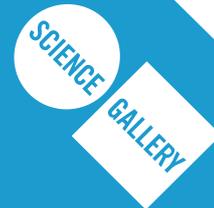
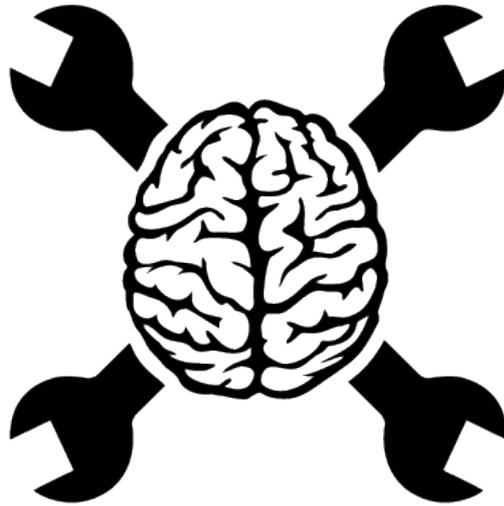


BrainHack



Ethical Oversight Article 1

Workpackage 3 // Deliverable 3.1



Brain**Hack**

DELIVERABLE

Project Acronym: BrainHack

Grant Agreement number: 686987

Project Title: BrainHack: Bringing the arts and sciences of brain and neural computer interface together

Deliverable reference number and title

D3.1 Ethical Oversight article 1

Revision: v1.0

Authors:

Prof. dr. Sabine Roeser (TUD)
Dr. Veronica Alfano (TUD)
Prof. dr. Caroline Nevejan (TUD)

Project co-funded by the European Commission within

H2020-FETOPEN-2015-CSA (Coordination and Support Activities)

Dissemination Level

P **Public** ✓

C Confidential, only for members of the consortium and the
Commission Services

Revision History

Revision	Date	Author	Organisation	Description
1.0	30 June 2017	Sabine Roeser, Veronica Alfano and Caroline Nevejan	TU Delft	Final

Statement of originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Disclaimer

'This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement **No 645791**'.

'The opinions expressed in this document reflect only the author's view and reflects in no way the European Commission's opinions. The European Commission is not responsible for any use that may be made of the information it contains.'

Table of Contents

1. Introduction.....	5
2. Risk, Art, and Moral Emotions: building blocks for a new theoretical framework.....	7
3. BNCI Technologies, Art, and Moral Reflection.....	13
4. Conclusion	23
References	24

The Role of Art in Emotional-Moral Reflection on Risky and Controversial

Technologies: The Case of BNCI

Sabine Roeser, Veronica Alfano and Caroline Nevejan, Faculty of Technology, Policy and Management, TU Delft

1. Introduction

New technologies promise thrilling new advances, but they usually have potentially negative side-effects or risks as well, which means that societal controversies often surround these technologies. Conventional approaches to risk are quantitative and technocratic, calculating potential positive and negative outcomes and comparing them on an aggregate level, as in a risk-cost-benefit analysis. These approaches have been criticized over the last decades by social scientists, psychologists, and philosophers, because they oversimplify the societal and ethical implications of risks and uncertainty (Shrader-Frechette 1991, Krimsky and Golding 1992, Slovic 2000, Hansson 2004, Roeser et al. 2012). There is now broad consensus that approaches to decision-making about risk and uncertainty have to be informed by the values and concerns of stakeholders (Asveld and Roeser 2009, Roeser et al. 2012). This also means that emotions should be explicitly addressed in debates about risky technologies, as they can draw our attention to important moral values (Roeser 2006). For example, compassion, indignation, and feelings of responsibility and care can point to ethical considerations such as autonomy, fairness, and justice. These ethical considerations are not included in technocratic approaches to risk, and they might be easily overlooked if emotions and their underlying moral concerns are not explicitly addressed in debates about risk (Roeser 2012).

However, stakeholders' emotions are also shaped by personal and cultural perspectives and might in that sense be biased (cf. Kahan 2012, Haidt 2012, Greene 2013). It can be hard to overcome such biases, which means that debates about risky and controversial technologies might not be fruitful. Works of art provide one possible way to broaden people's emotional and moral outlooks. By exploring scenarios in unexpected ways, art can challenge the imagination, involve people emotionally, and provide them with alternative viewpoints. It can exaggerate certain aspects of reality in order to bring them to the fore and scrutinize their complexities, thereby fuelling public debates (Nevejan 2007). Furthermore, art can investigate tendencies that are latent in society, and can make these tendencies explicit, by sketching futuristic scenarios. This is why artists, such as Leonardo da Vinci, are often among the first to envision new technologies. Indeed, today artists are on the cutting edge of various digital technologies, bio-technologies, and brain technologies. The way in which we perceive such works of art can broaden and challenge our moral and emotional outlooks on new, potentially risky and controversial technologies, as scholars from the field of cultural studies have shown (Da Costa and Philip 2008). Yet philosophers have not yet studied art as a catalyst for new perspectives on public deliberation about emerging technologies that are potentially risky or controversial. Hence, this constitutes an entirely new field of research for scholars of risk and moral theory (Roeser 2016).

In the pages that follow, we will explore this idea in the context of so-called BNCI (Brain Computer Neural Interface) technologies. These are particularly interesting from a philosophical standpoint because they engage with people's brains; thus they come especially close to the source of our thoughts, feelings, and personal identity, shedding new light on what it means to be human. We will draw on the experiences of artists, scientists, and engineers who are directly involved with BNCI technologies. These experts, whom one of us (Veronica Alfano) interviewed in-depth, took part in a Hackathon at Amsterdam's Waag Society in June 2016. We will discuss their views in light of philosophical and aesthetic theories, exploring relevant ethical and conceptual issues as well as topics for further investigation.

2. Risk, Art, and Moral Emotions: building blocks for a new theoretical framework

Philosophers have long reflected on the role that art can play in contributing to ethical reflection and political discussions (e.g. Carroll 2001, Nussbaum 2001, Gaut 2007, Adorno et al. 1980, Rorty 1989, Butler and Sontag 2007, Kompridis 2014). These philosophers have discussed works of art mainly to the extent that such works provide insights into interpersonal and societal relationships, as well as into the development of personal virtues. In his seminal work *A Theory of the Good and the Right*, Richard Brandt (1979) offers the utilitarian argument that a person's life is going well to the extent that she has what she would want if she were wellinformed and engaged both emotionally and imaginatively with possible ways of living. Art allows for precisely this kind of emotional and imaginative engagement, which cues ethical reflection.

However, neither Brandt nor the aforementioned philosophers who have written about art and ethical and political reflection have extended their discussions into deliberation about artworks that engage with technologies. Zwijnenberg (2009, 2014) is one of the few philosophers who studies bio-art, but he focuses neither on other forms of technology-engaged art nor on risk and emotions. Social scientists, literary critics, and scholars of cultural and media studies have considered technology-engaged art (Kember 1998, Ede 2000, Wilson 2002, Pinsky 2003, Zylinska 2002, Anker and Nelkin 2004, Da Costa & Philip 2008, Weichmann 2008, Reichle 2009, Munster 2013), yet they have not focused specifically on the interrelation among such works of art, moral emotions, and risk. While there are empirical studies of the role of journalistic images, narratives and emotions in climate change risk perception (Leiserowitz 2006, Spence & Pidgeon 2010) and studies of the contribution of narrated scenarios to moral reflection on biotechnology (Boenink et al. 2010), these studies do not focus on art.

Hence, while art that engages with risky and potentially controversial technologies is a highly dynamic field that requires critical philosophical reflection to theorize its potential impact on emotions and on moral and political deliberation, this area of research has hitherto remained largely uncharted. In this paper, we will explore a few philosophical issues that merit further consideration.

First we will take a look at several artists and writers who have explicitly or implicitly reflected on the moral values embedded in their work. One might think that even if artists have a duty to engage with ethical and political questions, they should do so obliquely and indirectly lest they risk sacrificing beauty and other aesthetic qualities on the altar of the quotidian or polemical. Artists do not need to resort to naïve preachiness or take on an explicitly didactic role in order to communicate aesthetic and moral values. Literature, for instance, can simultaneously embrace self-justifying beauty and lay claim to moral authority. The end of the nineteenth century saw the rise of a prominent *l'art pour l'art* movement, with Oscar Wilde proudly proclaiming in *The Picture of Dorian Gray* that “all art is quite useless.” All the same, considering the case of William Morris – poet, novelist, artist, designer, and political activist – demonstrates that such pure Wildean beauty can, paradoxically, itself take on social relevance and utility. Rather than devoting himself to the freewheeling amorality often associated with the late-century Decadents, Morris seeks instead to aestheticize morality. So the harmoniously symmetrical structure of Morris’s poetic stanzas, echoed in his patterned textiles and wallpaper designs, can signify concord and fellowship. Elaine Scarry links balanced aesthetic proportions to “distributive justice and fairness” in social relations (Scarry 1999, p. 97); Michelle Weinroth, too, proposes that for Morris, the beautiful “constitutes a symbol of societal commonweal in which social iniquities would be erased” (Weinroth 1996, p. 64).

Furthermore, Morris believes that art should be “the expression by man of his pleasure in labour” (“The Art of the People”). It therefore adumbrates a humanistic civilization in which all work is purposeful, in which people are free from unjust burdens, and in which everyone has leisure time to enjoy creations that reflect the individuality of skilled and satisfied makers. The

existence of such art can be seen as an attempt to establish this cultural context, and so as intrinsically worthwhile. A just society, as Kant might put it, is a condition for the possibility of art. In an ideal society, art and literature would not need to justify themselves via overt didactic usefulness, and the utility of an artwork could derive solely from its sensual beauty and other aesthetic qualities. The American poet George Oppen claims in his essay “The Mind’s Own Place” that “the definition of the good life is necessarily an aesthetic definition” rather than an explicitly political one; instead than making his or her work rhetorical, the poet should bear in mind that “a world without art is simply and flatly uninhabitable” (Oppen 2003, 181-2). Ezra Pound offers a variation on the theme of *l’art pour l’art*, and echoes the Aristotelian idea that art reveals one way in which to live a good life, when he declares that beauty produces “data for ethics” by providing insight into human nature and the sources of human happiness (Pound 1954, 46).

Morris, an avowed socialist who idealized the medieval era, execrated mechanized mass production on the grounds that it allowed for capitalist abuses of laborers and represented the crushing of the individual artisan’s creative spirit. Morris’s reservations about it aside, however, technology raises new questions about the ways in which art allows for emotional-moral reflection.

New technologies are often highly contested by society, because they are experienced as risky and are thought to have undesirable effects on our wellbeing, environment, and way of life. More and more artists are experimenting with these technologies; some take up an explicitly critical stance, but others prefer to play with the promises the technologies hold, even exploring possibilities that go beyond current scientific practice. For digital technology in the last three decades, a series of conferences has been instrumental to the emergence of the so-called art and technology paradigm: that is, Siggraph, CHI, Ars Electronica, Doors of Perception, Transmediale, Medialabs, and Hackathons in different parts of the world. Furthermore, the European Commission has included artists in several technology research funding calls.

When artists are among the first to explore new technologies, which has happened often during the last few decades of rapid technological change, their work sheds light on potential societal applications of these technologies. Dave Eggers's novel *The Circle* (Eggers 2013) provides a gripping narrative that at the same time contributes to critical deliberation on the risks of a society that is dominated by social media. The same could be said for "Nosedive," a recent episode of the popular British TV series *Black Mirror*. A very different type of artwork with a similar message is the "Web 2.0 Suicide Machine", which uses software technology as its medium. The suicide machine, which allows users to erase their data from social networks, resulted in a dispute with Facebook that has caused the artwork to be banned.

Artworks that engage with risky and (potentially) contested technologies can heighten awareness of urgent societal issues that these technologies might produce (Gessert 2003). They can provide for critical reflection on technological and scientific developments (Reichle 2009, Zwijnenberg 2009). When they critically engage with technologies in an exploratory phase, works of art do not qualify as *l'art pour l'art* in the Wildean sense. They also do not comment on society only implicitly or symbolically. Rather, they explicitly touch on the societal implications of developments that take place primarily outside the world of art, within the realm of scientific and technological practice. This gives rise to the question of how to assess such artworks. Do conventional criteria, such as beauty and thought-provoking self-reflexivity, suffice? Are these aesthetic criteria irrelevant? Moving beyond aesthetics could mean that a work that does provide for interesting ethical insights might not be recognizable as an artwork – or that we need new aesthetic criteria.

But art that engages with technologies also generates different ethical questions than do conventional artworks. This fact complicates the claims of the aforementioned philosophers who theorize art's contributions to ethical and political reflection. Crucially, artworks that engage with new and potentially risky technologies do not primarily focus on (inter)personal relationships and individual virtues, but rather on the impact that technology can have on those relationships, virtues, and constituents and of wellbeing and their distribution – as well as (for instance) on the natural environment. Sven Ove Hansson (2012) has argued that the

topic of risk requires new ethical theories, as conventional ethical theories are not well-equipped to deal with uncertain outcomes. Similarly, we need new theories, frameworks and perspectives concerning the relationship between art and morality in the context of risky and controversial technologies. We also need hermeneutical studies that provide for in-depth interpretations of such artworks. These theories and studies should focus on (for example) which aspects of risky and potentially contested technologies are examined in a given artwork, which ethical issues are addressed, and in which way the artwork offers a fresh perspective on these issues.

Arguably, artworks have the potential to reach people in a more direct way than scientific studies or theoretical ethical argumentations (cf. Zwijsen 2009). Artworks offer sensory experiences and affect people's emotions, trigger their imagination, and provide them with concrete examples and unexpected perspectives – all of which can prompt emotional and moral reflection much more directly than abstract reasoning. And because they often exaggerate societal issues, artworks can entice and oblige people to ask questions that are scientifically and morally complex.

This means that artworks could play an important role in public debates about risky and contested technologies. Traditional approaches to risk communication focus on disseminating information; this favours those with technological expertise. It obstructs genuine ethical discussion, stakeholder participation, and the creation of educational experiences for all involved parties (Roeser and Pesch 2016). Over the last decades, scholars have proposed various approaches to participatory technology assessment or participatory risk assessment, which aim for more democratic decision-making about risky and contested technologies (cf. van Asselt & Rijkens-Klomp 2002). However, as noted above, public debates are still often not very fruitful. It can be hard for people to grasp the scientific, technological, and ethical complexities of technologies. Moreover, it can be difficult to overcome one's own emotional-moral inclinations and be open to the perspectives of others. Artworks that engage with risky, contested technologies can make an important contribution to participatory technology assessment. By appealing to the imagination, they

can augment moral awareness and intellectual flexibility. While traditional risk communication focuses on technical information, as well as on narratives and images that are meant to convey a specific message to the public, artworks instead engage people in critical and open-minded reflection. Works of art typically explore ambiguities rather than providing clear-cut answers, which means that they can be a powerful trigger for public deliberation, explicitly inviting different perspectives and thus breaking deadlocks in debates. Of course, inviting art into the arena of public deliberation about risky technologies is not a guaranteed recipe for success. Public deliberation will always be a complex matter, but artworks allow for a different kind of exploration of new themes and can provide an additional way to open up a constructive dialogue (Reichle 2009, Castells 2012).

A complicating factor is that the way in which an artwork can work as a catalyst for deliberation on risky and controversial technologies is defined not only by the artwork itself, but also by its positioning at a specific time and in a specific context or community. Hence, how the artwork will be used, displayed or exhibited, by whom, at what kind of location, at what point during a societal debate, can make a considerable difference in how it will be interpreted by the public.

For example, an artwork engaging with nuclear energy displayed by scientists at a nuclear research institute will attract a different audience and hold different connotations than that same artwork displayed in an interactive installation in a public space. It can also make a difference whether this installation has been placed there by a public body, by a private hightech company, or by an art gallery. In other words, the meaning of the artwork is mediated by many contextual factors that can be perceived as more or less neutral, strategic, or even potentially manipulative (Benjamin 2008, Barthes 1972). This means that using art to contribute to public debates on risky technologies is itself a potentially risky activity, requiring careful consideration of how to ensure that it invites and stimulates a genuinely open dialogue.

3. BNCI Technologies, Art, and Moral Reflection

Having discussed the contribution that art could make to emotional-moral deliberation about risky and contested technologies more generally, we would now like to focus on a specific technology and artworks that engage with it, namely so-called BNCI technologies and related artworks.

BNCI technologies

There are currently important technological developments that allow computers to connect to people's brains; these are so-called Brain Neural Computer Interface Technologies, or BNCI technologies for short (or alternatively, BCIs: Brain Computer Interfaces). BNCI technologies allow to control external devices by translating activities of the central nervous system (Wolpaw and Wolpaw 2012). Such technologies offer possible treatments for people suffering from brain defects (cf. e.g. Riccio et al. 2016), and they allow scientists to study processes inside the brain. BNCI technologies have great promise, but they also give rise to ethical concerns. For example, there are concerns about privacy, because information from people's brains could become publically available; about responsibility and personal identity, because people's decision-making capacities could be interfered with; about safety and security as well as responsibility and liability, because people could be harmed and because individuals could manipulate one another's brain-states; about informed consent, for example when using BNCI technologies with patients who suffer from locked in syndrome (for discussions of ethical aspects of B(N)CIs, cf. Tamburrini 2009, Haselager et al. 2009, and Nijboer et al. 2013).¹ BNCI technologies may have a significant impact on our understanding of what it means to be human, and on our sense of self. Thus they prompt and require profound ethical reflection.

BNCI Art

¹ Thanks to Angela Riccio for providing us with these references.

The question of what new experiences are being addressed by BNCIs cues us to draw connections to other traditions of human imagination. If ethical maturity should involve decisions informed by empathy and imagination, and if such imaginative and emotional capacities can be assisted by works of art, as we argued in the previous sections, then perhaps it is also the case that art provides us with a means to imagine both the promises and the risks of BNCI technologies – that is, the ways in which they might bring pleasure, provide people with profound aesthetic experiences and improve the quality of life of for example patients with (severe) brain defects, as well as the ways in which they might give rise to serious ethical concerns.

In this section we focus primarily on what we call ‘BNCI art,’ by which we mean art that raises issues related to Brain Neural Computer Interface technology. Thus broadly defined, BNCI art can engage directly with BNCIs and can use these technologies, but it can also employ other media and other strategies to address issues related to BNCI. BNCI technology can be used to make certain brain processes visible. In BNCI art, this can be performed aesthetically and can include active audience participation. An example of this is Karen Lancel and Hermen Maat’s “EEG Kiss.” An EEG (electroencephalogram) is a test that uses electrodes to detect electrical activity in the brain. In “EEG Kiss,” the EEG data of two people who are kissing is projected onto a screen or in circles on the ground. Here the audience experiences a wholly new portrait of a kiss, of a relationship, of intimacy, namely in the form of a visual representation of brain activity. Seeing a graphic representation of the brain activity of two people who are kissing can create a sense of discomfiting voyeurism – but can also give rise to curiosity and fascination. It can lead people to engage in deliberation about such an interface and to what it can lead. If a portrait of an intimate kiss can be made, what other experiences (or characteristics) of brain activities can be visualized via BNCIs and BNCI art? Digital technologies introduced the concept of Interface, a space-in-between that allows human beings to communicate across different places and times; BNCI technologies have the potential to collapse this threshold space into a single physical location while

offering intimate experiences that are triggered by the art piece directly. “EEG Kiss” exemplifies such potential developments.

Performance artists have used various forms of technology to experiment with direct intervention into the physical integrity of a person, resulting in a loss of control over his or her body. As early as the 1990s, Arthur Elsenaar connected electrical sensors directly to the muscles of his face and triggered these with a computer system he designed. In the artwork “Face Shift” (2005), his facial expressions are controlled by this system as he loses control over his own muscles. In 1995, during a Telepolis festival in Luxembourg, the artist Stelarc linked his neuromuscular system to the Internet in the performance “Ping Body.” Audiences at festivals in various European cities could move his muscles remotely and make him dance (Massumi 2002). Because there was also a sound connection, this became an uncanny experience. Stelarc could not control his own body anymore, and he became exhausted by all the remotely-triggered dance moves he was forced to make. This gives rise to the question of responsibility. If Stelarc had died then and there, who would have been responsible? The maker of the sensors, the internet service providers who allowed the signals to be sent, the members of the audience, the technicians in Luxembourg, or Stelarc himself? This case illustrates the classical ethical problem of many hands. It is also directly relevant for ethical reflection on BNCI technologies, illustrating possible complexities in a very tangible way: when for example a locked-in patient is aided by a BNCI technology, how can we ensure they really consent with what is happening to them? Our communication with the patient is (partially) mediated by the technology. The Stelarc example illustrates that, artistic performance aside, at least in a medical context, there should for example be agreements under which conditions to stop using the technology.

There are many additional examples of BNCI art that give rise to interesting ethical issues. Marina Abramovic’s “Mutual Wave Machine,” for instance, involves an interactive neurofeedback installation that connects one person to another through brainwave synchronization. This raises questions relating to privacy and personal identity. Guy Ben-Ary has developed CellF, “the world’s first neural synthesizer,” based on biological neural

networks that grow in a Petri dish and control various synthesizers that play with human musicians. This work casts a new light on creativity and artistic intentions.

In health care, scenarios for introducing BNCI technologies are considered more and more as solutions to help people recover (Riva 2005). The arts also focus on other aspects of these technologies, including their risks. But BNCI technologies can at the same time provide exciting new repertoires for aesthetic experience. Can we experience beauty and sense-making in ways we do not know about yet? BNCI technologies can allow for interventions in the brain of the artist, and perhaps ultimately even in the brain of the audience, that trigger direct expressions and experiences that can no longer be controlled by the person who is participating in the BNCI. What happens if the technology becomes invisible and only the changed behaviour can be observed? As with all new technologies, scenarios for risk and trouble and scenarios for beauty and salvation alternate.

It is important to note that many of the aforementioned artworks are in a sense transparent. The artist shows the technology that is being used – it is, as it were, “raw” technology – and the audience can experience the possibilities that such technology creates. Such “living labs,” as they are called today (by, for example, the EU Commission) allow for the exploration of new technologies. Art happens in public spaces; art is made for an audience to perceive; artists are specialists in designing experiences for others; living labs, then, are cutting-edge examples of art’s capacity for social engagement. We will now turn to the Hackathon in Amsterdam as a prime example of such a living lab.

The BrainHack Hackathon in Amsterdam: Experiences of Participants and Philosophical Reflection

In June of 2016, a group of BNCI artists, scientists, programmers, and technologists gathered at Waag Society in Amsterdam for a so-called “Hackathon” with the provocative title “Hack the Brain.” This was an intensive three-day workshop meant to encourage brainstorming, innovation, and networking. Such Hackathons are increasingly popular events that provide unique opportunities for collaboration. One of us conducted in-depth interviews

with several of the participants.² The Amsterdam meeting allowed us to gain insight not only into the participants' projects but also into their ideas about the ethical aspects of BNCI technologies, particularly regarding the role of art in such questions.

Conversations with Hackathon participants thus furnished an illuminating overview – and several concrete illustrations – of issues that have arisen in scholarly debates about BNCI ethics as mentioned above (cf. e.g. Haselager et al. 2009). For instance, what about the question of privacy? Who should be granted access to BNCI data, and who owns it? If brain-states could be influenced or manipulated in addition to being measured, if consciousness itself could be hacked, then what would become of human intellectual autonomy and free will? How would our conception of selfhood have to be adjusted? In the future, will the notion of human identity ineluctably shift, or should we say to brain technology “this far and no farther”? We will examine these issues based on the interviews with Hackathon participants in what follows. Furthermore, we will raise an issue that has not yet been discussed in the literature on art, ethics and technology, namely concerning potential ethical and legal constraints on artists.

Paul R. Wolpe has considered the ramifications of “information technology [...] integrated into our neural tissue.” He concludes that although humans “have evolved for over 100,000 years with the brain isolated in the skull, inviolate,” brain technology means that it is “inviolate no more” (Wolpe 2007, p. 130). At the Hackathon, some participants had already accepted this shift as a reality. Nowadays, they said in interviews, many people's identities are already radically multiple; thanks to social media, for example, we can be many different selves in different contexts. BNCI interventions would merely reflect this state of affairs. It seems that ideas of collective identity, of membership in various social networks, are displacing the notion of a single essential self. Yet this process need not be dehumanizing. Why would one have to make the seat of selfhood obscure or singular or inaccessible in order to consider it sacred? Emotions may appear to comprise a secure fortress of authentic self – but can't they

² We would like to thank the following Amsterdam Hackathon interviewees for their time and reflections: Jennifer Canary Nikolov, Luis Miguel Girao, Jason Farquhar, Aniela Hoitink, Olof van der Werf, Peter Friess, and Marcel van Brakel.

always be manufactured or manipulated? If reading or watching *Romeo and Juliet* makes us cry even though we have not experienced any of its tragic events ourselves, how is this instance of factitious sentiment different from (say) a feeling of sadness brought on by brain stimulation? What's more, added some Hackathon participants, perhaps it is time to re-evaluate the status of privacy. Why, exactly, is it valuable? Why is it considered a right? It may be the case that both privacy and autonomy are inevitably compromised by our burgeoning addiction to connectivity, to smart systems, and to the free flow of information. In fact, our sensitivity to invasions of privacy is constantly changing; many these days hardly blink an eye at the idea that their photograph and location are readily available to anyone with an internet connection.

This line of argumentation relates to “technological determinism” (cf. e.g. Smith and Marx 1994), i.e. the idea that technological developments are unavoidable and that we therefore need to accept them as a reality. However, the fact that something happens does not mean that it cannot be avoided. Furthermore, this argument is an example of the naturalistic fallacy (Moore 1903): the fact that something happens does not necessarily mean that it is morally acceptable. Indeed, contemporary ethicists of technology are developing approaches to responsible innovation and “designing for values” that are meant to steer technological developments into directions that are ethically and societally acceptable (cf. Van den Hoven et al. 2015). And as argued above, this is exactly where art can make an important contribution: by explicitly stimulating ethical reflection and exploring possible scenarios and their societal implications, thereby contributing to critical reflection and decision making of the public as well as to that of technology developers and policy makers.

Like science fiction, BNCI art can ask “what if?,” and can thus become an instrument for imagining potential futures. Silvia Casini writes that it “evokes the dimension of possibility rather than that of actuality” in order to create “unexplored patterns among words, things, and images” (Casini 2010, n.p.). To paraphrase Alfred Hitchcock, drama is life with the dull bits cut out; similarly, BNCI art can strip away extraneous detail and expose the essential core of an issue. It can function as a societal thought-experiment conducted in the imaginative

realm. In exaggerating certain aspects of day-to-day reality, art that engages with technologies lets us temporarily experience the potential consequences of technological developments (Zwijnenberg 2009, Reichle 2009, Gessert 2003). It calls forth possible worlds for us to live in.

One Hackathon participant remarked that one should know what the consequences of a technology are before one starts developing it, but without starting it is hard to see how one could find the consequences out, which seems paradoxical. This relates to the so-called Collingridge dilemma (Collingridge 1980). This dilemma states that risks or negative side effects of technologies only emerge clearly once the technology is already in use in society, but then ethical restrictions and regulations come too late to be effective, and people may already have suffered from the negative side effects. The alternative is to require further research until we have sound knowledge that a technology is safe, as expressed by the precautionary principle. However, this principle is both ambiguous and problematic; it raises the question of “how safe is safe enough?,” and it could mean that people have to use a less beneficial technology than is the current status quo (cf. Ahteensuu and Sandin 2012 for an extensive discussion of the precautionary principle). Another way to avoid the Collingridge dilemma has been proposed by Ibo van de Poel. He argues that we should interpret the introduction of new and potentially risky technologies as a social experiment, analogous to the experimental stages in medical research. This approach means that when there are well-grounded ethical concerns, we should develop stopping rules in advance as we introduce a technology into society (Van de Poel 2013). Nevertheless, even in such a case, our interventions might come too late to stop people from experiencing the technology’s negative side effects. Here art that engages with risky and controversial technologies can provide a fruitful reflective resource. Because artists often take technologies to extremes and explore scenarios in tangible ways, technology-engaged art can transcend the limitations of the lab. Furthermore, such artworks usually do not introduce genuine risks, as they are still confined to a restricted setting. Thus art that engages with emerging technologies can contribute to ‘responsible innovation’ by providing technology developers, policy makers, and the broader

public with a 'safe haven' to explore and reflect on the possible impact of these technologies (cf. Damm et al. 2013).

Indeed, during Hackathon collaborations, artists often found themselves pushing forward into new realms of possibility while scientists, who felt more bound by pragmatic limitations, pulled back. This may sound like a destructive and oppositional relationship, but it can actually be a form of cooperation in which consensus is reached by means of conflict. As mentioned above, artworks that engage with risky technologies open up an additional reflective space before the release of a technology into society. Artists can explore the promises and risks of new technologies in a more playful way than can scientists and engineers. This enables them to explicitly consider societal and ethical complexities and ambiguities. They are not bound by strict scientific demands and can therefore take technologies to greater extremes, well before those technologies are ready to be introduced into society. What's more, though art might not be universally accessible, it is often more accessible than the work of scientists.

Even when art is not overtly didactic, it can become a 'prosthesis for the imagination', as one Hackathon participant phrased it. And when it elicits a strong emotional reaction, when it inspires awe or anger or disgust or hope or fear, it can inspire crucial debates (Roeser 2016, cf. Roeser 2011 on the role of emotions for moral decision making). Perhaps there is an ethical imperative to use the aesthetic realm in precisely this way. Yet this creates a dilemma: how does one balance artistic freedom with the expectation that certain artists can contribute to public conversations about new and potentially risky and controversial technologies? It is not a simple matter to ensure that artists have the freedom to explore their own ideas while also sparking public discussion and doing justice to ethical, legal, and scientific complexities and constraints.

Relatedly, should artists working with risky technologies have to undergo reviews by ethics committees, just as scientists do? An example from another area of technology-inspired art can help to illustrate this. The international art, science, and ethics project "Trust Me, I'm an Artist" initiated an ad hoc ethics committee at the Transmediale Festival in Berlin

in February 2016.³ At that event, a panel of ethicists (of which Sabine Roeser was a member) discussed the work “Cellular Propeller” by bioartist Howard Bolland. Bolland initially worked on test animals in a scientific lab to create floating organic propellers, which would use the heart cells of rats. However, to avoid issues concerning animal ethics he eventually turned to using his own human material. This resulted in a very different artwork given the more limited possibilities; it became a race among human sperm cells, questioning norms of masculinity. Ethically the new artwork was less controversial, but aesthetically the original one may have been more compelling.

This gives rise to the question of whether artists should obey the same ethical and legal rules as scientists who develop technologies. One could argue that the benefits deriving from artworks are less clear than those deriving from new technologies, and that technology developers should therefore have more freedom. They have the power to improve societal wellbeing, despite the potential downsides or risks associated with their work. This could apply, for example, to the use of life-stock and animal experiments – as in the case of Howard Bolland, who imposed stricter conditions on himself than the technology developers at whose side he was working. Alternatively, one could argue that because artworks will not be released into society for wider use, artists should be granted more freedom than technology developers. Or perhaps artists and scientists deserve the same room for exploration as long as they do not harm the public or the environment, which would mean that they should be bound by the same ethical and legal constraints. These questions have not yet been discussed in the academic literature. However, since more and more artists are now working with potentially risky technologies, the time is ripe for philosophers to open this debate.

Hackathon artists were asked to meditate on these issues, questioning the function of their work and the values underlying it. Some simply believed that their task was to create, to

³ <http://trustmeimanartist.eu/projects/cellular-propeller-howard-bolland-project/> (website accessed October 28th 2016).

make something new, and that this value – in contrast with technology developers’ implicit duty to improve the quality of human life – was entirely amoral; some extended this line of thinking by linking true innovation to provocation, and thus to sparking discussion of controversial issues; some remarked that, paradoxically enough, art is primarily meant to teach people to think for themselves. It prompts them to fine-tune their capacity for imagination, to become more curious, to realize that the world is not fixed or static. Perhaps art, then, can help those who encounter it to live a good life precisely because it is liberated from specific moral obligations. Here again is a way to reconcile Wildean uselessness with utility.

It is a glaring problem for contemporary ethicists that most definitions of “the human” exclude or marginalize disabled individuals. A topic deserving further exploration is how BNCI art may help philosophers recalibrate these definitions. What’s more, if such art has the capacity to inspire developers of BNCI technology that can aid the disabled and improve their quality of life, then perhaps one can associate viewing and creating such art with moral virtue. Jennifer Canary Nikolov, who participated in the Amsterdam Hackathon, has pioneered an artistic research project called *Labyrinth Psychotica* that uses augmented reality systems to simulate psychosis. Nikolov hopes that her project will serve as a tool both of reflection and of destigmatization. It will make psychosis less alien, creating common feelings and experiences that will lead to sensitive and informed conversations. And ultimately, it will posit a more holistic and less isolated view of selfhood.

There is, then, a universalizing aspect to BNCI art. In externalizing or directly tracing or seeking to make concrete the impulses of the brain, such artworks have the ability to bypass or nullify the traits that threaten to divide us. People of different genders, ages, races, religions, sexualities, nationalities, political affiliations, and socio-economic backgrounds all have brains in common. Thus BNCI art can create a broad community of viewers or participants based in shared humanity.

This universalizing power might seem to give BNCI art access to pure and unmediated representation – after all, it appears to tap directly into the brain. But pure representation is

always inevitably a fantasy. What larger issues of authenticity and mediation might this open? There is an ethical dimension to this line of thought as well. Standpoint epistemology, a field of inquiry associated with feminist philosophy, maintains that there is no such thing as a value-free description (cf. eg. Harding 2004, Harraway 1985). An implicit perspective, a set of biases, underlies any description; in the same way, the standpoint of the viewer and the effects of the artistic medium shape any aesthetic representation (Berger 2008). Moreover, the perspective of the developers of BNCI technologies is not value-free, and BNCI art makes the implicit priorities and points of view that are involved in these technologies more explicit, facilitating ethical reflection on desirable ways to develop and use them. Art, once again, provides raw material for ethical questions.

Finally, BNCI art may also shed fresh light on much-debated aesthetic issues. Artists of many stripes tend to be concerned with expanding or recalibrating ideas of the human, enhancing and/or virtualizing experience, making thoughts and dreams and motions of the mind concrete and external, and aligning the perspectives of those who view their work. Perhaps, then, BNCI art distills or literalizes these fundamental artistic aspirations. It provides a fertile testing ground for issues surrounding both representation and the mind.

4. Conclusion

In this article, we have argued that public debates about risky and controversial technologies can profit from including artworks as triggers for emotional-moral reflection. We first sketched the initial building blocks of a new philosophical framework that can shed light on these issues. We then focused in particular on BNCI technologies and artworks that engage with them. Such artworks are an especially interesting case in point, as BNCI technologies can directly interfere with our brains, consciousness, and identity, raising questions about what it means to be a unique person. They can also bring up issues relating to privacy, informed consent, and personal integrity. We discussed a few salient artworks that illustrate these moral issues, engaging the audience more directly in emotional-moral reflection than abstract ethical and scientific theorizing might be able to do. We then went on to consider the

experiences of participants in a Hackathon. These scientists, technologists, and artists provided a useful overview of ethical challenges surrounding BNCI technologies and their links to art. We discussed these challenges in light of ethical and aesthetic theories. Important issues that arose from this discussion concerned privacy, personal identity, and technological determinism. We asked how artworks can contribute to ethical reflection about risky, contested technologies and to responsible innovation, even when they do not explicitly take on an ethical stance. By being inherently open and non-purposive, artworks can inspire more profound thinking. We also asked whether and in what way there should be ethical and legal ramifications for artworks that engage with risky technologies – bearing in mind that it is crucial not to undermine the autonomy of art, which paradoxically becomes useful to society by having no specific use and by flouting predetermined restrictions. All these points merit further investigation; our aim in this article has been to highlight possible avenues for future research regarding the role of art in emotional and moral reflection on risky and controversial technologies, illustrated in particular by BNCI art.

Acknowledgements

The authors wish to thank the EU-funded FET-Open project “BrainHack” for making this research possible. This project also co-sponsored the “Hack the Brain” Hackathon that we discuss in this article. We would like to thank the Hackathon participants for their input during the interviews that we discuss in this article. Furthermore, we would like to thank Angela Riccio, Jurre Ongering and Lucas Evers for their comments on earlier versions of this article.

References

- Asveld, Lotte, and Sabine Roeser (eds.) (2009), *The Ethics of Technological Risk*, London: Routledge / Earthscan
- Adorno, Theodor, Walter Benjamin, Ernst Bloch, Bertold Brecht, Georg Lukacs (1980), *Aesthetics and Politics*, New York: Verso

Barthes, R. (1972 [1957]), *Mythologies*, Trans. Annette Lavers. New York: Hill and Wang, 302-06

Benjamin, W. (2008 [1936]), *The Work of Art in the Age of Mechanical Reproduction*. Penguin UK

Berger, J. (2008), *Ways of Seeing*, Penguin UK

Brandt, Richard (1979), *A Theory of the Good and the Right*, Oxford: Clarendon Press

Butler, Judith, & Sontag, Susan (2007), 'Torture and the ethics of photography', *Environment and Planning D: Society and Space*, 25, 951-966

Casini, Silvia (2010), 'The Aesthetics of Magnetic Resonance Imaging (MRI): From the Scientific Laboratory to an Artwork', *Contemporary Aesthetics* 8

Castells, Manuel (2012), 'Art and Meaning', in Caroline Nevejan (ed.), *Witnessing You: on trust and truth in a networked world*, Participatory Systems Initiative, Delft University of Technology

Collingridge, David (1980), *The Social Control of Technology* (New York: St. Martin's Press; London: Pinter

Da Costa, Beatriz and Kavita Philip (eds.) (2008), *Tactical Biopolitics: Art, Activism, and Technoscience*, Cambridge (MA): The MIT Press

Damm, Ursula, Bernhard Hopfengärtner, Dominik Niopek & Philipp Bayer (2013), 'Are artists and engineers inventing the culture of tomorrow?' *Futures* vol. 48, 55-64

Eggers, Dave (2013), *The Circle*, Vintage Books, Random House New York

Gessert, George (2003), 'Notes on the Art of Plant Breeding', in *L'Art Biotech Catalogue*, exhibition catalog, Nantes: Le Lieu Unique, 47

Gaut, Berys (2007), *Art, Emotion and Ethics*, Oxford University Press

Greene, Joshua (2013), *Moral Tribes*, New York: Penguin

Haidt, Jonathan (2012), *The Righteous Mind: Why Good People Are Divided by Politics and Religion*, New York: Vintage Books

Hansson, Sven Ove (2012), 'A Panorama of the Philosophy of Risk', in Sabine Roeser,

Rafaela Hillerbrand, Martin Peterson and Per Sandin (eds.) (2012), *Handbook of Risk Theory*, Springer, pp. 27-54

Hansson, S. O. (2004), 'Philosophical perspectives on risk', *Techné*, 8, 10–35

Harding, Sandra (ed.) (2004), *The Feminist Standpoint Theory Reader*. New York and London: Routledge

Haraway, D. J. (1985), 'A Manifesto for Cyborgs: Science, Technology, and Socialist Feminism in the 1980s', *Socialist Review* 80: 65-107

Haselager, Pim, Rutger Vlek, Jeremy Hill, and Femke Nijboer (2009), 'A note on ethical aspects of BCI', *Neural Networks* 22: 1352–1357

Kahan, Dan (2012), 'Cultural Cognition as a Conception of the Cultural Theory of Risk', in Sabine Roeser, Rafaela Hillerbrand, Martin Peterson and Per Sandin (eds.) (2012), *Handbook of Risk Theory*, Springer, pp. 725-759

Kompridis, Nikolas (ed.) (2014), *The Aesthetic Turn in Political Thought*, London: Bloomsbury Academic

Krimsky, S., and D. Golding (1992), *Social Theories of Risk*, Westport, CT: Praeger

Massumi, B. (2002). *Parables for the virtual: Movement, Affect, Sensation*. Duke University Press

Moore, G.E. (1988) [1903], *Principia Ethica*, Buffalo, NY: Prometheus Books

Nevejan, C. I. M. (2007), *Presence and the Design of Trust*, Dissertation. University of Amsterdam

Nijboer, Femke, Jens Clausen, Brendan Z. Allison & Pim Haselager (2013), 'The Asilomar Survey: Stakeholders' Opinions on Ethical Issues Related to Brain-Computer Interfacing', *Neuroethics* 6:541–578

Nussbaum, Martha (2001), *Upheavals of Thought*, Cambridge: Cambridge University Press

Oppen, George (2003), *Selected Poems*, ed. Robert Creeley. New York: New Directions Books

Pound, Ezra (1954), *Literary Essays of Ezra Pound* (ed. T. S. Eliot). London: Faber and Faber Limited

Reichle, Ingeborg (2009), *Art in the Age of Technoscience: Genetic Engineering, Robotics, and Artificial Life in Contemporary Art*, Vienna, New York: Springer

Riccio, A. F Pichiorri, F Schettini, J Toppi, M Riseti, R Formisano, M Molinari, L Astolfi, F Cincotti, D. Mattia (2016), 'Interfacing brain with computer to improve communication and rehabilitation after brain damage', *Progress in Brain Research* 228: 357–387

Riva, G. (2005). Virtual reality in psychotherapy: review. *Cyberpsychology & behavior*, 8(3), 220-230

Sabine Roeser, Rafaela Hillerbrand, Martin Peterson and Per Sandin (eds.) (2012), *Handbook of Risk Theory: Epistemology, Decision Theory, Ethics, and Social Implications of Risk*, Springer

Roeser, Sabine (2011), *Moral Emotions and Intuitions*, Palgrave Macmillan

Roeser, Sabine (2012), 'Risk Communication, Public Engagement, and Climate Change: A Role for Emotions', *Risk Analysis* 32, 1033-1040

Roeser, Sabine and Udo Pesch (2016), 'An Emotional Deliberation Approach to Risk', *Science, Technology and Human Values* 41: 274-297

Roeser, Sabine (2016), 'How Art Can Contribute to Ethical Reflection on Risky Technologies', Liisa Janssens (ed.), *The Art of Ethics in the Information Society*, Chicago / Amsterdam: University of Chicago Press / Amsterdam University Press

Rorty, Richard (1989), *Irony, Contingency, and Solidarity*, Cambridge: Cambridge University Press

Ahteensuu, Marko and Per Sandin (2012) 'The Precautionary Principle', in Sabine Roeser, Rafaela Hillerbrand, Martin Peterson and Per Sandin (eds.) (2012), *Handbook of Risk Theory:*

Epistemology, Decision Theory, Ethics, and Social Implications of Risk, Springer, pp. 961-978

Scarry, Elaine (1999), *On Beauty and Being Just*. Princeton: Princeton University Press

Shrader-Frechette, K. (1991), *Risk and Rationality: Philosophical Foundations for Populist Reforms*, Berkeley: University of California Press.

Slovic, P. E. (2000), *The Perception of Risk*, London, UK: Earthscan.

- Smith, Merritt Roe, Leo Marx (eds.) (1994) *Does Technology Drive History?: The Dilemma of Technological Determinism*, Cambridge, MA: The MIT Press
- Tamburrini, Guglielmo (2009), 'Brain to Computer Communication: Ethical Perspectives on Interaction Models', *Neuroethics* 2(3): 137-149
- Van Asselt, M. B., & Rijkens-Klomp, N. (2002), 'A look in the mirror: reflection on participation in Integrated Assessment from a methodological perspective', *Global Environmental Change*, 12(3), 167-184
- Van de Poel (2013), 'Why New Technologies Should be Conceived as Social Experiments', *Ethics, Policy & Environment* 16: 352-355
- Van den Hoven, Jeroen, Pieter Vermaas and Ibo van de Poel (eds.) (2015), *Handbook of Ethics, Values, and Technological Design*, Dordrecht: Springer
- Zwijnenberg, Robert (2009), 'Preface', in Reichle, Ingeborg (2009), *Art in the Age of Technoscience: Genetic Engineering, Robotics, and Artificial Life in Contemporary Art*, Vienna, New York: Springer, xiii-xxix
- Weinroth, Michelle (1996), *Reclaiming William Morris: Englishness, Sublimity, and the Rhetoric of Dissent*. Montreal: McGill-Queen's University Press
- Wolpaw, J.R., and E.W. Wolpaw (eds.) (2012), *Brain-Computer Interfaces: Principles and Practice*, 1st Edn. New York, NY: Oxford University Press
- Wolpe, Paul R. (2007), 'Ethical and Social Challenges of Brain-Computer Interfaces', *Virtual Mentor, AMA Journal of Ethics* 9.2:128-131.

www.hackthebrainhub.com



<https://github.com/hackthebrain>



<https://twitter.com/HackTheBrainHub>



<https://vimeo.com/hackthebrainhub>



https://www.youtube.com/channel/UCKIak_GxRqYMRYefHgApuDw/featuredwww.vim



<https://www.facebook.com/hackthebrainhubproject/>



"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 686987"