ABSTRACTS BOOKLET

V INTERNATIONAL COLLOQUIUM ON PHILOSOPHY OF NEUROSCIENCE

DECISION MAKING AND RESPONSIBILITY

X MEETING OF THE ANPOF WORK GROUP ON PHILOSOPHY OF NEUROSCIENCE

RIO DE JANEIRO, APRIL 1-5 , 2024
HYBRID EVENT

V International Colloquium on Philosophy of Neuroscience: The Organizers, Themes, and Nature of the Event

This event is organized by the Working Group on Philosophy of Neuroscience, X-PHI, AI, and Neuroethics of ANPOF (National Association of Postgraduate Studies in Philosophy). The group brings together professors and postgraduate students who produce philosophical research based on empirical and scientific evidence.

The V International Colloquium on Philosophy of Neuroscience is an international meeting event for the exchange of knowledge among researchers conducting cutting-edge research, at the interface between philosophy of neuroscience, laboratory neuroscience, neuroethics, neurolaw, neuroeconomics, experimental philosophy, philosophy of science, and philosophy of mind. The ethical and legal relations surrounding the theme of Artificial Intelligence are also addressed, contemplating absolutely contemporary and pressing issues.

The event aims at developing the debate on the philosophy of neuroscience on Brazilian soil, generating an increase in research production and publication in the area at a qualitative and quantitative level, and aims to build partnerships with leading international institutions, enabling the attraction of financial resources and equipment from such partnerships, in addition to gaining academic excellence in research. Even though there are notable research efforts in the area in Brazil, an event of this magnitude allows us not only to expose the quality of these diligent investigative efforts to the scrutiny of internationally renowned researchers, who can collaborate with constructive criticisms of the national production, but also to provide national research with the gain of knowledge at a level of excellence from some of the world's greatest authorities in the area in question. ln this way, technical-scientific partnerships and international collaborations can also be enhanced, generating access to resources, equipment and funds. The contact between the national research community in the area and the leading international research community allows for the enhancement of theoretical and practical research opportunities in the area. Furthermore, researchers at the beginning of their careers and in development benefit from the discovery of new possibilities and international collaborations. ln addition to a general development of academic production in the area of philosophy of neuroscience, the specific themes of decision-making and responsibility and their ethical, moral and legal implications, pillars of Western civilization since the advent of monotheistic religions and which permeate the history of philosophy as classic problems, bring gains to the creation of public and private policies in the judicial, economic and public health areas, allowing an update of the debate on the rationality of action, from a well-informed point of view of the potentialities and limitations of human action involved in decision-making processes.

This year, the talks revolve around the issue of decision making and responsibility, a debate with profound practical, ethical, moral, and legal repercussions. The colloquium will feature speakers of international renown and great prestige, including Martha Farah, Dana Kay Nelkin, Jennifer Chandler, Robert Sapolsky, Julian Savulescu, Walter Sinnott-Armstrong, Eddy Nahmias, Thomas Nadelhoffer, Peter Tse, Roger Crisp, Samuel Murray, and Jonathan S. Phillips. With the insights of these eminent speakers, the colloquium aims to explore various aspects of responsibility and decision making. By fostering a multidisciplinary dialogue, the event seeks to advance our grasp of these complex issues, contributing to more nuanced and informed public and private policy-making in the face of contemporary challenges.

INDEX OF ABSTRACTS

ABSTRACTS

CARDOSO, RENATO CÉSAR (Universidade Federal de Minas Gerais)

TITLE – NEURORIGHTS AND THE NEUROSCIENCE OF FREE WILL

ABSTRACT – Due to the advent of modern neuroscience, several scientific disciplines have developed entirely new theories, perspectives, and methodologies. The substantial advances and discoveries made in this field over the last decades, especially those concerned with human cognition and behavior, have steered the course of many traditional research areas and given rise to others, like neuroethics and neurolaw. We will take a look at some of the general characteristics of the growing field of neurolaw, an interdisciplinary field that dwells on the intersection of law and neuroscience. We then discuss the neuroscience of free will, one of the most impacting and pressing topics in the neurolaw debate. Further attention will be directed to the neurorights debate and the growing support it is amassing in many countries. Finally, we will add some critical considerations on current legal propositions in this area - especially about the controversial "neuroright to free will".

CHANDLER, JENNIFER (University of Ottawa)

TITLE – THE ETHICS OF INFERRING MENTAL STATES FROM BRAIN DATA

ABSTRACT – The development and improvement of methods to collect and analyze data regarding brain activity is galvanizing discussion about the validity and consequences of inferring mental states from brain data. While we have long drawn inferences about mental states by observing each other, the collection and interpretation of brain data is newer. Among the concerns raised is whether this threatens "mental privacy" – a proposed novel "neuroright" now being considered by the United Nations Human Rights Council. Mental privacy is imperilled by access to brain data only to the extent that the data supports inferences about the mind (or is treated as if it does). The purpose of this

presentation is to look more closely at the nature of these inferences between brain data and mental states. The objective is to consider the actual and potential cases in which societies may find it useful to use brain data to access mental states, to set out the inferential steps between them, and to try to identify ethical considerations regarding these inferences, as well as the refusal to draw inferences that are imperfect but nonetheless weakly informative.

CRISP, ROGER (University of Oxford)

TITLE – REDUCTIONISM AND WHAT MATTERS IN SURVIVAL

ASBTRACT – Many current philosophers, often influenced by the arguments of Derek Parfit, incline towards a reductionist view of personhood. This paper, focusing on Parfit's famous case of My Division, discusses three related questions. First, given that an individual's relation to some future individual is most often a matter of degree, how should we understand what matters in the light of decreasing connectedness or continuity over time? Second, since well-being is good for a person, whose well-being is at stake in My Division? Finally, how do differences between accounts of well-being affect views on what matters in survival?

FARAH, MARTHA (University of Pennsylvania)

TITLE – THE NEUROSCIENCE OF SOCIOECONOMIC STATUS (SES): WHAT DO WE KNOW? WHAT ARE THE ETHICAL IMPLICATIONS?

ABSTRACT - In this talk I will address each of the subtitle questions. Recognizing that it is still early days for SES neuroscience, I will offer my best "educated guesses" to each question, and consider the sources of uncertainty that remain.

GOMES, GILBERTO (Universidade Estadual do Norte Fluminense)

TITLE – DETERMINISM, DECISION-MAKING AND RESPONSIBILITY.

ABSTRACT – Although determinism is popular among scientists and scientifically-minded philosophers, I will argue that, according to its standard definition, it is neither a

scientific fact, nor a scientific hypothesis, nor a scientific presupposition, but rather a pseudoscientific doctrine, or myth. Determinism may be contrasted with probabilistic causality, which will be argued for. The Epicurean argument that determinism is self-defeating will be revisited and supported. Finally, it will be argued that probabilistic causation offers a more fruitful basis for a concept of bounded free will, and for understanding decision making and responsibility, although it is insufficient for explaining them.

GOUVEIA, STEVEN S. (Universidade do Porto)

TITLE – THE RESPONSIBILITY GAP IN MEDICAL AI

ABSTRACT – The rapid evolution of Artificial Intelligence (AI) in medicine is reshaping today's healthcare landscape and how decision-making processes are done in medicine. With the aim of enhancing reliability, accuracy, efficiency, and cost-effectiveness, AI technologies are increasingly employed to augment or even replace human decision-making processes in medical contexts.

However, the reliance on complex and multifaceted data in AI systems often renders them as "black-boxes," where the internal workings remain opaque to practitioners. While these systems provide inputs and outputs, the inner mechanisms remain inaccessible, creating epistemic opacity and ethical concerns.

This opacity gives rise to a "Responsibility gap" in the patient-medical expert relationship, leading to uncertainty and ambiguity regarding the nature of the medical practice itself and who is accountable from a moral point of view.

In this talk, we will explore three approaches to addressing the challenges posed by black-box AI systems in healthcare for responsibility:

(1) Restricting or constraining the development and use of AI systems in healthcare;

(2) Acknowledging the benefits of black-box AI systems while disregarding the ethical consequences on responsibility;

(3) Advocating for the adoption of "Explainable/Transparent" Artificial Intelligence.

By examining the merits and drawbacks of each approach in bridging the "Responsibility gap," we aim to identify key considerations essential for reconciling responsibility and algorithmic processes in healthcare.

MAOZ, URI (Chapman University)

TITLE – WHY NEUROSCIENCE AND PHILOSOPHY NEED EACH OTHER IN THE STUDY OF VOLITION

ABSTRACT – How thoughts, ideas, and plans turn into actions in the real-world is one of the longest-standing discussions of human scholarship, especially in relation to the lingering debate on free will. In this talk I will discuss the contribution of neuroscience to this discussion and debate, starting from the Libet experiment. Though a main focus will be on how collaboration between neuroscientists and philosophers was central to understanding key problems in the Libet experiment and its interpretation and to moving the field beyond that experiment. In that I will also highlight the potential benefits and costs of such an interdisciplinary collaboration.

MOGRABI, GABRIEL (Universidade Federal do Rio de Janeiro)

TITLE – GLOBAL ARBITRATION IN THE BRAIN: VARIOUS LEVELS OF DECISION MAKING AND RESPONSIBILITY

ABSTRACT – In this talk, I contend that decision making must be understood in naturalistic terms. I propose a very brief philosophical account of “decision making” that has support from empirical data. I will mainly focus on the two issues: levels of self-control and selfhood in decision making. In a consilient manner, I defend that “Ecological Relevance” (the capability of an experiment to mimic, emulate or simulate, in a lab setting, situations closer to our contexts in daily life) is mandatory to address the costs

and values of human ethically and practically relevant decisions. Experimental designs are taken into consideration to make this point. I also want to put forward the concept of “Global Arbitration in the Brain” which points to the set of many types of selective and disjunctive mechanisms the brain has to cope with it tasks.

MOLL NETO, JORGE (Instituto D'Or de Ensino e Pesquisa)

TITLE - NEURAL BASIS OF MORAL JUDGMENT, ALTRUISTIC MOTIVATIONS AND IMPLICATIONS FOR AI

ABSTRACT – The neurological basis of moral judgments, social values and their cognitive and affective dimensions are inherently intertwined, thus challenging an often-perceived dichotomy between reason and emotion. This creates important needs and opportunities for ethical systems. The brief history of neuroscience and morality will be reviewed, as well as recent developments on modification of cognitive-emotional processes in healthy volunteers and patients. Finally, in a provocative exercise, we will ask whether there are fundamental human aspects of human morality and judgment that will remain opaque to AI.

NADELHOFFER, THOMAS (College of Charleston) **& MURRAY, SAMUEL** (Providence College)

TITLE – COMMONSENSE MORALITY AND THE BEARABLE AUTOMATICITY OF BEING

ABSTRACT – There is a large body of research which shows that moral behavior can be strongly influenced by trivial features of the environment of which we are completely unaware. Consider, for instance, the Watching Eye Effect. Multiple studies (both in the laboratory and in the field) have found that priming people with images of eyes can make them, among other things, not only more likely to be charitable (Kelsey, Vaish, & Grossmann, 2018; Fathi, Bateson, & Nettle, 2014), cooperative (Ernest-Jones et al., 2011), and generous (Burnham & Hare, 2007; Haley & Fessler, 2005), but it can make people less likely to litter (Ernest-Jones, Nettle, and Bateson, 2011) or steal a bicycle (Nettle, Nott, & Bateson, 2012). Evidence for the Watching Eye Effect has even been found in young children (Kelsey, Grossmann, & Vaish, 2012).

There is a lively ongoing debate among psychologists and philosophers about whether findings like these challenge or undermine our commonsense understanding of agency and responsibility. Some draw skeptical conclusions from the findings, arguing that the unwitting influence of environmental stimuli sems to undermine our autonomy and diminish the role of practical reasoning in action. While responses have been offered to these arguments, both sides have proceeded on the basis of untested assumptions about commonsense conceptions of agency and responsibility, such as the importance of consciousness in practical deliberation. Thus, rather than take a stand in the debate about the implications of the scientific evidence, we instead discuss the results of four vignette-based studies designed to investigate how laypersons think about the metaphysical and moral implications of scientific findings like the ones mentioned above. These results show that most participants do not draw skeptical conclusions about agency and responsibility when presented with evidence for phenomena like the Watching Eye Effect. Before discussing our studies—which we believe are the first to explore this issue—we will set the stage with an overview of the relevant scientific findings and the associated philosophical arguments.

NAHMIAS, EDDY (Georgia State University)

TITLE – What Robots can Teach us about Free Will: Why Consciousness Matters

ABSTRACT – Many philosophers and scientists, and most ordinary people, think that consciousness is essential for free will. Few tell us why, and when they do, it's often mysterious or unclear what kind of consciousness matters or why. I will explore these questions by considering what might convince us that robots have free will—what "flips our switch" to see them as persons, not mere programs. The answer suggests a plausible account of why consciousness matters for free will. We will also discuss whether or not we ever could make such robots and whether or not we should.

NELKIN, DANA KAY (University of California, San Diego)

TITLE – THRESHOLD DEONTOLOGY AND THE DISTRIBUTION OF RISK: ASKING THE RIGHT QUESTIONS

ABSTRACT – In this presentation, I set out some of the UC San Diego Moral Judgments Project's recent and new work on implicit moral theorizing. Systematic psychological experiments have been constructed that purport to sort instances of reasoning into an implicit consequentialist moral theory (according to which only consequences are morally relevant) or a non-consequentialist moral theory (according to which other considerations, such as rights, always override the maximization of good consequences). In our work (Ryazanov et al (2020) and (in preparation)), we show that asking participants questions in which we vary the ratio and probability of outcomes reveals that people may be appealing to more subtle non-consequentialist theory known as threshold deontology. According to that theory, rights matter and can override the maximization of consequences in moral decision-making, but they do not always do so. In this presentation, I set out some of that work, and then present several unpublished studies that ask participants comparative questions that involve not just comparing acting in such a way that there is a risk of killing vs. allowing to die, but how the risk is distributed. As we show, asking this kind of comparative question offers intriguing results that do not emerge when the standard sorts of questions are asked.

OLIVEIRA, NYTHAMAR (Pontifícia Universidade Católica – RS)

TITLE – RECONCILING CAUSAL DETERMINISM AND MORAL RESPONSIBILITY: RECASTING THE COMPATIBILIST DECISION-MAKING THESIS

ABSTRACT – Since the times of Hobbes, Spinoza, and Kant, philosophers have been trying to make sense of compatibilism, broadly understood as the belief that free will (entailing moral responsibility) and causal determinism are mutually compatible and that it is possible to believe in both without being logically inconsistent. This metaphysical thesis has been recast nowadays as the Compatibilist Decision-Making Thesis, supported by neuroscientific experiments and findings. Following Michael Shadlen and Adina Roskies (2012) I would like to argue that the neural mechanisms that underlie decision-making might help us not only distinguish agents from each other, but also realize that their multiple agential states conditioned by physical, biochemical, and social states allow indeed for various aspects of decision-making processes that could bridge the neurobiology of decision-making (NBDM) and the philosophical problems in ethics and

metaphysics. By revisiting this compatibilist thesis, I aim to fill in what I have dubbed the phenomenological deficit of normativity and naturalism, to avoid both normativism (in moral philosophy and metaphysics) and reductionist approaches to neuroscience, human nature, and the natural sciences. Insofar as it unveils the irreducibility of first-person, attitudinal accounts, including moral beliefs and belief in free will, the Compatibilist Decision-Making Thesis thus recast allows for justifying enactive, embodied, and situated approaches to neuroscience and artificial intelligence within a Spinozan-inspired supervenience monism and a Rawlsian-like wide reflective equilibrium, without any specific substantive commitment to moral, metaphysical or comprehensive doctrines overall.

PHILLIPS, JONATHAN S. (Dartmouth College)

TITLE – THE ROLE OF CONSCIOUS CONTROL IN OPTION GENERATION

ABSTRACT – Empirical studies of decision-making typically utilize paradigms where people are given a highly constrained set of options (usually two) to choose between. However, most of the actual choices we make, from mundane ones like what to have for lunch to profoundly important ones like the choice of a career, do not work like this. Instead they require us to choose from poorly defined and often overwhelmingly large sets of options (e.g., every possible combination of foods one could eat for lunch or every possible career). This is a critical aspect of real-world decision making that has been vastly understudied in both the free will and decision-making literatures. I will discuss some ongoing work investigating the role of conscious control in generating sets of options for problems of real-world complexity.

SAPOLSKY, ROBERT (Stanford University)

TITLE – LIFE WITHOUT FREE WILL (FROM A BIOLOGIST'S PERSPECTIVE)

ABSTRACT – While people have been debating the free will question for millennia, the biological sciences have only recently entered into the debate. An extensive biological literature now shows that each of us is nothing more than the sum of our biology, over which we had no control, and its interactions with environment, over which we also had no control. Prof. Sapolsky will present the case for this conclusion; just as importantly, he will argue that it will be a wonderful thing for society if people rejected the notion of free will.

SAVULESCU, JULIAN (University of Oxford)

TITLE – RESPONSIBILITY AND LARGE LANGUAGE MODELS

ABSTRACT – I will outline the concept of moral responsibility and apply it to the use of Large Language Models, particularly within health care. I will outline the normative conditions for praise and blame relating to the decision to employ or not employ them, as well as when harm results. I will review empirical research of ordinary people's ascription of praise and blame towards the use of such models. I will also consider impact of personalized Large Language Models for moral responsibility and ascription of praise and blame.

Large Language Models, and AI generally, offer the opportunity for great progress but also existential harm. We need to get clear who is responsible and this will require new norms and regulations.

SINNOTT-ARMSTRONG, WALTER (Duke University)

TITLE – MORAL AI AND HOW WE GET THERE

ABSTRACT – Artificial intelligence (AI) is beginning to be used for many life-changing decisions in medicine, law, transportation, the military, and other areas. Critics object that using AI in these areas is too likely to lead to harm, unfairness, and other moral wrongs. In response, I will argue that these decisions can be made safer and more ethical by building human moral values into the AI decisionmaker. But how can we do that? I will discuss problems for some proposed ways to build morality into AI from the top down and from the bottom up. Then I will explain our lab's novel hybrid alternative, which surveys human moral judgments and then corrects for ignorance, confusion, and partiality. Because our approach is based on idealized observer theories in ethics, it minimizes substantive assumptions about what is morally right or wrong, and it can be used in a wide variety of contexts. I will report initial empirical results using our method and discuss potential applications to kidney allocation, dementia, criminal law, and the military. I will also show how our method can contribute to moral psychology as an academic field by leading to a deeper understanding of the computations behind human moral judgments and decisions.

TSE, PETER ULRIC (Dartmouth College)

TITLE – FREE IMAGINATION, SECOND-ORDER FREE WILL AND ACTS OF WILD SELF-CREATION

ABSTRACT – I will argue that free will is not a matter of action primarily, but instead one of mental operations, particularly volitional mental operations that take place in working memory, colloquially known as imagining. Even someone with locked-in syndrome could choose to attend to the sound of a radio or a TV or an internal recollection. Some mental operations allow us to imagine a different set of circumstances than are currently the case that hold in the imagined future, which we can then set about trying to make happen. Some of these acts of willing involve imagining a new self and then setting about instantiating that new self, say, one that can speak Swahili, or even one with a different character. I will talk about work in my Cognitive Neuroscience lab that tries to uncover how such acts of self-re-creation happen in the brain at the neural level, whether in terms of language learning, changing how we perceive the world, or automatizing previously volitionally effortful mental or motoric acts. This work hearkens back to the virtue ethics of Aristotle, in that mastery of some domain, whether the piano or the self, is not a matter of virtue, but is instead a matter of habit formation and automatization, in the sense of presently willed losses of future willing.