



ELSEVIER

Contents lists available at ScienceDirect

International Journal of Law and Psychiatry

journal homepage: www.elsevier.com/locate/ijlawpsy

Legal liabilities of BCI-users: Responsibility gaps at the intersection of mind and machine?



Christoph Bublitz^{a,*}, Andreas Wolkenstein^b, Ralf J. Jox^c, Orsolya Friedrich^d

^a Faculty of Law, Universität Hamburg, Rothenbaumchaussee 33, 20148 Hamburg, Germany

^b Institute of Ethics, History and Theory of Medicine, Ludwig-Maximilians-Universität München, Lessingstr. 2, 80336 Munich, Germany

^c Centre Hospitalier Universitaire Vaudois (CHUV), Avenue Pierre-Decker 5, CH-1011 Lausanne, Switzerland

^d Institute of Ethics, History and Theory of Medicine, Ludwig-Maximilians-Universität München, Lessingstr. 2, 80336 Munich, Germany

Introduction

*But when a real signal came through for the first time, a complete and strong signal, without any assistance – that was a killer feeling. That was “Fuck. Great”. I had a huge grin on my face and thought “This is unbelievable. That is Back-to-the-Future style ... A pie in the sky that is coming true, and I will have been a pioneer of the field.” That was cool, literally, that was really cool.*¹

Brain-Computer-Interfaces (BCIs) connect brains with computers. They measure and process brain signals and translate them into commands which control further applications or devices such as software programs, neuro- or speech prostheses, wheelchairs, and even drones or planes (for an introduction cf. [Wolpaw & Wolpaw, 2012](#)). The striking feature of BCIs is that they couple the human mind-brain system with hard- and software systems. Computer code and algorithms become functionally and structurally interwoven with the central nervous system, supplementing and sometimes replacing its functions.² Especially BCIs with intracortically implanted electrodes appear as epitomes of the perhaps unsettling but apparently inevitable technological trajectory merging minds with machines ([Brunner et al., 2015](#); [Clausen, 2009](#)). Although in many aspects still a vision of the future, the pace of convergence is remarkable, and the technology available today already raises intriguing questions for the law. Many stem from the fact that normative rules and principles that apply to humans (to their bodies, minds, actions) categorically differ from those that apply to physical

objects, machines, or software. Once these entities become physically inseparable and functionally interwoven, and jointly generate outputs in which the contributions of each can no longer be discerned, it becomes unclear which legal categories do – and should – apply.³ This paper seeks to provide answers to one aspect of this problem, namely the legal liability of persons who cause harm to third-parties through devices or machines controlled by BCIs (henceforth: “controlled devices” or “applications”).

The relevance of the legal allocation of liability may extend beyond the BCI-context. It might serve as a case-study for broader discussions about responsibility gaps caused by automation, AI, and algorithms. Such gaps, it is worried, may emerge in interactions of persons with autonomous machines.⁴ Whereas persons might be exonerated from responsibility due to the “autonomous” inputs of machines, the machines are no proper targets for the attribution of blame and liability.⁵ As a result, it may appear that no one bears liability; liability seems to vanish (for different takes on this with respect to BCIs see [Clausen, 2009](#); [Grübler, 2011](#); [Holm & Voo, 2011](#); [O’Brolchain & Gordijn, 2014](#); [Tamburrini, 2009](#); [Weinberger & Greenbaum, 2016](#)). In fact, several such chasms might emerge in the multipolar relations between producers and programmers of devices, operators and users, injured-parties and other stakeholders. The following discussion concerns one specific instantiation, the supposed gap between users and parties who sustain harms from the operation of BCIs. It has received considerably less

* Corresponding author.

E-mail address: christoph.bublitz@uni-hamburg.de (C. Bublitz).

¹ These remarks are by tetraplegic patient in his twenties who controls software through a BCI. They and subsequent remarks stem from a yet unpublished study by Kögel (p. 35). Some of these remarks were translated (from German); and some remarks were stylized (deletion of repetitions, fillers, etc) by the authors.

² Usually, we are no friends of the undifferentiated term “mind-brain system”, but it seems to be the best fit in a BCI-context. Focusing on either brain or mind would leave out important aspects. We also wish to note that we use “responsibility” and “liability” interchangeably.

³ For instance, a permanent prosthesis can be conceived as an ordinary object which can be possessed under a property right, or alternatively, as part of the body to which it is connected and that cannot be owned. This is, in a sense, the reverse of the question whether one can have property in body parts that are separated or easily separable, such as bio specimen, human tissue, blood or sperm, DNA. See [Goold/Maslen/Auckland \(unpublished manuscript\)](#). For further vagaries in the categories of human and machine see ([Schermer, 2009](#)). Also, many implantable medical devices are physically and functionally inseparably connected to the human body, the peculiarity of BCIs is that they encompass the mind.

⁴ A classic statement of the gap is ([Matthias, 2004](#)); also see the recent resolution of the [EU parliament \(2017\)](#), which states that “traditional rules will not suffice to give rise to legal liability for damage caused by a robot” which operates autonomously (at AF). For a dismissive view of such gaps, see [Sauer et al., \(2017\)](#).

⁵ One might alternatively understand the gap proportionally: The greater the degree of autonomy of the machine, the less liable the person instructing or supervising it.

<https://doi.org/10.1016/j.ijlp.2018.10.002>

Received 4 April 2018; Received in revised form 30 September 2018; Accepted 11 October 2018

Available online 16 November 2018

0160-2527/ © 2018 Published by Elsevier Ltd.

attention than the one between producers and harmed-parties in terms of product liability (Beck, 2016; Chinen, 2016; Grübler, 2011; Pagallo, 2013; Palmerini et al., 2014).

It might be tempting to regard BCIs as simply another tool and deny that they raise (sufficiently interesting) novel questions for theories of responsibility (Clausen, 2008). But this a matter of abstraction. In the rather broad perspectives many current contributions to the debate take, BCIs may indeed appear as no more. But this view misses relevant aspects. A more fine-grained look at specific cases reveals several open questions about liability arising from the unique way in which BCIs mediate the relation between user and world (Weinberger & Greenbaum, 2016). Getting these idiosyncrasies and peculiarities in focus requires descending to the level of case discussions. To our knowledge, this is the first paper that does so. It introduces a dozen simple case vignettes that exemplify typical BCI-uses and brings normatively salient features into sharper relief. The cases systematically vary key conditions of legal liability such as the act requirement, *mens rea*, foreseeability and permissible risks.

At several stages, these stylized and sterile vignettes based on speculative technologies are contrasted with the practical realities of restorative BCI-use. For this, we draw on one of the first qualitative interview studies with (the still few) users of restorative BCIs (Kögel, unpublished manuscript). These users, have abilities that differ from normal-functioning and face a range of severe difficulties in motor control of their bodies (ALS, "locked in syndrome", paralysis). They are pioneers of BCI-use, often in the context of basic research. Their often adverse experiences remind us that the technology is still a long way from realizing its promises.

The paper proceeds as follows: The first section introduces some basic distinctions between civil and criminal forms of responsibility. The second provides a brief overview of BCI-technology. The third and main section addresses relevant conditions of liability by discussing a dozen vignettes. In *nuce*, we argue that the law should view many BCI-mediated movements as actions, even though they may not qualify as such under current legal conceptions. However, some movements mediated by (passive) BCIs still fall short of constituting actions in the legal sense. Users might nonetheless be responsible for resulting harms because, so we claim, they are under a duty to *avert* harms arising from BCI-operations. In addition, we suggest that users might incur liability for *deploying* BCIs if harms result from foreseeable malfunctions. The fourth section seeks to define limits to (negligence) liability, mainly through the idea of permissible, socially acceptable risks. We propose that lawmakers should allow users the creation of risks through *restorative* BCIs and exempt them from criminal liability because all countries are obliged by international law to advance the interests of people with disabilities.⁶ This paradigmatically includes abilities to move one's body. We also like to suggest to the BCI-community to start discussion and formulating standards of care in anticipation of coming debates among regulators. The fifth section, by contrast, calls for imposing strict liability on BCI-users in civil matters. The sixth and seventh section highlight remaining responsibility problems: Although the worried responsibility gap does not exist, lacking criminal liability for harms resulting from use compliant with safety standards leads to a *deterrence gap*, and the peculiar form of interaction between mind and machine through BCIs leads to an evidential or *epistemic gap*.

Our interest lies in legal responsibility; "rights", "duties", "responsibility" thus refer to their legal (not moral) forms. Such discussions are associated with a caveat: Legal systems and jurisdictions differ; they recognize different types of liability under different conditions. The argument therefore remains on a more general level and largely abstains from discussing norms or case-law of specific jurisdictions⁷ Although arguments may not

apply to every jurisdiction to the same extent, the main distinctions and outcomes presumably apply to most western jurisdictions.

1. Forms of responsibility

BCI-technology is currently in a translational phase. Recently, leading BCI-researchers identified liability for harms resulting from BCI-use as one of the yet unresolved questions to be addressed before they can be deployed more widely clinically or released on consumer markets (Consortium, 2015, p. 40). The European Parliament addressed neuroprosthesis in its resolution on robotics and calls on the Commission to develop a legislative instrument that comprehensively addresses legal issues such as liability in the next decade (2017, at Z seq.) The topic affects all stakeholders: Users have to be aware of the duties and liability risks they incur by using BCIs. Potentially harmed third-parties and the public at large have to know which damages they have to bear themselves and for which they may seek redress. Manufacturers need to assess, and possibly insure, the liability risks to which they become exposed by releasing BCIs on markets. Lawmakers and regulators need to review liability rules in light of fairness, rights of users, and incentives for innovative developments as well as define conditions for market release and terms of real-life applications. Finally, issues of responsibility should influence engineering and technical BCI-designs.

The present inquiry has strong parallels to other contemporary debates about responsibility and new technologies – autonomous machines, robots, self-driving cars. For instance, BCIs regularly deploy self-learning algorithms, i.e., those capable of changing the rules of decision-making by themselves (Alpaydin, 2004; Domingos, 2012; Müller, Krauledat, Dornhege, Curio, & Blankertz, 2007). Such systems are regularly not error-free because adapting to novel environments through trial and error is part of the logic of the learning feature (Matthias, 2004). There is also overlap with ethical debates about autonomous vehicles and robotic devices.⁸ However, these fields are so wide that specific technologies such as BCI deserve and require special treatment.

There is also an important dissimilarity: Most autonomous robots and cars are designed to operate largely independent from the input of persons. Many strive to operate so independently that treating them as distinct legal personalities ("e-persons") appears worth discussing (EU Parliament, 2017, at 59 F; Palmerini et al., 2014). The aim of self-learning BCI systems, by contrast, is close functional embeddedness with the mind-brain system of users. Responsibility gaps in the former cases arise partly because of the growing distance, both physically and functionally, between autonomous machines and persons whereas BCIs bring persons and machines closer together. These spatial and temporal differences are relevant for responsibility ascriptions.

A frequent shortcoming of debates about responsibility gaps is insufficient attention to different types of responsibility. Suppose that a user (henceforth U, female) seriously harms the victim (V, male) by hitting him with her wheelchair. Questions about her responsibility can refer to different normative aspects: Did U have a duty to prevent the wheelchair from hitting V? Should U apologize? Should she compensate V for material damage such as destroyed clothes or pay for immaterial damage such as pain? Should she be put in jail, be banned from driving, or her wheelchair be confiscated? These questions concern different

(footnote continued)

With respect to tort law, it draws upon the two leading works of harmonizing EU law: The Common Draft Frame of Reference, DCFR (von Bar et al., 2009), and the Principles of European Tort law, EUTL (European Group on Tort Law, 2005). For criminal law, the model is the US Model Penal Code (American Law Institute, 1962).

⁸ For the European Parliament, characteristics of a robot are, inter alia, data-exchanges with the environment, self-learning from experience and adaption to the environment, which some of the discussed BCIs exhibit (2017, at 1). They are thus included by the Robolaw study referenced several times here (Palmerini et al. (2014).

⁶ For an overview of BCI technology for neurological rehabilitation see Daly/Wolpaw (2008).

⁷ In fact, the paper mainly refers to what one could call "ideal" legal codes.

types of legal responsibility which may have different prerequisites. Some questions might be answered affirmatively, others dismissively. Speaking of “responsibility” *tout court* regularly obfuscates its many facets and is a source of disagreements about responsibility gaps.

Legal responsibility splits into civil (contract, tort, manufacturer) and criminal responsibility. Civil liability obtains between private parties and seeks to compensate injured parties for damages resulting from norm violations or breaches of duties, frequently according to the idea of *restitutio ad integrum*: Victims have to be put in the position they would have had in the absence of the wrongful deed. Some (mainly US) jurisdictions allow for additional punitive damages, an idea alien to most Continental legal orders left aside here. Legal duties can arise from different sources. For instance, persons might incur them voluntarily by contracting. Relevant for present purposes, however, are non-contractual duties and liabilities between persons who do not stand in any specific legal relation to one another. They are the general duties people owe to each other.

Criminal responsibility, by contrast, is ascribed by the state to punish offenders rather than compensate victims and pursues further penological aims such as retribution or deterrence. Requirements for criminal are regularly stronger than for civil liability, i.e., thresholds are higher and conditions narrower.

As BCIs can be used to control almost every device, from neuroprostheses to airplanes, they can cause all kinds of damages. Whether causing a specific harm gives grounds for legal action depends on the respective conditions of liability in each jurisdiction. Although they differ between types and jurisdictions, there are some common features. A general distinction runs between fault and no-fault – or strict – liability (cf. Chapter 4 and 5 EUTL). Fault liability requires that users breach a duty at least negligently, with *mens rea*. No-fault liability, by contrast, neither requires fault nor specific mental states, it can be imposed even on persons who acted correctly. It often covers damages that result from the permissible creation of risks. For instance, keepers of automobiles or people deploying hazardous goods are regularly liable for damages resulting from these activities, regardless of whether they made a mistake or acted wrongly (and insofar no one else was at fault). The rationale justifying strict liability is that people who derive economic or other advantages from activities which put others at risk should bear the burdens and compensate for damages. It is often the corollary to a regulatory permit to engage in dangerous activities at all. Notably, in many jurisdictions strict liability only concerns *civil* liability as punishing people who acted faultless is considered unfair.⁹ With this sketch of different types and conditions of responsibility, let us turn to BCIs.

2. BCI technology

According to a standard definition, a BCI is a “system that measures central nervous system (CNS) activity and converts it into artificial output that replaces, restores, enhances, supplements, or improves natural CNS output and thereby changes the ongoing interactions between the CNS and its external or internal environment.” (Wolpaw & Wolpaw, 2012, p. 3). Less technically, BCIs connect the brain with a computer by detecting, extracting and interpreting brain signals, and transmitting them to an external device. This enables interactions between a user, a computer and further applications or devices (Lebedev, 2014; Lebedev & Nicolelis, 2006). The normative debate requires only a basal technical understanding. Most importantly, several BCI designs exist. They can be invasive (intracortically implanted electrodes, “BrainGate”) or non-invasive (EEG cap). BCIs detect electric, magnetic, or metabolic brain signals. Most current systems detect electric signals, but novel ones utilizing functional near-infrared spectroscopy (fNIRS) measuring the blood-oxygen-level dependent (BOLD) signal may become available soon (Wolpaw & Wolpaw, 2012, ch. 4–6).

The connection between brain and computer can be one- or bi-

directional. In the former, data flows in one direction only, e.g., brain signals are decoded by a computer.¹⁰ Of course, users may get some kind of feedback, they may, for instance, watch the movement of the device controlled by the signals. But this feedback is only indirect. In truly bi-directional systems, a response to the brain signals is fed back directly into the brain via e.g., electric stimulation that stimulates arousal when the BCI detects tiredness (AlZu'bi, Al-Nuaimy, & Al-Zubi, 2013; O'Doherty, 2009; O'Doherty et al., 2011). In such bi-directional systems, the user is “out of the loop”, as her (active) participation is not required (Kellmeyer et al., 2016). One-directional systems usually keep the user “in the loop” as they require some participation. This paper focuses on one-directional BCIs for rehabilitative motor purposes which neither intervene into brains, nor alter preferences or influence decision-making of users. Users remain “in the loop” and are legally competent as well as psychologically capable of forming desires and intentions in the ordinary way.¹¹

In a bit of engineering jargon, one might say that “BCIs introduce a second adaptive controller that can ... change to ensure that the organism's goals are realized. BCI usage therefore depends on the effective interaction of two adaptive controllers: the user's CNS and the BCI” (Wolpaw & Wolpaw, 2012, p. 8). For questions of liability, the precise interplay of the two is relevant. One important aspect is the *timing* of interaction. BCIs can detect brain signals from different stages in a movement sequence, typically ranging from unconscious formation of desires over intentions to movement execution, and they can provide various inputs to the “where, when and how” of movements of controlled devices. Moreover, the *mode* of interaction, primarily, the way users control BCIs, is relevant. The central distinction runs between systems controllable via conscious mental actions of users (active BCIs), systems that measure reactions to stimuli which users can modulate through directing their attention (reactive BCIs), and those that monitor brain activity not willfully controllable by users (passive BCIs). In active BCIs, users modulate their brain activity mostly by consciously performing a mental task which is associated with a distinct brain activation pattern that is well-detectable by BCIs. A key example are sensorimotor rhythms in the sensorimotor cortex. Specific brain waves are suppressed if people think about moving particular parts of their body (motor imagery). The resulting activation patterns seem to serve well as inputs to the BCI (for a review, Yuan & He, 2014).

Spectacular experiments used motor imagery in a different technical design. Patients in a minimally conscious state were asked to imagine typical motor tasks such as playing tennis or navigating through their house (Monti et al., 2010; see also the work of Birbaumer, 2006; Chaudhary et al. 2016). Imagining these activities produces particular activation patterns in the motor cortex. Thereby, some patients were able to generate distinct brain signals that could then be used as codes for communication, re-enabling (a minimal form of) communication.

Control of many current BCIs is strenuous and effortful. Users have to concentrate to generate good signals, which can only be sustained over a limited time-span. Moreover, many technical setups work by controlling individual parameters, which means that even complex tasks are executed by one command at a time. This may soon turn into a long and exhaustive process with frustrating failures. A user of a BCI-speller (by which he can select letters on a screen through attending to visual stimuli) expresses his experiences like this: “Well, you have to

¹⁰ The cochlear implant exemplifies the reverse direction, with signals being transferred into the brain-mind.

¹¹ Bi-directional BCIs “closing the loop” raise additional problems of liability insofar as they intervene into decision-making processes. BCI-systems that intervene into decision-making processes or place the person entirely “out of the loop” appear promising for psychiatric treatments and raise additional problems of responsibility revolving around the authenticity of desires or the history of volitions which lie outside the scope of this paper. Many parallels can be found in discussions over effects of pharmaceutical interventions into minds on autonomy, see e.g. Bublitz and Merkel (2009).

⁹ A good introduction to the debate and the many finer distinctions we have to leave out here can be found in Simester (2005).

truly fight for every letter actually” (Kögel, p.15).

Present BCIs still have significant error rates, the accuracy of signal detection and correct interpretation is far from perfect. Use often requires intensive training. A significant percentage of users fail to use BCIs, mainly for physiological reasons (called “BCI illiterates”, for a critique of the term see Thompson, 2018). As they are subject to many technical factors, it is not possible to provide precise numbers for error-rates. Intracortical signals are significantly better, but this requires surgically implementing electrodes. Accuracy also depends on computation time. Restorative BCIs which provide a somewhat good user experience need responsive designs, which means short reaction times and less accuracy. To provide an illustrative order of magnitude, we assume that error rates are no better than 5%, i.e., at least five out of a hundred brain signals are classified inaccurately. This is an optimistic estimate at present, but rates will likely improve in the future.¹²

Errors can have many sources: Users can give incorrect signals, the BCI can fail to detect or misinterpret them. Even if users perform the mental task correctly, they might not evoke the right or sufficiently distinct brain signals because of noise or interferences by other physiological processes. The BCI may fail to detect and process signals correctly for hardware or software reasons. As there is not direct feedback about errors, these sources are hard to keep apart in practice. More broadly, one may say that BCI control requires a two-way adaptation between user and device. A user who directs a robotic arm through mental imagery describes this adaptive process like this (Kögel, p.13):

“Well, I felt like I needed to adapt to the computer, but then the technical team assured me that very often, they needed to adapt the computer to me. Well, it ended up being both. More often than not, they had to figure out what was wrong with the training, or what was wrong with the hardware or software because my brain was doing what it was supposed to do.”

So much for the technical side. Let us now introduce some case vignettes as a basis for further discussions. They are simple illustrations, and nothing hinges on their actual feasibility. What we wish to emphasize are the structural features of the relations between the user, the BCI and the harmful event. For illustrative reasons, they concern the control of wheelchairs, but the normative problems equally apply to any other device, from prosthetics to robots or software.

Case 1: *The BCI monitors the mental state of U. When she becomes inattentive or tired, the BCI plays a loud warning sound. U cannot consciously or willfully control her attentiveness or tiredness. One day, U repeatedly becomes inattentive, triggering the sound. The repeated sounds keep neighbor V from sleeping.*

Variation: *U habitually relies on the system. In the absence of a signal, she deems herself fit to drive a heavy-duty truck. One day, the BCI falsely interprets brain signals and therefore fails to warn her. Because of lacking attention, U causes an accident harming V.*

The BCI in these vignettes is a passive one, it detects brain signals not under U's willful control. Compare this to an active BCI:

Case 2: *Paralyzed U controls her wheelchair through mental actions. By imagining “left” or “right”, she directs the wheelchair to turn in that direction to a certain degree, by imagining “forward”, “stop”, or “backwards”, she controls its movement.¹³ Through training, she has mastered control of the wheelchair. One day, she meets V on the street, whom she seriously dislikes. She directs the wheelchair to hit him.*

Let us call the type of active-BCI control “process control” because the BCI signal controls details of the operation of the application, in this case, the movement of the wheelchair. It can be contrasted with less specific forms of control such as “goal selection”, in which applications compute major parts and less user input is required, more like an autopilot (we

borrow these terms from Wolpaw and Wolpaw, 2012, p. 9). Of course many systems fall somewhere between. A user describes the control of a robotic arm via “goal-selection” mode in the following way (Kögel, p. 18):

“[I]f you're going to reach for, you know, your coffee cup on your table, you don't think, oh I need to lift my shoulder up, unbend my arm and reach with my hand and then, and then I close my hand, you just think, you know, I'm going to go grab my cup and that's just how it works”

3. Fault-liability

3.1. Agency & disembodied action

Is U responsible for the harms sustained by V in the two cases? It is worthwhile to take a broader look at the requirements of civil and criminal fault liability. Legal responsibility is regularly grounded in agency (“voluntary act requirement”). Persons are responsible for (the results of) their actions. This does not imply that persons are *only* responsible for actions, they may also be for failures to act, omissions. However, and in contrast to popular philosophical positions, legal conditions of responsibility for acts and omissions usually differ (Steinert et al., 2018). Accordingly, determining whether persons acted in the legal sense is usually a necessary step in responsibility assessments.

When does a person act? The law has developed its own concept of action in light of considerations that sometimes differ from those of philosophy of action. This is not surprising given the different objectives of the inquiries: philosophy seeks to describe a supposedly ontologically distinct category – action – mainly in light of paradigmatic instances. The law seeks a concept of action for normative purposes, the ascription of responsibility, and is often concerned with grey areas and exceptional cases. According to many jurisdictions a voluntary act is a “willed bodily movement” (Steinert et al., 2018; Moore, 2010). This concept has two conditions, both pose problems with respect to BCI-mediated movements. The first is that movements have to be “willed”. Irrespective of the notoriously problematic concepts of “will” or “volitions” (Moore, 2010; Roskies, 2010), the key idea is that the movement has to be controlled by the (reason-responsive) decision-making capacities of the person. Accordingly, involuntary or non-controlled movements such as knee-jerk reflexes or sneezing are no actions proper. Therefore, the outputs generated by *passive BCIs* which monitor brain activity not willfully controlled by the person are *no actions*. This allows a first assessment of case 1: The BCI has detected brain signals associated with a particular mental state of the user (attentiveness). Because U could not willfully control her attentiveness, she was not performing an action, she was not triggering the warning sign and is thus not responsible for keeping N up through her action (she might be responsible on different grounds, *infra*).

The second peculiar feature of BCI-mediated movements is their *disembodied nature*. They enable persons, for the first time, to affect the world without moving parts of their bodies, through “thoughts alone” (more precisely, through brain signals, correlates of thoughts, that causally produce external effects via technical devices). The muscular system and large parts of the body are bypassed and become obsolete. Some celebrate this as the liberation of the mind from the body (Nicoletis, 2011). Note that this does not imply falling into mind-brain dualisms. The only parts of the body still necessary for executing an intention are those that realize and sustain the brain signals detected by the BCI.¹⁴

¹⁴ Surely, brain signals require the CNS and the rest of the body to support it with blood, oxygen etc. But this is only a contingent factor. The key idea, however is that the *execution* of the movement (not the forming of the intention) proceeds on ways bypassing large parts of the body. Furthermore, one may conceive of the devices that are controlled as replacements of the body. This is especially plausible when the BCI restores or mimics bodily functions (prostheses, wheelchairs). One could thus speak of a re-embodiment through these artefacts. This view becomes less appealing in cases in which the BCI controls software (such as a cursor in an app). Speaking of the cursor or a

¹² (Haselager, 2013) speaks of “classification accuracy rates of around 75–80%” and refers to Pfurtscheller et al. (2010).

¹³ Prototypes of such BCI-controlled wheelchairs have been build (Galán et al., 2008).

BCIs thus create a new output channel of the CNS that no longer relies on the body in the traditional way (Wolpaw and Wolpaw, 2012, p. 6). This new way to affect the world is remarkable and does not easily map on existing legal categories. The law draws an (at least implicit) distinction between mental actions (such as imagining, calculating, or thinking) and bodily movements, and considers only the latter relevant. But the brain signals controlling BCIs arise from brain activity or mental actions, not from bodily movements, and bypass the muscular systems. They are not of the kind that the law understands as willed *bodily* movements and thus, as actions.¹⁵ Consider case 2: U controls her wheelchair through mental actions. However, since she is not driving the wheelchair through movements of her limbs (she is paralyzed), she is not performing a willed *bodily* movement and thus is not acting in the traditional legal sense.

In our view, the legal focus on the body leads to implausible results in cases like that. Elsewhere, we have called for reforming the legal concept of action to comprise “mental actions with external effects”. We cannot rehearse the argument here, but wish to note that expansions of legal concepts are often not as innocuous as they may appear as they apply to a wide variety of unforeseeable cases. For instance, in a spectacular experiment, a monkey at Duke University in North Carolina learned to control the movements of a robot in Kyoto via a BCI (Nicolelis, 2011). One might say that the monkey acted in North Carolina and in Japan. But then, the agent becomes spread across half of the planet. This spatial distance is a novel aspect of BCI-mediated agency. It differs from the familiar fact that actions may have consequences in faraway places (e.g., shooting a rocket from Durham to Kyoto). It may raise questions as to who has jurisdiction. But without delving into such issues, we shall assume in the following that disembodied movements via BCIs qualify as legal actions if they are controlled by consciously performed mental actions. Thus, in case 2, U acted as she directed her wheelchair in this way.

A bit more precision with respect to the kind of control necessary for action is required. We suggest that for movements to constitute legal actions, they have to be either *initiated* by an intention to act (an executory *go-command*), or consciously controlled during execution (*guidance control*). Conversely, someone who lacks both executory and guidance control, does *not* act. Execution and guidance control loosely map on the BCI-control modes outlined above (“process control” vs “goal selection”).

Case 3: *U uses an active BCI controlled by mental actions with a “goal selection” mode. She does not direct every turn of her wheelchair, she only sets the destination (“living room”). The wheelchair is equipped with sensors and fully controlled by the BCI, which calculates the best route and further details of driving.*¹⁶ *U sees V standing on the route which the BCI will likely take and realizes a high chance of hitting him if she gives the go-command. So she does, as she seriously dislikes him.*

Despite her lack of guidance control, U acted because she initiated the driving of the wheelchair by a go-command. She is thus liable for hitting V. Not all BCI-mediated movements are initiated by such a pronounced user command. Some systems may initiate movements on their own. This is an important feature, especially in restorative BCIs, as commanding can be cumbersome and tiring. A somewhat smooth user experience that meets hopes of users requires “smart” and predictive assistive technology. To achieve this, research focuses on decoding intentions in the sensorimotor cortex and other areas of the brain, such as the posterior parietal cortex associated with plans, goals and intentions, that go beyond current sensorimotor devices detecting motor imagery. Some studies managed to read out imagined goals, trajectories, and types of movements (Aflalo et al., 2015). If this line of research proves

(footnote continued)

computer as re-embodiments seems to stretch the concepts too far, at least for legal purposes.

¹⁵ Again, for a thorough exposition of the problem and argument to this end, see Steinert et al. (2018).

¹⁶ Haselager (2013).

successful, BCIs might be able to predict what users *will* intend and initiate movements before users form a motor-intention or become aware of it. This may lead to scenarios like this:

Case 4: *The BCI is designed for a smooth user experience for tetraplegic patients. It predicts, with high accuracy, routinely recurring action-intentions of users. Suppose U desires to get a drink from the fridge, but has neither formed a motor-intention nor given a go-command. The BCI detects her desire, predicts her intention and directs the prosthesis to get a drink from the fridge. U neither initiated the movement, nor controlled it during execution. She did not act.*

In the absence of executory commands or guidance control, there is no action. The potential power of BCIs to predict intentions from (non-conscious) brain signals and initiate movement in the absence of go-commands has worried scholars because of cases like the following (Nakar et al., 2015; Yuste et al., 2017):

Case 5: *U is angry at V and feels the desire to punch him. Suddenly, her prosthesis moves into V's direction and punches him. The BCI detected her desire and executed it. Although U desired to punch V, she neither initiated the punching, nor controlled the movement of the prosthesis and thus did not act.*

3.2. Mens rea

In addition to the act requirement, criminal and civil *fault* liability require a corresponding subjective side, a psychological relation between mind of the person, her movement and the resulting consequences in the external world. Especially criminal law draws fine distinctions between different mental states. Responsibility requires that agents have a “guilty mind” (*mens rea*), further distinguished in several categories. The strongest is purpose or intent, satisfied if a person acts because she wants to attain a specific end. The second and weaker category is knowledge, requiring that persons foresee the consequences of their actions (though not necessarily desire them). The third category is the conscious creation of risks (called “recklessness” in many jurisdictions)¹⁷: The agent is aware of the substantial risks posed by her conduct but acts nonetheless (e.g., because she is indifferent to results). The last and weakest category is negligence. A person acts negligently if she is factually not aware of the risks of her conduct, but if she could and should have been so, and if the risks were substantial and unjustified. These are the four mental categories recognized by the US Model Penal Code (§ 2.02).¹⁸ Although precise conditions may vary in legal systems, the basic picture is the same. To be clear: there is another possible mental state which we may call “non-negligent unawareness” in which persons neither foresaw the results of their action, nor could have reasonably done so. It is insufficient to ground *fault* liability.

Case 6 (variation of case 3): *U uses an active BCI to drive a wheelchair which she controls by performing mental actions. Rather than directing every turn, she only sets the destination (“living room”). The BCI computes route and further details. This time, U does not realize that there is a chance of colliding with V, nor could she have foreseen it. Upon giving the BCI the go-command, it directs the wheelchair to the destination and collides with V en route.*

U acted because she had executory control. However, she lacked *mens rea* as she did not foresee the collision, nor could have done so. Consequently, she has not incurred fault liability for the collision (but might be liable for deploying the BCI, *infra*).

Furthermore, the agent must be in the requisite mental state at the moment of action, preceding or subsequent *mentes reae* are irrelevant (“concurrency principle”).¹⁹

Case 7: *U commands the BCI to cross the road, not foreseeing that V is also about to do so. The wheelchair hits V. Shortly before impact, U realizes*

¹⁷ For the differences between US and UK criminal law in recklessness, see Fletcher (1998).

¹⁸ Nakar et al. (2015) note that conditions for *mens rea* differ in US tort and criminal law.

¹⁹ See Ashworth (2013).

that V will be hit. While she could not stop the collision, it occurs to her that it is not a bad thing to hit V, he actually deserves to be hit. U thus intended to hit V at the moment of the crash. However, for lack of control, she did not act at that time. Her last action was commanding the BCI. As she did not have intent then, she did not harm V intentionally.

With these conditions of liability, we are in a position for a final assessment of cases 2 and 3. In case 2, U controls her wheelchair by mental actions, she has guidance control. In case 3, she has executory control. She acts in both cases. As she also hit V purposefully in both cases, she is liable, civil and criminal, for harming V.

3.3. Omissions

3.3.1. Liability for omissions

What about cases in which users do not act in the legal sense (such as cases 1, 4 and 5)? Person can also be responsible for failures to act – omissions. Liability for omissions will become increasingly relevant in the context of autonomous machines and BCIs. As a general legal principle, persons are responsible for omissions if and only if they had a duty to act. This duty differs from the general duty to not harm others through one's actions, because it requires to *prevent* impending harms to third-parties which are not proximally caused by one's own actions. Whereas *neminem laedere* is a negative duty that prohibits actions, liability for omissions requires failing to discharge a positive duty which commands performing an act. Notably, there is no general legal duty to avert harms threatening to befall others. Some jurisdictions stipulate so called “Good Samaritan” or “rescue clauses” that demand assistance in cases of accidents or life-threatening situations. But the relevant question here is whether BCI-users bear a different and more specific duty to avert harms to others that arise from the operation of the BCI.

We wish to submit that BCI users are regularly under such a duty to avert harms resulting from operations of their BCIs, even if not initiated or controlled by them. This is justified by the following considerations: Everyone bears the legal duty to avert harms emanating from one's body. A simple example is sneezing. For lack of control, it is not an action. Still, a person affected by a transmittable disease is obliged to take precautionary measures to not harm others through sneezing. The reason for the duty is that one's body is the source of the danger. The duty to prevent harm emanating from one's body should extend to BCIs. As a result, BCI-users have a duty of care to stop BCIs and controlled applications and machines from harming third-parties. This positive duty is a central aspect of BCI responsibility. Many more detailed questions about BCI-responsibility are about rendering scope and limit of this duty more precise.

Imposing this duty on BCI-users surely limits their freedom and requires justification. Several reasons speak in its favor: One might claim that BCIs and similar devices should be legally treated as body parts.²⁰ The general duty to stop harms emanating from one's body would then directly encompass BCIs. However, considering (external, removable) BCIs as part of the body is controversial. We wish to remain uncommitted. But even if one rejects this, BCIs connect bodies to devices that directly react to bodily signals. If one has the duty to avert harms emanating from X, and if X is coupled with Y, the duty to avert harms may well expand to those emanating from Y in its interaction with X, broadly conceived. Further arguments speak in favor of imposing responsibility on BCI-users. They have most (though limited) control over the source of the danger – the mind-brain-BCI-system –, at least much more than harmed parties have. This speaks in favor of placing the duty to oversee and control this system with them. Moreover, users are regularly the beneficiaries from deploying the BCI. And as those who benefit from an activity should bear its burdens, placing the duty on users is the best option. Accordingly, BCI-users owe a duty of care for the dangers and risks emanating from BCIs and controlled applications and devices. It exists independently of fault or evil

intentions, e.g. in cases of malfunctioning. Whenever BCIs are about to cause harm, users are obliged to intervene and avert it. Failing to discharge this duty gives rise to criminal and tort liability for omissions. On the subjective side, omissions can be intentional or negligent. When users foresee harm and refrain from intervening because they approve of the result or are indifferent to it, they omit intentionally, knowingly or recklessly. If they fail to foresee the ensuing harm, but could have done so, they might omit negligently.

Against this backdrop, let us revisit cases 1, 4 and 5 in which BCI-users did not act. The question is whether they had a chance to prevent the harmful event. If they had, but did not try, they might be liable for omissions. Accordingly, when the BCI in case 5 reads out U's desires and acts on them by punching V, the question is not so much about U's control over her non-conscious processes, but whether U was at any point in a position to avert the punching. As soon as she realizes that the BCI might have picked up her desires and is about to act on them, to punch V, she is obliged to intervene. If she fails to do so although she could have, she is liable for harming V by omission.

3.3.2. Built-in emergency stop

The duty of users to prevent harm is limited by their factual abilities to do so (*ought implies can*). If users cannot stop movements of controlled devices or avert harms through other means, such as warning victims, they cannot be blamed for omitting (discharging their duty was impossible). The only remaining charge is then negligently deploying a risky BCI-system in the first place (which we address in a moment). This leads to a general point about veto-options and technical BCI-design. We wish to urge that BCI designs should provide users with the opportunity to immediately halt further execution of operation of applications or devices, at every moment and with low effort, possibly through an additional channel (e.g., through blinking with an eye) or a “kill-switch”. The law may consider deploying BCI-systems without such emergency-stop options as negligence, for lack of means to avert foreseeable damages to others. This may even prompt regulatory agencies to ban unstoppable BCI-systems if they are potent to cause severe damage (spelling devices, for instance, may not qualify).

3.4. Transformations of agency

3.4.1. Mental control & effects on self-relations

The duty to avert harms combined with the presence of a stop option may have users constantly supervise the operation of BCIs and controlled devices, not unlike overseeing other machines, animals or kids. Some devices such as prostheses may move rapidly, which leaves users only a brief moment to intervene and requires quick reactions. Supervising such devices might soon become stressful. As consequence, users may find themselves in a somewhat transformed agential role: Instead of actively planning and initiating actions, they become more passive observers of devices. This is a peculiarity BCIs share with other “smart” technologies that let agents somewhat disappear. It ties in with a supposed loss of true agency through technology lamented at least since Heidegger (1977). One of the immanent dangers of this development is overreliance on technology, as the variation of case 1 shows. It also nicely illustrates that persons have the duty to *not* rely on technology if they cannot ensure its proper functioning.

Mental control of BCIs and the duty to monitor their operation may have further interesting psychological consequences: BCI-users will naturally try to generate brain signals that the device recognizes well. They learn to give better commands by generating “better” neural signals. This is remarkable. It is, in a sense, the adaption of one's mental operations to *technical parameters*. Usually, humans use their minds without being aware or caring about the underlying brain activity. But here, generating a particular neural signal becomes the aim of the mental task, and its quality and condition of success is decided by its measurability, its electro-magnetic properties. This is a novel form of *mind-to-brain* interaction (the dualist undertones here do not undermine the point): The performance of mental

²⁰ For a discussion see [Goold/Maslen/Auckland \(unpublished manuscript\)](#)

actions for the sake of creating specific brain signals. Moreover, the way in which persons learn to generate such brain signals is interesting. As they lack any kind of direct access to brain signals, they learn to modulate their brain activity through trial and error, by trying out different mental strategies or styles of thinking. The reaction of the BCI provides a form of neurofeedback. A user who controls an arm-prosthesis through an active BCI describes this adaptive process. He “has to try around and sometimes, [the BCI] is expecting a signal from my brain that looks slightly different” to previous ones. But he learns to generate this signal (Kögel, p. 13):

I “need to think in one direction, a little stronger than I did the first day we did it. That kind of stuff where, you know, instinctively, not even consciously adapt to what the computer ends up expecting. You just develop a rhythm for that”

These learning processes unfold intuitively, through nonconscious mechanisms. Too much conscious control can actually be counterproductive. Here is what another active BCI-user (who moves a prosthetic arm) describes (Kögel, p. 17):

“In the beginning, I was concentrating very hard. Then I was thinking, ‘Move right, move right, move right.’ As I stopped concentrating so hard, I just trusted that I knew how to do it, and started to just do it without thinking about how I was doing it, then it became much easier. I realized I didn’t have to try as hard as I was trying”. The control worked best “when I let instinct take over”

These intuitive methods of controlling devices resemble other automatic actions, and their conceptualization raises intriguing questions for action theory. For present purposes, the question is whether such intuitive, nonconscious forms of control qualify as action. Many legal systems view automated behavior, such as driving a car, as actions, although they are largely performed non-consciously, because they can be brought under conscious control if necessary. However, it is not clear that this reasoning applies to BCI-control because, as just described, consciousness may impair control. Non-conscious control is insufficient for action (users may still bear responsibility for omissions).

There is yet another side to this new form of acting. Users may have to suppress unwanted thoughts passing through their minds, or stop their minds from wandering, so that brain signals are not altered or distorted. Psychological research has shown how hard it is to intentionally avoid thinking about something, described as “ironic processes of mental control” (Wegner, 1994). Equally, emotional reactions may have to be suppressed as they may distort brain signals. This is experienced by a user who controls virtual avatars by the BCI in this way (Kögel, p. 5):

“I was not allowed to be upset, because that would have been a new signal, and that would have distorted the training ... You are not allowed to have emotions, none, completely, you must be complete dead in this sense”

These drastic description underlines that cognitive demands on users can be considerable and require extraordinary mental skills. Their exercise might be demanded by law. The expansion of the concept of action to the mental realm, as suggested here, brings with it the creation of mental duties, usually alien to the law. But insofar as averting harm to others requires suppressing thoughts or emotions, BCI-users are obliged to do so, to the best of their abilities, and may incur liability for failures. It is at least noteworthy that in this (limited) sense, “emotional coldness” might become a legal duty.

A final intriguing aspect of this interaction is that users will have to make predictions about the operations of the controlled devices, to ensure that commands are correctly interpreted and to prevent harmful movements. The user thus begins to interpret the BCI. The BCI, at the same time, is interpreting brain data of the user – a peculiar heuristic circle sets in.

3.4.2. Sense of agency

A psychological facet of acting through BCIs concerns the feeling of

acting or the *sense of agency* (Gallagher, 2007, 2013; Limerick et al., 2014). In ordinary actions, persons ascribe bodily movements to themselves and experience them as their actions, they feel a sense of agency. Research has shown that this sense arises under specific conditions and can be misleading (Haggard & Eitam, 2015; Wegner, 2002). Whether persons experience movements of their bodies as their own actions seems to depend on matching of information and feedback. If, for instance, a person desires to X and shortly after receives feedback that her arm is X’ing, she likely experiences the arm movement as her action, as the result of her volition which causally made her arm rise (Evans et al., 2015; Chambon et al., 2014). But strikingly, this sense of agency can be manipulated (Wegner, 2002; Moore et al., 2009). People then experience a sense of agency even though they were not, in fact, causally producing the movement. Especially in BCI-controlled prosthesis, mismatches are to be expected (Haselager, 2013; Vlek et al., 2014). Users will experience movements as their own although they neither initiated, nor controlled them (Bashford & Mehring, 2016). And conversely, for lack of feedback, persons may not realize that they are in fact causally directing the movement of a device such as a prosthesis. Kögel’s interviews with BCI-users revealed another interesting phenomenon. If movements were successful, users attributed their initiation and control to themselves, whereas they ascribed it to the BCI when they were unsuccessful. This peculiarities raise the question about the relevance of the sense of agency for responsibility ascriptions.

Case 8 (a variation of case 5): *U is angry at V and desires to punch him. Suddenly, her prosthesis moves into V’s direction and punches him. The BCI had detected U’s desire and executed it. U experiences the punching of V as her own action, as initiated by her, because her antecedent desire coupled with the immediately following execution gives her the feeling that she initiated and executed the punching. Asked what happened, she answers, in her view truthfully: “I hit him intentionally.”*

This is a mistaken ascription of agency. U falsely believes she acted by giving a go-command while in fact she did not. False beliefs about agency cannot ground legal liability. It is not imposed for having immoral desires or false beliefs, but only for causally generating adverse events with *mens rea*. Furthermore, BCIs might also lead to the reverse phenomenon, false rejections of agency:

Case 9: *Using an active BCI, U performs a mental action commanding the wheelchair to drive to the bathroom. The BCI reacts slowly, with significant delays, but drives to her destination. U does not experience the driving as an action of herself. Rather, she feels watching the act of a different entity (a computer). Asked, she would truthfully say: It wasn’t me, it was the BCI.*

Although U lacks a sense of agency, she initiated the movement via a mental action and thus acts. The missing sense of agency due to the delay between initiation and execution does not negate this. These two cases illustrate potential mismatches, illusions of agency. They lead into a peculiarity of BCI responsibility: Persons can be held responsible for acts, although they do not experience movements as their actions, and *vice versa*.

3.5. Foreseeability

A further condition of fault liability merits attention—foreseeability. It requires that users can foresee harmful events resulting from the operation of the BCI. One may say that the horizon of foreseeability marks the realm of fault liability. It can be judged by two standards: the point of view of the user, or a normatively infused objective standard (reasonable person). Both are relevant. Legal liability for intent, knowledge or recklessness regularly refers to the former, whereas negligent liability refers to the latter. Moreover, in a world with few certainties and many possibilities, foreseeability is a matter of degree (ranging from remote possibility to almost certainty). As jurisdictions define relevant thresholds differently, we only briefly address peculiarities with respect to BCIs.

3.5.1. Machine-learning algorithms and foreseeability

Lacking foreseeability is one source of responsibility gaps and gave rise to its original formulation (Matthias, 2004). Autonomous machines

run on algorithms which change the rules of decision-making by themselves, based on their “experiences” in diverse environments (“organic algorithms”). This is the basis for their capacity to learn, but also the grounds for the worry that algorithms become complex and their outputs unpredictable (EU Parliament, 2017). Here, responsibility gaps may emerge. As neither programmers of algorithms, nor producers of devices are able to foresee the evolution of code and resulting operations of devices, they might not be responsible for outputs. And as autonomous machines do not have operators, it appears that no one is left to bear liability (if not the machine itself). This marks the general responsibility gap problem.

In BCIs, machine-learning algorithms allow adaption to individual brain characteristics and changing environments. Depending on technical designs and controlled applications, algorithms are involved at several stages in the computing process that translates raw signals into the desired output, from signal extraction over classification to controlled applications (Müller et al., 2007; Rao, 2013). Insofar as outputs are unpredictable, the problem of responsibility gaps applies to BCIs as well.

Case 10 (malfunction): *U is a proficient BCI user, and the BCI software has evolved in specific adaption to her brain signals. One day she visits V, with whom she once had a love affair. Unfortunately, the BCI misinterprets the unfamiliar brain signals evoked by the exceptional encounter with V and directs her prosthesis to punch him. U did not have a chance to stop the rapid punching movement. U did not act. Producers and manufacturers, so we assume, cannot be held liable either because the evolution of the software was unforeseeable.*

This vignette characterizes cases of malfunctioning. Acquiring, processing and classifying brain signals is complex and susceptible to errors, especially with unusual or exceptional brain signals (Clausen, 2008). Even well-trained systems need periodic recalibration as brain activation patterns are not stable over time (Jarosiewicz et al., 2015). Users may not be able to foresee malfunctions, nor stop the device or prevent harmful events. At this point, insofar as producers are not liable, a responsibility gap seems to emerge.

3.5.2. Failures & liability for deployment

This supposed gap concerns all cases in which users are not liable for acting or omitting at the moment of the harmful event, e.g., in cases of malfunctioning. There are two main sources: Brain signals can be inadequate because users fail to generate good enough ones or they are distorted by neurophysiological noise. BCIs may also misinterpret signals. In practice, users are often not aware whether mistakes are due to failures of themselves, such as lacking mental effort, or to the machine. As one user notes: “Well, either I ... did not concentrate hard enough. But it could just as well be that perhaps the measurement was not optimal.” And another remarks: “That’s quite difficult for me, because sometimes I thought I could really focus and concentrate and it was quite a surprise, for instance ... I was thinking to the right and it went to the left”. So even users may be unaware of why a movement or command failed. And as long as it not clear that the failure originated from within their control, it might appear that they do not incur fault liability.

However we wish to call attention to the fact – also in regard to the broader debate about responsibility gaps — that in addition to producers or operators of machines, there is often a further party that may bear liability: The person who puts the autonomous machine to service, who arranges its running, determines the details of its performance, and reaps its benefit. With respect to cars or animal liability, this person is often called the “keeper”. Keepers might be responsible for damages caused by machines uncontrollable in the moment of harmful events, precisely because they brought an unpredictably moving entity into service and thereby put others at risk. Current debates on responsibility gaps tend to oversee this form of liability.

With respect to BCIs, this person is regularly the user herself. She deploys the BCI and thereby creates risks of malfunctioning. Users might hence be held responsible for their conduct on three related but separate charges: For actions that directly cause the harmful event, for omitting to avert it, or for deploying the BCI in the first place. For lack of a better term,

we call this *deployment liability*. It is grounded in the general duty to not harm others, which entails to refrain from deploying machines or tools that put others at risk, unless sufficient precautions are taken. Accordingly, people are held liable for harms resulting from deploying insufficiently safe machines. Deployment liability requires that users foresee, at the time of, or during, deploying the BCI, that it may malfunction and cause harm to third-parties. Defining the requisite degree of certainty more precisely is tricky and impossible in abstract. However, at present, given technical limitations, considerable error-rates (>5%), or unpredictable outputs from machine-learning algorithms, “reasonable persons” can foresee malfunctions of current BCI. Accordingly, the EU parliament notes that users of robotic prosthesis have to be made aware “of the perceptual, cognitive and actuation limitations” of these devices and the ensuing risks to others.²¹ Therefore, if BCI operations harm third-parties, users may often incur tort or criminal liability simply for deploying insufficiently safe BCIs. Then, no responsibility gap emerges (Holm & Voo, 2011, come to a similar result by different reasoning).

4. Risks, rights and disabilities

4.1. Negligence & permissible risks

This leads to questions about the degree to which putting others at risk is reasonable and permissible. Let us reformulate this in the language of negligence law (where legal systems show a lot variation).²² The gist of negligence, as understood here, is the creation of a substantial risk to others which is not justified, and the subsequent objectively foreseeable realization of this risk.²³ Negligent agents breach their general duty to not harm others by creating a risk, even if they fail to foresee its realization. It suffices that it was foreseeable by an objective standard, so that a “reasonable person” had taken precautions. Although one may call into question that such deviations from a standard of care suffice to ground criminal responsibility,²⁴ we assume with many jurisdictions that this is case.²⁵ As argued, reasonable persons have to expect that BCIs in their current state malfunction and thereby misdirect applications or machines. This creates a substantial risks to third-parties. But is it justified?

The creation of risks is not necessarily unjustified, there are what one may call permissible risks. If people had to avoid creating any risks, they could hardly act at all. Many beneficial activities entail risks as their flipside, and some of them realize with statistical certainty. And if benefits of actions outweigh risks, societies often permit them. Autonomous cars provide a good illustration. A duty to constantly supervise them to intervene in dangerous situations would run counter to many of their benefits, as their aim is to render drivers superfluous. Autonomous machines can realize their full (economic) potential only if they do not have to be constantly overseen by humans. This raises the question under which conditions societies are prepared to drop the duty of overseeing dangerous machines. This may foreseeably increase the rate of harmful incidents, but allows to reap the benefits of automation (Matthias, 2004). Technologies with favorable risk/benefit ratios should plausibly be introduced as long as a basic levels of safety are ensured. This reasoning underlies regulation of ordinary road traffic. In the EU, it kills almost 30.000 people annually and leaves another

²¹ See the Annex to EU 2017 – License for users.

²² See Raz (2010).

²³ See Fletcher (1998, Ch. 7). This is also the essence of negligence according to the US Model Penal Code. § 2.02 (d) defines negligently action when he should be aware of a “substantial and unjustifiable risk that the material element exists or will result from his conduct. ... The actor’s failure to perceive it, considering the nature and purpose of his conduct and the circumstances known to him, [has to involve] a gross deviation from the standard of care that a reasonable person would observe in the actor’s situation.”

²⁴ Alexander/Ferzan (2009).

²⁵ Cf. Hart (2008, Ch. 6 & 7).

100.000 permanently disabled. Although well foreseeable,²⁶ driving is permitted under strict conditions (licenses, traffic rules). The risks of road traffic are not eliminated but managed to benefit from mass transportation. The same is true for other hazardous technologies. Managing instead of eliminating risks is a feature of what sociologists have dubbed “risk societies” (Beck, 1992). Accordingly, legal orders allow the creation of some (“reasonable” or “permissible”) risks. In the absence of expressive codes, the realm of permissible risks and standards of care are assessed *ex post* by courts, taking into account a range of factors such as the risk/benefit ratio, the seriousness and probability of potential damages, burdens and costs of precautionary measures.²⁷ However, stipulating permissible risks *ex ante* provides better orientation for all parties. This requires a code that defines the standard of care and permissible risks. A consequence is that users who observes these standards regularly do not incur liability if harm occurs nonetheless. For instance, a driver usually does not bear fault liability for those crashes that happen even though everyone observes traffic rules.

Safety standards and permissible risks for BCIs have not been set (McGee, 2007).²⁸ In the absence of specific regulations, they fall under general ones. In the EU, this is the Medical Device Directive, which is concerned with technical matters of safety and efficacy, but is un-specific with respect to particular technologies. The Directive does not address duties of care tailored to BCIs. This current situation leads to a regulatory paradox: Because of weak oversight, there is a wide permission to use BCIs *de iure*, but at the same time, the scope of liability is not limited. This may deter BCI-use *de facto*.²⁹ A liberal regulatory regime thus has the effect that people make less use of the liberally regulated activity. More specific standards should thus be developed (Weinberger & Greenbaum, 2016). As it is unlikely that regulatory agencies initiate such processes in the near future, we encourage the well-connected BCI community to consider initiating a process of drafting safety standards itself. This would be significant step in defining the realm of permissible risk of BCI-use,³⁰ and prospects are good

²⁶ Predictions of the EU for road accidents beyond 2020 can be found at their website: www.ec.europa.eu/roadsafety

²⁷ Art. 4:102 EPTL exemplifies relevant factors: “The required standard of conduct is that of the reasonable person in the circumstances, and depends, in particular, on the nature and value of the protected interested involved, the dangerousness of the activity, the expertise to be expected of a person, the foreseeability of the damage, the relationship or proximity or special reliance between those involved as well as the availability and the costs of precautionary or alternative methods”

²⁸ In the EU, BCIs fall und the recently reformed regulation of medical devices. The new EU regulation (EU) 2017/745 which came into effect in 2017 demands “clinical evaluations” of risk-benefit ratios of devices (Art. 61). Pursuant to Art. 1 of Annex I, devices “shall be safe and effective and not compromise ... the safety and health of users or ... other persons, provided that any risks which may be associated with their use constitute acceptable risks when weighed against the benefits to the patient and are compatible with a high level of protection of health and safety, taking into account the generally acknowledged state of the art.” This standard demands weighing the benefits to users against harms. The benefits of restorative BCIs are potentially large and outweigh many harms to third-parties. They might be limited by the general requirement of a “high level of safety”. However: these standards are too broad and un-specific to demarcate the scope of permissible risks.

²⁹ Leenes et al. (2017), note in the same vein that “the fixation *ex ante* of high technical standards producers have to conform to before the product can be released onto the market” may provide solutions to unclear safety issues and therewith facilitate development and market release of novel technologies.

³⁰ These guideline should explicitly state the duty of users to prevent harms to others emanating from the operation of BCIs. The following additional considerations might be relevant: As machine learning algorithms likely shows weaknesses in environments different to those in which they were trained, users should take special caution in them. Perhaps regular training might significantly reduce misinterpretation rates. Further, users should report to technicians for recalibration, and debugging every time they have encountered a major security issue (e.g., uncontrollable dangerous movements of prostheses).

that regulators and courts will orient future assessments of care in the light of such expert-written guidelines.³¹

4.2. Rights of disabled persons

However, due care can reduce, but cannot eliminate risks. The degree to which creating remaining risks is permissible is ultimately a political decision of lawmakers and regulators. Concerning *restorative BCIs*, we wish to draw attention to an important factor to be accommodated. Advancing restorative BCIs is an important *public interest*. Insofar as they alleviate impairments due to disability, governments are positively obliged to advance them. This duty might be established under the domestic law. In the EU, it follows from Art. 26 of the Charter of Fundamental Rights, the right to integration of person with disabilities. In the global context, the UN Convention of Rights of Persons with Disabilities (CRPD), which entered into force in 2008, is crucial, clear and comprehensive.³² It obliges the (currently 160) signatory states to promote “full and effective participation and inclusion in society” (Art. 3c), and to research, develop and make available assistive technologies for disabled persons (Art. 4 g). The aim is to improve an independent live (Art. 9). Restorative BCIs are prime examples of the technologies to which the Convention refers. In the spirit of the Convention, we therefore suggest that developing, providing access to, and promoting the use of restorative BCIs in real life environments should be deemed as an important public interest.³³ This includes the creation of a regulatory framework that supports innovation and market access.³⁴ And this ties back to the previous discussion as it calls for defining a scope of permissible risks that facilitates the use of restorative BCIs, allowing for a margin of error greater than in non-therapeutic uses. This does not mean that disabled persons are allowed to *harm* others, but that they are allowed to create *risks*, some of which will realize. BCI-users still have the duty to supervise devices and take precautionary measures. But avoiding any risks is often impossible. In these cases, the restorative aim should negate the charge of negligent deployment liability.

5. Strict liability of users

The permission to create such risks will increase harms to third-parties. They should be able to seek redress and receive compensation for damages. To counter the permission of risky conduct, we thus suggest imposing strict civil liability on BCI-use. It ensures that no responsibility gap arises and that third-parties are compensated. This is a fair distribution of the risks of BCIs. The harmed third-party is the one that is least involved in creating or managing risks, it can usually neither prevent, nor insure damages. According to the (non-binding) Principles of European Tort Law, strict liability can be imposed for dangerous activities which create “a foreseeable and highly significant risk of damage even if all due care is exercised”.³⁵ For reasons of unreliability (*supra*), algorithm-driven BCIs fall

(footnote continued)

Perhaps, self-learning algorithms can be tested and audited regularly or improved by regular training routines. Protocols for specific fields of application might reduce risks. Measures as such may reduce risks of BCIs.

³¹ From a democracy perspective, such a making of soft-law through private actors is a bit dubious, but not uncommon especially in tech-regulation. At some stage, the law has to be open to factual developments and standards as they emerge in real life practice in business and industries. Lawmakers are, of course, free to adopt their own, overriding regulations.

³² United Nations 2007. Convention on the Rights of Persons with Disabilities. Adopted by the General Assembly 24 Jan 2007; A/RES/61/106.

³³ Just a brief note on human rights to the use of non-therapeutic BCIs: We suggest that users can invoke a different human right for non-therapeutic uses, the right to mental self-determination. Unfortunately, it is not (yet) recognized in international human rights law (Bublitz & Merkel, 2014).

³⁴ See (Bertolini, 2015).

³⁵ Art.5:101 para 1 and 2. PETL. In the DCFR, book VI deals with strict liability and enumerates a list of sources of dangers, from immovable and cars to

in this category. Furthermore, as users benefit from BCIs, they should bear its burdens. Finally, having users compensate third-parties provides an incentive to take due care. A strict liability regime is thus warranted and should be implemented according to specificities of domestic law.³⁶ In addition, to ensure that injured parties can recover damages, strict liability should be complemented by a mandatory insurance scheme.³⁷ Governments should encourage insurers to offer respective packages and should step in if they fail to do so. Moreover, we wish to mention at least *prima facie* attractive idea of “no-fault insurance” schemes that cover all damages regardless of responsibility. Their advantage is that liability does not have to be allocated between parties, overcoming costly litigation.³⁸ But as insurance systems and regulatory landscapes vary, we reserve further comments.

6. Deterrence gap

Strict liability bridges responsibility gaps. But many jurisdictions allow it only for civil compensation of losses, not for criminal liability, as punishment requires culpability. We agree with this position. This may leave a gap in *criminal* responsibility. Whereas BCI-users are liable for intentionally or negligently causing harms to third-parties, there is no room for negligent liability as long as they exercise due care. As an illustration, consider another malfunctioning case:

Case 11: *U uses her BCI-controlled prosthesis according to safety guidelines. The machine-learning BCI works well, but malfunctions occur randomly every now and then. One day, the BCI misinterprets brain signals, raises the prosthesis and hits V who happens to pass by. The movement was too quick for U to intervene.*

U neither intentionally harmed V, nor could she have stopped the prosthesis. If at all, she can only be blamed for deploying the BCI. But by observing guidelines, she discharges her duty of care. Absent further unlawful behavior, she is not criminally liable. The civil law damages can be recovered because of strict liability, as just argued. Here we encounter a responsibility gap in cases in which harms result from behavior which was pursuant to safety standards and within the realm of permissible risks.

One may ask whether it is fair that BCI users put others at risk, without them or anyone else bearing criminal responsibility if harms realize. Injured parties, even though their pecuniary damages are compensated by users because of strict liability, may complain about being exposed to risks and wish to be better protected against injuries by the legal order. This protection is usually attained through the deterring force of criminal law. Its absence may lead into, as we may say, a *deterrence gap*, the lack of punitive incentives to avoid negligently harming others.³⁹ However, legal deterrence of risky BCI-use would run counter to the permission to do so within the scope of permissible risks. For reasons of internal coherence, the law cannot permit a type of behavior and consider it a breach of a duty at the same time. If regulators follow our suggestion and create standards defining permissible risks of

(footnote continued)

animals and emissions. Malfunctioning BCI devices do not fall into these categories. Thus, national legislation would have to be passed (VI – 3:207).

³⁶ Strict liability also covers damages resulting from defective products. Of course, users should be able to take recourse to manufacturers in that case. The extent to which this is possibly depends on product liability rules, exceeding the scope of this paper. In this respect, the authors of the Robolaw study provide an remarkable idea. Palmerini et al. (2014) ponder whether to exempt producers from product liability to facilitate development. The state should then compensate for damages (p. 142 seq.).

³⁷ The EU Parliament (2017) has also adopted the call for mandatory insurance for robotics, at para 57.

³⁸ Nakar et al. (2015). Cf. Bertolini (2015) for further relevant remarks on liability distribution.

³⁹ It is related to Danaher's (2016) retribution gap, which emerges when people consider autonomous machines as proper targets of blame whereas the law does not.

BCI-use, a deterrence gap for behavior observant of these standards necessarily follows.

7. Epistemic gap & burden of proof

The gap in criminal responsibility for negligence is the flipside of the permission to use risky BCIs. Further worrisome gaps in the distribution of responsibility do not arise. However, BCIs may lead to an intriguing *epistemic* gap in criminal responsibility.

Case 12: *Suppose U intentionally hits V with her BCI-controlled prosthesis as she seriously dislikes him. U is responsible for intentionally inflicting physical harm, civil and criminal. However, to avoid charges, she denies having hit V intentionally and suggests the BCI must have misinterpreted brain signals.*

This raises an evidentiary problem. From the outside, it is in principle undetectable whether a BCI acts according to users' wishes or on its own accord, as neither mental states, nor brain signals of users or decision-procedures of BCIs are directly epistemically accessible. Recall that even for users, it often remains opaque whether their mental performance, some physiological fact, the BCI-detection, or the software was the cause of a malfunction. Thus, from the third-person perspective, there are two black-boxes: the mind-brain of the user and the BCI. This is a noteworthy epistemic inaccessibility that follows from the disembodied nature of BCI-movements. Usually, interactions between persons and tools are visible. Allegedly malfunctioning tools can be forensically analyzed. Moreover, if bodies move, it can be assumed that it is under control of persons unless indications point to the contrary (e.g., intoxication). Interactions between minds, brain signals and BCIs are different because none of these elements is perceptible from the outside. Thereby, an *epistemic gap* emerges at the intersection of person and BCI.

Some authors propose storing BCI-data in a way analogous to black-boxes in aviation (O'Brochain & Gordijn, 2014). This raises privacy concerns as intimate personal information might be extractable. Without storage, however, retrospective analyses of past operations of devices are impossible. Evidentiary problems call for storing BCI-data, which seems also necessary for purposes such as debugging. But a range of privacy and data security questions needs to be addressed before (rights to access brain data, self-incrimination). However, even such stored data may often be inconclusive or insufficient. Self-learning algorithms may have become too complex to understand.⁴⁰ Courts would need algorithm experts to reconstruct the decision-logic at the moment of action, and even if this is possible in theory, it is likely not in everyday legal practice.⁴¹

Apart from that, users can still claim that their mental states were different to what is suggested by brain signals. To falsify such testimonies, mental states would have to be decoded from brain data. But precisely that, so called reverse inference, is not possible at the moment, even with much richer fMRI data, and it is not likely that it will become so in the near future (Hutzler, 2014; Poldrack, 2006). This evidentiary problem is aggravated by the rules of criminal procedure that place the burden of proof on the prosecution. Accordingly, in the absence of corroborative evidence, reasonable doubts about the user having given the command detected by the BCI will often remain, leading to acquittals of defendants. Thus, although BCIs-actions do not create responsibility gaps, their disembodied nature and the opaque interplay between mind and machine generate epistemic gaps with practical consequences. If people avail themselves of the possibility to hide behind this technical veil of non-traceability, a practical responsibility gap arises. Unfortunately, we fail to conceive an adequate remedy. There might be technical solutions and we urge BCI engineers and regulators to address them.

⁴⁰ See Mittelstadt et al. (2016); Burrell (2016); Lepri et al. (2017).

⁴¹ Other potential solutions are discussed in Mittelstadt, Allo, Wachter & Floridi 2016; see also Wolkenstein et al. (2018).

8. Conclusion

Responsibility of BCI-users for damages to third-parties follows regular rules of liability, mainly tort and criminal law. BCI users can be liable on three separate counts: For actions, mediated by BCIs; for omissions to prevent harms arising from operations of BCIs; for deploying BCIs, provided harmful events were foreseeable. Assessing whether BCI-mediated movements constitute actions can be challenging. We suggest the law expands its concept of action to include the category of mental actions with physical consequences, and therewith, willfully initiated BCI-mediated movement. Most active and reactive BCI-users then qualify as actions, passive BCIs not. In addition, BCI-users might be liable for omissions, for failing to avert harms that BCIs and controlled devices inflict on others. Notably, we argue that, as a general rule, users have a duty to prevent harms to others arising from their BCIs, i.e., to oversee the operations of BCIs and controlled devices and intervene to prevent damages to others. This puts them in the role of supervisors of operations performed by BCIs. Discharging this duty might be stressful and lead into an interesting heuristic circle: BCI-users are interpreting the operations of a device which is interpreting their brain signals. Oversight of BCIs also calls for technical implementation of an emergency-stops.

However, given the short period of time users may have to stop unpredictable movements of devices, as well as various sources of malfunctioning and signal distortion, some harmful events are hard or impossible to prevent, even for dutiful, observant and bona fide users. Such cases are discussed in the context of responsibility gaps in automation and AI. However, according to the law as it stands, deploying BCIs would regularly be negligent, as error-rates are high and harms to third-parties foreseeable. Then, there is responsibility gap.

Virtually every country in the world is obliged to promote living conditions and social integration of people with disabilities, often by domestic law, and internationally by the CRPD. Restorative BCIs are a paradigmatic technology potentially improving the lives of the beneficiaries of these duties. To facilitate development, market release and use of BCIs, we suggest that regulators define permissible risks of BCI-use by stipulating more specific technical safety standards and user guidelines, detailing safety procedures, training phases, recalibration, auditing and more. As such activity by regulatory agencies is unlikely, we suggest that the BCI community and professional bodies consider drafting such guidelines and define standards of care themselves. All relevant stakeholders should be invited to participate. Then, risks created by BCI-use in conformity with these standards might be deemed as justifiable and permissible because of the overall societal benefits from restorative BCI-use. Harms resulting from other uses still provide grounds for negligent liability claims. While such standards may relieve BCI-users from both criminal and civil liability, we also call for imposing strict civil liability on BCI-users to cover foreseeable and non-foreseeable damages to third-parties. As users benefit from BCIs, they should also bear the burdens, possibly with a mandatory insurance.

Does BCI-use open up responsibility gaps? By and large, our answer is dismissively. However, the new disembodied mode of acting may lead to an *epistemic* gap in criminal responsibility. Because of the intimate interplay between person and BCI, it becomes hard to identify the root cause of a BCI-movement, whether it was caused by an intention of users or the BCI. This epistemic gap is deepened by the presumption of innocence. If it cannot be proven by other corroborating evidence and with a high level of certainty (beyond reasonable doubt), that a user and not the BCI caused a movement, users have to be acquitted *in dubio pro reo* from criminal charges. This might lead to an unacceptable practical gap in criminal responsibility. One countermeasure suggests itself: Recording and storage of BCI-data, including brain data, also for purposes of debugging. This, in turn, raises intriguing privacy concerns.

After all, current legal frameworks seem *cum grano salis* suitable to adequately allocate responsibility for BCI-mediated movements, though more detailed rules have to be stipulated by lawmakers or regulatory agencies. With a view to the future, our conclusion might turn out

transient. Once the merging of minds and machines progresses to another level and preferences, choices and decision-making are altered by machines operating on self-learning algorithms, current frameworks drawing on residual distinctions between minds and machines may become inadequate (von Bar et al., 2009).

Acknowledgements

This paper was written in the interdisciplinary research project INTERFACES, funded by the German Federal Ministry of Education and Research (BMBF) and ERA-NET Neuron, the European Research Area Network for Neuroscience Research. We like to thank Dr. Johannes Kögel, Munich, for allowing us to quote extensively from his unpublished manuscript.

Funding

Project has been funded by the German Ministry of Education & Research (BMBF) under a grant in the ERA-Neuron funding scheme.

References

- Aflalo, T., Kellis, S., Klaes, C., Lee, B., Shi, Y., Pejsa, K., ... Liu, C. (2015). Decoding motor imagery from the posterior parietal cortex of a tetraplegic human. *Science*, 348(6237), 906–910.
- Alexander, L., & Ferzan, K. K. (2009). *Crime and culpability: A theory of criminal law*. Cambridge University Press.
- Alpaydin, E. (2004). *Introduction to machine learning*. Cambridge, Mass: MIT Press.
- AlZu'bi, H. S., Al-Nuaimy, W., & Al-Zubi, N. S. (2013). EEG-based driver fatigue detection. In 2013 Sixth international conference on developments in systems engineering. Abu Dhabi: IEEE111–114. <https://doi.org/10.1109/DeSE.2013.28>.
- American Law Institute (1962). *Model Penal Code*.
- Ashworth, A., & Horder, J. (2013). *Principles of criminal law*. Oxford University Press.
- von Bar, C., Clive, E., & Schulte-Nölke, H. (2009). *Principles, definitions and model rules of European private law: Draft Common Frame of Reference (DCFR)*. Walter de Gruyter.
- Bashford, L., & Mehring, C. (2016). Ownership and agency of an independent super-numerary hand induced by an imitation brain-computer interface. *PLoS ONE*, 11(6), e0156591.
- Beck, S. (2016). The problem of ascribing legal responsibility in the case of robotics. *AI & Society*, 31(4), 473–481. <https://doi.org/10.1007/s00146-015-0624-5>.
- Beck, U. (1992). *Risk society: Towards a new modernity*. London; Newbury Park, Calif: Sage Publications.
- Bertolini, A. (2015). Robotic prostheses as products enhancing the rights of people with disabilities. Reconsidering the structure of liability rules. *International Review of Law, Computers & Technology*, 29(2), 116–136.
- Birbaumer, N. (2006). Breaking the silence: Brain computer interfaces (BCI) for communication and motor control. *Psychophysiology*, 43(6), 517–532. <https://doi.org/10.1111/j.1469-8986.2006.00456.x>.
- Brunner, C., Birbaumer, N., Blankertz, B., Guger, C., Kübler, A., Mattia, D., ... Müller-Putz, G. R. (2015). BNCI Horizon 2020: Towards a roadmap for the BCI community. *Brain-Computer Interfaces*, 2(1), 1–10. <https://doi.org/10.1080/2326263X.2015.1008956>.
- Bublitz, C., & Merkel, R. (2009). Autonomy and authenticity of enhanced personality traits. *Bioethics*, 23(6), 360–374.
- Bublitz, J. C., & Merkel, R. (2014). Crimes against minds: On mental manipulations, harms and a human right to mental self-determination. *Criminal Law and Philosophy*, 8(1), 51–77.
- Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 1–12.
- Chambon, V., Sidarus, N., & Haggard, P. (2014). From action intentions to action effects: How does the sense of agency come about? *Frontiers in Human Neuroscience*, 8. <https://doi.org/10.3389/fnhum.2014.00320>.
- Chaudhary, U., Birbaumer, N., & Ramos-Murguialday, A. (2016). Brain-computer interfaces for communication and rehabilitation. *Nature Reviews Neurology*, 12(9), 513.
- Chinen, M. A. (2016). The Co-evolution of autonomous machines and legal responsibility. *Va. J.L. & Tech.*, 20, 338.
- Clausen, J. (2008). Moving minds: Ethical aspects of neural motor prostheses. *Biotechnology Journal*, 3(12), 1493–1501. <https://doi.org/10.1002/biot.200800244>.
- Clausen, J. (2009). Man, machine and in between. *Nature*, 457.
- Consortium Future BNCI (2015). A roadmap for future directions in brain/neuronal computer interaction. Graz. Available at: http://bnci-horizon-2020.eu/images/bncih2020/FBNCI_Roadmap.pdf.
- Daly, J. J., & Wolpaw, J. R. (2008). Brain-computer interfaces in neurological rehabilitation. *The Lancet Neurology*, 7(11), 1032–1043.
- Danaher, J. (2016). Robots, law and the retribution gap. *Ethics and Information Technology*, 18(4), 299–309.
- Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55(10), 78–87. <https://doi.org/10.1145/2347736.2347755>.
- European Group on Tort Law (2005). *Principles of European tort law (EUTL): Text and*

- commentary. Wien: Springer.
- European Union, Parliament (2017). *Resolution on "Civil Law Rules on Robotics" (P8_TA(2017)0051, 16 February 2017)*.
- Evans, N., Gale, S., Schurger, A., & Blanke, O. (2015). Visual feedback dominates the sense of agency for brain-machine actions. *PLoS ONE*, 10(6), e0130019.
- Fletcher, G. P. (1998). *Basic concepts of criminal law*. New York: Oxford University Press.
- Galán, F., Nuttin, M., Lew, E., Ferrez, P. W., Vanacker, G., Philips, J., & Millán, J. D. R. (2008). A brain-actuated wheelchair: Asynchronous and non-invasive Brain-computer interfaces for continuous control of robots. *Clinical Neurophysiology*, 119(9), 2159–2169. <https://doi.org/10.1016/j.clinph.2008.06.001>.
- Gallagher, S. (2007). The natural philosophy of agency. *Philosophy Compass*, 2(2), 347–357.
- Gallagher, S. (2013). Ambiguity in the sense of agency. In A. Clark, J. Kiverstein, & T. Vierkant (Eds.), *Decomposing the will* (pp. 118–135). Oxford: University Press.
- Goold/Maslen/Auckland, Damages to Prostheses and Compensation for Harm (unpublished manuscript).
- Grübler, G. (2011). Beyond the responsibility gap. Discussion note on responsibility and liability in the use of brain-computer interfaces. *AI & Society*, 26(4), 377–382. <https://doi.org/10.1007/s00146-011-0321-y>.
- Haggard, P., & Eitam, B. (Eds.). (2015). *The sense of agency* Oxford University Press <https://doi.org/10.1093/acprof:oso/9780190267278.001.0001>.
- Hart, H. L. A. (2008). *Punishment and responsibility: Essays in the philosophy of law*. Oxford University Press.
- Haselager, P. (2013). Did I do that? Brain-computer interfacing and the sense of agency. *Minds and Machines*, 23(3), 405–418. <https://doi.org/10.1007/s11023-012-9298-7>.
- Heidegger, M. (1977). *The question concerning technology and other essays*. London: Garland.
- Holm, S., & Voo, T. C. (2011). Brain-machine interfaces and personal responsibility for action - maybe not as complicated after all. *Studies in Ethics, Law, and Technology*, 4(3), <https://doi.org/10.2202/1941-6008.1153>.
- Hutzler, F. (2014). Reverse inference is not a fallacy per se: Cognitive processes can be inferred from functional imaging data. *NeuroImage*, 84, 1061–1069. <https://doi.org/10.1016/j.neuroimage.2012.12.075>.
- Jarosiewicz, B., Sarma, A. A., Bacher, D., Masse, N. Y., Simeral, J. D., Sorice, B., ... Hochberg, L. R. (2015). Virtual typing by people with tetraplegia using a self-calibrating intracortical brain-computer interface. *Science Translational Medicine*, 7(313), <https://doi.org/10.1126/scitranslmed.aac7328> (313ra179-313ra179).
- Kellmeyer, P., Cochrane, T., Müller, O., Mitchell, C., Ball, T., Fins, J. J., & Biller-Andorno, N. (2016). The effects of closed-loop medical devices on the autonomy and accountability of persons and systems. *Cambridge Quarterly of Healthcare Ethics*, 25(04), 623–633. <https://doi.org/10.1017/S0963180116000359>.
- Kögel, J. Brain-computer-interfaces: Ergebnisse einer qualitativen Interviewstudie (unpublished manuscript).
- Lebedev, M. (2014). Brain-machine interfaces: An overview. *Translational Neuroscience*, 5(1), <https://doi.org/10.2478/s13380-014-0212-z>.
- Lebedev, M. A., & Nicolelis, M. A. L. (2006). Brain-machine interfaces: Past, present and future. *Trends in Neurosciences*, 29(9), 536–546. <https://doi.org/10.1016/j.tins.2006.07.004>.
- Leenes, R., Palmerini, E., Koops, B.-J., Bertolini, A., Salvini, P., & Lucivero, F. (2017). Regulatory challenges of robotics: Some guidelines for addressing legal and ethical issues. *Law, Innovation and Technology*, 9(1), 1–44.
- Lepri, B., Oliver, N., Letouze, E., Pentland, A., & Vinck, P. (2017). Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy & Technology*. <https://doi.org/10.1007/s13347-017-0279-x>.
- Limerick, H., Coyle, D., & Moore, J. W. (2014). The experience of agency in human-computer interactions: A review. *Frontiers in Human Neuroscience*, 8, 643. <https://doi.org/10.3389/fnhum.2014.00643>.
- Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, 6(3), 175–183.
- McGee, E. M. (2007). Should there be a law-brain chips: Ethical and policy issues. *24 TM Cooley L Rev*, 81.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 1–21.
- Monti, M. M., Vanhudenhuysse, A., Coleman, M. R., Boly, M., Pickard, J. D., Tshibanda, L., ... Laureys, S. (2010). Willful modulation of brain activity in disorders of consciousness. *New England Journal of Medicine*, 362(7), 579–589.
- Moore, J. W., Wegner, D. M., & Haggard, P. (2009). Modulating the sense of agency with external cues. *Consciousness and Cognition*, 18(4), 1056–1064.
- Moore, M. S. (2010). *Act and crime: the philosophy of action and its implications for criminal law*. New York: Oxford University Press.
- Müller, K.-R., Krauledat, M., Dornhege, G., Curio, G., & Blankertz, B. (2007). Machine learning and applications for brain-computer interfacing. In M. J. Smith, & G. Salvendy (Vol. Eds.), *Human interface and the management of information. Methods, techniques and tools in information design*. Vol. 4557. *Human interface and the management of information. Methods, techniques and tools in information design* (pp. 705–714). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-73345-4_80.
- Nakar, S., Weinberger, S., & Greenbaum, D. (2015). Legal and social implications of predictive brain machine interfaces: Duty of care, negligence, and criminal responsibility. *A JOB Neuroscience*, 6(4), 40–42.
- Nicolelis, M. (2011). *Beyond boundaries: The new neuroscience of connecting brains with machines—and how it will change our lives*. Macmillan.
- O'Brolchain, F., & Gordijn, B. (2014). Brain-computer interfaces and user responsibility. In G. Grübler, & E. Hildt (Vol. Eds.), *Brain-computer-interfaces in their ethical, social and cultural contexts*. Vol. 12. *Brain-computer-interfaces in their ethical, social and cultural contexts* (pp. 163–182). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-8996-7_14.
- O'Doherty, J. E. (2009). A brain-machine interface instructed by direct intracortical microstimulation. *Frontiers in Integrative Neuroscience*, 3. <https://doi.org/10.3389/fneuro.07.020.2009>.
- O'Doherty, J. E., Lebedev, M. A., Ifft, P. J., Zhuang, K. Z., Shokur, S., Bleuler, H., & Nicolelis, M. A. (2011). Active tactile exploration using a brain-machine-brain interface. *Nature*, 479(7372), 228.
- Pagallo, U. (2013). *The Laws of Robots*. Dordrecht: Springer Netherlands <https://doi.org/10.1007/978-94-007-6564-1>.
- Palmerini, et al. (2014). RoboLaw: guidelines on regulating robotics. Available at the website www.robotlaw.eu.
- Pfurtscheller, G., Allison, B. Z., Brunner, C., Bauernfeind, G., Solis-Escalante, T., Scherer, R., et al. (2010). The hybrid BCI. *Frontiers in Neuroscience*, 4(30), 1–11.
- Poldrack, R. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, 10(2), 59–63. <https://doi.org/10.1016/j.tics.2005.12.004>.
- Rao, R. P. N. (2013). *Brain-computer interfacing: An introduction*. New York: Cambridge University Press.
- Raz, J. (2010). Responsibility and the negligence standard. *Oxford Journal of Legal Studies*, 30(1 (2010)), 1–18.
- Roskies, A. L. (2010). How does neuroscience affect our conception of volition? *Annual Review of Neuroscience*, 33(1), 109–130. <https://doi.org/10.1146/annurev-neuro-060909-153151>.
- Sauer, H., Roughley, N., & Köhler, S. (2017). Technologically blurred accountability? Technology, responsibility gaps and the robustness of our everyday conceptual scheme. *Moral agency and the politics of responsibility* (pp. 65–82). Routledge.
- Schermer, M. (2009). The mind and the machine. On the conceptual and moral implications of brain-machine interaction. *NanoEthics*, 3(3), 217–230. <https://doi.org/10.1007/s11569-009-0076-9>.
- Simester, A. P. (2005). *Appraising strict liability*. Oxford University Press.
- Steinert, S., Bublitz, C., Jox, R., & Friedrich, O. (2018). Doing things with thoughts: Brain-computer interfaces and disembodied agency. *Philosophy & Technology*. <https://doi.org/10.1007/s13347-018-0308-4>.
- Tamburrini, G. (2009). Brain to computer communication: Ethical perspectives on interaction models. *Neuroethics*, 2(3), 137–149. <https://doi.org/10.1007/s12152-009-9040-1>.
- Thompson, M. C. (2018). Critiquing the concept of BCI illiteracy. *Science and Engineering Ethics*. <https://doi.org/10.1007/s11948-018-0061-1>.
- Vlek, R., van Acken, J.-P., Beursken, E., Roijendijk, L., & Haselager, P. (2014b). BCI and a user's judgment of agency. In G. Grübler, & E. Hildt (Vol. Eds.), *Brain-computer-Interfaces in their ethical, social and cultural contexts*. Vol. 12. *Brain-computer-Interfaces in their ethical, social and cultural contexts* (pp. 193–202). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-8996-7_16.
- Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review*, 101(1), 34–52. <https://doi.org/10.1037/0033-295X.101.1.34>.
- Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge, Mass: MIT Press.
- Weinberger, S., & Greenbaum, D. (2016). Are BMI prosthetics uncontrollable Frankensteinian monsters? *Brain-Computer Interfaces*, 3(2016), 149–155.
- Wolkenstein, A., Jox, R. J., & Friedrich, O. (2018). Brain-computer interfaces: Lessons to be learned from the ethics of algorithms. *Clinical Neuroethics Cambridge Quarterly of Healthcare*, (4), 27.
- Wolpaw, J., & Wolpaw, E. W. (Eds.). (2012). *Brain-computer interfaces: Principles and practice*. USA: OUP.
- Yuan, H., & He, B. (2014). Brain-computer interfaces using sensorimotor rhythms: Current state and future perspectives. *IEEE Transactions on Biomedical Engineering*, 61(5), 1425–1435. <https://doi.org/10.1109/TBME.2014.2312397>.
- Yuste, R., Goering, S., Arcas, B. A. Y., Bi, G., Carmena, J. M., Carter, A., Fins, J. J., ... Wolpaw, J. (2017). Four ethical priorities for neurotechnologies and AI. *Nature*, 551(7679), 159–163.