomenological and experimental perspectives, research in schizophrenia has emphasized deficits in "higher" cognitive functions, including attention, executive function, as well as memory. In contrast, general consensus has viewed dysfunctions in basic perceptual processes to be relatively unimportant in the explanation of more complex aspects of the disorder, including changes in self-experience and the development of symptoms such as delusions.

In the lecture, I will summarize our recent work with Magnetoencephalography (MEG) which has examined the role of neural oscillations and event-related fields (ERFs) in sensory processing in schizophrenia. These results highlight a pronounced impairment in high-frequency activity in both chronic and unmedicated patients as well as the potential contribution of impaired prediction processes as revealed by the analysis of the magnetic mismatch negativity field (MMF) towards perceptual impairments in the disorder. The pattern of dysfunctional gamma-band activity and aberrant ERF-responses in schizophrenia are consistent with the effects of ketamine in healthy volunteers, highlighting the central role of aberrant NMDA-receptor functioning for the understanding of abnormal circuit functioning in schizophrenia.

In addition to investigations into the adult cortex, our recent work has employed MEG to understand the developmental trajectory of neural oscillations during adolescence and the possibility to develop a biomarker for early detection and diagnosis of ScZ. We found marked changes in the amplitude of high-frequency oscillations and synchrony that were particularly pronounced during the transition from adolescence to adulthood. These findings may be relevant for the understanding of schizophrenia as a disorder of late brain maturation with implications for early intervention and diagnosis.

I will describe experiments exploring the relationship between prefrontal interneurons, gamma oscillations, and cognition in mice. Fast-spiking interneurons (FSINs), a class of GABAergic interneurons that can be specifically identified based on either their electrophysiological properties or their expression of the calcium-binding protein parvalbumin (PV) (1, 2), generate synchronized gamma frequency (~30-120 Hz) oscillations that are hypothesized to enhance cortical information processing (3, 4). Both prefrontal FSINs (5, 6) and gamma oscillations (7-14) are abnormal in schizophrenia. This has led to hypothesis that prefrontal interneuron dysfunction leads to deficiencies in task-evoked gamma oscillations that cause cognitive deficits which represent the core of schizophrenia (6, 15). However, until now, it has not been possible to demonstrate a causal role for interneuron-driven gamma oscillations in cognition, and the hypothesized link between prefrontal interneuron dysfunction, abnormal gamma oscillations, and cognitive deficits in schizophrenia has remained tentative.

To address these questions, we have studied $Dlx5/6^{+/-}$ mice, which are deficient in transcription factors necessary for normal FSIN development (16). We find that in $Dlx5/6^{+/-}$ mice, the post-adolescent maturation of FSINs is impaired. Furthermore, this coincides with the post-adolescent appearance of behavioral, cognitive, and electrophysiological endophenotypes of schizophrenia. Specifically, we studied cognition using a task in which mice must choose between two bowls to find a food reward. Each bowl is filled with a different textured digging medium and scented with a different odor, such that one cue (either an odor or digging medium) indicates the presence of food reward. Once mice learn an initial rule, they are tested on their ability to learn a new rule. $Dlx5/6^{+/-}$ mice perform normally when learning an initial rule, or a new rule that is the reverse of a previously learned one. However, $Dlx5/6^{+/-}$ mice are markedly and selectively impaired when required to learn a new rule in which a cue that was previously irrelevant to the outcome of each trial becomes associated with food reward – e.g. when required to switch from an odor based rule to a texture based one or vice-versa. Moreover, $Dlx5/6^{+/-}$ mice exhibit deficient gamma oscillations during this rule switching task as well as during other PFC-dependent cognitive tasks, e.g. social exploration. Interestingly, they also exhibit elevated gamma oscillations at baseline, matching the complex pattern of changes (decreased task-evoked gamma oscillations but increased baseline gamma oscillations) observed in schizophrenia.

Consistent with a critical role for interneuron-driven gamma oscillations in cognitive flexibility, restoring these oscillations by optogenetically stimulating PFC interneurons rescues cognitive deficits in $Dlx5/6^{+/-}$ mice. Moreover, these pro-cognitive effects are both frequency-specific and long-lasting. $Dlx5/6^{+/-}$ mice that receive gamma-frequency (40 Hz) optogenetic stimulation on one task day are able to perform rule switching normally up to one week later, whereas stimulation using a combination of higher and lower frequencies fails to rescue cognition, even partially. Conversely, optogenetically inhibiting prefrontal interneurons is sufficient to induce similar rule switching deficits in wild-type mice.

These results illustrate how the post-adolescent arrest of FSIN maturation could drive key cognitive and electrophysiologic endophenotypes of schizophrenia. They also demonstrate, for the first time, a causal relationship between gamma frequency activity in prefrontal interneurons and PFC-dependent cognition, such that the acute restoration of interneuron-driven gamma rhythms may represent a novel therapeutic strategy for cognitive enhancement in schizophrenia and related disorders. In addition, the fact that $Dlx5/6^{+/-}$ mice are selectively impaired when learning rules based on previously irrelevant cues (but not when learning reversals of previously learned rules) suggests that prefrontal interneuron-driven gamma oscillations may be particularly critical for shifting attention between sets of cues. Furthermore, optogenetically inducing prefrontal gamma oscillations may lead to a "eureka" moment, in which mice learn the requirements of the task, leading to long-lasting improvements in their performance.

- 1. Kawaguchi Y, Kubota Y. Correlation of physiological subgroupings of nonpyramidal cells with parvalbumin- and calbindinD28kimmunoreactive neurons in layer V of rat frontal cortex. J Neurophysiol. 1993;70(1):387-96.
- 2. Cauli B, Audinat E, Lambolez B, Angulo MC, Ropert N, Tsuzuki K, et al. Molecular and physiological diversity of cortical nonpyramidal cells. J Neurosci. 1997;17(10):3894-906.
- 3. Cardin JA, Carlen M, Meletis K, Knoblich U, Zhang F, Deisseroth K, et al. Driving fast-spiking cells induces gamma rhythm and controls sensory responses. Nature. 2009;459(7247):663-7.
- 4. Sohal VS, Zhang F, Yizhar O, Deisseroth K. Parvalbumin neurons and gamma rhythms enhance cortical circuit performance. Nature. 2009;459(7247):698-702.
- 5. Benes FM, Berretta S. GABAergic interneurons: implications for understanding schizophrenia and bipolar disorder. Neuropsychopharmacology. 2001;25(1):1-27.
- 6. Lewis DA, Hashimoto T, Volk DW. Cortical inhibitory neurons and schizophrenia. Nat Rev Neurosci. 2005;6(4):312-24.
- 7. Spencer KM, Nestor PG, Perlmutter R, Niznikiewicz MA, Klump MC, Frumin M, et al. Neural synchrony indexes disordered perception and cognition in schizophrenia. Proc Natl Acad Sci U S A. 2004;101(49):17288-93. PMCID: 535363.
- 8. Gallinat J, Winterer G, Herrmann CS, Senkowski D. Reduced oscillatory gamma-band responses in unmedicated schizophrenic patients indicate impaired frontal network processing. Clin Neurophysiol. 2004;115(8):1863-74.
- 9. Symond MP, Harris AW, Gordon E, Williams LM. "Gamma synchrony" in first-episode schizophrenia: a disorder of temporal connectivity? Am J Psychiatry. 2005;162(3):459-65.
- 10. Wynn JK, Light GA, Breitmeyer B, Nuechterlein KH, Green MF. Event-related gamma activity in schizophrenia patients during a visual backward-masking task. Am J Psychiatry. 2005;162(12):2330-6.
- 11. Cho RY, Konecky RO, Carter CS. Impairments in frontal cortical gamma synchrony and cognitive control in schizophrenia. Proc Natl Acad Sci U S A. 2006;103(52):19878-83. PMCID: 1750867.
- 12. Ford JM, Roach BJ, Faustman WO, Mathalon DH. Synch before you speak: auditory hallucinations in schizophrenia. Am J Psychiatry. 2007;164(3):458-66.
- 13. Ford JM, Roach BJ, Faustman WO, Mathalon DH. Out-of-synch and out-of-sorts: dysfunction of motor-sensory communication in schizophrenia. Biol Psychiatry. 2008;63(8):736-43. PMCID: 2330266.
- 14. Minzenberg MJ, Firl AJ, Yoon JH, Gomes GC, Reinking C, Carter CS. Gamma Oscillatory Power is Impaired During Cognitive Control Independent of Medication Status in First-Episode Schizophrenia. Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology. 2010.
- 15. Uhlhaas PJ, Singer W. Abnormal neural oscillations and synchrony in schizophrenia. Nature reviews Neuroscience. 2010;11(2):100-13.
- 16. Wang Y, Dye CA, Sohal V, Long JE, Estrada RC, Roztocil T, et al. Dlx5 and Dlx6 regulate the development of parvalbumin-expressing cortical interneurons. J Neurosci. 2010;30(15):5334-45. PMCID: 2919857.

There is considerable research on the neural correlates of self-awareness which was boosted by the emergence of the neuroimaging revolution approximately two decades ago. Initial enthusiasm in terms of these investigations was merited as questions previously unanswerable were soon addressed and at least partially answered. Understanding the neural correlates of self-awareness was, and remains, a scientifically perfect question as it bridges philosophy, psychology, religion, and neuroscience. Ultimately, self-awareness is a trait, no different than any other. While there is a temptation to elevate it as a non-biological entity, we have discovered that its existence is best viewed as an inherited adaption with benefits outweighing costs in certain populations. Psychology, philosophy, religion, and even neurology often fail to account for evolutionary factors and in terms of self-awareness research, such failures have resulted in an incomplete understanding of the neurological underpinnings. It is suggested that blindness to selective principles is often seen in traits that are expressed in a unique manner in humans as if they exist outside the normal laws of biology. It is argued here in no uncertain terms that self-awareness provides significant benefits in apes, in particular humans, and that such benefits are translated into a reproductive advantage.

Specifically, both the enlargement of the cortex and the increased lateralization of the brain in homo sapien is a direct result of self-awareness and its relationship to self-deception and language. It is suggested that critical to evolution is self-deception and that those organisms that engaged in self-deception secured a significant advantage over those that did not. We further suggest that self-deception encouraged both the increase of neural underpinnings and its psychological correlates including verbal language. This theory is based on research that links self-deception to self-awareness and to similar neural underpinnings. Furthermore, the evidence suggests that Homo sapiens, unlike Pan paniscus or troglodyte have had a mating and social system that encouraged self-deception which aided in the growth of cerebral regions that are unique to humans. In terms of molecular genetics, little is known in terms of self-awareness. It appears that structural differences between humans and other great apes. Variants in allelic frequency would be unlikely to account for the differences between humans and other great apes. While purely speculative, we predict that expression of particular genes form the behavioral differences observed, likely due to the underlying neuronal anatomical differences created by such gene expression. Put another way, transcription factors are likely differentiating humans from apes and focus should be placed on next generational sequencing and RNA. It is suggested that subtle differences in DNA exist, but the major difference in expression underlies the persistent finding that differences in degree rather than kind exist between the species.

Critical to this theory is presenting a case that self-awareness a) is linked to self-deception and b) self-deception provides a significant advantage. The temptation is to simply demonstrate that humans have 'big brains' and we lie to ourselves all the time. The establishment of these links is difficult in a scientific manner, but we believe there is enough evidence to bring us beyond the suggestive level. From the early work of Trivers to recent findings that clinical depression may be somewhat thwarted by a healthy false ego, a case could be made that the human psyche is built heavily on an unreliable story teller. The work of Libet and his followers is clearly relevant as it suggests that humans clearly exist in a world in which intentionality is entrenched in non-reality. Starting with social psychology and some studies from the late 1920s, it has been consistently found that self-appraisals are horribly inaccurate and that even before, with the work of psychoanalysis, there were suggestions that the entirety of the human mind is false.

Therefore, while initially odd, the suggestion that self-awareness is based evolutionarily on deception appears consistent with centuries of research. The link that self-deception is beneficial is equally valid. This is not to suggest that other correlates of self-awareness (e.g., theory of mind) are not critical. Instead, we imagine that without self-deception, theory of mind would not be possible.

There are major gaps in our evolutionary understanding of almost everything related to humans, and a superficial reminder that most behaviors don't fossilize is only a reminder of the work that remains. It still remains unknown which ribonucleic sequences correlate with self-awareness. It remains unknown when humans became self-aware, or even how best to test self-awareness. Furthermore, it remains unclear, and virtually unknown, if and how other apes self-deceive. Basic questions related to the brain remain, such as which neuronal sequences are critical for producing the sense of the self.

What the neuroimaging revolution provided, however, was the bravado to ignore the gaps and to ask bold questions. These gaps remain, as they did 20 years ago, but they have narrowed and they have become less daunting. That is, it is speculated that fMRI and TMS allowed us to discover many truths that did not come directly from the data, but rather came from us asking, finally, the right questions. The inevitable collision course that psychology, philosophy, and religion are on with functional genetics will forever change the course of these disciplines as only hinted at by the neuroimaging revolution. Solving the mysteries of self-awareness starts, in my opinion, by treating it no differently than any other biological entity, and as such, an inherited trait that provided a reproductive advantage.

- The elements of a scientific theory of self-deception. Trivers R. Ann N Y Acad Sci. 2000 Apr; 907:114-31.
- A neural correlate of consciousness related to repression. Shevrin H, Ghannam JH, Libet B. Conscious Cogn. 2002 Jun;11(2):334-41.
- Overclaiming and the medial prefrontal cortex: A transcranial magnetic stimulation study. Amati F, Oh H, Kwan VS, Jordan K, Keenan JP. Cogn Neurosci. 2010 Dec;1(4):268-76.
- Evolutionarily dynamic alternative splicing of GPR56 regulates regional cerebral cortical patterning. Bae BI, Tietjen I, Atabay KD, Evrony GD, Johnson MB, Asare E, Wang PP, Murayama AY, Im K, Lisgo SN, Overman L, Šestan N, Chang BS, Barkovich AJ, Grant PE, Topçu M, Politsky J, Okano H, Piao X, Walsh CA.
- The mental self. Lou HC, Nowak M, Kjaer TW. Prog Brain Res. 2005;150:197-204

The concepts of self-awareness is a loosely defined entity and self-reports and introspection based discussions represents a large part of the research. Placebo manipulations provide a tool whereby these subjective reports may be manipulated in a systematic fashion and thereby allow for a deconstruction on some of the factors influencing the subjective report. Self-awareness and and the description of e.g. state of mind has three major modifiers: exteroception, interception and top-down regulation. Shifts in self-awareness and self-regulation induced by systematic manipulations of all three factors will be discussed as to illustrate the volatility of the selfawareness.

In a series of experiments we have manipulated the perception for nociceptive input and demonstrated how contextual mechanisms influence both the neurophysiological evoked responses to pain as well as the self report on severity of a nociceptive input. In large these types of experiments allow for an interpretation whereby the subjective report of a pain provocation is influenced by a manipulation of top down mechanisms.

In a recent experiment we have extended this type of experimentation towards the un-aware domain. We used a backward masking paradigm in which stimuli can be presented for conditioning outside the conscious awareness of the subject. By associating a face with a stronger stimulus than expected and vice versa with a another face it was possible to adjust the response to a nociceptive input by conditioning towards adverse stimuli irrespective if the subject was aware or not about the conditioned stimulus. This illustrates the notion that a final conscious experience represents a final composite of many sources that may be influenced by many mechanisms. The role of the insula, lateral orbitofrontal cortex and adjacent regions as neurophysiological correlates for several qualities of self-awareness/ conscious experience will be discussed.

Hallucinogens

Hallucinogens are a group of compounds that can be divided into three categories: psychedelics (alter cognition and perception), deliriants (causing delirium), and dissociatives (feelings of detachment). The common feature of these psychoactive drugs is that they can cause subjective changes in perception, thought, emotion and consciousness. Unlike other psychoactive drugs, they induce experiences that are qualitatively different from those of ordinary consciousness. They do not induce significant changes in intellectual or memory capabilities, autonomic nervous system side effects are minimal, and addictive craving is absent.

Classical psychedelics are serotonin 2A receptors (5-HT2AR) agonists and they include LSD, psilocybin ("magic mushrooms"), mescaline (the active constituent of peyote), and DMT (the active constituent of ayahuasca). Apart from naturally occurring psychedelic compounds, a number of synthetic 5-HT2AR agonists have evolved (Ettrup et al., 2011). The 5-HT2AR is an excitatory postsynaptic receptor that has been implicated in the regulation of mood and consciousness (Nichols, 2004). Stimulation of the 5-HT2AR via administration of a 5-HT2AR agonist is known to produce marked changes in consciousness, such as are seen with the psychedelic drugs LSD, psilocybin and mescaline (Glennon et al., 1984). 5-HT2AR stimulation with, e.g., psilocybin has been shown to result not only in profound changes in consciousness but also in lasting improvements in well-being (Griffiths et al., 2011). Some describe a psychedelic drugs for the 5-HT2AR is positively correlated with their hallucinogenic potency. LSD is the most potent psychedelic, producing its characteristic subjective effects in doses as low as 20-50 µg. Psilocybin 2 mg IV is well-tolerated and produces typical psychedelic effects that begin within seconds of infusion (Glennon et al., 1984) and pre-treatment with ketanserin, a non-selective 5-HT2AR antagonist, markedly reduces the major psychedelic effects of psilocybin in humans (Vollenweider et al., 1998). Likewise, treatment with the 5-HT2AR inverse agonist pimavanserin has been found effective for treatment of levo-dopa induced psychosis, particularly hallucinatory delusions (Meltzer et al., 2010).

Neuroimaging studies of hallucinogen effects

Hallucinogens have a considerable potential as a pharmacological tool to study consciousness, but there are only a few neuroimaging studies that investigate their exact effects on the brain. Like research on psychedelics (Sessa, 2012), human consciousness research is currently in the ascendancy. Psychedelics are characterized by markedly perturbing consciousness without compromising arousal (Muthukumaraswamy et al., 2013) and thus offer a unique opportunity to study aspects of consciousness without being confounded by changes in arousal. Recent neuroimaging studies with the classic psychedelic drug, psilocybin, have given rise to the suggestion that when in the psychedelic state, the brain exhibits more characteristics of criticality in the psychedelic state than are apparent during normal waking consciousness (Carhart-Harris et al., 2014). The clinical effects of psilocybin 2 mg iv are associated with significant decreases in cerebral blood flow (CBF) in the medial prefrontal cortex (mPFC), consistent with the mPFC being hyperactive in depression, as reviewed in (Jaworska et al., 2014). Also brain areas such as the posterior cingulate cortex (PCC) shows decreased CBF in response to psilocybin (Carhart-Harris et al., 2012). The PCC has attracted particular interest due to its importance as a cortical connector hub (Hagmann et al., 2008) and integration center (Leech et al., 2012). In a recent case study of an epilepsy patient, it was reported that direct electrical stimulation of PCC lead to a transient behavioral unresponsiveness with loss of external connectedness and the patient reported to have had a dreamlike experience (Herbet et al., 2014).

Based on extensive medicinal and radiochemistry work, we have developed the first 5-HT2AR agonist PET radioligand, 11C-Cimbi36, for in vivo brain imaging of humans (Ettrup et al., 2014). Anecdotal reports from users suggest Cimbi36 (also called 25B-NBOMe) to be an active hallucinogen at a dose of as little as 250-500 μ g (Wikipedia), making it a similar potency to other phenethylamine derived hallucinogens such as bromo-dragonfly. Duration of effects lasts about 12–16 hours. However, when administered in radiolabeled form for brain imaging, the maximum doses given are about 1.5 μ g and thus without any subjective effects (Ettrup et al., 2014). With the recent technical development of simultaneous acquisition of PET and MRI neuroimaging data, one is now in a position to study the effects of psychedelics to a much greater extent.



5-HT2AR agonist radioligand 11C-Cimbi36 distribution in the brain of a healthy volunteer, before (left) and after blocking with the 5-HT2AR antagonist ketanserin. Modified from (Ettrup et al., 2014).

The therapeutic use of hallucinogens

Canadian researchers Humphry Osmond, Abram Hoffer, and Alfred M. Hubbard developed psychedelic therapy in the mid to late 1950s. They made use of the powerful psychedelic experience to induce changes in the patients attitude, perspective, and behavior. The treatment schedule consisted of an intensive course of psychotherapy, leading to a single high-dose LSD session (Oram, 2014). By the start of the 1960'ties it was a well-established therapy and it has been estimated that more than 40,000 patients were treated with psychedelics in the 1950s and 60s; most of these received LSD. LSD was most commonly used in the treatment of mood disorders and addiction and a recent meta-analysis of the use of LSD in the treatment of alcohol dependence found evidence of an impressive efficacy (Krebs and Johansen, 2012). Depression is a another major public health problem of which brain disorders are prevalent (Wittchen et al., 2011). Medication with a selective serotonin reuptake inhibitor (SSRI) is the first line treatment but about 50% of patients remain symptomatic after a single treatment. Psilocybin shifts the emotional bias across various psychological domains and emotional face recognition in healthy subjects (Kometer et al., 2012) which may have important implications for the pathophysiology of dysfunctional emotional biases as seen in, e.g., major depression.

In conclusion, there is good evidence that research with psychedelics has considerable potential for developing aspects of psychoanalytic theory, for therapeutic effects in brain disorders, and for studying human consciousness more generally.

- Carhart-Harris RL, Erritzoe D, Williams T, Stone JM, Reed LJ, Colasanti A, Tyacke RJ, Leech R, Malizia AL, Murphy K, Hobden P, Evans J, Feilding A, Wise RG, Nutt DJ. 2012. Neural correlates of the psychedelic state as determined by fMRI studies with psilocybin. Proc Natl Acad Sci U S A 109:2138-2143.
- Carhart-Harris RL, Leech R, Hellyer PJ, Shanahan M, Feilding A, Tagliazucchi E, Chialvo DR, Nutt D. 2014. The entropic brain: a theory of conscious states informed by neuroimaging research with psychedelic drugs. Front Hum Neurosci 8:20.
- Ettrup A, da Cunha-Bang S, McMahon B, Lehel S, Dyssegaard A, Skibsted AW, Jørgensen LM, Hansen M, Baandrup AO, Bache S, Svarer C, Kristensen JL, Gillings N, Madsen J, Knudsen GM. 2014. Serotonin 2A receptor agonist binding in the human brain with [11C]Cimbi-36. J Cereb Blood Flow Metab In press.
- Ettrup A, Hansen M, Santini MA, Paine J, Gillings N, Palner M, Lehel S, Herth MM, Madsen J, Kristensen J, Begtrup M, Knudsen GM. 2011. Radiosynthesis and in vivo evaluation of a series of substituted 11C-phenethylamines as 5-HT (2A) agonist PET tracers. Eur J Nucl Med Mol Imaging 38:681-693.
- Glennon RA, Jacyno JM, Young R, McKenney JD, Nelson D. 1984. Synthesis and evaluation of a novel series of N,N-dimethylisotryptamines. J Med Chem 27:41-45.
- Griffiths RR, Johnson MW, Richards WA, Richards BD, McCann U, Jesse R. 2011. Psilocybin occasioned mystical-type experiences: immediate and persisting dose-related effects. Psychopharmacology (Berl) 218:649-665.
- Hagmann P, Cammoun L, Gigandet X, Meuli R, Honey CJ, Wedeen VJ, Sporns O. 2008. Mapping the structural core of human cerebral cortex. PLoS Biol 6:e159.
- Herbet G, Lafargue G, de Champfleur NM, Moritz-Gasser S, le BE, Bonnetblanc F, Duffau H. 2014. Disrupting posterior cingulate connectivity disconnects consciousness from the external environment. Neuropsychologia 56:239-244.

- Jaworska N, Yang XR, Knott V, Macqueen G. 2014. A review of fMRI studies during visual emotive processing in major depressive disorder. World J Biol Psychiatry.
- Kometer M, Schmidt A, Bachmann R, Studerus E, Seifritz E, Vollenweider FX. 2012. Psilocybin biases facial recognition, goal-directed behavior, and mood state toward positive relative to negative emotions through different serotonergic subreceptors. Biol Psychiatry 72:898-906.
- Krebs TS, Johansen PO. 2012. Lysergic acid diethylamide (LSD) for alcoholism: meta-analysis of randomized controlled trials. J Psychopharmacol 26:994-1002.
- Leech R, Braga R, Sharp DJ. 2012. Echoes of the brain within the posterior cingulate cortex. J Neurosci 32:215-222.
- Meltzer HY, Mills R, Revell S, Williams H, Johnson A, Bahr D, Friedman JH. 2010. Pimavanserin, a serotonin(2A) receptor inverse agonist, for the treatment of parkinson's disease psychosis. Neuropsychopharmacology 35:881-892.
- Muthukumaraswamy SD, Carhart-Harris RL, Moran RJ, Brookes MJ, Williams TM, Errtizoe D, Sessa B, Papadopoulos A, Bolstridge M, Singh KD, Feilding A, Friston KJ, Nutt DJ. 2013. Broadband cortical desynchronization underlies the human psychedelic state. J Neurosci 33:15171-15183.
- Nichols DE. 2004. Hallucinogens. Pharmacol Ther 101:131-181.
- Oram M. 2014. Efficacy and enlightenment: LSD psychotherapy and the Drug Amendments of 1962. J Hist Med Allied Sci 69:221-250.
- Sessa B. 2012. Shaping the renaissance of psychedelic research. Lancet 380:200-201.
- Vollenweider FX, Vollenweider-Scherpenhuyzen MF, Babler A, Vogel H, Hell D. 1998. Psilocybin induces schizophrenia-like psychosis in humans via a serotonin-2 agonist action. Neuroreport 9:3897-3902.
- Wittchen HU, Jacobi F, Rehm J, Gustavsson A, Svensson M, Jonsson B, Olesen J, Allgulander C, Alonso J, Faravelli C, Fratiglioni L, Jennum P, Lieb R, Maercker A, van OJ, Preisig M, Salvador-Carulla L, Simon R, Steinhausen HC. 2011. The size and burden of mental disorders and other disorders of the brain in Europe 2010. Eur Neuropsychopharmacol 21:655-679.

Introduction

Biological evolution is a complex process allowing species and organisms to try to resist the second law of thermodynamics. Over time, this has led to the evolution of increasingly sophisticated brains allowing organisms to adapt and survive longer by learning from experience and predicting future events. Yet, this increase in adaptability also leads to an increasingly difficult resource allocation problem of how to best select between competing wants and likes. Self-awareness is of exceptional value for prioritizing. Therefore human consciousness can be thought of as one particular successful solution to this taxing problem (1). In recent years advances have been made in defining not only the neural correlate for self-awareness, but also how it may be instrumental in its regulation, and the severe consequences of its default.

The paralimbic network of self-awareness

Interestingly, part of this progress has been made possible by studying the introspective states involved in meditation. Studies of highly experienced Yoga-nidra practitioners have shown that there are significantly different brain networks involved in the mind-wandering states related to relaxation meditation (Yoga-nidra) and normal resting state (i.e., at rest, without intended motor, or mental function), and that these networks vary greatly with the changing contents of consciousness during the two conditions. In contrast to this variability of activity patterns, sustained activity in a paralimbic network is common to both states and traits. The network includes midline frontal regions (anterior cingulate and medial prefrontal cortices), parietal regions (primarily precuneus), and striatum (1)).



Sketch of the paralimbic network.

Principal component analysis was used to identify brain regions common to conscious experiences during yoga-nidra meditation and the normal resting conscious state. Two major clusters were found, explaining 25 and 18% of the variability. The paralimbic network consists of a set of three regions which contributed to both clusters: Striatum, medial prefrontal, and medial parietal cortices (precuneus). Being active in a wide variety of conscious experiences, we hypothesized that this paralimbic network represents self-reference, as a common denominator for conscious experiences to account for a sense of unity of consciousness.

Paralimbic regions are anatomically located at the interface between the limbic and the neocortical brain, the former providing information on the bodily state and emotions, while the heteromodal neocortical association regions provide extrapersonal sensory and mnemotic information on the highest level of integration available for attention and awareness.

This discovery led to the hypothesis that the paralimbic network contributes to the sense of unity of conscious experience by acting as a common reference of self-perspective for changing conscious states and traits). Such a role for the paralimbic network would be consistent with ontogenetic considerations of brain development (2).

Activation of paralimbic system is linked to transition from subliminal to supraliminal stimuli. When visual words are presented serially at 28 ms duration they are perceived as meaningless flickering light. In contrast, when the appearance of each word is presented for a longer period of time (57ms) the words are clearly perceived. This change from meaningless subliminal perception to perception which is meaningful to the observer (the self) is linked paralimbic activation revealed by positron emission tomography (PET), even if the total exposure time during the scan is counterbalanced by reduced number of word exposures (3)

Paralimbic activity is linked to the minimal self-awareness accompanying any first person experience of ownership as evidenced by comparison of paralimbic activity induced by first person perspective compared to third person in a study using avatars (4).

Paralimbic activity depends on degree of self-content in extended self-awareness. This has been demonstrated in a number of studies using tasks ranging from mentalizing own features vs features of others (5), and retrieval of personal judgment of one-self vs other (6).

The paralimbic system is instrumental in self-awareness

According to David Chalmers, the "hard problem" of consciousness is why and how subjective experience arises from a physical basis (8). He considered the hard problem to be un-accessible with current technology at the time (8). With few exceptions, the position has been so influential that contemporary research has been stymied by focusing on what David Chalmers considered to be the "easy" problem of consciousness, i.e. its neural correlates. This limitation has the inherited risk of leading to a dualistic concept of human nature with two different worlds, the physical and the mental, without understanding their relationship. It also impedes our understanding of the biological function of self-awareness and conscious experience, and how its disturbance in pathology may account for major symptoms in self-regulatory disorders like autism, ADHD, and schizophrenia. However, recent methodological and conceptual advances justify a paradigm shift to focus on causality in consciousness research.

Importantly, causality of the network for self-awareness has now been established using single pulse transcranial magnetic stimulation (TMS) to examine the behavioral consequences of disturbing normal function of a given neural assembly by means of magnetic induction of non-physiologic neural activity (6,7,9), by granger causality analysis, a mathematical tool to ascertain causal influence of one neural assembly on another, and by pharmacological manipulation, for instance dopaminergic activation (10).

Pathophysiology

The paradigm shift from studying mere correlation to causality of self-awareness has already given promising results in a better understanding of pathophysiology of some of the many disorders of self-awareness and conscious selfregulation.

Self-awareness is tightly linked to conscious self-monitoring and self-control, and recent data show that these functions are regulated by dopaminergic activity, linked to activity in the medial prefrontal/anterior cingulate cortices (10). Illustrating the importance of this regulatory mechanism, we have recently shown that it is deficient in gambling disorder (11). More profound pathology is present in a number of disorders, for instance autism, adhd, borderline personality disorder, schizophrenia, and the dementias.

ADHD is an example. It is characterized by poor self-regulation, including conscious self-regulation, here mainly in the form of poor attention and impulse control. Conscious self-regulation involves coordination and organizing self-awareness during planning, execution and monitoring goal-directed behaviors, resulting in impulse control. This capacity is one of the defining ontological features of human development, and children with ADHD are often regarded as immature in this respect. ADHD has shared traits with other disorders of impulse control, such as borderline personality disorder. fMRI studies have shown decreased activity in the paralimbic network of self-awareness and self-monitoring (12) in agreement with the concept of causality of dysfunctional paralimbic self-awareness in dysfunctional conscious self-control.

Conclusion

The paradigm shift in consciousness research from mere correlation to instrumental neural mechanisms is opening new vistas in pathophysiology and its remediation of important disorders of self-awareness and self-monitoring.

- 1. Lou, Joensson, Kringelbach (2011) Yoga lessons for consciousness research. Front Psychol 2, article 366.
- 2. Gallagher (2000) Philosophical conceptions of the self: Implications for cognitive science. TICS 4, 14-21.
- 3. Kjaer TW et al. Consc Cogn (2001)
- 4. Vogeley K et al. (2004) Neural correlates of first-person perspective as one constituent of human self-consciousness. J Cogn Neurosci 16,817.
- 5. Kjaer TW et al (2002) Reflective self-awareness and consciousness: PET evidence for a common midline parieto-frontal core. Neuroimage 17, 1080.
- 6. Lou HC et al.(2004) Parietal cortex and representation of the mental self. PNAS 101, 6827
- 7. Lou HC (2012) Paradigm shift in consciousness research: the child's self-awareness and abnormalities in autism, adhd, and schizophrenia. Acta Paediatrica 101, 112-119.
- 8. Chalmers (1996). The conscious mind: In search of a fundamental theory. Oxford University Press
- 9. Luber B et al. (2012) Self-enhancement processing in the default network: a single pulse TMS study. Exptl Brain Res 223, 177-543.
- 10. Joensson M et al. (2014) making sense: Dopamine regulates metacognition and self-awareness with its paralimbic network (submitted).
- 11. Rømer Thomsen K, Joensson M, Lou HC, et al. (2013) Altered paralimbic interaction in behavioral addiction PNAS 110, 4744-4749
- 12. Castellanos et al. (2008) Cingulate-precuneus interactions: A new locus of dysfunction in adult attention deficit/hyperactivity disorder. Biol. Psychiatry 63, 332-337.

NOTES

