Agency, Morals & the Mind

Event Report

26-27 September 2017, London

The sense of agency – the feeling that we are in control of our thoughts and actions – is a central feature of the human mind. The experience of agency influences the conscious selection and avoidance of courses of action, our sense of responsibility, interaction with other people and the way in which we address societal challenges. It also has crucial implications for what we deem to be right and wrong in human behaviour.

How can we define the relation between agency, moral responsibility and the brain? *Can cognitive explanations shed light on the subjectivity and voluntariness of action? How can the science of evolution help us understand the nature of ethical constructs, and address the possibility of moral progress? What turns the mere control of bodily movements into conscious acts of morality or immorality?*

Authors

Dr Sofia Bonicalzi

Dr Sofia Bonicalzi is a postdoctoral researcher currently based in the School of Advanced Study at the University of London, where she is part of the team of *The Human Mind Project*. A trained philosopher, and a Honorary Research Associate of the "Action & Body Group" at the Institute of Cognitive Neuroscience (UCL), she works on aspects of human voluntary actions, sense of agency, and responsibility, drawing on methods and insights from the humanities, as well as the psychological and brain sciences.

sofia.bonicalzi@sas.ac.uk

Dr Uri Hertz

Postdoctoral Research Fellow at the Sackler Centre for Consciousness Science, University of Sussex.

Dr Uri Hertz is a cognitive neuroscientist interested in the psychological and neural underpinnings of social influence and collective decision making. He is now part of the team of the Project and an associate researcher at the Crowd Cognition group at the Institute of Cognitive Neuroscience (UCL). He can also be found on his Website.

urihertz@gmail.com

This report from *The Human Mind Project* aims to draw out key themes from our event for future discussion and exploration. As well as providing a material resource that allows us to take discussion started at events forward, event reports form part of our Grand Challenges; an attempt to define the major intellectual challenges in understanding the nature and significance of the human mind.

The notes on each talk provide a summary and reflection on key themes, with questions raised in discussion highlighted in the notes on each roundtable.

Table of Contents

Morals, Culture and Society

Will to Fight: Devoted Actors and the Spiritual Dimension of Human Conflict

Scott Atran

Director of Research, CNRS; and Research Fellow, University of Oxford

The (unpredictable?) rise of global terrorism

In September 2014, President Obama acknowledged the U.S.'s mistaken underestimation of the rise of the Islamic State in Iraq and Syria. But was Isis' ability to fight imponderable and unpredictable? What general lesson might we learn from the peculiar evolution of the Islamic State? Anthropologist Scott Atran presented the results of fieldwork jointly conducted with experts of different disciplines, ranging from psychology to biology and foreign politics. Atran warned against interpreting Isis simply as a nihilistic form of violent terrorism, and introduced his explanatory hypothesis (the *Devoted Actors Hypothesis*), grounded on interviews with political and military leaders, lab experiments and surveys in Iraq and Syria with members of the Islamic State and their opponents. According to the hypothesis, global terrorism is partially led by Devoted Actors, who are willing to jointly defend the transcendental values of the social group they belong to, and whose personal identities are fused within the collective one. When these two conditions (adhesion to sacred values and identity fusion) are satisfied, people are prone to extreme actions and sacrifices, which are resistant to any material trade-off, and normative influence. Such collective empowerment must be taken into account as part of the explanation of the rise of global terrorism.[1] The interaction between these factors largely determines who is likely to become a devoted actor: the ability to see one's own identity as fused to that of the group is the best predictor of the willingness to fight.

The privilege of absurdity

Among living creatures, only men are subjected to the privilege of absurdity. Atran quoted Thomas Hobbes' acknowledgment of the power of the absurd as a conceptual tool to understand contemporary social phenomena. During the history of mankind, humans accepted inherently absurd concepts as natural. As an example, we defend human rights and equality, turning apparently absurd ideas into fundamental pillars of modern western societies.

We can start from this assumption in order to interpret some characteristics of global terrorism. The Islamic State represents an incredibly efficient war machine, with no real competitors in the rest of the Islamic world. Despite losing men and territories and being surrounded by enemies, it remains the largest volunteer fighting force – mainly composed by self-organised networks or radicalised friends and kin, and including women and foreign individuals – able to continuously attract supporters who are willing to die for it. But why do people ultimately go to the front line and refuse any political compromise or exit strategy? In writing to the British evolutionist A. R. Wallace, Charles Darwin expresses his sense of astonishment at the human attraction to heroism and martyrdom, which seem to go beyond what is taught by the *Golden Rule* - which prescribes that one should treat others as one wishes to be treated oneself.

[1] Scott Atran, The Devoted Actor: Unconditional Commitment and Intractable Conflict across Cultures. Current Anthropology 57, no. S13 (June 2016).

Does this undeniable characteristics of humans constitute evidence against the idea that we can understand evolution only in materialistic terms? And how to explain what Atran calls "Jihad's fatal attraction for violence"? It is a visceral and physical phenomenon, steeped in emotion and identity, far from the nature of current modern ideologies, which would rather favour reason and logical argumentation. Indeed, in many cultures, violence against other groups has been considered as a sacred and sublime form of moral virtue, which cannot be exchanged for money. As an example, Atran recalled that, after the promulgation of the Nuremberg Laws, the Nazi party chancery received thousands requests for *Aryanization* from wealthy Jewish families, able to offer conspicuous sums of money to be "reclassified". Receiving over 2,100 applications in 1939, Hitler granted only 12.

The inscrutable logic of religion

Beside the lack of field experience of most experts, a major problem in the interpretation of global terrorism is connected to the difficulty of reducing the phenomenon to the logic of utilitarian and rational thinking. We should not underestimate the peculiar role of religion as a facilitator and a multiplier of large-scale cooperation, Atran claimed. How can we impart reason to an inscrutable belief, which is able to galvanise in-group solidarity and to strengthen the social bond? Paradoxically perhaps, fully reasonable social contracts have been proven more liable to collapse than religious ties.

However, while economic-decision making has been extensively investigated, we seem to lack deep knowledge about morally motivated behaviour. Sacred values appear to be so difficult to scrutinise because they are insensitive to quantities – the number of lives to be sacrificed, the prospect of success, the costs and consequences –, temporal discounting, and material trade-off. As a matter of fact, in many contexts, sacred values have been proven to defy cost-benefit logic and real-politik.

Such a mentality offers extraordinary proof of the well-known *backfire effect*, according to which, in the face of contradictory evidence, established beliefs do not change but actually get stronger. In particular, introducing material incentives that go against sacred values is likely to strengthen the opposition to any form of compromise. Atran tested the force of the backfire effect in different populations, including Palestinians, Israeli settlers, Indonesians, Indians, Afghans and Iranians. A clear example is offered by a study conducted in the West Bank and Gaza: a group of Palestinian refugees holding a "Right of Return" to their previous homes in Israel would have been willing to abandon their right in response to an apology from Israel. In contrast, a group of refugees who were offered material incentives, in the absence of apologies, in order to abandon their right were violently opposed to any form of compromise.[2]

The Islamic State and the construction of democracy

It would be also mistaken to interpret people's adhesion to global terrorism as the effect of brainwashing. For many people, the Caliphate constitutes a real and powerful attractor, the ending point of the history of salvation. In order to understand the rise of violent extremism, we have to interpret it as the most influential and politically creative force on the world scene. In 1940, in his review of *Mein Kampf*, George Orwell acknowledged that "Hitler […] knows that human beings don't only want comfort, safety, short working-hours, hygiene, birth-control and, in general, common sense; they also, at least intermittently, want struggle and self-sacrifice". Hitler understood that sometimes people need to feel the sense of the transcendental. Because of the current geopolitical situation, the Islamic State might have better chances of surviving than the Third Reich. However, as it happens to the Nazi Germany – the best fighting force during World War II – the Islamic State will probably collapse due to number of his enemies. However, in the attempt to export democracy, the

[2] Atran, S. & Ginges, J. (2013). Religious and sacred imperatives in human conflict. Science, 336:855-857.

Western world should not underestimate differences between historical contexts. In Europe, the construction of liberal democracy took hundreds of years and its values were relatively easy to re-establish after World War II, but nothing like that existed in Syria before the Islamic State. How can you replicate democracy? The U.S. historically has a very poor history of successes, including only Japan and, again, Germany. Only in those two cases, the U.S. were able to "flip" the culture. Beyond historical differences, we must also interrogate ourselves about disturbing similarities. As in the twenties and thirties, the values of democracy are rapidly losing ground. Even beyond our need to face the Islamic State, this should be a key point in our future political agenda.

Responsibility as a Social Construction

Catherine Wilson
Anniversary Professor of Philosophy, University of York

Responsibility. Old and new threats

The problem of responsibility and control has been extensively treated both from a scientific and from a philosophical point of view. We have been told that we are different from non-human animals because, unlike them, we are not slaves of our desires and inclinations, said philosopher Catherine Wilson. At least according to the mentality of Western culture, we have free will, and we believe in the afterlife and in the existence of God. Historically, the existence of an omniscient God created problems for how to understand individual responsibility. During the last few decades, the threat to responsibility came from a different perspective. How do we assign and how should we assign responsibility to living beings? How do we control what we do? Following Darwin, humans are organisms with an internal system able to represent what is happening in the outside world. This seems to be the basis for the uniquely human ability for self-control and free will. However, this very traditional view has been deeply challenged. From the one side, researchers are willing to ascribe awareness also to other, non-human, living beings. From the other side, it has been shown that much of what humans can do can be accomplished even without awareness. We show many automatic behaviours that do not reach the threshold of consciousness. Of course there are many tasks that require a very specific form of awareness, including interacting with conspecifics, attending unexpected events, dealing with unpredictable situations, or mentally processing instructions. The "invention" of consciousness expands widely the number of activities we can do, and permits more plastic responses to new situations.

Body ownership, sense of self, and responsibility

People have a very specific sense of 'mineness' with respect to their own body. Different pathologies, including neglect, could alter or destroy one's sense of self. In contrast, individuals born without a limb can show a predisposition to feel the ownership of the phantom one. The sense of self is important to distinguish what happens to me from what I did. It is the feedback from the environment that tells us what we can and cannot do. However, it seems hard to give a definition of what is "mine" or what is "me" that goes beyond the level of physiology. We seem to obtain a very illusive notion of those concepts. Am I responsible, for example, for occupying the role I have now? Are we responsible for what we do if everything is caused by the multiple influences surrounding me? What if my feeling of making a decision is what my nervous system decides I have to do? Reasons we create a-posteriori appear to be the causes of our actions. As Thomas Nagel wrote, if we decided to do something else, a different justifying reason would suddenly appear. Why should I be blamed or rewarded if I am the victim of the circumstances? Regarding this topic, there seems to be no real fact of the matter that philosophy or science can teach or discover.

Responsibility in the social domain

Despite its lack of success in reforming behaviour, and its tendency to increase the recidivist rate, traditional punishment has been often invoked as an educational tool. Other justifications for punishment are needed – claims Wilson. We often ascribe responsibility on the basis of irrational parameters. How do we judge, for example, accidental mistakes? Is Jocasta to be blamed for the incest with her son Oedipus given that they were both unaware of their bond? How to explain our tendency to blame the inanimate objects, like murder weapons, that were involved in crimes? Why, as Adam Smith reminds us, do we judge on the basis of outcomes and not of intentions? The tendency to blame the victim – and oneself in particular, or that woman with a short skirt in a dodgy part of the town – is a subtle and dangerous habit. When we perform very simple actions – for example when we pick up a glass – it is relatively easy to determine if we were in control of our behaviour or not. The social realm is much more complicated than this, and philosophy has always struggled in the attempt to offer respectable criteria for responsibility. Those criteria remain inevitably vague. It is only by doing things, in the practical domain, that we will progressively understand what we are able to do and what we are not. In relation to punishment, we might take a more empirical approach: it might work in some circumstances, but this cannot be assumed a priori.

How does the Behaviour of Others Influence what we do?

Emma Flynn
Professor of Developmental and Comparative Psychology, Durham University

Becoming a social being

A developmental psychologist working with children and animals, Emma Flynn is interested in how children become social beings. We are born embedded in different cultures and social worlds: what are the specific mechanisms through which we adopt social behaviours? What is the importance of culture beyond the mere possibility of being part of a group? We have to navigate the physical space, constantly receiving information from society, symbolic language and the social structures surrounding us. We are able to deal with the complex problems posited by the environment not only because we are clever, but also because culture helps us deeply. Flynn recalled Franklin's lost expedition to the Arctic, and Burke and Wills' disastrous attempt to cross Australia: in those cases, the price of lacking knowledge of the local culture and environment was extreme. Moreover, far from being a static item, cultural evolution allows us to build up: it is largely due to the efforts of our predecessors – to incremental discoveries and big changes – if we are able to produce extraordinary technological breakthroughs.

The mechanisms of social learning

Social mechanisms are also extremely complex, and culture can be opaque. How do children become able to deal with this? Flynn's hypothesis is that, in contrast to non-human animals, children tend to over-imitate, by reproducing series of complex, and task irrelevant, actions. In sets of experiments in which children and chimpanzees were presented simple demonstrations with many irrelevant details, Flynn and colleagues discovered that, while monkeys rarely copy irrelevant elements, children (3 and 5-year-old) copied redundant actions, as if they did not have the ability to filter irrelevant details. For example, children were shown a video or a live model illustrating how to obtain a reward from a clear or an opaque box. Crucially, only some of the actions in the operational sequence were causally relevant to get the reward, whereas others were irrelevant. The clear box made the causally irrelevant actions visible, whereas the opaque box prevented them from

being seen. With some differences depending on age, both 3- and 5-year-old children imitated the irrelevant actions regardless of the availability of causal information. To explain over-imitation, Flynn and colleagues suggest that imitation "develops to be such an adaptive human strategy that it may often be employed at the expense of task efficiency".[3] Interestingly, children seem to copy irrelevant intentional actions, in particular if presented by an adult, more than accidental ones. The effect is also amplified if the demonstrator is still present in the room where the child is performing the task. And it does not disappear when children are required to be quick and to compete with others, or if they think that the experiment is over. In general, both children and adults tend to adopt the opaque behaviour of others even when the imitative behaviour implies a cost. Such adaptive behaviour is supported both by social and by cultural motivations, including the need to act appropriately in a new social context.

When do children stop imitating useless causal information in order to produce novel solutions? Flynn introduced a study where children aged 4–9 years were presented a puzzle box, the Multiple-Methods Box (MMB), and social demonstrations of the tools, access points and exits they could employ to extract a reward from the box. The results showed that children tend to imitate quite irrespective of the efficacy of the method and that innovation is a rarity. Indeed, only 12.4% of children innovated by discovering at least one novel reward exit.[4]

The mechanisms of cultural transmission

What sort of mechanism allows the transmission of social behaviour? Who is learning from who? Where do the rules we employ come from? Studies show that children copy the methods they witness. Flynn and her team ran a series of "broken telephone" experiments. The first child was taught a skill, for example getting a treat out of a puzzle box, by the experimenter. In series the children were introduced to the skill and were able to see a demonstration offered by the previous child. Flynn examined how faithfully the skill was transmitted, and whether at some point redundant actions were omitted, or novel actions were introduced. She found that imitation was remarkably faithful, with only few innovators emerging over time. In a follow up study in a natural setting (a kindergarten), two popular children, one from each group, were taught two different types of behaviour. In the following weeks, Flynn and her team examined how these behaviours were transmitted. Once again, imitation was found to be faithful within the social circle of each popular kid. Kids that had friends in both social groups played the roles of innovators as they allowed one behaviour to leak to the other team.

Variance and deviations from the standard are developed by microcultures or small-scale traditions inside the group.[5] Cultural transmission can take different forms, ranging from direct observation to teaching. If copying is such a common occurrence, could we use its mechanisms for transmitting positive normative behaviour? Quite independently from the characteristics of the task, children appear to be very good at coordinating with each other: cooperation is infectious. But how does the behaviour of others influence what we do? We tend to copy more in specific circumstances, including situations of uncertainty or sedimented habits. Society seems to need innovators to survive, but in relatively small numbers. Given the way in which we are influenced by others, is agency a myth? Why is the influence of others so powerful? How does this vary across cultures? Who is able to innovate and in which contexts?

---

[3] McGuigan, N., Whiten, A., Flynn, E. & Horner, V. (2007). Imitation of causally-opaque versus causally-transparent tool use by 3- and 5-year-old children. Cognitive Development, 22:353-364.

[4] Carr, K., Kendal, R.L. & Flynn, E.G. (2015). Imitate or Innovate? Children's Innovation is Influenced by the Efficacy of Observed Behaviour. Cognition, 142:322-332.

[5] Whiten, A. & Flynn, E. (2010). The Transmission and Evolution of Experimental Microcultures in Groups of Young Children. Developmental Psychology, 46(6):1694-1709.

Q&A: Scott Atran, Catherine Wilson, Emma Flynn

Q: What are the ingredients that make the recent revolutionary movements possible?

Atran: Every revolutionary situation is characterised by a cascade of events breaking down the social order. However, a revolution requires something more, namely a new moral world that can replace the old one. These moral worlds are quasi-religious, unverifiable by nature but adaptive, and nurtured by transcendental ideas pervading large-scale societies.

Q: But is it true that transcendental ideas operate irrationally? – If the Caliphate were established, in the long historical trajectory we would probably be willing to define as perfectly rational those concepts and drives that now we define as irrational.

Atran: In line with the Darwinian tradition: there is no difference, at the moral level, between the French Revolution and the Islamic one. The difference is in their success rate. The lack of rationality characterising the Devoted Actors is not to be found at a moral level: transcendental ideas grounding the Islamic State are not reducible to any form of utilitarian theory. In other words, they violate every axiom about how a rational actor should behave.

Wilson: However, a moral difference might exist at a deeper level, we can distinguish between what is genuinely moral and what is not. The measure of immorality is how much the weak are exploited and pushed by the strong.

Atran: If the Nazi regime had won, the conversation would be very different today. Every culture has its own conception of what is morally wrong. When I asked the White Supremacist Leader what was evil for him, the answer was: “Evil is to refuse the existence of races”.

Q: But why are Devoted Actors willing to sacrifice themselves even when they know that there is no hope? Is the ability to envisage yourself as an individual in the society the main driver? For some people, Western society represents a totally unacceptable compromise. Adhesion to contemporary terrorist movements might be fuelled not just by passion and emotions, but by the unacceptability of the current political situation.

Atran: The Islamic State is supported by people who would live at the margins in the Western world, but most of its supporters belong to cultural and political élites. They are well educated people with a wealthy lifestyle. A good comparison might be with some of the leaders of the American Revolution. In purely economic terms, Washington and Adams had everything to lose in fighting against Britain, but they were willing to die for something they hoped was more important. They were the richest people in the world, far from belonging to marginal societies. Revolution requires such willingness to die. As Mahatma Gandhi said: “Rivers of blood may have to flow before we gain our freedom, but it must be our blood”.

Q: A crucial distinction in psychology is the one between exploration and exploitation, which are thought to be part of individual decision-making. What kind of impact might blind copying or over imitation have in society? What would happen if everyone imitates every innovation I introduce? It seems that we need a more stable society.

Flynn: People have to understand how innovations work before starting imitating them. Innovation does not spread as quickly as we might imagine. It is competition between cultures that helps innovation. When societies are in competition and you have good ideas, you will find a place where your ideas are accepted and implemented.

Atran: And innovation is also promoted under conditions of life and death: you innovate or you die.

Q: But is individual non conformity a social virtue?

Flynn: Innovation is extremely important, despite our bias towards conformity.

Q: Who are the innovators? Are they idiosyncratic individuals standing against the crowd?

Tuesday, 27th September

Agency and Subjectivity

Chair: Robyn Repko Waller

Lecturer in Philosophy, King's College London

Volition and Value

Patrick Haggard

Professor of Cognitive Neuroscience, University College London

Neuroscientific basis of volition

Patrick Haggard discussed the neuroscientific basis of volition – the process of generating voluntary movements. A crucial turning point in the scientific literature about voluntary actions is represented by Benjamin Libet's work on the neural antecedents of endogenous actions. Famously, Libet's experiments showed increased brain activity prior to the generation of a voluntary movement, called *Readiness Potential* (RP). Since their appearance, Libet's studies have been criticized in two ways: an ontological critique has been moved mainly from a neuroscientific perspective (is RP just averaged noise? RP looks like a brain signal causally involved in voluntary actions but it might be something else). On the other side, philosophers have questioned the ecological validity of Libet's model of voluntary actions (poor approximation, minimal laboratory abstraction, action processes and experiences must be reasons-responsive). In the light of the former critique, Haggard introduced a line of research dealing with the following question: what is RP? What is its causal role? Two main hypotheses regarding brain signals underlying voluntary movements have been outlined in the literature. According to the first view, endogenous actions are preceded by a gradual buildup of neuronal activity. In this case, RP is thought to reliably precede voluntary self-initiated movements. The competing hypothesis is that voluntary actions are preceded by stochastic fluctuations in neural activity. The precise moment in which an action occurred depends on when those spontaneous fluctuations surpass a threshold: if you average the random brain activity preceding the action, you have the impression of a signal, but what you really have are just stochastic fluctuations.[6]

[6] Schurger, A., Sitt, J.D., and Dehaene, S. (2012). An Accumulator Model for Spontaneous Neural Activity prior to Self-Initiated Movement. Proceedings of the National Academy of Sciences 109, no. 42:E2904–13.

Causal role of readiness potential

To test these hypotheses, Haggard and his team used an experimental design in which voluntary movements and cued movements were embedded in one task, keeping experimental context constant. On every trial in this experiment participants had to complete a visual detection task. On some trials the visual stimuli appeared after a long latency. Participants could wait for the stimuli to appear and make their decision, and have the opportunity to gain a high reward, or opt to skip the trial and settle for a lower reward. As the experiment lasted a fixed amount of time, to maximize your rewards it was worthwhile to skip some trials and cut the waiting time. Importantly, 'skip' responses were voluntary and self-initiated, while detection responses were cued. By using EEG, researchers measured brain activity locked in time to cued and uncued movements. They showed that RP, an elevation in signal prior to action associated with preparation period, was displayed before voluntary actions, but not before cued actions. In addition, before voluntary, but not cued, actions brain signal showed a reduction in variability. This indicated a consistent preparatory brain activity, supporting the hypothesis of preparation rather than fluctuation as the source of voluntary actions.

Causality, learning and the sense of agency

Haggard and his team were also interested in the sense of agency, the feeling that you are in control of the external world through your actions. In the lab environment, sense of agency is usually measured by referring to explicit (such as self-report) and implicit measures (which do not rely on participant's explicit ratings). Among implicit measures, the so-called *intentional binding* has been considered a reliable tool for assessing individual sense of agency. Intentional binding refers to a widely observed phenomenon, characterised by the compression of the subjective temporal interval between a voluntary action and its effects (temporal attraction of the action towards the effect and of the effect towards the action).[7]

To examine how the sense of agency developed over time in a specific context, Haggard and colleagues used a learning paradigm in which participants learned about the probability of getting a reward in an elaborate and stochastic environment. In a two stage decision task, participants could choose whether to go left or right in a maze. The decisions were stochastic in the sense that, after choosing 'right', participants still had 20% probability of ending up on the left. After two decisions you were given a reward. How is sense of agency – measured by intentional binding – modulated by learning? Participants were asked to report when an outcome was presented to them after the second choice in the maze paradigm, using a clock presented on a screen. Participants showed a boost in binding, i.e. perceived the outcome as appearing earlier, on trials following an error. This finding suggests that learning enhances the sense of agency: people use feedback to make reasons-responsive choices. People who learn more tend to show more post error boost in action binding.

Haggard suggested a unified model of voluntary actions, in which intentions to act are followed by preparation and action, the process of volition, and movement is followed by a perceived effect, which is mediated by the sense of agency. As actions, outcomes and perception follow each other, the link between volition and agency becomes stronger.

[7] Haggard P., Clark, S. & Kalogeras, J. (2002) Voluntary action and Conscious Awareness. Nature Neuroscience, 5: 382-385.

Getting out of your Head - Addiction and the Motive of Self-Escape

Lucy O'Brien
Professor of Philosophy, University College London

Liking and pleasure in addictive behaviour

Philosophy is always concerned with what is like to be me. How do the ways in which we apprehend ourselves enter into our behaviour? Presenting her joint work with Daniel Morgan, Lucy O'Brien suggested that, in many cases, consumption of an addictive substance can alter our self-perception, allowing us to act in a manner that we would not approve while sober. Addictive behaviour is contrary to a person's view of how he should act. But what underlies addictive desires? Work by Berridge and Robinson on rats suggests that elevated dopamine levels were associated with addictive behaviour. [8] Interestingly, dopamine levels seemed to be dissociated from liking the substance and the pleasure it brings. O'Brien suggested a model, by which substance S is desirable as it is good for wellbeing, it brings relief from pain, or it fixes or fulfils desires. According to Berridge and Robinson's data, addictive substance can bypass the pleasure system, elevating dopamine levels directly. Lucy presented a hypothetical situation, in which two substances 'balcohol' and 'calcohol' have different features of alcohol. 'Balcohol' has taste and feel of beer, but does not affect dopamine levels, whereas 'calcohol' has no taste but effects dopamine level. The dopamine assumption will suggest that only 'calcohol' will be addictive. If dopamine assumption is taken to its limit, why are we not all addicts?

Addiction as self-escape

O'Brien pointed out that we should assume a pluralist perspective regarding addiction. Indeed, also personal history and social factors play a crucial role in addictive behaviour. Especially in situations in which – being in a costly emotional state such as the feeling of being oppressed by others or depression – one needs an escape from negative self-consciousness.[9] In those situations, addictive substances tend to secure 'self-escape', a relief from one's emotional state. This suggests an additional motivation for addictive behaviour, having to do with the experience addictive substances promote, rather than with their subsequent increased dopaminergic level. In this sense 'balcohol' will be addictive as it provides the 'self-escape' experience of alcohol.

The Heart of Human Sociality

Keith Jensen
Lecturer of Psychology, University of Manchester

Prosocial behaviour

[8] Berridge, K.C., Robinson, T.E. (1998). What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? Brain Research Reviews, 28 (1998): 309-369.
[9] Hull, J. G., & Reilly, N. P. (1983). Self-awareness, self-regulation, and alcohol consumption: A reply to Wilson. Journal of Abnormal Psychology, 92:514-519; Randles, D., & Tracy, G.L., Nonverbal Displays of Shame Predict Relapse and Declining Health in Recovering Alcoholics, Clinical Psychological Science April 2013 vol. 1 no. 2:149-155

Keith Jensen examined the complexities of human sociality, with humans displaying "ultra-cooperative" behaviour on the one hand, and "hyper competitive" behaviour on the other hand. He focused on pro-sociality,[10] the phenomenon of helping others in order to increase their well being. Does pro-social behaviour truly exist? And what are its evolutionary and developmental origins? Jensen first suggested that empathy – the capacity to understand or feel what another person is experiencing – is not enough for pro-social behaviour. Indeed, there are number of cases in which others happiness and our happiness are not aligned, such as in jealousy and glee. However, we can display other-regarding concerns, for example sympathizing with others, or displaying empathic distress. These behaviours, Jensen suggested, bridge present and future benefits.

Development of pro-social behaviour in humans

Jensen continued to examine how pro-social behaviour develops in children. In a number of studies, he showed that children learn to share with others, and later on learn about fairness.[11] They start off by preferring equal split of the loot, and later learn to differentiate between more fair and less fair non-equal splits, preferring 30-70 split for example, over 10-90 split. Jensen also showed that young toddlers (less than 24 months old), display helping behaviour under a variety of scenarios, and pay more attention to people and agents that display helping behaviour compared with non-helpers. Children were also willing to punish someone for misbehaving - for example for snatching food from someone else.

Evidence of pro-social behaviour in primates

Jensen also examined pro-social behaviour in primates, specifically in chimpanzees. While chimpanzees display social behaviour such as grooming, it is not clear if they display helping behaviour or a preference for fairness like young human children. Jensen surveyed a number of studies that showed little evidence for pro-social behaviours in chimpanzees. They were less likely to help someone, and only in one case helped an experimenter. They did not display preference for fair split, and were willing to accept all splits of loot in an ultimatum game. They were not willing to punish another chimpanzee for stealing, unless they could get the stolen goods themselves.

Pro-social behaviour, and human morality, seems therefore to emerge later in evolution, and follow a developmental process shaped by society and culture.

Q&A: Patrick Haggard, Lucy O'Brien, Keith Jensen

Q: While an addict is sometimes – between substance effects and self-escape – "sober", the devoted warrior described by Scott Atran is continuously self-escaping, or under the influence.

O'Brian: It is questionable if addicts really have 'sober' intervals in terms of their desire for self-escape. It may be the case that these two conditions are more similar than we appreciate.

Q: Is a drug facilitating self-escape a bad thing?

[10] Jensen, K., Vaish, A., & Schmidt, M. F. (2014). The emergence of human prosociality: aligning with others through feelings, concerns, and norms. Frontiers in Psychology, 5:822.

[11] Fehr, E., Bernhard, B., Rockenbach, B. (2008). Egalitarianism in young children, Nature 454:1079-1083; Hamlin, J.K., Wynn, K., Bloom, P. (2007). Social evaluation by preverbal infants, Nature 450:557-559; Warneken F., Tomasello M. (2006). Altruistic helping in human infants and young chimpanzees, Science, 311(5765):1301-3; Wittig, M., Jensen, K, Tomasello M. (2013), Five-year-olds understand fair as equal in a mini-ultimatum game. J Exp Child Psychol., 116(2):324-37.

O'Brian: No, alcohol is good for self-escape, which is sometimes a desirable thing. When we avoid returning to ourselves, addictive self-escape is bad.

Q: Is addiction a "pathology" of free will?

Q: Some people report feeling more "themselves" under the influence of alcohol. Is it self-escape necessarily?

O'Brian: the escape is from self-conscious. We regularly inhibit our actions. Sometimes people want to escape from these inhibitions, and become more engaged.

Q: What are the arguments against pro-social behaviour?

Jensen: Such behaviour is considered irrational, and is evolutionarily unstable. Altruism is often explained by providing some utility to the sharer, either by enhancing one's reputation and social status, or by providing good feelings by tapping into our reward mechanism.

Towards a (Neuro)Science of Morality

Chair: Sofia Bonicalzi

Postdoctoral Researcher, The Human Mind Project, School of Advanced Study, University of London

Morality, Self-Control and the Brain

Richard Holton

Professor of Philosophy, University of Cambridge

Willpower and self-control

Richard Holton started by examining the idea of agency and free will, and the notion that the experience of free will has to do with thinking about what drives us to action. He then focused on the notion of will power – our control over our action and the ability to constrain ourselves. He cited Walter Mischel's famous marshmallow experiment, in which young children were offered a marshmallow right now, or two marshmallows if they could wait for a number of minutes. Crucially, they were left in the room with the single marshmallow while waiting for the time to pass. An interesting observation from this and follow up studies, was that the value of two marshmallows in the future seemed to decline as time passes, until it reaches the same value of one marshmallow at the point at which the participant decides to opt for the immediate reward.

Mechanisms of dissolving resolutions

Holton suggested that we constantly adapt our preferences to what is currently available. Counterintuitively, thinking more about future rewards increases the rate of adaptation, making us less likely to wait for them. Bringing future rewards to mind allows a process of re-evaluation. It is when we don't think about the future option and focus on the action of abstinence that we are less likely to succumb. Holton suggested that we need a resolution to follow through, and that resolutions should be strong enough to keep you going, but allow responsiveness in the face of changes in the environment. When president Obama declared "we do not torture", he was making a normative and not a descriptive claim. He was establishing a rule for the future, not

describing a matter of fact. He made a resolution. Once reasons for a resolution are considered, it is no longer effective. When one is on a diet, giving reasons why one must avoid a specific food leads to finding counter reasons why this food is actually ok.

Resolution and flexibility in ethics

Finally, Holton underlined the importance of balance between resolution and flexibility in ethics. The only way to constrain armies, for example, is through rules. These should be made outside of the context of military action, where they can be evaluated and shaped. But in the context of action, these rules are non-negotiable. This final argument relates to the predominance of reasons or sentiments in ethics. In the history of philosophy there was huge debate about what the source of moral motivation is. If it comes from reason (Kant) or from sentiments (Hume, Scottish Sentimentalists). Holton claimed the debate is somehow misguided, arguing for a pluralist view according to which both reason and sentiments work as a source for morality.

The Price of Principles: Experiments in Moral Decision-Making

Molly Crockett
Associate Professor of Experimental Psychology, University of Oxford

Immoral Profits

Molly Crockett opened by citing Adam Smith in 'The Theory of Moral Sentiments', claiming that there is something immoral about profiting from others' harm.[12] She than proceeded to ask how much people value profits resulting from an immoral action, and how we might examine such a question in the lab.

Harming others vs. Harming oneself

In a study from 2014, Crockett and her team examined the willingness of participants to give others painful shocks in exchange for monetary reward.[13] In a series of trials participants had to choose between two options: more money and more shocks, or less money and less shocks. Money was always allocated to the participant, but in two different experimental conditions the shocks could be delivered either to the participant (harm oneself) or to another anonymous participant (harm to other). Anonymity of the other participant was crucial, to exclude effects such as reputation maintenance, reciprocity, or retaliation. The researchers examined the choices made by participants. They hypothesised that choice behaviour was affected by how much pain each option entails, and how much money can be gained from each option. Using model fitting, they estimated how much each participant weighted these outcomes, pain and money, when making a choice. Having low weight on shocks meant that the participant always opted for the higher monetary reward option. Having high weight on shocks meant that the participant always opted for the minimum shocks option. The researchers found that the weight assigned to shocks in the 'harm other' condition was higher than the weight of shocks in the 'harm self' condition. Participants were more willing to hurt themselves for money than to profit from other's distress. This finding is in line with Adam Smith's

[12] Inbar, Y., Pizarro, D.A., & Cushman, F. (2012). Benefiting from misfortune: When harmless actions are judged to be morally blameworthy. Personality and Social Psychology Bulletin, 38:52-62.
[13] Crockett M. J., Kurth-Nelson Z., Siegel J. Z., Dayan P., Dolan R. J. (2014). Harm to others outweighs harm to self in moral decision making. Proc. Natl. Acad. Sci. U.S.A. 111:17320–17325.

conclusions, and is not trivial from a rational, economic point of view, which suggests that people will try to maximize their own monetary gain, even on the expense of others.

Neural correlates of the value of immoral profits

In a follow-up experiment, Crockett and her team used functional magnetic resonance imaging (fMRI) to examine participants' brain activity while performing the task. They showed that immoral profit – gaining monetary rewards while harming others – elicited less activity in the ventral striatum, a brain structure associated with processing reward, than profiting whilst harming oneself. Another brain system associated with pain and conflict perception, including the Insula and the anterior cingulate cortex (ACC), was more responsive to the prospect of pain caused to others, compared with prospect of pain caused to oneself. These neural mechanisms can mediate the observed behavioural effects of refraining from profiting from others' misfortune.

Is a "Psychocivilized Society" Possible?

Steve Fuller

Auguste Comte Professor of Social Epistemology, University of Warwick

Remote controlling the brain

Steve Fuller opened with a video featuring Yale professor of physiology Jose Delgado, demonstrating remote controlling the behaviour of a bull by stimulating its brain. Starting in the sixties, this line of research caused a stir in public opinion, as it echoed with the notion of "brainwash" introduced during the cold war. Especially unsavoury was the use of remote control, which was the basis of many conspiracy theories.

Un-constraining the human experience

However, Fuller suggested that the motivation of Delgado and others' research is crucial to scientific progress. They wanted to find means to un-constrain the human experience. At the end of his life, Delgado said that we could make considerable progress in enhancing human capacities if we had the possibility to conduct unrestricted experiments on humans. In Delgado's early years, others also shared a more liberal attitude towards the brain. They defended the idea that the brain is an open terrain, and were willing to explore its potential using different means, including LSD and other psychoactive drugs. The psychoanalyst Lawrence Kubie[14] claimed that normal life is restraining us from expressing the things we usually express. Michael Polanyi[15] argued that we know more than we can say, but are unable to access this knowledge – are we fully using our potentialities? T.H. Huxley[16] said that human beings belong to a different species because they resist natural selection and desire transcendence. These and others supported the idea that the mind has to be opened up.

As these endeavours fell out of favour, these days' neuroscience only tries to understand how the mind works, avoiding grander schemes of experimental exploration into mind enhancement.

[14] Kubie, L. S. (1958). Neurotic distortion of the creative process, University of Kansas Press.
[15] Polanyi, M. (1966). The Tacit Dimension, University of Chicago Press.
[16] Huxley, T. H. (1863). Evidence as to mans place in nature by Thomas Henry Huxley, Williams and Norgate.

Fuller suggested that we should re-evaluate the ways in which research programs are developed, and the scientific questions we allow ourselves to ask. We should not avoid exploring new paths for the human mind, beyond studying the biological mechanisms underlying established behaviour and observations.

Q&A Richard Holton, Molly Crockett, Steve Fuller

Q: Are there rules for flexibility? Does the plasticity of decisions keep people from destructive consequences?

Holton: We operate very close to optimum in our balance between flexibility and consistency. We know that making public our resolutions is often a replacement for actually following them, and that the more we dwell on the end result of resolutions, the less likely we are to follow them.

Q: How do you make decisions about rules?

Holton: "From the rule all authority derives" – When creating rules, we encapsulate feasible behaviour patterns that helped and guided us before. We draw a distinction between what we actually do and the rules that govern us. It is important to note that most rules are appropriate and helpful, and it is therefore preferable not to question them too often to make sure they are followed most of the time.

Q: Did you consider not shocking the anonymous participant?

Crockett: We were very careful not to deceive our participants, making sure that they knew that there was someone else there, but did not know their identity. This is an especially important experimental demand in behavioural economics, and we wanted to make sure that this study would reach this standard in order to ensure its robustness.

Q: Is hypnosis another case of remote controlling?

Fuller: Hypnosis, and the notion that everyone is suggestible, do indeed appear to be very intimidating ideas. People are averse to hypnosis because they ask: what happens if we are not aware of it?

Roundtable: The Future of Research on Agency, Morals & the Mind

Chair: Colin Blakemore

Professor of Neuroscience and Philosophy; Project Leader, The Human Mind Project, School of Advanced Study, University of London

All events of The Human Mind Project conclude with an Open Roundtable on the Future of Research. Discussion involves speakers, chairs and the public, and aims to draw out of the workshop key issues for the future of interdisciplinary research. Topics raised in the discussion form part of our Grand Challenges, an open consultation aiming to identify key questions facing new research on the mind.

How can questions related to agency and morality be tested in the laboratory setting? Is this enterprise too artificial to capture the complexity of the moral landscape? How to go beyond the lab to interpret our experiences in the real world? And is philosophy any better suited, given the intrinsic limitations of thought

experiments (the trolley problem). Are there intrinsic differences, beyond methodology, between computer simulations and mental experiments?

As Catherine Wilson argued, when asked to express a judgment about a trolley problem scenario, many people tend to behave less morally than they would do in real life situations. According to Steve Fuller, the main objection to the scientific and philosophical approach to agency and morality is not to be formulated in terms of ecological validity, but rather depends on the limits to what we can explore and develop. On the other side – claimed Molly Crockett – intrinsic limitations and artificiality are not good reasons to label this line of research as not serious enough. Our cognitive and social abilities have been proven to be powerful. They should not be so fragile as to break down as soon as we enter the lab. If we applied the same logic to the study of vision – where stimuli are artificially construed for studying specific properties of the visual system – we would not be able to say much about how the system works.

While we have evidence allowing us to say that these experiments work with vision, morals and the agency seem to be more complicated to address. It might seem that everything that is important disappears in the lab - added Patrick Haggard: very simple rules can explain extremely complex behaviours. An example is offered by the discussion surrounding Libet's experiments: there are no real choices, only about "when" to make an action. Despite this, it seems excessive to say that those experiments do not say anything relevant outside the lab. They have the merit of being able to explore sources of endogenous actions. Science has to simplify and we have to be careful in critiquing such simplicity. When Libet published his well-known article about "Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action",[17] many critics replied that he was unable to capture what we really mean by "free will". This is a clear example of how operational definitions might be useful: Libet's experiments provide an operational definition of endogenous action. Philosophers could certainly say that what we need from a theory of voluntary action is not offered by such operational definition. However, understanding the operational definition guiding an experiment allows us to capture what a scientist is trying to study.

What could be the role of philosophy be in this arena? We could draw parallels between conceptual analysis and experimental design. Philosophy could also provide support in the elaboration of relevant operational definitions, claimed Mattia Gallotti. And philosophers have devoted considerable work to science, through interpreting the large scale changes in scientific processes, as in the tradition of Thomas Kuhn. Is this enough? Is this role too ancillary? Does philosophy have to say something about why I should be moral and what the justification for morality is?

As Blakemore added, the real issue often comes from the theoretical interpretation of the experimental results rather than from the results by themselves. Speculations have to be followed by experimental tests, but how we can make the transition from the lab to real life is often unclear. Philosophy has an important role to play in the interpretation and criticism of the scientific results, and in interdisciplinary discussion. As Richard Holton argued, it is the rigid separation between philosophy and science that is rather artificial. During its long history, philosophy has been always interested in science, except for a limited period during the XIX century.

Could some discoveries about agency and morals change our attitudes or behaviour in the world? Some studies have shown that by telling people that free will does not exist you may elicit a change in behaviour.18

[17] Libet, B. (1985). Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action. Behavioral and Brain Sciences, 8:529-566.
[18] Vohs, K.D., Schooler, J.W. (2008). The value of believing in free will: encouraging a belief in determinism increases cheating. Psychol Sci., 19(1):49-54.

How much should we use this knowledge and its potential power? It may very difficult to answer this question, especially if the evidence goes against our moral assumptions - argued Lucy O'Brien.

Is it true, as Scott Atran claimed, that only field studies can provide us with a real knowledge of how people behave? What is added by the environment? What is the potential role of historians in understanding current political transformations? What can we learn from previous experience?

Are human beings special when it comes to morals and agency? As Haggard claimed, there are parts of our brain that deal with morality. Since these parts are continuous with those of non-human animals, we should not think that everything we usually recognise as belonging to agency and morality is specifically human. For example, we share with other animals the ability to perform endogenous actions, sense of agency, reinforcement learning, and the capacity to explore the environment.