Advances in Neuroethics
Series Editors: V. Dubljević · F. Jotterand · R.J. Jox · E. Racine

Orsolya Friedrich · Andreas Wolkenstein
Christoph Bublitz · Ralf J. Jox
Eric Racine *Editors*

Clinical Neurotechnology meets Artificial Intelligence

Philosophical, Ethical, Legal and Social Implications

Springer

Advances in Neuroethics

Series Editors

Veljko Dubljević
North Carolina State University
Raleigh, NC
USA

Fabrice Jotterand
Medical College of Wisconsin
Milwaukee
USA

Ralf J. Jox
Lausanne University Hospital and University of Lausanne
Lausanne
Switzerland

Eric Racine
IRCM, Université de Montréal, and McGill University
Montréal, QC
Canada

Advances in neuroscience research are bringing to the forefront major benefits and ethical challenges for medicine and society. The ethical concerns related to patients with mental health and neurological conditions, as well as emerging social and philosophical problems created by advances in neuroscience, neurology and neurotechnology are addressed by a specialized and interdisciplinary field called neuroethics.

As neuroscience rapidly evolves, there is a need to define how society ought to move forward with respect to an ever growing range of issues. The ethical, legal and social ramifications of neuroscience, neurotechnology and neurology for research, patient care, and public health are diverse and far-reaching — and are only beginning to be understood.

In this context, the book series “Advances in Neuroethics” addresses how advances in brain sciences can be attended to for the benefit of patients and society at large.

More information about this series at http://www.springer.com/series/14360

Orsolya Friedrich • Andreas Wolkenstein
Christoph Bublitz • Ralf J. Jox • Eric Racine
Editors

Clinical Neurotechnology meets Artificial Intelligence

Philosophical, Ethical, Legal and Social Implications

Springer

Editors
Orsolya Friedrich
Institute of Philosophy
FernUniversität in Hagen
Hagen
Germany

Christoph Bublitz
Faculty of Law
University of Hamburg
Hamburg
Germany

Eric Racine
University of Montreal and McGill University
Institut de recherches cliniques de Montréal
Montréal, QC
Canada

Andreas Wolkenstein
Institute of Ethics, History and Theory of Medicine
Ludwig-Maximilians-Universität (LMU) München
Munich
Germany

Ralf J. Jox
Institute of Humanities in Medicine and Clinical Ethics Unit
Lausanne University Hospital and University of Lausanne
Lausanne
Switzerland

ISSN 2522-5677 ISSN 2522-5685 (electronic)
Advances in Neuroethics
ISBN 978-3-030-64589-2 ISBN 978-3-030-64590-8 (eBook)
https://doi.org/10.1007/978-3-030-64590-8

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book has a somewhat longer history and many helping hands were needed to realize it. Therefore, we would like to thank all persons who contributed their share so that this volume could finally be published. Since most of the articles collected in this book originate from a conference entitled “Neurotechnology meets Artificial Intelligence. Ethical, social and legal implications of neurotech and AI” (held in Munich, May 8–10, 2019), we wish to thank all those who made the conference the refreshing, inspiring, thought-provoking, and enjoyable event that it was. The conference brought together a wide range of scholars with various disciplinary backgrounds (philosophy, law, social science, cognitive sciences, medicine) to discuss the multidimensional implications of neurotechnology and AI. It was mainly the outcome of Johannes Kögel’s impressive organizing capabilities that the participants were able to experience a great conference, both academically and socially. The organizing team was supported by Nicola Williams, Natalie Kopczewski, and Armin Gruber who tirelessly helped in the background. Further, we would like to thank Georg Marckmann and the Institute of Ethics, History and Theory of Medicine at LMU Munich. Georg is head of the institute and continuously supported the conference and all activities around the INTERFACES project. Most importantly, we extend our gratitude to all speakers and authors without whose inspiring talks, smart contributions to the debates, and interest in the many facets of the conference’s subject no such event would have been possible.

With regard to the realization of this book, we are deeply grateful to Meliz Kaygusuz and Bernadette Scherer. Due to their efforts we were able, among others, to overcome so many technical hurdles in the manuscript preparation and find all those tiny sources of potential errors that a book project usually hides. We are thankful for the proofreading services that Dorothea Wagner von Hoff has provided us with. Dorothea found many interesting, but clearly erroneous combinations of words that would have rendered some parts of the book extremely hard to read.

Finally, we would like to thank Sylvana Freyberg and the Springer team for their interest in publishing our book with them. They had to put a lot of patience in the project, so we are especially grateful for their continuous trust and interest. Moreover, the editors of the Springer series “Advances in Neuroethics”—Veljko Dubljevic, Fabrice Jotterand, Ralf J. Jox, and Eric Racine—accepted the inclusion of our book in the series, for which we are also very thankful.

Work on the book was funded by the Federal Ministry of Education and Research (BMBF) in Germany (INTERFACES, 01GP1622A) and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, 418201802), which we highly appreciate.

We now hope that readers find many important insights, points to consider, food for thought, and inspirations for their own work in the pages to come.

Hagen, Germany — Orsolya Friedrich
Munich, Germany — Andreas Wolkenstein

Contents

1 **Introduction: Ethical Issues of Neurotechnologies and Artificial Intelligence** 1
Orsolya Friedrich and Andreas Wolkenstein

2 **Actions, Agents, and Interfaces** 11
Tom Buller

3 **Skilled Action and the Ethics of Brain-Computer Interfaces** 25
Sebastian Drosselmeier and Stephan Sellmaier

4 **Augmenting Autonomy Through Neurotechnological Intervention à la Kant: Paradox or Possibility?** 39
Anna Frammartino Wilks

5 **Can BCIs Enlighten the Concept of Agency? A Plea for an Experimental Philosophy of Neurotechnology** 55
Pim Haselager, Giulio Mecacci, and Andreas Wolkenstein

6 **Brain-Computer Interfaces: Current and Future Investigations in the Philosophy and Politics of Neurotechnology** 69
Andreas Wolkenstein and Orsolya Friedrich

7 **Pragmatism for a Digital Society: The (In)significance of Artificial Intelligence and Neural Technology** 81
Matthew Sample and Eric Racine

8 **Brain-Computer Interface Use as Materialized Crisis Management** 101
Johannes Kögel

9 **The Power of Thoughts: A Qualitative Interview Study with Healthy Users of Brain-Computer Interfaces** 117
Jennifer R. Schmid and Ralf J. Jox

10 **Diffusion on Both Ends: Legal Protection and Criminalisation in Neurotechnological Uncertainty** 127
Susanne Beck

11 Data and Consent Issues with Neural Recording Devices 141
Stephen Rainey, Kevin McGillivray, Tyr Fothergill, Hannah Maslen, Bernd Stahl, and Christoph Bublitz

12 Ethical Implications of Brain-Computer Interface and Artificial Intelligence in Medicine and Health Care 155
Ralf J. Jox

13 Practical, Conceptual and Ethical Dimensions of a Neuro-controlled Speech Neuroprosthesis 169
Stephen Rainey

14 The Emperor's New Clothes? Transparency and Trust in Machine Learning for Clinical Neuroscience 183
Georg Starke

15 The Security and Military Implications of Neurotechnology and Artificial Intelligence 197
Jean-Marc Rickli and Marcello Ienca

16 Connecting Brain and Machine: The Mind Is the Next Frontier 215
Mathias Vukelić

Contributors

Susanne Beck Institute for Criminal Law and Criminology, Leibniz University Hanover, Hanover, Germany

Christoph Bublitz Faculty of Law, University of Hamburg, Hamburg, Germany

Tom Buller, PhD Department of Philosophy, Illinois State University, Chicago, IL, USA

Sebastian Drosselmeier Munich Graduate School for Ethics in Practice, Ludwig-Maximilians-Universität (LMU) München, Munich, Germany

Tyr Fothergill De Montfort University, Leicester, UK

Orsolya Friedrich Institute of Philosophy, FernUniversität in Hagen, Hagen, Germany

Pim Haselager, PhD Donders Institute for Brain, Cognition, and Behaviour, Department of Artificial Intelligence, Radboud University, Nijmegen, The Netherlands

Marcello Ienca, PhD Department of Health Sciences and Technology (D-HEST), Swiss Federal Institute of Technology, ETH Zurich, Zurich, Switzerland

Ralf J. Jox Institute of Humanities in Medicine and Clinical Ethics Unit, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland

Johannes Kögel Institute of Ethics, History and Theory of Medicine, Ludwig-Maximilians-Universität (LMU) München, Munich, Germany

Hannah Maslen University of Oxford, Oxford, UK

Kevin McGillivray University of Oslo, Oslo, Norway

Giulio Mecacci, PhD Department of Artificial Intelligence, Radboud University, Nijmegen, The Netherlands

Department of Values, Technology and Innovation, Faculty of Technology, Policy and Management, TU Delft, Delft, The Netherlands

Eric Racine, PhD Pragmatic Health Ethics Research Unit, Institut de recherches cliniques de Montréal, Montreal, QC, Canada

Department of Neurology and Neurosurgery, McGill University, Montreal, QC, Canada

Department of Experimental Medicine (Biomedical Ethics Unit), McGill University, Montreal, QC, Canada

Department of Medicine and Department of Social and Preventive Medicine, Université de Montréal, Montreal, QC, Canada

Stephen Rainey, PhD Oxford Uehiro Centre for Practical Ethics, University of Oxford, Oxford, UK

Jean-Marc Rickli, PhD Geneva Centre for Security Policy (GCSP), Geneva, Switzerland

Matthew Sample, PhD Pragmatic Health Ethics Research Unit, Institut de recherches cliniques de Montréal, Montreal, QC, Canada

Department of Neurology and Neurosurgery, McGill University, Montreal, QC, Canada

Center for Ethics and Law in the Life Sciences, University of Hannover, Hannover, Germany

Jennifer R. Schmid Institute of Ethics, History and Theory of Medicine, Ludwig-Maximilians-Universität (LMU) München, Munich, Germany

Stephan Sellmaier Research Center for Neurophilosophy and Ethics of Neurosciences, Ludwig-Maximilians-Universität (LMU) München, Munich, Germany

Bernd Stahl De Montfort University, Leicester, UK

Georg Starke Institute for Biomedical Ethics, University of Basel, Basel, Switzerland

Mathias Vukelić, PhD Fraunhofer Institute for Industrial Engineering IAO, Stuttgart, Germany

Anna Frammartino Wilks, PhD Department of Philosophy, Acadia University, Wolfville, NS, Canada

Andreas Wolkenstein Institute of Ethics, History and Theory of Medicine, Ludwig-Maximilians-Universität (LMU) München, Munich, Germany

Introduction: Ethical Issues of Neurotechnologies and Artificial Intelligence

1

Orsolya Friedrich and Andreas Wolkenstein

Contents

1.1 Neurotechnology + Artificial Intelligence = Intelligent Neurotechnologies (INT)......... 1
1.2 Novel Philosophical, Ethical, Legal, and Sociological Approaches to INT: An Overview.......... 3
References.......... 7

Abstract

In this introduction to the volume, we present an overview of existing research on intelligent neurotechnologies, i.e., the combination of neurotechnologies with Artificial Intelligence (AI). Further, we present the ideas behind this volume and an overview of each chapter.

1.1 Neurotechnology + Artificial Intelligence = Intelligent Neurotechnologies (INT)

Imagine that the coffee machine in your kitchen starts brewing your urgently needed morning coffee as soon as you *think* the command "start the coffee machine" while you are still in bed. Is that realistic? Is it desirable? Using neurotechnologies, i.e.,

O. Friedrich (✉)
Institute of Philosophy, FernUniversität in Hagen, Hagen, Germany
e-mail: orsolya.friedrich@fernuni-hagen.de

A. Wolkenstein
Institute of Ethics, History and Theory of Medicine, Ludwig-Maximilians-Universität (LMU) München, Munich, Germany
e-mail: andreas.wolkenstein@med.uni-muenchen.de

O. Friedrich et al. (eds.), *Clinical Neurotechnology meets Artificial Intelligence*, Advances in Neuroethics, https://doi.org/10.1007/978-3-030-64590-8_1

technologies that lead to understanding, changing or interacting with the brain, combined with artificial intelligence (AI) might allow for such an application, even though many scientists doubt that technologies such as this one could be available in the near future. However, basic principles of brain-computer interfacing (BCI) have become reality and are currently the subject of intense research efforts [1–4]. BCIs measure brain activity and convert brain signals into computer commands, e.g., moving a cursor or a wheelchair [5, 6]. The most common way to measure brain activity is with non-invasive electroencephalography (EEG). BCIs use the power of thought or of focusing on a signal in order to give computational commands and require no neuromuscular innervation.

At the same time, BCIs and other neurotechnologies stand in relation with another emerging technology: AI. AI is already being used in many technologies to solve problems, which usually require human intelligence, such as reasoning, planning, and speech perception [7]. It is not a technology designed for a specific task, but cuts across all societal domains [8, 9] and comprises several technologies such as machine learning and artificial neural networks. The term "AI" thus denotes a variety of converging technologies that are used across many platforms and technologies. Kellmeyer [10] lists five different aspects: ubiquitous data collection, storage and processing of large amounts of data (big data), high performance analysis, machine learning, and interfaces for human-AI interaction.

AI is used in a number of ways in neuroscience and neurotechnology in the medical domain [11]. For example, computer vision capacities are being applied to detect tumors in magnetic resonance imaging (MRI) [12] or to detect anomalies in other kinds of data [13], e.g., EEG data [14–16]. These capacities lead to an improved diagnosis, prediction, and treatment of clinical pictures in a variety of medical domains [10]. In psychiatry, researchers have recently used AI to reach a biomarker-based diagnosis and determine therapy in patients with dementia, attention deficit hyperactivity disorder (ADHD), schizophrenia, autism, depression, and posttraumatic stress disorder (PTSD) [17–20]. AI that is used for speech recognition, in addition to many available data sources on the internet, helps researchers predict mental illness, for example [21].

Beyond its application in clinical research and therapy, AI is being used in combination with neurotechnologies. Big data and deep learning, for example, are promising trends that will influence the development of BCIs [22]. Among many other uses, these devices can be used by patients who suffer from amyotrophic lateral sclerosis (ALS) or severe paralysis in order to restore communication capacities and mobility, or in rehabilitation to facilitate the recovery process of patients after stroke [23–25]. With the help of AI, important BCI features such as signal processing and feature extraction can be improved [22]. Outside the strictly medical arena, EEG-based BCIs and other forms of AI-based neurotechnology are sold for entertainment purposes [26]. Facebook famously works with a typing-by-brain technology, which allows for a seamless social media experience [27]. Research behind this technology was already capable of showing how algorithms could decode speech in real time with a high amount of reliability [28]. Similarly, progress has been made in terms of facial recognition in EEG data [29]. BCIs, as well as other applications of (AI-enhanced) neurotechnology can also be found in military research. Warfighter enhancement is one motivation, but others include enhancing military equipment or deception detection [30–33].

In addition to technological development and progress, the number of articles, books, and events such as workshops or conferences that deal with the neuroethics of AI and neurotechnology is steadily increasing. Generally speaking, AI raises a host of original problems that can most aptly be summarized as "black box"-problems: It becomes increasingly difficult to supervise and control an AI's operation, because it manages its decision-making logic all by itself [34–37]. The combination of neurotechnologies and AI raises a host of further pressing problems. Yuste and colleagues [38] mention four broad areas of ethical concern: privacy and consent, agency and identity, augmentation, as well as bias. They propose various measures to address these issues, ranging from technological safeguards to legislation. For medical neurotechnology, a number of articles also emphasized problems regarding data protection and privacy as important issues to consider [39]. Moreover, questions of responsibility and shared agency are repeatedly brought up when it comes to neurotechnologies [40]. How BCIs affect agency and autonomy is another topic that drew attention to philosophers and ethicists [41, 42]. This body of research adds to more general approaches that examine the ethical quality of algorithms per se [9, 43]. Articles on issues such as hackability and problems derived from unwanted access to brain data [44] complement work that looks at specific forms of neurotechnology, e.g., in the medical, military, or consumer area [32, 33, 45, 46]. In addition, neurotechnology becomes increasingly interesting for political philosophers and others who approach INT with an eye on regulation questions and broader democratic worries [39, 47].

1.2 Novel Philosophical, Ethical, Legal, and Sociological Approaches to INT: An Overview

As this brief overview shows, many questions have already been addressed in the emerging literature both on technical issues and the normative implications of INT. Some of these questions have not been sufficiently or satisfyingly answered. Scholars from philosophy, sociology, and the law continue to exchange arguments and ideas while medical researchers, engineers, and computer scientists keep exploring new technologies and improve existing ones. The aim of this book is to provide a forum for the continuous exchange of these arguments and ideas. From a philosophical and ethical perspective, normatively relevant notions such as agency, autonomy, or responsibility have to be analyzed if humans interact with INT. This volume also asks, in a descriptive manner, how the reality of using INT would look like. It sheds light on the legal dimensions of INT. In addition, it explores a number of specific use cases, in that these concrete scenarios reveal more about the various domains of human agency in situations where technology and human-machine interaction play a distinctive role.

Accordingly, the methods used in this book vary considerably. They range from philosophical analysis, sociologically inspired descriptions, legal analysis, and socio-empirical research. This provides the book with the capacity to address a wide range of philosophical, normative, social, legal, and empirical dimensions of neurotechnology and AI. Most of the papers of this volume are the result of a conference that was held in Munich, in which the ethics of (clinical) neurotechnologies and AI were intensely discussed.[1]

The *first section* of the book reflects on some philosophically relevant phenomena and implications of neurotechnology use. From a philosophical and ethical perspective, it must be asked how normatively relevant notions such as action, agency, autonomy, or responsibility can be conceptualized if humans act and interact with neurotechnologies. The most basic question is if BCI effects are actions at all and if there are normatively relevant differences between paradigmatic bodily actions and BCI-mediated actions. If there is no action or agency to be claimed, subsequent issues of autonomy and responsibility are affected, as well. Therefore, philosophical analyses of BCI use that focus on action-theoretical implications have emerged recently [41]. Two articles in this first section take this path.

Tom Buller analyzes the implications of BCI use for the nature of action. He claims that present BCI-mediated behavior fails to meet the necessary condition of intentional actions, namely the causation of an event and thus of bodily movement that is directly related to relevant beliefs and desires. Furthermore, he states that current BCI-mediated changes in the world do not qualify as non-deviant causal processes.

Sebastian Drosselmeier and *Stephan Sellmaier* also address the issue of action. However, they focus on the acquisition of a skill while using BCIs, which allows the user to make BCI-mediated changes in the world without performing a mental act. This would result—according to their argumentation—in the ability to perform BCI actions as basic actions. They also conclude that BCI users are able to differentiate between having a thought and an action relevant intention. Therefore, skilled users should be seen as competent and able to voluntarily control the BCI effects, which they cause in the world.

The concepts of action and agency are closely connected to the concept of autonomy. Therefore, this suggests that some authors have recently also addressed the implications of BCI use on autonomy [42]. The first section of this volume also deals with this issue. Realizing the ability to act autonomously might be hampered or enhanced by using neurotechnologies.

Anna Wilks takes a closer look at the question of whether it would be a paradox or a possibility, following Kant, to augment autonomy through neurotechnologies. The paradox seems obvious at first hand: someone claims to augment autonomy with BCI use, but is able to perform self-legislation, whereas autonomous agency in a Kantian understanding requires that the person is not affected by external factors. Wilks, however, suggests that operating with a broader Kantian framework would

[1] https://neurotechmeetsai.wordpress.com/

allow integrating external components of BCIs into the understanding of self-legislation and thus avoid the paradox.

Pim Haselager, *Giulio Mecacci,* and *Andreas Wolkenstein* argue that BCIs, especially passive BCIs, shed new light on the traditional question of agency in philosophy. More precisely, they argue that the notion of ownership of action ("was that me?") might be affected by closely examining the action-theoretical implications of passive BCIs. If BCIs register intentions without the user being aware of this, and if they consequently act on them, then subconscious brain states may influence one's actions in a technology-mediated way. This observation serves as the basis for their plea to use passive BCIs, or what they call symbiotic technology, in experimentally guided thought experiments aimed at the study of the notion of agency. The authors suspect that by doing so, symbiotic technology may give new answers to how we must understand ownership of action and what consequences we have to expect.

Andreas Wolkenstein and *Orsolya Friedrich* contribute to the first section of the volume by summarizing the philosophical and ethical analysis that they described in their BCI-use analyzing project (Interfaces) and suggest some future directions for research and regulation of BCI development and use. They show that relevant results have been produced in recent philosophical, ethical, social, and legal reflections of BCI use. However, concluding results that could profoundly advise technology-regulating institutions or engineers are not present yet. Nevertheless, the development of AI-driven neurotechnologies are emerging and therefore, some preliminary ethically based regulatory framework is necessary. They suggest using procedural criteria as a first step.

Neurotechnology and AI also have broad social implications. These social implications not only include societal issues in general; certain areas of society, like research and medicine, are affected in a specific way. The *second section* of this volume focuses on some social implications of neurotechnology and AI use.

Matthew Sample and *Eric Racine* recall in their article that other emerging technologies, e.g., genomics or nanotechnology, have been promoted in ethics research in the past similar to the way that neural technologies are now. They address the question of how ethics researchers should deal with such research developments and question the significance of digital society for ethics research. They show how the significance of artificial intelligence and neural technologies, as examples of digital technologies, is affected by both sociological and ethical factors. They conclude that ethics researchers have to be careful in attributing significance and to reflect their own function in the process of attribution.

Johannes Kögel also focuses on BCI use from a sociological perspective. He shows that the BCI laboratory is not only a place to train this novel technology, but also a place of crisis management. The aim to discuss BCI use also as crisis management is to understand this social process and to increase sensitivity for the user experience. He argues that users currently experience BCI training and tasks as tedious and exhausting, because they have to make many "back-to-back decisions" for a long period of time and under immense time pressure, which is not common to activities in everyday life. His focus emphasizes the importance of developing BCI applications that allow for a more routine way of acting.

Jennifer R. Schmid and *Ralf J. Jox* further highlight the relevance and implications of the training process for the user experience in BCIs. They report on a qualitative interview study with healthy BCI users, e.g., neuro-gamers or pilots. The interviews show that the success of BCI use strongly depends on the motivation as well as the duration of training and that the time-consuming procedure of use results in discomfort and cognitive exhaustion.

This *second section* of this volume also approaches intelligent neurotechnologies from a legal perspective. The legal system faces the need to update some of its notions and regulatory action is needed to cover these new, neurotechnology-based forms of acting and acting together. BCIs also raise the question about mental privacy as well as data and consent issues.

Susanne Beck focuses on criminal law issues that result from neurotechnology use. She shows how neurotechnologies might lead to diffusion on the end of the victim, as well as the offender. Such diffusion would be important for criminal law, in that in traditional criminal law the roles of offender and victim are very clear. Therefore, criminalizing might lose some of its legitimacy. Another problematic diffusion in criminal law might occur, if there are no clear borders between the body and the mind.

Stephen Rainey et al. address further legally relevant issues, namely those related to data and consent in neural recording devices. They discuss whether current data protection regulation is adequate. They conclude that brain-reading devices present difficult consent issues for consumers and developers of the technology. They are also a potential challenge for current European data protection standards. Their use might become legally problematic, if the nature of the device results in an inability for the user to exercise their rights.

Finally, in the *third section* the book takes a closer look at neurotechnologies in their contexts of use. This section covers both the introduction of using neurotechnologies in various domains and an explication and discussion of their deeper philosophical, ethical, and social implications.

Ralf J. Jox discusses the ethical implications of the use of neurotechnologies and AI in the domain of medicine. He shows that such technology use challenges not only the patient–physician relationship, but also the whole character of medicine. He further highlights the potential threats to human nature, human identity, and the fundamental distinction between human beings and technological artifacts that could arise when AI technology with certain features is closely connected with the human brain.

The next contribution highlights one of these close connections of AI-neurotechnology and the human brain. *Stephen Rainey* discusses neuro-controlled speech neuroprosthesis from an ethical perspective. A speech neuroprosthesis picks out linguistically relevant neural signals in order to synthesize and realize, artificially, the overt speech sounds that the signals represent. The most important question in this special neurotechnology application is whether the synthesized speech represents the user's speech intentions and to what extent he can control the speech neuroprosthesis.

Georg Starke's contribution addresses another field of clinical neuroscience, namely the application of ML to neuroimaging data and the potential challenges of this application with regard to transparency and trust. He shows why transparency and trustworthiness are not necessarily linked and why transparency alone won't solve all the challenges of clinical ML applications.

Another field of application of neurotechnology and AI is their use in the military. *Jean-Marc Rickli* and *Marcello Ienca* discuss the security and military implications of neurotechnology and AI with regard to five security-relevant issues, namely data bias, accountability, manipulation, social control, weaponization, and democratization of access. They show that neurotechnology and AI both raise security concerns and share some characteristics: they proliferate outside supervised research settings, they are used for military aims, and they have a transformative and disruptive character. They highlight that it is extremely difficult to control the use and misuse of these technologies and call for global governance responses that are able to deal with the special characteristics of these technologies.

Finally, *Mathias Vukelić* directs our attention to a new research agenda for designing technology. Given the increasingly symbiotic nature of neurotechnology, where humans and technology closely interact, he emphasizes the need for a human-centered approach that puts human needs at the core. He attests that the detection of brain states, such as emotional or affective reactions, are of great potential for the development of symbiotic, interactive machines. Beyond assistive technology, this research leads to neuroadaptive technologies that are usable in a broad variety of domains. Vukelić argues that the primary goal of such an undertaking is the alignment of increasingly intelligent technology with human needs and abilities. While this could itself be viewed as following an ethical imperative, the author also stresses the wider ethical and societal implications of such a research agenda.

This short overview of existing research on intelligent neurotechnologies and of the articles in this volume offers a first insight into the emerging philosophical, ethical, legal, and social difficulties that we will have to face in the future and which require further conceptual as well as empirical research.

Acknowledgments Work on this paper was funded by the Federal Ministry of Education and Research (BMBF) in Germany (INTERFACES, 01GP1622A) and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—418201802. We would like to thank Dorothea Wagner von Hoff for *proof reading* the article, Meliz-Sema Kaygusuz and Bernadette Scherer for formatting.

References

1. McFarland DJ, Wolpaw JR. EEG-based brain-computer interfaces. Curr Opin Biomed Eng. 2017;4:194–200.
2. Clerc M, Bougrain L, Lotte F, editors. Brain-computer interfaces 1. Foundations and methods. London: Wiley; 2016.
3. Clerc M, Bougrain L, Lotte F, editors. Brain-computer interfaces 2. Technology and applications. London: Wiley; 2016.

4. Graimann B, Allison B, Pfurtscheller G, editors. Brain-computer interfaces. Revolutionizing human-computer interaction. Berlin: Springer; 2010.
5. Shih JJ, Krusienski DJ, Wolpaw JR. Brain-computer interfaces in medicine. Mayo Clin Proc. 2012;87(3):268–79.
6. Graimann B, Allison B, Pfurtscheller G. Brain-computer interfaces: a gentle introduction. In: Graimann B, Allison B, Pfurtscheller G, editors. Brain-computer interfaces: revolutionizing human-computer interaction. Berlin: Springer; 2010. p. 1–27.
7. Luxton DD, editor. Artificial intelligence in behavioral and mental health care. Amsterdam: Elsevier; 2016.
8. Mainzer K. Künstliche Intelligenz. Wann übernehmen die Maschinen? Berlin: Springer; 2016.
9. Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L. The ethics of algorithms: mapping the debate. Big Data Soc. 2016;3(2):1–21.
10. Kellmeyer P. Artificial intelligence in basic and clinical neuroscience: opportunities and ethical challenges. e-Neuroforum. 2019;25(4):241–50.
11. Gunes O, Gunes G, Seyitoglu DC. The use of artificial intelligence in different medical branches: an overview of the literature. Med Sci. 2019;8(3):770–3.
12. Pereira S, Pinto A, Alves V, Silva CA. Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Trans Med Imaging. 2016;35(5):1240–51.
13. Litjens G, Sánchez CI, Timofeeva N, Hermsen M, Nagtegaal I, Kovacs I, et al. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. Sci Rep. 2016;6(1):26286.
14. Tabar YR, Halici U. A novel deep learning approach for classification of EEG motor imagery signals. J Neural Eng. 2016;14(1):016003.
15. Schirrmeister RT, Springenberg JT, Fiederer LDJ, Glasstetter M, Eggensperger K, Tangermann M, et al. Deep learning with convolutional neural networks for EEG decoding and visualization. Hum Brain Mapp. 2017;38(11):5391–420.
16. Schirrmeister RT, Gemein L, Eggensberger K, Hutter F, Ball T. P64. Deep learning for EEG diagnostics. Clin Neurophysiol. 2018;129(8):e94.
17. Fakhoury M. Artificial intelligence in psychiatry. In: Kim Y-K, editor. Frontiers in psychiatry: artificial intelligence, precision medicine, and other paradigm shifts. Singapore: Springer Singapore; 2019. p. 119–25.
18. Vieira S, Pinaya WHL, Mechelli A. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: methods and applications. Neurosci Biobehav Rev. 2017;74:58–75.
19. Meyer-Lindenberg A. Künstliche Intelligenz in der Psychiatrie – ein Überblick. Nervenarzt. 2018;89(8):861–8.
20. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44–56.
21. Eichstaedt JC, Smith RJ, Merchant RM, Ungar LH, Crutchley P, Preoţiuc-Pietro D, et al. Facebook language predicts depression in medical records. Proc Natl Acad Sci. 2018;115(44):11203.
22. Lin C, Liu Y, Wu S, Cao Z, Wang Y, Huang C, et al. EEG-based brain-computer interfaces: a novel neurotechnology and computational intelligence method. IEEE Syst Man Cybernet Mag. 2017;3(4):16–26.
23. Chaudhary U, Birbaumer N, Ramos-Murguialday A. Brain–computer interfaces for communication and rehabilitation. Nat Rev Neurol. 2016;12(9):513–25.
24. Salisbury DB, Parsons TD, Monden KR, Trost Z, Driver SJ. Brain–computer interface for individuals after spinal cord injury. Rehabil Psychol. 2016;61(4):435–41.
25. McFarland DJ, Daly J, Boulay C, Parvaz MA. Therapeutic applications of BCI technologies. Brain Comput Interfaces. 2017;4(1–2):37–52.
26. Blankertz B, Acqualanga L, Dähne S, Haufe S, Schultze-Kraft M, Sturm I, et al. The Berlin brain-computer interface: progress beyond communication and control. Front Neurosci. 2016;10:1–24.

27. Robertson A. Facebook just published an update on its futuristic brain-typing project. The Verge 2019 July 30.
28. Moses DA, Leonard MK, Makin JG, Chang EF. Real-time decoding of question-and-answer speech dialogue using human cortical activity. Nat Commun. 2019;10(1):3096.
29. Nemrodov D, Niemeier M, Patel A, Nestor A. The neural dynamics of facial identity processing: insights from EEG-based pattern analysis and image reconstruction. eNeuro. 2018;5(1):ENEURO.0358-17.2018.
30. Evans NG, Moreno JD. Neuroethics and policy at the National Security Interface: a test case for neuroethics theory and methodology. In: Racine E, Aspler J, editors. Debates about neuroethics. Cham: Springer; 2017. p. 141–57.
31. Tennison MN, Moreno JD. Neuroscience, ethics, and national security: the state of the art. PLoS Biol. 2012;10(3):e1001289.
32. Kotchetkov IS, Hwang BY, Appelboom G, Kellner CP, Connolly ES. Brain-computer interfaces: military, neurosurgical, and ethical perspective. Neurosurg Focus. 2010;28(5):E25.
33. Munyon CN. Neuroethics of non-primary brain computer interface: focus on potential military applications. Front Neurosci. 2018;12:696.
34. de Laat PB. Algorithmic decision-making based on machine learning from big data: can transparency restore accountability? Philos Technol. 2018;31(4):525–41.
35. Wachter S, Mittelstadt B, Floridi L. Transparent, explainable, and accountable AI for robotics. Sci Robot. 2017;2:eaan6080.
36. Pasquale F. The black box society. The secret algorithms that control money and information. Cambridge: Harvard University Press; 2015.
37. O'Neil C. Weapons of math destruction. How big data increases inequality and threatens democracy. New York: Crown; 2016.
38. Yuste R, Goering S, Agüera y Arcas B, Bi G, Carmena JM, Carter A, et al. Four ethical priorities for neurotechnologies and AI. Nature. 2017;551(7679):159–63.
39. Kellmeyer P. Big brain data: on the responsible use of brain data from clinical and consumer-directed neurotechnological devices. Neuroethics. 2018. https://doi.org/10.1007/s12152-018-9371-x.
40. Bublitz C, Wolkenstein A, Jox RJ, Friedrich O. Legal liabilities of BCI-users: responsibility gaps at the intersection of mind and machine? Int J Law Psychiatry. 2019;65:101399.
41. Steinert S, Bublitz C, Jox R, Friedrich O. Doing things with thoughts: brain-computer interfaces and disembodied agency. Philos Technol. 2019;32(3):457–82.
42. Friedrich O, Racine E, Steinert S, Pömsl J, Jox RJ. An analysis of the impact of brain-computer interfaces on autonomy. Neuroethics. 2018. https://doi.org/10.1007/s12152-018-9364-9.
43. Wolkenstein A, Jox RJ, Friedrich O. Brain–computer interfaces: lessons to be learned from the ethics of algorithms. Camb Q Healthc Ethics. 2018;27(4):635–46.
44. Ienca M, Haselager P. Hacking the brain: brain-computer interfacing technology and the ethics of neurosecurity. Ethics Inf Technol. 2016;18(2):117–29.
45. Ienca M, Haselager P, Emanuel EJ. Brain leaks and consumer neurotechnology. Nat Biotechnol. 2018;36:805.
46. Fiske A, Henningsen P, Buyx A. Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. J Med Internet Res. 2019;21(5):e13216.
47. Wolkenstein A, Friedrich O. Brain-computer interfaces: current and future investigations in the philosophy and politics of neurotechnology. In: Friedrich O, Wolkenstein A, Bublitz C, Jox RJ, Racine E, editors. Clinical Neurotechnology meets Artificial Intelligence: Philosophical, Ethical, Legal and Social Implications. Heidelberg: Springer; 2021.

Actions, Agents, and Interfaces

2

Tom Buller

Contents

2.1 Introduction.......... 12
2.2 BCIs and the Decoding of Movement Intention.......... 14
2.3 Basic and Non-basic Actions.......... 15
2.4 Action, Belief, and Reliability.......... 17
2.5 Action, BCIs, and Identification.......... 21
2.6 Conclusion.......... 22
References.......... 22

Abstract

Ideally, a brain-computer interface (BCI) would enable bodily movement that is functionally and phenomenologically similar to "ordinary" behavior. One important element of this desired functionality is that the user would be able to control movement through the same types of mental activity that are used in "ordinary" behavior. For example, arm movement is caused by neural activity that underlies the conscious intention to move the arm. At present, however, the BCI-user has to learn to control movement by consciously imagining the movement, or by controlling neural activity that is only indirectly related to the intended movement. According to the standard account of action, a bodily movement qualifies as an action if its proximate cause is the conscious or unconscious intention to perform that movement. Since it can be argued that this condition is not met in the case of BCI-mediated behavior, an important question to ask is whether this type of behavior qualifies as intentional action.

T. Buller (✉)
Department of Philosophy, Illinois State University, Chicago, IL, USA
e-mail: tgbulle@ilstu.edu

O. Friedrich et al. (eds.), *Clinical Neurotechnology meets Artificial Intelligence*, Advances in Neuroethics, https://doi.org/10.1007/978-3-030-64590-8_2

2.1 Introduction

A brain-computer interface (BCI) is a neuroprosthetic device that enables the control of bodily movement or an external device through the detection and decoding of neural activity. As the following case illustrates, significant progress in the development of BCI technology over the past years has helped increase the physical autonomy of individuals who have suffered a loss of motor function.

> [BK] has had electrical implants in the motor cortex of his brain and sensors inserted in his forearm, which allow the muscles of his arm and hand to be stimulated in response to signals from his brain, decoded by computer. After eight years, he is able to drink and feed himself without assistance. [1]

BCIs have been described as devices that translate thought into action [2–5]. This description seems appropriate since the movement of BK's arm and hand, for example, is neither a reflex nor did it just happen to occur; rather, the BCI detected and decoded BK's movement intentions and thereby effected the intended bodily movement. Accordingly, we might view BCIs as functional replacements for the damaged parts of the motor system, as novel realizers of the agent's movement intentions. In this regard BCIs present us with the latest—and most advanced—instance of replacement technology.

According to an influential and widely held view, physical actions are intentionally caused bodily movements. More precisely, the Causal Theory of Action (CTA) can be stated in the following way.

> (CTA) Any behavioral event A of an agent S is an action if and only if S's A-ing is caused in the right way and causally explained by some appropriately nonactional mental item(s) that mediate or constitute S's reasons for A-ing. ([6], p. 1)

The movement of BK's arm and hand counts as an action, therefore, because he wants to take a drink from his cup and the desire (and the attendant belief) causes the bodily movement. In this regard, actions are distinguished from "mere happenings"—bodily movements that lack this specific etiology. To say that a person's physical behavior is intentional is to say that it is causally related to their beliefs and desires. Tripping and falling over does not, therefore, count as an action since we can assume that the person did not have the belief and desire to trip and fall.

The matter is complicated, however, by the fact that not just any causal connection between intention and bodily movement will do. For we can imagine cases in which we would be reluctant to conclude that the person has acted even though bodily movement is causally related to the person's intentions.

> Bob desires and intends to shoot the sheriff, but this makes him nervous and causes his finger to cramp, which in turn causes the trigger to be pulled, resulting in the gun being fired and the sheriff being shot. ([7], p. 12)

Since the trigger being pulled was caused by Bob's nervousness, and his nervousness was caused by his intentional states, his bodily movement was causally related to his intentions. However, if we suppose quite plausibly that his nervousness was not itself intentional, then we can doubt that Bob intentionally shot the sheriff. To put the point in more theoretical terms: although the bodily movement matches Bob's original intention, it is not a *function* of his intention. As a consequence, we cannot exhaustively explain the trigger-pulling in terms of his beliefs and desires.

The above suggests that the causal process in physical action is of the right type if the intended bodily movement is a function of the person's beliefs and desires to perform that movement. If we adopt a broadly physicalist framework, then this is to say that an arm-raising, for example, qualifies as a physical action if it is brought about by the neurophysical state(s) that realizes the person's intention to move their arm. Unfortunately, this revised framework does not solve all our problems. For we can imagine cases in which our intuition is that the person has acted even though movement is not brought about by the appropriate neurophysical states.

> After suffering a severe spinal cord injury LC has lost a substantial degree of motor function. By concentrating on directional symbols displayed on a computer screen, LC is able to control the movement of a robotic limb with the aid of a BCI.

If we assume for the sake of argument that the neural activity underlying conscious attention is distinct from the neural activity underlying movement intention, then the causal process in this case is not of the right type. Nevertheless, LC would appear to be performing a physical action.

LC's case raises a number of important issues regarding the nature of physical action. First, we might ask whether, and under what conditions, the robotic limb counts as part of the body. Presumably, our answer to this question will depend in considerable part on the degree of functional and phenomenological similarity between control of the robotic limb and of "ordinary" arm movement—the greater the similarity, the greater the reason to conclude that that LC's robotic limb is "incorporated." If we conclude that robotic limb is not part of the body, then we can ask whether in moving the limb LC has performed a mental, rather than a physical, action or whether the movement is merely the effect of action.

Second, it can be claimed that the proximate cause of movement in LC's case is not *intention* to move the limb but *concentration* on a specific symbol. It is true, of course, that this event is not a mere happening since it is part of an intended causal process (unlike Bob's nervousness), but importantly the movement is not directly brought about by the neurophysical state(s) that realizes LC's intention to move the arm. Third, if it is the case that control of the robotic limb is substantially dissimilar to ordinary behavior, then it is not clear what is the nature and content of LC's beliefs and intentions which bring about movement. Ordinarily speaking, to intend to move my arm I must believe that my intention will bring the movement about, and that this movement is an arm movement. If the connection between intention and movement is unreliable, or the resultant movement is not of the right kind, then it is difficult to say exactly what I am intending.

2.2 BCIs and the Decoding of Movement Intention

An injury or disease that has damaged the spinal cord and has caused the loss of motor function may leave higher brain functions substantially intact. Motor function can be restored through an interface that bypasses the injury and connects the intact motor centers to an external device, robotic limb, or even the person's own body as in BK's case. A BCI designed to restore motor function decodes neural signals to extract voluntary motor commands that reflect the person's movement intentions, and then uses the process signal to control the external device or limb (robotic or "natural"). Typically, a BCI is composed of three components: a sensor to detect neural signals, a signal processor that converts neural activity into a command related to the desired action, and a device to effect action [10]. The BCI is able to detect motor commands from neural signals due to established correlations between neuronal firing rates and motor parameters like arm position, velocity, and joint torque [8]. Neuronal recording of motor commands has focused primarily on the primary motor cortex, although higher-level movement intentions and imagery, for example, imagined goals, trajectory and types of movement can also be decoded from the posterior parietal cortex [2, 8, 9].

BCIs can be categorized in a number of different ways [10]. First, we can distinguish the devices in terms of their level of invasiveness—whether they are placed on top of the scalp (EEG), subdurally on top of the cortex (ECoG), or inserted into the brain. Second, BCIs can be differentiated in terms of the type of signal recorded: field potentials (the summed electrical current from multiple neurons) from multiple recording sites (EEG and ECoG), or action potentials ("spikes") from single neurons or small groups of neurons. Although the less invasive devices can detect brain activity that correlates with visual stimuli and voluntary intention, for example, more specific and accurate details of action are obtainable from spiking activity. For instance, hand velocity and position, and movement goals can be detected from single cells in the motor cortex [11]. Third, we can distinguish between *direct* devices that enable the control of an external device through neural events that underlie, that is to say, are intrinsically related to the intended movement, and *indirect* devices that co-opt neural events that are not intrinsically related. A device that enabled the control of a robotic limb by neural signals that correlate with arm movement would be an example of a direct device; in contrast, one that controlled this movement through the suppression of cortical rhythms, or by the detection of amplitude differences between attended and non-attended stimuli, would be an indirect device.

An alternative way to categorize BCIs is to distinguish among *active*, *reactive*, and *passive* devices:

> In *active* BCIs, the user intentionally performs a mental task that produces a certain pattern of brain activity, which the BCI system detects for processing. A commonly deployed mental strategy in active BCIs is motor imagery. The user imagines moving parts of her body, without actually performing the movement. The imagination of the movement of different body parts corresponds to different activations of the primary somatosensory and motor cortical areas.