Aus dem Institut für Ethik, Geschichte und Theorie der Medizin
Institut der Ludwig-Maximilians-Universität München
Vorstand: Prof. Dr. Georg Marckmann

Autonomie und neue biomedizinische Technologien

Dissertation
zum Erwerb des Doktorgrades der Medizin
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München
vorgelegt von

Johannes Pömsl
aus Rosenheim
2021

Mit Genehmigung der Medizinischen Fakultät der
Universität München

Berichterstatter:	Prof. Dr. Georg Marckmann
Mitberichterstatter:	Prof. Dr. Fuat S. Oduncu Prof. Dr. Claudia Bausewein Prof. Dr. Dr. Ernst Pöppel
Mitbetreuung durch die promovierte Mitarbeiterin:	Dr. Dr. Orsolya Friedrich
Dekan:	Prof. Dr. med. dent. Reinhard Hickel

Tag der mündlichen Prüfung: 01.07.2021

Eidesstattliche Versicherung

durch

Johannes Sebastian Pömsl
geboren am 07.08.1987
in Rosenheim

Hiermit versichere ich an Eides statt, dass ich die eingereichte Dissertation selbständig und ohne unzulässige fremde Hilfe verfasst, andere als die in ihr angegebene Literatur nicht benutzt und dass ich alle ganz oder annähernd übernommenen Textstellen kenntlich gemacht habe. Außerdem versichere ich, dass die vorgelegte elektronische mit der schriftlichen Version der Dissertation übereinstimmt und die Abhandlung in dieser oder ähnlicher Form noch nicht anderweitig als Promotionsleistung vorgelegt und bewertet wurde.

München, 05.07.2021, Johannes Pömsl

Ort, Datum, Unterschrift

INHALTSVERZEICHNIS

1.) Abkürzungsverzeichnis

AE	Autonomy Enhancement
BCI	Brain-Computer Interface
cLiS	complete Locked-in-Syndrome
CE	Cognitive Enhancement
DBS	Deep-Brain-Stimulation
EEG	Elektroenzephalographie
GCE	Genetic Cognitive Enhancement
OCD	Obsessive-Compulsive Disorder
PCE	Pharmaceutical Cognitive Enhancement
SSRI	Selective Serotonin Reuptake Inhibitor / selektiver Serotonin Wiederaufnahme Hemmer

2.) Publikationsliste

Friedrich, O., E. Racine, S. Steinert, J. Pömsl, and R. J. Jox. 2018. An Analysis of the Impact of Brain-Computer Interfaces on Autonomy. *Neuroethics*. doi:10.1007/s12152-018-9364-9.

Pömsl, J., and O. Friedrich. 2017. Why Enhancing Autonomy Is Not a Question of Improving Single Aspects of Reasoning Abilities through Neuroenhancement. *Neuroethics* 10 (2): 243-254. doi:10.1007/s12152-016-9299-y.

3.) Bestätigung der Ko-Autoren

Erklärung zur Autorenschaft

Artikel:

Pömsl, J., and O. Friedrich. 2017. Why Enhancing Autonomy Is Not a Question of Improving Single Aspects of Reasoning Abilities through Neuroenhancement. *Neuroethics* 10 (2): 243-254. doi:10.1007/s12152-016-9299-y.

Hiermit bestätige ich, , dass oben genannte Veröffentlichung in einer Zusammenarbeit zu jeweils gleichen Anteilen mit Johannes Pömsl entstand. Sowohl konzeptionelle Vorüberlegungen, als auch Recherche, argumentativer Aufbau, Schlussfolgerungen und Revisionen wurden gemeinsam erarbeitet. Es besteht eine geteilte Erstautorenschaft. Diese Publikation wird und wurde nicht von mir im Rahmen einer anderen Dissertation verwendet.

[redacted]

Ort, Datum, Unterschrift

Confirmation of authorship

Article:

Friedrich, O., E. Racine, S. Steinert, J. Pömsl, and R. J. Jox. 2018. An Analysis of the Impact of Brain-Computer Interfaces on Autonomy. *Neuroethics*. doi:10.1007/s12152-018-9364-9.

I, [redacted] hereby confirm that the publication mentioned above is a result of a collaboration between myself and the authors mentioned above.
As first author, Orsolya Friedrich was responsible for the underlying hypothesis and the methodological approach of the paper. Furthermore, all the authors as co-authors contributed equally to research, construction of arguments, conclusions and revisions. This publication is not being used by me and was not being used by me as part of another dissertation.

[redacted]
Place, Date, Signature

Confirmation of authorship

Article:

Friedrich, O., E. Racine, S. Steinert, J. Pömsl, and R. J. Jox. 2018. An Analysis of the Impact of Brain-Computer Interfaces on Autonomy. *Neuroethics*. doi:10.1007/s12152-018-9364-9.

I, [redacted], hereby confirm that the publication mentioned above is a result of a collaboration between myself and the authors mentioned above.
As first author, Orsolya Friedrich was responsible for the underlying hypothesis and the methodological approach of the paper. Furthermore, all the authors as co-authors contributed equally to research, construction of arguments, conclusions and revisions. This publication is not being used by me and was not being used by me as part of another dissertation.

[redacted]
Place, Date, Signature

Confirmation of authorship

Article:

Friedrich, O., E. Racine, S. Steinert, J. Pömsl, and R. J. Jox. 2018. An Analysis of the Impact of Brain-Computer Interfaces on Autonomy. *Neuroethics*. doi:10.1007/s12152-018-9364-9.

I, [redacted] hereby confirm that the publication mentioned above is a result of a collaboration between myself and the authors mentioned above.
As first author, Orsolya Friedrich was responsible for the underlying hypothesis and the methodological approach of the paper. Furthermore, all the authors as co-authors contributed equally to research, construction of arguments, conclusions and revisions. This publication is not being used by me and was not being used by me as part of another dissertation.

[redacted]

Place, Date, Signature

Erklärung zur Autorenschaft

Artikel:

Friedrich, O., E. Racine, S. Steinert, J. Pömsl, and R. J. Jox. 2018. An Analysis of the Impact of Brain-Computer Interfaces on Autonomy. *Neuroethics*. doi:10.1007/s12152-018-9364-9.

Hiermit bestätige ich, [redacted], dass oben genannte Veröffentlichung in Zusammenarbeit mit Dr. med. Dr. phil. Orsolya Friedrich, Eric Racine, Ph.D., Johannes Pömsl und Prof. Dr. med. Dr. phil. Ralf J. Jox entstand. Als Erstautorin war Orsolya Friedrich federführend in Bezug auf die grundlegende Fragestellung und methodische Ausrichtung der Arbeit. Weiter wurden Recherche, argumentativer Aufbau, Schlussfolgerungen und Revision gemeinsam innerhalb der Autorengruppe erarbeitet. Für die weiteren genannten Autoren besteht der Status einer Ko-Autorenschaft. Diese Publikation wird und wurde nicht von mir im Rahmen einer anderen Dissertation verwendet.

[redacted]

Ort, Datum, Unterschrift

4.) Einleitung

4.1) Der Begriff der Autonomie im Kontext neuer medizinischer Entwicklungen

Überlegungen zum Begriff der Autonomie stellen einen Grundpfeiler moderner (Medizin-) Ethik dar [1]. In der Regel wird Autonomie als hohes moralisches und intrinsisch wertvolles Gut beschrieben, das es unbedingt zu bewahren gilt. Im Kontext medizinischer Fragestellungen geht es häufig um ein Verständnis von (Patienten-)Autonomie, das sich auf konkrete Handlungen bezieht: unter welchen Voraussetzungen kann davon ausgegangen werden, dass sich ein Patient oder eine Patientin selbstbestimmt für oder gegen eine Behandlung entscheidet? In einer einflussreichen Analyse haben etwa Beauchamp und Childress Bedingungen wie Verständnis für Art, Nutzen und Risiken eines Eingriffes, eine bewusste Entscheidungsfindung sowie Freiheit von steuernder Einflussnahme durch Dritte formuliert, die situationsgebunden eine Entscheidung als hinreichend autonom erscheinen lassen [1, 2]. Klassische Fragestellungen betreffen, neben der Zustimmung zu oder Ablehnung von Behandlungen, Themen wie Wünsche von Patientinnen und Patienten nach aktiver Sterbehilfe, Inanspruchnahme genetischer Diagnostik oder Lebendorganspenden [1].

Im Zuge neuer biowissenschaftlicher oder technologischer Entwicklungen etwa in der Psychopharmakologie, der Genetik oder im Bereich der Mensch-Maschine-Interaktionen scheint sich der Fokus in den Diskussionen um Autonomie innerhalb der letzten Jahre und Jahrzehnte verschoben zu haben: Weg von der Fragestellung, wann eine Entscheidung in einem bestimmten Kontext als autonom gelten kann, hin zu Überlegungen, welche Auswirkungen im weitesten Sinne medizinische Eingriffe auf die Autonomie oder Fähigkeit zur Autonomie sowohl von Patientinnen und Patienten als auch von potentiellen „gesunden“ Nutzerinnen und Nutzern haben können. Diese Fragen werden umso drängender, je spektakulärer etwa die Meldungen zu psychopharmakologischem „Enhancement“ aller Art, zu Möglichkeiten zielgerichteter Eingriffe in menschliches Erbmaterial oder direkten Gehirn-Computer-Schnittstellen werden, und je invasiver, unveränderbarer und potentiell unkontrollierbarer diese Interventionen erscheinen.

Diese Promotionsarbeit bündelt zwei Veröffentlichungen, die sich intensiv mit Analysen zu Autonomie im Kontext solcher aktueller medizinischer und technologischer Entwicklungen beschäftigen: Einerseits im Rahmen der Debatte um sogenanntes „Neuroenhancement“, also die Verbesserung kognitiver Fähigkeiten, in unserem Falle mittels pharmakologischen und

genetischen Enhancements. Andererseits als Analyse potentieller Auswirkungen sogenannter Brain-Computer Interfaces, bisher zumeist therapeutisch genutzter Schnittstellen zwischen Gehirn und Computern, auf menschliche Autonomie. Zwangsläufig werden dabei verschiedene Ansätze und Möglichkeiten beschrieben, sich einer unverändert bestehenden Kernfrage aktueller Debatten in der Medizinethik zu widmen: Was genau ist „Autonomie“, was genau meint „Selbstbestimmung“? Je stichhaltiger man die Auswirkungen neuer biomedizinischer Eingriffe sowie Technologien oder psychotroper Medikamente auf menschliche Autonomie analysieren möchte, desto genauer muss dieser Begriff in seinen konzeptionellen Bedeutungen und Verästelungen verstanden werden. Dabei stößt die bioethische und im spezielleren die neuroethische Debatte immer wieder an ihre Grenzen.

4.2) Pharmakologisches und genetisches Neuroenhancement und Brain-Computer-Interfaces

In dieser Arbeit werden Überlegungen zu möglichen Auswirkungen neuer medizinischer und technologischer Entwicklungen auf die Möglichkeiten von Patientinnen und Patienten oder Nutzerinnen und Nutzern, Autonomie zu verwirklichen, versammelt. Veröffentlichung I beschäftigt sich mit sogenanntem „pharmakologischem und genetischem Neuroenhancement“ [3]. Der Begriff „Enhancement“, abgeleitet von englisch „to enhance“, meint dabei zunächst nur ein „Steigern, Verbessern“ unterschiedlichster Fähigkeiten, in der Regel, im Gegensatz zu medizinischer Therapie, über das durchschnittliche Maß hinaus [4]. Der Gedanke, menschliche Fähigkeiten oder gar Lebensbedingungen mittels biomedizinischer Technologien zu verbessern, lässt sich bis ins 16. Jahrhundert und den Beginn der neuzeitlichen Wissenschaft zurückverfolgen [4]. Die aktuelle Debatte um sogenanntes „Neuroenhancement“ meint zumeist die Verbesserung kognitiver Fähigkeiten durch die Einnahme von Psychopharmaka oder, noch deutlich spekulativer, durch gezielte Veränderungen am menschlichen Genom. Bei genauerer Betrachtung bestehen auch diese Diskussionen zumindest in Teilen bereits seit der ersten Hälfte des 20. Jahrhunderts [5].

Veröffentlichung II analysiert mögliche Auswirkungen sogenannter Brain-Computer Interfaces (BCI), Schnittstellen zwischen Gehirn und Computern, auf menschliche Autonomie [6]. Vereinfacht ausgedrückt handelt es sich bei BCIs um technische Vorrichtungen, die Gehirnaktivität zum Beispiel mittels Elektroenzephalographie (EEG) aufnehmen können, über ein Interface an einen Computer weitergeben, wo die Daten nach bestimmten Algorithmen

analysiert werden und zu einem Effekt, beispielsweise das Bewegen eines Cursors auf einem Bildschirm, führen [7]. Potentiell können damit ohne jegliche Muskelaktivität Handlungen durchgeführt werden. An BCI-Technologie wurde mit wechselnder Intensität seit den 1960er Jahren geforscht, etwa seit Beginn des 21. Jahrhunderts erlebt das Feld eine Blüte mit mittlerweile knapp 200 Forschungsgruppen weltweit und vielfältigsten Anwendungen und Fortentwicklungen [7, 8].

Die Notwendigkeit genauer ethischer Diskussionen und Analysen sowie normativer Einschätzungen solcher Entwicklungen scheint auf der Hand zu liegen – schließlich handelt es sich in beiden Fällen potentiell um Technologien, die tiefgreifende Veränderungen am einzelnen Menschen, der Gesellschaft, sowie, besonders deutlich im Falle bestimmter genetischer Manipulationen, an heutigen und folgenden Generationen und letztlich der „conditio humana“ [4] hervorrufen können.

Gleichzeitig erfolgt in vielen Debattenbeiträgen notwendigerweise eine Vermengung zwischen ethischer und philosophischer Theorie auf der einen Seite sowie aktueller, hochspezialisierter biomedizinischer und technischer Forschung auf der anderen Seite, die eine stichhaltige und präzise Argumentation häufig erschwert. Ein Beispiel von vielen sei hier genannt: im Rahmen der derzeit ebenfalls vehement geführten Debatte um „moralisches Enhancement“ führt deGrazia in einem Artikel von 2014 das Antidepressivum Citalopram aus der Gruppe der selektiven Serotonin Wiederaufnahme Hemmer (SSRI) als möglichen psychopharmakologischen „moral enhancer“ an, da sich in einer Studie gezeigt habe, dass Probanden unter dem Einfluss von Citalopram weniger wahrscheinlich anderen Menschen aggressiv gegenübertreten würden [9]. In einer direkten Antwort auf diesen Artikel äußert sich die Autorin der Studie Crockett und fordert einen deutlich differenzierteren Umgang mit ihrer Arbeit: So sei das Studiendesign hochartifiziell und spezifisch gewesen, mit der Aufgabe, in hypothetischen Szenarien moralische Dilemmata zu bewerten, die Ergebnisse beträfen alleine erkrankte Patientinnen und Patienten mit bestimmten Persönlichkeitsstörungen, Daten zu gesunden Probandinnen und Probanden seien nicht verfügbar, potentielle Nebenwirkungen einer dauerhaften SSRI-Einnahme würden nicht bedacht [10] – die Idee, SSRIs seien eine vielversprechende Option für allgemeines moralisches Enhancement, erscheint offensichtlich übereilt und reduktionistisch.

Diese Arbeit versucht, dieser konzeptionellen Komplexität in jeder Veröffentlichung mit jeweils einer knappen, jedoch fundierten und auf die jeweiligen Fragestellungen zugeschnittenen Einführung in die biomedizinischen und technologischen Grundlagen zu

begegnen. Dass diese Bemühungen keinesfalls allumfassend und in Details vereinfachend ausfallen müssen, lässt sich kaum verhindern.

4.3) Konzeptionen von Autonomie / Methodik

Veröffentlichung I ist zunächst eine ausführliche Antwort auf einen Debatten-Beitrag von Schaefer et al. [11]. Die Autoren versuchen das Argument auszuhebeln, zukünftig vielleicht mögliches genetisches Enhancement durch Manipulationen am Genom von Embryonen verletze die Autonomie dieser zukünftigen Menschen, da sie sich zum Zeitpunkt des Eingriffes nicht für oder gegen diesen Eingriff entscheiden könnten. Ihre Überlegung geht folgendermaßen: Wenn eine Manipulation am Genom eines Embryos dazu dient, den später entstehenden Menschen autonomer werden zu lassen, und wenn man davon ausgeht, dass Autonomie intrinsischen Wert besitzt und ein autonomeres Leben wünschenswert ist, kann eine Verletzung von Autonomie zum Zeitpunkt des Eingriffes nicht mehr als Argument gegen einen solchen Eingriff dienen. Neben verschiedener normativer Einwände, die dieses Argument hervorrufen kann, stellt sich im Anschluss die Frage, wie „Autonomie“ in der Praxis über biomedizinische Eingriffe gesteigert werden könnte. Schaefer et al. schlagen, zur Operationalisierung von Autonomie, die Methode eines „overlapping consensus“ vor. Eine Analyse verschiedener gängiger Konzepte von Autonomie sowie einiger paradigmatischer Fälle von Verletzungen von Autonomie zeigt ihrer Ansicht nach, dass „reasoning abilities“ oder kognitive Fähigkeiten als kleinster gemeinsamer Nenner all dieser Ansätze gelten können. Folglich würden mehr kognitive Fähigkeiten mehr Autonomie bedeuten, genetisches kognitives Enhancement wäre also gleichzusetzen mit genetischem Autonomie Enhancement. Veröffentlichung I soll eine tiefergehende, detailliertere Analyse dieser Konzeption von Autonomie anbieten, und zeigen, dass die Methode eines kleinsten gemeinsamen Nenners im Kern einem „philosophischen Isolationismus“ folgt, der Teilkomponenten von Autonomie ihres jeweiligen Kontextes beraubt und damit keine ausreichend genaue Argumentation erlaubt. Kognitive Fähigkeiten als Voraussetzung für Autonomie müssen im Rahmen der jeweiligen Konzeption inklusive ihrer normativen Forderungen betrachtet und gewichtet werden. Dass in einer solchen Argumentation darüber hinaus Fallstricke wie in der oben geschilderten Diskussion bei deGrazia und Crockett lauern, soll ebenfalls deutlich werden.

Veröffentlichung II erarbeitet eine Heuristik zur Analyse möglicher Auswirkungen von Brain-Computer Interfaces auf die Verwirklichung von Autonomie. Es werden drei Fähigkeiten

dargestellt, die für ein autonomes Leben als wichtig gelten können: die Fähigkeit, Informationen und Wissen zu nutzen, um Gründe für das eigene Handeln zu entwickeln; die Fähigkeit, Kontrolle über die Durchführung intendierter Handlungen auszuüben; die Fähigkeit oder Möglichkeit, diese auch in konkreten Beziehungen und sozialen Kontexten verwirklichen zu können. Diese Heuristik wird auf die Nutzung von Brain-Computer Interfaces angewendet. Die Arbeit soll eine detaillierte Analyse bieten, ohne den Anspruch, pauschale, allgemeingültige Aussagen zur Auswirkung einer Technologie auf autonomes Leben zu machen. Es wird kein neues, allgemeingültiges Konzept von Autonomie vorgestellt. Vielmehr sollen einige mögliche Teilkomponenten von Autonomie zu ersten analytischen (und aufgrund der sich im Entstehen befindenden Technologie teils spekulativen) Überlegungen herangezogen werden.

4.4) Ergebnisse der Veröffentlichungen

Die erste Veröffentlichung zeigt, wieso eine Verbesserung von kognitiven Fähigkeiten oder „reasoning abilities“ nicht automatisch einen höheren Grad an Autonomie bedeutet. Es werden verschiedene Vorstellungen von Autonomie und verschiedene Szenarien einer potentiellen Anwendung neuer „Neuroenhancement“-Verfahren diskutiert, gleichzeitig soll eine allgemeine Kritik an Schaefer et al.’s Herangehensweise erfolgen. Deren Vorschlag eines gemeinsamen Nenners verschiedener Autonomie-Konzeptionen birgt die Gefahr eines philosophischen Isolationismus – andere jeweils relevante Aspekte einer Theorie, konzeptionelle Rahmenbedingungen und mögliche Auswirkungen in der Realität werden dadurch in den Hintergrund gedrängt oder bleiben gänzlich unbeachtet. Besonders Forderungen nach dem Einsatz sozialer und finanzieller Ressourcen zur Förderung und Popularisierung etwa von Neuroenhancement erscheinen auf dieser Basis fragwürdig.

In der zweiten Veröffentlichung wird eine Konzeption von Autonomie entwickelt, die mehrere Teilkomponenten eines autonomen Lebens und autonomer Entscheidungen beleuchtet. Schritt für Schritt erfolgt eine Analyse verschiedener BCI-Technologien mit ihren potentiellen Auswirkungen auf die jeweiligen Teilkomponenten. Es wird ein möglichst vielschichtiges und differenziertes Bild dieser Technologien und ihrer möglichen Konsequenzen gezeichnet. Ein abschließendes Urteil, ob BCI-Technologie in Bezug auf eine Förderung von Autonomie wünschenswert ist oder nicht, kann damit nicht gefällt werden. Es soll deutlich werden, dass der Zusammenhang zwischen spezifischem theoretischem Hintergrund, spezifischer

Technologie und möglicher Kontextfaktoren entscheidend ist für die philosophisch-moralische Einschätzung und Bewertung. Es wird gezeigt, dass auch Technologien wie BCIs, die zunächst den Anschein erwecken können, Autonomie durch eine Wiederherstellung oder Verbesserung von Fähigkeiten zu fördern, in bestimmten Kontexten problematische Auswirkungen haben können. Diese Ergebnisse bedeuten nicht, dass BCI-Technologie an sich problematisch sein muss. Kontextspezifische Analysen sollten jedoch ein Problembewusstsein schaffen, um auch mögliche negative Auswirkungen dieser Technologien rechtzeitig zu erkennen, Nutzerinnen und Nutzer zu informieren und bei Bedarf auch während Entwicklung oder Zulassung günstige Rahmenbedingungen zu erarbeiten.

4.5) Limitationen der vorliegenden Arbeit / Ausblick

Eine grundlegende Limitation der hier vorgelegten Arbeiten wird bereits durch die bisherigen Ausführungen deutlich: Die besprochenen Technologien befinden sich in der Entwicklung, hochspezialisierte Forschungsgruppen arbeiten an verschiedensten Details und Facetten, die in ihren Einzelheiten nur von Expertinnen und Experten auf dem jeweiligen Feld vollends verstanden werden können. Viele potentielle technische Erneuerungen, wie die Entwicklung sogenannter „affektiver" BCIs, die Gefühlszustände ihrer Nutzerinnen und Nutzer auslesen und bewerten sollen, bieten vielfältigste Ansätze zur Diskussion moralischer und ethischer Aspekte, scheinen jedoch von einer zuverlässigen technischen Umsetzbarkeit im Alltag noch weit entfernt [12]. Andere intensiv diskutierte Bereiche, wie etwa die Genomik, scheinen momentan hinter den Erwartungen vieler Forscherinnen und Forscher zurückzubleiben. Die Zusammenhänge zwischen Genotyp und Phänotyp stellen sich auch in Zeiten von Genomweiten Assoziationsstudien deutlich komplexer und schwerer nachvollziehbar dar als von vielen einst gedacht [13]. So spielt sich ein Teil der Debatten im Bereich des Spekulativen ab. Dieser Tatsache soll Rechnung getragen werden, indem einerseits betont wird, dass es sich vielfach um theoretisch-philosophische Diskussionen möglicher Zukunftsszenarien handelt, und andererseits Forderungen nach lebensweltlichen Konsequenzen wie etwa der Förderung potentieller Enhancement-Verfahren kritisch begegnet wird. Warnungen vor möglichen negativen Auswirkungen bestimmter Technologien sollen spezifisch für das jeweilige Szenario erfolgen.

Mindestens genauso komplex gestaltet sich die Diskussion des Begriffes „Autonomie". In beiden Veröffentlichungen werden wichtige philosophische Autonomie-Konzepte vorgestellt

und zur Analyse herangezogen. Dennoch kann das breite Feld möglicher theoretischer Annäherungen bei weitem nicht ausschöpfend behandelt werden[1]. Diese Einschränkung beinhaltet auch die Möglichkeit, dass im Rahmen anderer Autonomiekonzeptionen andere Ergebnisse hinsichtlich normativer Bewertungen neuer biomedizinischer Technologien entstehen. Auch aus dieser Tatsache ergibt sich der Bedarf an weiterer Forschung zu Autonomie und neuen biomedizinischen Technologien. Je analytisch klarer und differenzierter diese Beiträge gestaltet werden, desto facettenreicher können die moralischen Implikationen neuer Technologien dargestellt werden, und desto hilfreicher werden Diskussionsbeiträge aus dem Bereich der biomedizinischen Ethik für die Gestaltung einer Zukunft sein, in der die Grenzen zwischen Mensch und Technik mehr und mehr verschwimmen werden, sein.

4.6) Eigenanteil an den jeweiligen Veröffentlichungen

Veröffentlichung I entstand in einer Zusammenarbeit zu jeweils gleichen Anteilen mit Dr. med. Dr. phil. Orsolya Friedrich. Sowohl konzeptionelle Vorüberlegungen, als auch Recherche, argumentativer Aufbau, Schlussfolgerungen und Revisionen wurden gemeinsam erarbeitet. Es besteht eine geteilte Erstautorenschaft.

Veröffentlichung II entstand in Zusammenarbeit mit Dr. med. Dr. phil. Orsolya Friedrich, Eric Racine, Ph.D., Steffen Steinert, und Prof. Dr. med. Dr. phil. Ralf J. Jox. Als Erstautorin war Orsolya Friedrich federführend in Bezug auf die grundlegende Fragestellung und methodische Ausrichtung der Arbeit. Weiter wurden Recherche, argumentativer Aufbau, Schlussfolgerungen und Revision gemeinsam innerhalb der Autorengruppe erarbeitet. Für den Autor dieser Promotionsarbeit besteht für Veröffentlichung II der Status einer Ko-Autorenschaft.

[1] Für einen ersten Überblick zu Theorien der Autonomie s. [14, 15].

5.) Zusammenfassung

Überlegungen zum Begriff der Autonomie stellen einen Grundpfeiler moderner (Medizin-) Ethik dar. Während traditionell Diskussionen um die Handlungsautonomie von Patientinnen und Patienten in konkreten Situationen, beispielsweise bei Entscheidungen für oder gegen eine Behandlung, im Zentrum der Debatten standen, drängen sich in den letzten Jahren neue Fragestellungen auf. Mit teils rasant fortschreitenden Entwicklungen in der Forschung und teils auch medial hitzig geführten Debatten um neue biomedizinische Technologien wird immer häufiger gefragt, welche Auswirkungen solche Technologien auf die Autonomie von Patientinnen und Patienten oder auch „gesunden" Nutzerinnen und Nutzern haben könnten. Diese Arbeit versammelt Analysen zum Begriff der Autonomie in zwei dieser aktuellen Debatten: der Debatte um sogenanntes pharmakologisches und genetisches „Neuroenhancement" und derjenigen um „Brain-Computer Interfaces", Schnittstellen zwischen Gehirn und Computern. Veröffentlichung I ist zunächst eine detaillierte Analyse eines Argumentationsstranges, der Verletzungen von Autonomie eines noch nicht geborenen Menschen durch genetische Manipulationen am Embryo als Einwand gegen solche Manipulationen auszuhebeln versucht. Veröffentlichung II bietet eine eigenständige Konzeption von Autonomie als Heuristik zur Analyse von Brain-Computer Interfaces und deren mögliche Auswirkungen auf autonomes Leben. Dabei zeigen sich auch diskussionsübergreifende Resultate: so ist das Offenlegen von und der Umgang mit Limitationen der eigenen Arbeit, wie ein notgedrungen unvollkommenes Verständnis für Details der jeweiligen technologischen Fortentwicklungen oder das Argumentieren mit Konzeptionen von Autonomie, die nicht alle relevanten Überlegungen miteinschließen können, für eine Einordnung von Diskussionsbeiträgen entscheidend. Zudem stößt man bei der Vermengung ethisch-philosophischer Diskurse und Konstrukte und empirischer Forschung immer wieder auf Hürden, die deutlich problematisiert werden sollten, wenn man reduktionistische Fehlschlüsse vermeiden möchte. Je analytisch klarer und differenzierter Beiträge zum Begriff der Autonomie und neuer biomedizinscher Technologien gestaltet werden, desto facettenreicher können moralische Implikationen dargestellt werden, und desto hilfreicher werden Diskussionsbeiträge aus dem Bereich der biomedizinischen Ethik für die Gestaltung einer Zukunft, in der die Grenzen zwischen Mensch und Technik mehr und mehr verschwimmen, sein.

6.) Abstract

[Autonomy and new biomedical technologies]

To engage with notions of autonomy constitutes a cornerstone of modern (medical) ethics. Traditionally, discussions about the autonomy of patients in certain situations, i.e. when deciding for or against a treatment, have been at the centre of debates, but in the last years new questions have arisen. What kinds of consequences such new technologies could have on the autonomy of patients or "healthy" users has moved more and more centre stage in these discussion, partly due to rapid developments in scientific research and partly due to heated discussions and media-coverage on new biomedical technologies. The following work presents an analysis of the concept of autonomy in two of these debates: the debate on so-called pharmacological and genetic "neuroenhancement" and the debate on brain-computer interfaces. Publication I entails a detailed analysis of an argument that has been put forward to counter the notion that genetic manipulations on embryos should be considered a violation of the autonomy of these unborn humans. Publication II offers a heuristic conception of autonomy to analyse brain-computer interfaces and their possible consequences for autonomous living. This yields results which seem to be relevant across all debates on autonomy and new biomedical technologies: In order to put the contributions to relevant debates into perspective, it is necessary to disclose and to discuss certain limitations – such as an incomplete comprehension of every technical detail or such as building arguments around concepts of autonomy that cannot possibly include all relevant aspects. Furthermore, mixing ethical and philosophical discourses and concepts with empirical research produces obstacles that have to be addressed if reductionist fallacies are to be avoided. The more we can provide analytically clear and differentiated contributions to the debate on autonomy and new biomedical technologies, the more we can show moral implications in detail, and the more helpful contributions from biomedical ethics will be in a world where borders between humans and technology become increasingly blurred.

7.) Veröffentlichung I: Why Enhancing Autonomy Is Not a Question of Improving Single Aspects of Reasoning Abilities through Neuroenhancement

Pömsl, J., and O. Friedrich. 2017. Why Enhancing Autonomy Is Not a Question of Improving Single Aspects of Reasoning Abilities through Neuroenhancement. *Neuroethics* 10 (2): 243-254. doi:10.1007/s12152-016-9299-y.

Neuroethics (2017) 10:243–254
DOI 10.1007/s12152-016-9299-y

ORIGINAL PAPER

Why Enhancing Autonomy Is Not a Question of Improving Single Aspects of Reasoning Abilities through Neuroenhancement

Johannes Pömsl · Orsolya Friedrich

Received: 20 June 2016 / Accepted: 20 December 2016 / Published online: 3 January 2017

Abstract In a recent paper, Schaefer et al. proposed to enhance autonomy via improving reasoning abilities through (genetic) cognitive enhancement [1]. While initially their idea additionally seems to elegantly avoid objections against genetic enhancements based on the value of autonomy, we want to draw attention to several problems their approach poses. First, we will show that it is not at all clear that safe and meaningful methods to genetically or pharmaceutically enhance cognition will be feasible any time soon. Second, we want to provide a deeper discussion of the role of cognition and reasoning abilities in philosophical concepts of autonomy, as discussed in the mentioned paper. In doing so, we wish to demonstrate that using reasoning abilities as the common denominator in different accounts of autonomy in the context of enhancement does not do justice to the highly complex interrelations between cognition, reasoning abilities and autonomy. Neither should this way of arguing be accepted as a basis to call for practical outcomes, such as funding research into e. g. genetic cognitive enhancements, if the examined concepts of autonomy are taken seriously.

Keywords Autonomy · Genetic enhancement · Neuroenhancement · Reasoning abilities · Cognition

Johannes Pömsl and Orsolya Friedrich have contributed equally.

J. Pömsl · O. Friedrich (✉)
Institute of Ethics, History and Theory of Medicine, Ludwig-Maximilians-University Munich, Lessingstraße 2, 80336 Munich, Germany
e-mail: orsolya.friedrich@lrz.uni-muenchen.de

Introduction

In one of their latest contributions to the human enhancement debate [1], G. Owen Schaefer and his colleagues Guy Kahane and Julian Savulescu engage with one of the most prevalent objections to biomedical enhancement: possible violations of autonomy. This objection raises concerns about promoting cognitive enhancement (CE), particularly the genetic manipulation of embryos to increase a future child's capacities (e.g., [2–5]). Schaefer and colleagues not only aim to provide counterarguments to such autonomy-based objections to genetic manipulation, they also argue in favor of autonomy enhancement actively through (genetic) cognitive enhancement. They identify an overlapping consensus in different theories of autonomy as "cognition and reasoning capacity", which they argue is not equal with autonomy but significantly contributes to it, and could be enhanced genetically [1, p. 123].

Here, we try to show why *cognition and reasoning capacity* are indeed highly relevant aspects of autonomy, but that related justifications for genetic or pharmaceutical improvement of special aspects of these to enhance autonomy are not convincing. Before presenting our conceptual objections, we first draw attention to a number of concerns about the empirical basis and practical possibilities of biomedical CE that have recently been raised. We then propose that a form of isolationism is at risk of infiltrating philosophical discussions. We refer to this notion as *philosophical isolationism*, which is inherent in the concept of "overlapping consensus". Our primary aim is to identify the

limitations of philosophical isolationism when analyzing the enhancement of autonomy via CE.

The Reality of Cognitive Enhancement

There are many potential areas for human enhancement, including cognitive enhancement, mood enhancement, physical enhancement, lifespan extension, and moral enhancement [5]. Here, we focus on potential cognitive enhancements.

Schaefer et al. state that rapid scientific advances could enable important enhancements in various human capacities [1]. While this might be true for some capacities, such as doping in sports, it is unclear whether the same is true for cognitive capacities. Though the core argument of Schaefer and colleagues does not depend on the practical feasibility of such methods, we wish to draw attention to the importance of providing an overview of the scientific debate surrounding CE. In this article, we examine controversies regarding both pharmaceutical and genetic CE techniques to illustrate the extent to which they are already in use or thought to be potentially feasible.

Pharmaceutical Cognitive Enhancement (PCE)

Cognition encompasses various processes necessary to organize information, such as perception, attention, understanding and memory, and using information to guide behavior (reasoning or motor coordination) [6, 7]. Cognitive capacity could be enhanced in several ways, including enhancement of memory, wakefulness, concentration, and attention [8, 9]. Several substances have been investigated in this regard including stimulants, nutrients, hormones, amphetamines, peptides, cholinergic agonists, substances from the piracetam family, ampakines, and consolidation enhancers [6]. However, much of the ongoing discussion focuses on three pharmaceuticals: methylphenidate (i.e., Ritalin) and similar substances, mixed amphetamine salts (e.g., Adderall) and modafinil (i.e., Provigil) [10].

Nevertheless, some pharmacopsychologists doubt whether there is robust scientific evidence that such drugs actually improve specific cognitive domains like memory or executive function, claiming instead that they work as "vigilance or motivation enhancers" [11 p. 154]. Furthermore, research indicates inevitable tradeoffs when pharmaceutically manipulating cognition. Improvements in one cognitive area (e.g., working memory) relate to simultaneous decreases in another area (e.g., long-term memory functioning) [12]. Predicting the effects of stimulants in different individuals seems to be another problem. For example, the interplay between cognitive performance and non-cognitive aspects of motivation cannot be ignored [9]. Following a summary of systematic reviews and meta-analyses of potential cognitive enhancement drugs, Zohny recently concluded that "among healthy, non-sleep deprived individuals, whether it's Aricept, Ritalin, Adderall or Provigil, the evidence for enhancement [...] is not promising" [9 p. 260]. Other studies have been cited that also support this conclusion [11, 13–15]. Criticism has also been raised regarding misinterpretations and exaggerations of the interpretation of currently available data on the topic, including data on the prevalence of use of potential pharmaceutical enhancers [15–21]. Currently, discussions of effective cognitive enhancements seem to be in the realm of thought experiments.

Genetic Cognitive Enhancement (GCE)

By defending cognitive enhancement via embryonic genetic manipulation [1], Schaefer et al. are engaging with an even more controversial scientific debate, namely the debate about the genetics of intelligence. Genes as well as environmental factors play an important role in the development of intelligence [22]. Interrelations between genes and intelligence are characterized by several problems, from confirming an operational concept of general intelligence [23] to difficulties in genetic prediction [24, 25] and its many sociopolitical implications [26].

Since the 1990s, the means to analyze genes and later whole genomes has existed. Subsequently, there have been many attempts to determine which specific DNA sequences contribute significantly to our cognitive abilities. However, efforts to investigate "candidate genes", as well as genome-wide association studies, have not led to a breakthrough in understanding the genetics of intelligence [27]. Even polygenic scores and genome-wide complex trait analyses have explained only a few percentage points of the variation in intelligence [24], and effect sizes of so-called "Hits", i.e., single nucleotide polymorphisms, that correlate with certain behavioral characteristics are deemed "vanishingly small" [28 p. 27, 29]. A recent meta-analysis of genome-wide

association studies of genetic contributions to variation in general cognitive function showed very modest results [30].

Schaefer et al. cite several studies to support their assertion that the genetic conditions involved in developing cognitive abilities and therefore gaining the means to meaningfully influence these conditions are soon to be discovered [31–34]. However, in the larger context, these seem to show what might be a small influence rather than a real breakthrough. Indeed, in a recent article Philipp Rosoff calls genetic enhancement of complex behavioral traits, such as intelligence, a myth and a project doomed to fail [35]. He argues that the genetic enhancement project of complex psychological or behavioral traits such as intelligence does not make sense without an "underlying dependence on a strong form of genetic determinism" [35 p. 165]. According to Rosoff, if you propose to genetically enhance a complex trait like intelligence, you make two mistakes: first, you wrongly assume that behaviors that are given a name (such as intelligence) are things actually existing in the physical world. Secondly, you buy into a deterministic account of the relationship between genes and those traits: you will have to assume that altering a certain gene will have a foreseeable effect on the expression of a certain trait. This assumption seems to be incompatible with the "irreducible complexity of phenotypic expression" [35 p. 163], which means that it is highly doubtful you could ever make predictions (about phenotypic expression of complex traits) from a genome, because it is an extremely difficult "matter of figuring out how the individual elements combine with each other, with the environment, and with time to develop an organism" [24 p. S37]. All in all, Schaefer et al. use careful formulations surrounding the relationship between genes and intelligence, seemingly avoiding the pitfalls of over simplistic deterministic thinking. On the other hand they appear to be quite optimistic that GCE might be feasible sometime soon [1] – which, following Rosoff, would mean that they are necessarily deterministic.

Any predictions about the future of GCE are of course highly speculative. In responding to the euphoria that surrounds these prospects, we agree with the perspective recently expressed by Turkheimer when writing about the explanatory power of genetic correlations with human behavior like intelligence: "[A]ll I ask is that inevitable findings of weak genetic influence not be accepted as strong genetic explanations of complex human behavior while we wait for the progress of science to take its inevitable course." [28, p. 28].

In this section we outlined the doubtful scientific basis for PCE and GCE. Furthermore, we drew attention to several recent thinkers arguing that even though advocates of GCE like Schaefer et al. are careful to avoid explicitly oversimplifying the underlying causes of complex traits like intelligence, in promoting GCE they are engaging in a project that inherently relies on deterministic thinking. We do, however, note that Schaefer et al.'s core argument does not necessarily depend on this stance. We discuss their main arguments in the next sections.

The Method of Overlapping Consensus

Schaefer et al. claim that if a rejection to human enhancement is based on a certain value X, then upholding this rejection will become more difficult if a method is suggested that is likely to enhance this specific value X. If X is a significant factor in the enhancement of a particular human capacity, enhancing this specific factor should not be problematic. Indeed, one would be inclined to promote this specific enhancement. Considering the diversity of the various concepts of autonomy in the literature, Schaefer et al. implement a method they call *overlapping consensus* to identify ways of enhancing autonomy that appear to be broadly acceptable. They suggest that *overlapping consensus* is a common method in applied ethics [1, p. 125].[1] Initially, this method appears to be an elegant way of bringing ambiguous and controversial philosophical concepts under the scope of enhancement. One can consider a complicated philosophical concept to enhance (like autonomy or morality), scan through some of its most common accounts in the literature, identify a feature they all share and then find the means to biomedically enhance (or conversely reduce) this feature in (future) human beings.

In Schaefer's pivotal paper discussed here, they cite highly differing concepts of autonomy, including those

[1] Savulescu and Persson used this technique to identify altruism as a core feature of morality: aiming to later enhance morality by enhancing altruism [8]. To support their analysis, they refer to Christianity, Confucianism, Buddhism and the moral philosophies of Arthur Schopenhauer and David Hume, which all seem to rely on altruism as a core feature of morality. Sparrow [36] has highlighted the irony of this particular paper by Savulescu and Persson, which justifies the need for biomedical moral enhancement to address the potential dangers that may result from implementation of the cognitive enhancement Savulescu and colleagues usually demand.

of Kant and Habermas, hierarchical and life-plan conceptions from Frankfurt, Dworkin or Young, and action autonomy used by Beauchamp and Childress. Briefly examining these concepts of autonomy and some paradigmatic cases of violations of autonomy, Schaefer and colleagues identified cognition or reasoning ability as the overlapping consensus in discussions about the nature of autonomy. More specifically, they described three features of reasoning or cognitive abilities on which they focus: "deductive/logical competence, comprehension (including the avoidance of false beliefs), and critical analysis" [1, p. 126]. Hence, if you were able to genetically or pharmaceutically enhance these abilities, you could also enhance a future child's autonomy and chances of living a more autonomous (and therefore better) life.

We disagree with this stance, which we argue constitutes a particular form of *philosophical isolationism*. As noted above we propose that the method they deploy is problematic in a particular way. By identifying cognitive capacities and reasoning abilities from a cursory analysis of diverse conceptions of autonomy as an overlapping consensus, we argue that the authors strip each individual conception of the potential difficulties it might contain regarding three important points: 1) the way in which cognitive capacities and reasoning abilities are necessary for autonomy; 2) whether CE would actually confer an enhancement of autonomy (AE); and 3) whether CE would even be considered a permissible form of enhancement in each individual conception of autonomy. Based on our own analyses of the individual accounts, we argue that the overlapping consensus approach is inadequate when discussing enhancements of complex human abilities. Such capacities require complex philosophical conceptualization, especially if the results are used as justification for practical actions, such as promoting enhancement or financing research of enhancement techniques [1].

Competence Condition of Autonomy

In the following sections, we analyze the different accounts of autonomy discussed by Schaefer et al., and examine the way the authors construe the relation between reasoning abilities and autonomy. After giving a brief etymological definition of autonomy, we will discuss the ways that reasoning abilities are important for its various conceptions, starting with a competence condition account, then moving on to the hierarchical and unified-life plan theories of autonomy advocated by Kant and Habermas, and finally to externalist accounts. We argue that each of these accounts has different implications for the relationship between reasoning and autonomy, and, consequently, that the role of reasoning in creating autonomy cannot be separated in the way that an overlapping consensus approach suggests.

Autonomy has been a core feature of previous morality and normative debate, and therefore the issue of its enhancement through neuroscientific methods seems to pose ethical questions. "Autonomy" is derived from the Greek stems "self" and "law"; therefore "self-rule", "self-government" or "self-determination" are generally used synonyms (even if not correct in every case) [37]. The etymology of autonomy seems to point towards autonomy as an ability to make decisions, perform acts and live a life that refers to and is grounded in yourself. It remains questionable though what this self should be and if cognition or reasoning abilities are essential to govern or to determine oneself.

Schaefer et al. suggest that reasoning and deliberative capacity are not sufficient for, or identical to, autonomy, but that they contribute to autonomy in a relevant way [1]. Therefore, it follows that inhibiting someone's ability to reason and deliberate would inhibit autonomy, and conversely reasoning abilities should be enhanced to enhance autonomy. We agree that reasoning abilities are an important or even basic capacity for self-determination.

We consider reasoning ability in the first instance as a *competence condition for autonomy.* Competence conditions for autonomy usually refer to rationality as being capable of performing an informed means-ends analysis and transforming the gained reasons into actions [38]. Schaefer et al. speak of a similar definition: "The idea of competence, though vaguely defined, is clearly linked to people's ability to reason – their ability to properly comprehend the options ahead of them, evaluate different options, deduce appropriate courses of action, weigh consequences, etc." [1, p. 126].

The competence condition for autonomy typically refers to a special or singular decision or action, but its relevant capacities are also a feature of a person's overall autonomy [39]. Competence in bioethics is used as a threshold value concept, as well as a gradual condition. Schaefer et al. try to argue for autonomy enhancement using both conceptions [1]. First, we focus on the

possible enhancement of reasoning abilities to increase human capacities and competence above a certain threshold.

Competence Threshold

Being legally competent means to possess special legal powers as an all-or-nothing concept [37]. For example, in medicine the necessary capacities to count as legally competent to make treatment decisions can be tested by investigating deficits of mental capacity, such as being able to comprehend and retain treatment information, weigh the information and reach a decision, and communicate the decision [40, 41]. Even if mental capacity is decision-specific and there are grades of each capacity, a certain threshold exists, below which people are, in legal terms, equally incompetent to make a special treatment decision. Those below the competence threshold (at least for a broad range of decisions) are usually infants, those with severe learning difficulties or dementia, comatose patients, and those with certain severe mental illnesses. However, this does not mean that those above a competence threshold necessarily make the right decisions for themselves [37].

Schaefer et al. attempt to deal with a more general, conceptual threshold of autonomy. However, they do not provide a precise definition of the threshold they use. Though there can be differences in various conceptions of such a threshold, we assume that the necessary conditions of meeting a threshold as discussed above would be similar to those Schaefer et al. describe. The authors subsequently suggest that being "brought above the threshold for autonomy by improving one's reasoning abilities would count as an enhancement of autonomy, while being brought below the threshold by inhibiting someone's reasoning ability would be an inhibition of autonomy" [1, p. 127]. They later suggest genetic manipulation could improve intelligence. Schaefer et al. do not claim that people who are below a threshold necessarily have genetic defects, but they discuss the possibility that certain genetic alteration of cognitive abilities could bring them above the threshold. If we accept that reasoning ability is an important aspect of, or equates to, having competence for autonomy above a threshold that must be met in order to act autonomously, their conclusion seems to be uncontroversial. Thus, if it were technically possible to genetically or pharmaceutically increase e. g. intelligence, and thereby bring a person above the competence threshold, this would lead to a simultaneous enhancement of autonomy.

Does Anybody Need a Lift?

We argue however that most cases in which reasoning ability is below threshold cannot be repaired through GCE and PCE. The examples following below should illustrate that cognitive capacities like intelligence have little to no impact on being above or below a competence threshold in cases where a lack of competence is relatively obvious. Thus, autonomy enhancement through PCE or GCE would not be relevant, at least for some of the most clear-cut examples of people being below a competence threshold.[2]

Illustrating this point, *dementia* is not necessarily related to the level of intelligence prior to onset. Indeed, early-onset dementia is not thought to be connected to childhood cognitive ability, although late-onset dementia is correlated with lower premorbid intelligence [42, 43]. There are suggestions that the increased risk associated with lower premorbid intelligence may be in the development of vascular dementia and not Alzheimer's [44]. Similarly, the risk of falling into a *coma* most probably is not at all related to intelligence.

Intellectual disability (ID) is thought to be present in about 1% of the general population. It is estimated that about one third of ID cases are caused by environmental exposure to teratogens, viruses, radiation etc. Another 10% are due to malformation, and about 35% have genetic causes [45]. Notably, genetic causes are found most commonly in cases of severe ID (IQ <50), whereas cases of mild ID (making up about 85% of all cases)

[2] As mentioned above, mental capacity or competence can be seen as decision-specific. With this in mind, one could argue for a context-sensitive threshold account where a person who doesn't belong to any of the exemplary groups presented above still becomes momentarily incompetent, for example through distraction. Since many people get distracted sometimes, our rarity objection wouldn't hold up in an account like that. Nevertheless, the relation between, say, momentary forgetfulness or distraction in a special context and cognitive enhancement is hard to measure – a person could be competent at a given moment due to cognitive enhancement, but also potentially due to the fact that she had a good night's rest, she got helpful advice, or any other kind of circumstantial influence. We therefore want to present examples where links between general cognitive capacities (and their potential enhancements) and being above/below a competence threshold could be evaluated more clearly. We thank an anonymous reviewer for pressing us on this point.

seem to be heavily dependent on environmental factors, such as maternal education level and access to education, opportunity and healthcare [45, 46]. Potential enhancement through genetic manipulation would therefore apply in very few cases, in which the identified cause of ID is genetic, most likely in very severe cases of ID. Pharmaceutical manipulation can be used in general in plenty of conditions with environmental causes, but we think that PCE with substances like methylphenidat that are typically referred to as pharmaceutical enhancers can't help to change the consequences of reduced intellectual abilities due to exposure to teratogens, viruses, radiation or other such factors (actual medical treatment for certain conditions caused by environmental factors of course does exist).

In the development of *mental disorders*, the role of premorbid intelligence and therefore cognitive abilities is thought to differ significantly in different conditions. There is evidence that lower premorbid intelligence is a risk factor for some mental disorders (most convincingly schizophrenia, e.g., [47]). Yet, conversely, there are also hints that higher childhood IQ may be a risk factor for developing mania as an adult [48]. Mental disorders and their progress are highly complex genetic and social processes. Therefore, PCE and GCE to increase a person's competence threshold would, in our view, be of little use.

Thus, the low number of people likely to be below the competence threshold due to cognitive defects that could be enhanced through GCE or PCE suggests that there is little need for the kinds of autonomy enhancement discussed by Schaefer et al. to achieve competence above a certain threshold.[3] However, other therapeutic approaches to enabling people with mental disorders to gain competence above the threshold can be understood as meaningful and practical methods for autonomy enhancement. Even if it is not possible to achieve competence levels above a certain threshold in cases of severe ID, practical help to understand given information more easily in everyday situations could increase autonomy.

Even though the numbers of relevant cases for autonomy enhancement through PCE or GCE are low, increasing reasoning abilities above a necessary competence threshold could theoretically suggest an enhancement of autonomy. However, some conceptions of autonomy, such as the approaches of Habermas and externalist theorists argue that this conclusion is premature, because of its primary focus on reasoning abilities, which we discuss in more detail later in the article. In addition, many other normative arguments have previously been raised in bioethical discussions of cognitive manipulations that consider factors other than autonomy, such as the protection of authenticity [50–52], that might still prohibit such manipulations.

The More the Better? On the Relation between Reasoning Ability and Personal Autonomy

Here, we analyze whether a gradual improvement in reasoning abilities above a certain competence threshold through CE could be meaningfully interpreted as autonomy enhancement. If we use a threshold conception of competence, it could be argued that, above such a threshold, more intelligence is "simply an unused surplus" [37, p. 29 f., 53].[4] Intelligence can range from very low through average to high. Furthermore, judgment and other relevant capacities can differ strongly between individuals. But above an appropriate threshold, people with greater capability are not more qualified "… at the 'task' of living their own lives according to their own values as they choose" [37, p. 30].

Schaefer et al. seem to argue for the opposite conclusion when they ascertain that "[t]his relation between autonomy and cognition is generally positive – greater reasoning, deliberation and evaluation typically leads to greater autonomy" [1, p. 126]. This would mean – going back to our earlier etymological remark – that greater

[3] Schaefer et al. discuss this potential objection in their footnote 7: "One might reply that, as a matter of fact, relatively few people are below the relevant threshold, so autonomy enhancement is not relevant to the vast population. But such a view would implausibly imply that very few acts of deception, manipulation, restriction, and so on actually reduce autonomy." [1] We will assess deception and manipulation as relevant factors to inhibit autonomy in our last section. However, focusing on those as an objection to our point raised here disguises the problems contained in their approach related to the use of such widely accepted intuitive conceptions of autonomy. The statement that few people below a threshold might be helped through CE is not equal to the statement that few acts of deception reduce autonomy. Schaefer et al.'s argument depends on showing which forms of deception would be preventable through CE, and must refute the possibility that CE might pose an even greater risk of deception. Levin [49] has pointed out similar problems in Schaefer et al.'s reasoning.

[4] Feinberg refers here to different notions of competence mentioned by Daniel Wikler. The sense mentioned above refers more to natural abilities than to legal powers. A conception of natural competence ("minimal relevant capacity for a task") can also provide conditions to ascribe sovereign rights to individuals [37, p. 30]. Here we rather refer to a broader, or more natural understanding of a threshold conception of competence than to a legal one.

reasoning abilities help us to govern or to determine ourselves in a better or more effective way. Schaefer et al. support their argument using conceptions of autonomy (e.g., hierarchical theories, Kant's or Habermas' conceptions). We believe that the philosophical isolationism inherent in the method of overlapping consensus led them to draw the wrong conclusion about the relationship between greater reasoning abilities and personal autonomy. We elucidate this argument by reanalyzing the conceptions of autonomy Schaefer et al. used to reach their conclusions about overlapping consensus, and highlight the role of cognitive capacities and reasoning abilities in those conceptions.

Consequences from Hierarchical and Unified Life-Plan Theories

According to Schaefer et al., hierarchical theories of autonomy, such as Frankfurt's, are more or less concerned with establishing a coherence of higher-order and lower-order desires, which requires reflective and self-critical capacities [1, 54, 55]. Furthermore, Schaefer et al. suggest that unified life-plan theories, such as Young's or Richards's, require a coherent picture of one's life [1, 56, 57]. To gain such a unified life-plan, they again presuppose abilities such as "adequate internal reflection, accurate comprehension of the relevant facts of one's life, the ability to critically evaluate different plans and deduce a course of action from relevant premises" [1, p. 126], which they subsume within reasoning abilities.

We also agree that autonomy within hierarchical and unified life-plan theories is based on a certain level of competence (not only in legal terms), which also refers to reasoning abilities. The coherence between reasons for actions and long-term values, as well as coherence between different values and actions, are based on explicit reflection by an individual on their long-term values and life plans [38]. To gain such coherence requires reasoning abilities. However, above a necessary threshold for that competence, better or worse reasoning abilities seem not necessarily directly relevant for achieving stronger coherence or better determining oneself. Many intelligent people, who are able to deliberate perfectly, to explain their reasons to act, and for their difficulties in life, are still not able to overcome their difficulties or to achieve more coherence as mentioned above [58]. Therefore, a person would not necessarily be more autonomous in the sense of "[…] living their own lives according to their own values as they choose." [37 p. 30] From an empirical perspective, nothing certain can be said. Nonetheless, it seems that the critical and reflective processes of evaluation and endorsement required in hierarchical and unified life-plan conceptions of autonomy go beyond quasi-rational capacities and cognitive abilities, which are relevant for reasoning abilities. Critical self-reflection is a complex psychodynamic process, informed by cognitive processes and based on reasoning abilities. However, the improvement of cognitive abilities above a certain threshold does not make us more self-critical or resolute towards our evaluations. Furthermore to achieve greater coherence, we must also consider psychodynamics on an affective, (in the first step) unconscious level. Why we fail to achieve greater coherence despite profound deliberation has less to do with our cognitive abilities, but rather may be related to hidden, unconscious dynamics, such as resistance or defense mechanisms in our psyche, where not more cognitive or intellectual insight, but more emotional insight might be helpful, as provided for example in psychodynamic therapy [58].[5]

Besides insight into our own psychodynamic processes, self-control is also required for greater coherence. Some thinkers have proposed that self-control and self-regulation are important aspects of autonomy [37], influencing the ability of a person to regulate their emotions and passions. Thus, self-control can be considered an executive function that depends on reasoning abilities and therefore cognitive capacity.[6] However, there are many examples – where a person has reached a sufficient level of reasoning ability – demonstrating that intrapersonal dimensions other than reasoning abilities are involved in improving self-regulation. Failures of self-control are often described in the literature as cases of *akrasia.* Addiction provides an example of cases in which reasoning ability does not aid self-control, leading to situations in which an agent fails to do A, even if they decisively judge that it would be best to A [59, p. 103]. For example, consider the unwilling addict described by Frankfurt. She has the capacity for

[5] We do not want to claim that cognitive insight and therapies that provide such insight are not effective in helping patients but rather point to cases where the person has enough cognitive insight, yet does not reach the expected level of coherence.

[6] We thank an anonymous reviewer for asking us to explore self-control as an executive function and a cognitive capacity more closely. Furthermore, Levin [49] discusses whether it is convincing to presume that reason is the source of higher order desires if lower order desires are only non-cognitive, and examines the relationship between them.

autonomy and to be a person, because she forms a higher-order volition to not be an addict, but nevertheless she is not able to ensure that volition is followed through. To increase autonomy in this sense does not require a greater amount of reasoning ability above a necessary threshold, but the ability to regulate emotions or passions properly.

We do not mean to argue that reasoning ability has no effect on diachronic coherence above a certain threshold, but that it is not the most relevant factor when cognition is above such a threshold. Reasoning ability applies below and above the threshold, but if a person reaches a necessary threshold in terms of their ability to reason, other aspects affecting coherence become more important, and enhancing reasoning ability does not increase the likelihood of living a more autonomous life, and can even produce more difficulties, as we show below.

Reasoning Abilities in Habermas' and Kant's Conceptions of Autonomy

Schaefer et al. also oppose concerns about genetic manipulation raised by Habermas in *The Future of Human Nature* [2], which they call the "Communicative Objection". Habermas emphasizes that the process by which a person is formed should occur without instrumentalization of the person, and that communicative processes with critical engagement are central to this process, to avoid producing asymmetrical communicative situations [1, 2]. Schaefer et al. admit that according to Habermas genetic manipulation could violate autonomy because a future person would have no "communicative engagement" with the intentions of the parents [1]. However, they further suggest that Habermas should consider autonomy enhancements as useful since they improve the "capacity to personally engage with and provide input into the factors that shape one's life" [1 p. 133]. They propose that autonomy enhancement could help people to develop a non-distorted view and therefore to counter instrumentalization [1]. The authors attempt to support their argument by citing a concession made by Habermas for "genetic manipulation in order to prevent grave illnesses in future children" [1 p. 133], because a future child would presumably agree to such a manipulation. Schaefer et al. assume that consent could be presumed to also apply to genetic manipulation to enhance reasoning ability and therefore autonomy. "It seems to us similarly implausible to expect that a child would want a less-autonomous life." [1, p. 133] However, we propose that it would be a misinterpretation of Habermas' arguments to assume 1. that the asymmetry of the communicative situation produced by genetic manipulation would be eliminated through the concession Habermas made for severe illnesses; and 2. that Habermas' concession would also apply to enhancement of reasoning abilities via genetic manipulation.[7] The asymmetry already exists if the creation of a person results from the plans of others realized through genetic manipulation, whereas these others exist as ends in themselves [60]. Although Habermas makes a concession for severe illness without embryonic selection [2 p. 69 f.], it is not possible to resolve the communicative asymmetry involved, and does not imply that the concession can be extended to the potential for increased reasoning abilities to lead to enhanced autonomy in a genetically manipulated person [2 p. 62 f.]. Thus, being in a subordinate position via genesis is not compatible with enhanced autonomy in Habermas' conception, even if a future person might consent retrospectively. This misunderstanding may result from the underestimation of the relevance of Kantian autonomy in Habermas' writing [2, p. 53 ff.].

Schaefer et al. rightly admit that Habermas' argument follows a stronger Kantian understanding of autonomy than their own. Unfortunately, they seem to forget this fact when raising their conclusions outlined within this paper. They realize that the Kantian presuppositions in Habermas' argument suggest autonomy as an intrinsic rather than an instrumental value [1], but they do not consider the full normative consequences of a Kantian understanding of autonomy. As they correctly summarize at the beginning of their paper, Kantian autonomy is not normatively neutral. It demands moral reasoning according to principles that allow for universalizing the maxims of actions [1]. The condition of universality may be the strongest concept of autonomy in the philosophical literature, requiring an additional formal condition that usually includes reciprocity or an aptitude of one's maxims [38]. Kant's Categorical Imperative calls for identifying moral duties by generalizing the maxim (a subjective principle of willing) that underlies the action [61]. The basis of the ability to be a morally

[7] Pugh discusses some other misunderstandings of Habermas' arguments by supporters of enhancement with regard to "the arguments of negative freedom" and "the argument from natality" [37, p. 30].

autonomous agent cannot be produced by genetic manipulation and its effects on reasoning abilities. According to Kant, striving to perfect oneself could be seen as one of the imperfect duties that allow for different judgments about the morally right action to achieve a given aim (the test of universalization ends in a *contradiction in willing*) [61]. Improving reasoning abilities to enhance your autonomy could be interpreted as such an imperfect duty. We assume that this is the way Schaefer et al. understand their overlapping consensus regarding Kant. However, this understanding fails to conduct the universalization test thoroughly. Our interpretation of Kant suggests that genetic manipulation to improve reasoning abilities (as an important aspect of autonomy) would result in a *contradiction in thinking* (results in perfect duties: allow no exception). Therefore, the first step in the properly conducted universalization test could not be passed successfully [61]. Similarly, regarding what has been said in relation to Habermas' conception of autonomy, genetic manipulation would dispossess the person of the necessary grounds to develop moral autonomy. This would not be eliminated by greater reasoning abilities. Here, we are again confronted with the problem of philosophical isolationism. Similar objections can be raised when considering externalist perspectives of autonomy.

Objections from Externalism

So far we have assessed the problems that arise from an overlapping consensus approach to interrelations between cognition and autonomy, focusing mainly on internalist accounts of autonomy. In this final section, we turn to another problem that is posed by the argument of Schaefer et al., which surfaces when considering other usual requirements for autonomy.

Externalist or relational concepts of autonomy pay greater consideration to how personal rules, attitudes and wishes (and also self-reflection itself) are constructed through social interactions [62]. They are more concerned with external conditions, such as the lack of brainwashing or manipulation, which must be secured to establish autonomy for individuals [59]. For example, Beauchamp and Childress identified the necessary conditions for autonomous actions as intentionality, understanding, and lack of controlling influences from others [40]. The last condition focuses on limitation of autonomous action by social influences, which can be classified as coercion, persuasion, or manipulation in the form of gradual categories [40]. A similar condition for personal autonomy discussed by Christman focuses on the development of desires that a person would not have resisted if they had attended the process [63]. Equally, they should not be under the influence of factors that inhibit their self-reflection, while being without self-deception and being minimally rational [63].

Many other conditions could be mentioned resulting from externalist or relational conceptions of autonomy that support our conclusions [62]. Their common ground is to formulate the different (social or external) circumstances that enable autonomy. This leads to several important questions. Can neuroenhancement or genetic manipulation of capacities such as intelligence be claimed to consistently enhance autonomy, when we take such external or relational conditions into account? How can we claim to enhance autonomy if we simultaneously neglect the foundations of autonomy such as not having controlling influences from others? Can we genetically manipulate abilities that are said to be relevant for the enhancement of self-reflection and ensure that the resulting self-reflection is not self-deception? It is very difficult to answer these questions empirically or conceptually with certainty.

However, instead of engaging with these questions, Schaefer et al. deal with external circumstances that could be necessary for enabling autonomy by suggesting that CE could result in higher intelligence and therefore in "direct reduction of conformist impulses" [1, p. 128 f.]. The authors argue that most people would agree that being less conformist is related to having greater autonomy.[8] However, the empirical studies Schaefer et al. quote in support of their rationale are notably old, and more recent empirical data would be required to confirm their suggestion. Moreover, the authors do not deal with the abundance of real-world cases in which higher intelligence or reasoning ability has not helped to avoid conformity (e.g., intelligent individuals who nonetheless conformed to the Nazi regime in Germany in the 1940s). Schaefer et al. also imply, without further argumentation or evidence, that enhancing reasoning abilities via CE is not a threat to externalist accounts of autonomy, and that externalist claims do not threaten the overlapping consensus they describe. In the present article, we have

[8] It should be noted that Levin attacks this line of reasoning, arguing that Schaefer et al. "isolate particular threads, for instance 'deception and manipulation are bad,' on which wide-ranging concurrence exists," and thereby "they divert attention from highly controversial assumptions of their own." [49, p. 61].

raised a number of important questions about whether enhancing reasoning abilities through CE could be understood as a threat to externalist accounts of autonomy.

Conclusion

To counter concerns that biomedical enhancement of cognition may compromise autonomy, Schaefer et al. suggest promoting autonomy itself using biomedical means. To support their conclusions, they identify reasoning ability as an overlapping consensus of concepts of autonomy, claiming that it could be biomedically enhanced. Nonetheless, whether autonomy can be improved biomedically, and particularly through genetic manipulation, remains unclear. Based on our conceptual analyses, we conclude that reasoning abilities are indeed important for self-determination, as suggested in various internalist concepts of autonomy. However, when trying to identify the core commonalities among diverse conceptions of autonomy as suggested by Schaefer et al.'s overlapping consensus approach, we argue that one must be aware of the shortcomings of such *philosophical isolationism*. Excess reasoning ability above a competence threshold may not provide an increased ability to make more autonomous choices or to live a more autonomous life. The overlapping consensus approach used by Schaefer et al. is therefore inadequate to draw conclusions about enhancing complicated aspects of human beings or complex philosophical concepts. We accept that an exception may exist in situations where reasoning abilities are below a certain threshold. In such situations, increasing reasoning abilities above a necessary competence threshold may result in an enhancement of autonomy. However, analysis of paradigmatic situations in which people are below a cognitive threshold reveals that there are few situations in which the kind of autonomy enhancement through cognitive enhancement advocated by Schaefer et al. is required to achieve competence above a specified threshold. A number of other normative arguments have previously been raised in bioethical discussions of cognitive manipulation that consider factors other than autonomy in such situations, such as the protection of authenticity.

In the current paper, we did not discuss whether biomedical and genetic interventions are normatively justified in a more general sense. Retrospective consent to biomedical interventions (e.g. curing severe illnesses) might, for example, be justifiable. Greater reasoning abilities may also plausibly play a role in improved well-being and their improvement in whatever manner may therefore be justified, depending on the underlying moral theory. However, if we consider autonomy to have primary intrinsic value, close analysis of the theories discussed by Schaefer et al. suggests that improvement of a single aspect of autonomy, even if it is relevant, will not necessarily equal an improvement of autonomy that is ethically justified.

Acknowledgement We want to thank one of the anonymous reviewers who suggested to use the term "isolationism" instead of "reductionism", which can be confusing given its common philosophical use. Our intention is to describe a problem encountered by the overlapping consensus method used by Schaefer et al. By extracting reasoning abilities as a core feature of very different accounts of autonomy, other relevant properties, conceptual ramifications, and potential real life consequences of these accounts can easily be blurred and disregarded.

References

1. Schaefer, G.O., G. Kahane, and J. Savulescu. 2014. Autonomy and enhancement. *Neuroethics* 7(2): 123–136.
2. Habermas, J. 2003. *The future of human nature*. Cambridge: PolityPress.
3. Davis, D.S. 2009. The parental investment factor and the Child's right to an open future. *Hastings Center Report* 39(2): 24–27.
4. Davis, D.S. 1997. Genetic dilemmas and the Child's right to an open future. *Hastings Center Report* 27(2): 7–15.
5. Savulescu, J., R. ter Meulen, and G. Kahane. 2011. *Enhancing human capacities*. Chichester: Blackwell Publishing Ltd..
6. Bostrom, N., and A. Sandberg. 2009. Cognitive enhancement: methods, ethics, regulatory challenges. *Science & Engineering Ethics* 15(3): 311–341.
7. Housden, C.R., S. Morein-Zamir, and B.J. Sahakian. 2011. Cognitivie enhancing drugs: neuroscience and society. In *Enhancing human capacities*, ed. G. Kahane, J. Savulescu, and R. Ter Meulen, 113–126. Chichester: Wiley.
8. Persson, I., and J. Savulescu. 2008. The perils of cognitive enhancement and the urgent imperative to enhance the moral character of humanity. *Journal of Applied Philosophy* 25(3): 162–177.
9. Zohny, H. 2015. The myth of cognitive enhancement drugs. *Neuroethics* 8(3): 257–269.
10. Singh, I., I. Bard, and J. Jackson. 2014. Robust resilience and substantial interest: a survey of pharmacological cognitive enhancement among university students in the UK and Ireland. *PloS One* 9(10): e105969.
11. Quednow, B.B. 2010. Ethics of neuroenhancement. A phantom debate. *BioSocieties* 5(1): 153–156.
12. de Jongh, R., I. Bolt, M. Schermer, and B. Olivier. 2008. Botox for the brain: enhancement of cognition, mood and

pro-social behavior and blunting of unwanted memories. *Neuroscience & Biobehavioral Reviews* 32(4): 760–776.
13. Lucke, J.C., S. Bell, B. Partridge, and W.D. Hall. 2011. Deflating the neuroenhancement bubble. *American Journal of Bioethics Neuroscience* 2(4): 38–43.
14. Hall, W.D., and J.C. Lucke. 2010. The enhancement use of Neuropharmaceuticals: more Scepticism and caution needed. *Addiction* 105(12): 2041–2043.
15. Lucke, J.C., S. Bell, B. Partridge, and W.D. Hall. 2010. Weak evidence for large claims contribute to the phantom debate. *BioSocieties* 5(4): 482–483.
16. Nadler, R.C., and P.B. Reiner. 2010. A call for data to inform discussion on cognitive enhancement. *BioSocieties* 5(4): 481–482.
17. McCabe, S.E., C.J. Teter, C.J. Boyd, J.R. Knight, and H. Wechsler. 2005. Nonmedical use of prescription opioids among US College students: prevalence and correlates from a National Survey. *Addictive Behaviors* 30(4): 789–805.
18. Maher, B. 2008. Poll results: look Who's doping. *Nature* 452: 674–675.
19. Babcock, Q., and T. Byrne. 2000. Student perceptions of methylphenidate abuse at a public liberal arts college. *Journal of American College Health* 49(3): 143–145.
20. Franke, A.G., C. Bonertz, M. Christmann, M. Huss, A. Fellgiebel, E. Hildt, and K. Lieb. 2011. Non-medical use of prescription stimulants and illicit use of stimulants for cognitive enhancement in pupils and students in Germany. *Pharmacopsychiatry* 44(2): 60–66.
21. Wiegel, C., S. Sattler, A.S. Göritz, and M. Diewald. 2016. Work-related stress and cognitive enhancement among university teachers. *Anxiety, Stress, and Coping* 29(1): 100–117.
22. Bouchard, T.J., and M. McGue. 1981. Familial studies of intelligence: a review. *Science* 212(4498): 1055–1059.
23. Plucker, J.A., and A.L. Shelton. 2015. General intelligence (g): overview of a complex construct and its implications for genetics research. *Hastings Center Report* 45(S1): S21–S24.
24. Turkheimer, E. 2015. Genetic prediction. *Hastings Center Report* 45(S1): S32–S38.
25. Wray, N.R., J. Yang, B.J. Hayes, A.L. Price, M.E. Goddard, and P.M. Visscher. 2013. Pitfalls of predicting complex traits from SNPs. *Nature Reviews Genetics* 14(7): 507–515.
26. Tabery, J. 2015. Why is studying the genetics of intelligence so controversial? *Hastings Center Report* 45(S1): S9–S14.
27. McGue, M., and I.I. Gottesman. 2015. Classical and molecular genetic research on general cognitive ability. *Hastings Center Report* 45(S1): S25–S31.
28. Turkheimer, E. 2016. Weak genetic explanation 20 Years later: reply to Plomin et al. *Perspectives on Psychological Science* 11(1): 24–28.
29. Chabris, C.F., J.J. Lee, D. Cesarini, D.J. Benjamin, and D.I. Laibson. 2015. The fourth law of behavior genetics. *Current Directions in Psychological Science* 24(4): 304–312.
30. Davies, G., et al. 2015. Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N = 53949). *Molecular Psychiatry* 20(2): 183–192.
31. Burdick, K.E., T. Lencz, B. Funke, C.T. Finn, P.R. Szeszko, J.M. Kane, R. Kucherlapati, and A.K. Malhotra. 2006. Genetic variation in DTNBP1 influences general cognitive ability. *Human Molecular Genetics* 15(10): 1563–1568.
32. Deary, I.J., S.E. Harris, H.C. Fox, C. Hayward, A.F. Wright, J.M. Starr, and L.J. Whalley. 2005. KLOTHO genotype and cognitive ability in childhood and old age in the same individuals. *Neuroscience Letters* 378(1): 22–27.
33. Harris, S.E., H. Fox, A.F. Wright, C. Hayward, J.M. Starr, L.J. Whalley, and I.J. Deary. 2006. The brain-derived neurotrophic factor Val66Met polymorphism is associated with age-related change in reasoning skills. *Molecular Psychiatry* 11(5): 505–513.
34. Houlihan, L.M., S.E. Harris, M. Luciano, A.J. Gow, J.M. Starr, P.M. Visscher, and I.J. Deary. 2009. Replication study of candidate genes for cognitive abilities: the Lothian birth cohort 1936. *Genes, Brain and Behavior* 8(2): 238–247.
35. Rosoff, P.M. 2012. The myth of genetic enhancement. *Theoretical Medicine and Bioethics* 33(3): 163–178.
36. Sparrow, R. 2014. Better living through chemistry? A reply to Savulescu and Persson on 'moral enhancement'. *Journal of Applied Philosophy* 31(1): 23–32.
37. Feinberg, J. 1986. *Harm to self*. New York: Oxford University Press.
38. Friedrich, O., and J.H. Heinrichs. 2014. Autonomie als Rechtfertigungsgrund Psychiatrischer Therapien. *Ethik in der Medizin* 26(4): 317–330.
39. Quante, M. 2011. In Defence of personal autonomy. *Journal of Medical Ethics* 37(10): 597–600.
40. Beauchamp, T., and J. Childress. 2009. *Principles of biomedical ethics*. Oxford: Oxford University Press.
41. Stewart, C., C. Peisah, and B. Draper. 2011. A test for mental capacity to request assisted suicide. *Journal of Medical Ethics* 37(1): 34–39.
42. Whalley, L.J., J.M. Starr, R. Athawes, D. Hunter, A. Pattie, and I.J. Deary. 2000. Childhood mental ability and dementia. *Neurology* 55(10): 1455–1459.
43. Deary, I.J., M.C. Whiteman, J.M. Starr, L.J. Whalley, and H.C. Fox. 2004. The impact of childhood intelligence on later life: following up the Scottish mental surveys of 1932 and 1947. *Journal of Personality and Social Psychology* 86(1): 130–147.
44. McGurn, B., I.J. Deary, and J.M. Starr. 2008. Childhood cognitive ability and risk of late-onset Alzheimer and vascular dementia. *Neurology* 71(14): 1051–1056.
45. Szymanski, L., and B.H. King. 1999. Practice parameters for the assessment and treatment of children, adolescents, and adults with mental retardation and comorbid mental disorders. *Journal of the American Academy of Child & Adolescent Psychiatry* 38(12): 5S–31S.
46. Kaufman, L., M. Ayub, and J.B. Vincent. 2010. The genetic basis of non-syndromic intellectual disability: a review. *Journal of Neurodevelopmental Disorders* 2(4): 182–209.
47. Khandaker, G.M., J.H. Barnett, I.R. White, and P.B. Jones. 2011. A quantitative meta-analysis of population-based studies of premorbid intelligence and schizophrenia. *Schizophrenia Research* 132(2): 220–227.
48. Koenen, K.C., T.E. Moffitt, A.L. Roberts, L.T. Martin, L. Kubzansky, H.L. Harrington, R. Poulton, and A. Caspi. 2009. Childhood IQ and adult mental disorders: a test of the cognitive reserve hypothesis. *American Journal of Psychiatry* 166(1): 50–57.
49. Levin, S.B. 2016. Upgrading discussions of cognitive enhancement. *Neuroethics* 9(1): 53–67.

50. Elliott, C. 1998. The tyranny of happiness: ethics and cosmetic psychopharmacology. In *Enhancing human traits: ethical and social implications*, ed. E. Parens, 177–188. Washington D.C: Georgetown University Press.
51. Levy, N. 2011. Enhancing authenticity. *Journal of Applied Philosophy* 28(3): 308–318.
52. DeGrazia, D. 2000. Prozac, enhancement, and self-creation. *Hastings Center Report* 30(2): 34–40.
53. Wikler, D. 1979. Paternalism and the mildly retarded. *Philosophy and Public Affairs* 8(4): 377–392.
54. Frankfurt, H.G. 1971. Freedom of the will and the concept of a person. *Journal of Philosophy* 68(1): 5–20.
55. Dworkin, G. 1988. *The theory and practice of autonomy*. Cambridge: Cambridge University Press.
56. Young, R. 1989. Autonomy and the inner self. In *The inner citadel: essays on individual autonomy*, ed. J. Christman, 77–90. New York: Oxford University Press.
57. Richards, D.A.J. 1981. Rights and autonomy. *Ethics* 92(1): 3–20.
58. Shedler, J. 2010. The efficacy of psychodynamic psychotherapy. *The American Psychologist* 65(2): 98–109.
59. Mele, A. 2001. *Autonomous Agents*. New York: Oxford University Press.
60. Heilinger, J.C., and K. Crone. 2014. Human freedom and enhancement. *Medicine, Health Care and Philosophy* 17(1): 13–21.
61. Kant, I. 1997 [1785]. Grundlegung zur Metaphysik der Sitten [groundwork of the metaphysics of morals]. In *Ak IV*, ed. M. Gregor, 387–436. Cambridge: Cambridge University Press.
62. Oshana, M. 1998. Personal autonomy and society. *Journal of Social Philosophy* 29(1): 81–102.
63. Christman, J. 1991. Autonomy and personal history. *Canadian Journal of Philosophy* 21(1): 1–24.

8.) Veröffentlichung II: An Analysis of the Impact of Brain-Computer Interfaces on Autonomy

Friedrich, O., E. Racine, S. Steinert, J. Pömsl, and R. J. Jox. 2018. An Analysis of the Impact of Brain-Computer Interfaces on Autonomy. *Neuroethics*. doi:10.1007/s12152-018-9364-9.

Neuroethics
https://doi.org/10.1007/s12152-018-9364-9

ORIGINAL PAPER

An Analysis of the Impact of Brain-Computer Interfaces on Autonomy

Orsolya Friedrich · Eric Racine · Steffen Steinert · Johannes Pömsl · Ralf J. Jox

Received: 30 September 2017 / Accepted: 11 April 2018

Abstract Research conducted on Brain-Computer Interfaces (BCIs) has grown considerably during the last decades. With the help of BCIs, users can (re)gain a wide range of functions. Our aim in this paper is to analyze the impact of BCIs on autonomy. To this end, we introduce three abilities that most accounts of autonomy take to be essential: (1) the ability to use information and knowledge to produce reasons; (2) the ability to ensure that intended actions are effectively realized (control); and (3) the ability to enact intentions within concrete relationships and contexts. We then consider the impact of BCI technology on each of these abilities. Although on first glance, BCIs solely enhance self-determination because they restore or improve abilities, we will show that there are other positive, but also negative impacts on user autonomy, which require further philosophical and ethical discussions.

Keywords Brain-computer interfaces · Brain-machine interfaces · Autonomy · Neuroethics

O. Friedrich (✉) · S. Steinert · J. Pömsl · R. J. Jox
Institute for Ethics, History and Theory of Medicine at Ludwig Maximilian University (LMU), Lessingstr. 2, 80336 Munich, Germany
e-mail: orsolya.friedrich@lrz.uni-muenchen.de

E. Racine
Neuroethics Research Unit, Institut de recherches cliniques de Montréal, Montréal, QC, Canada

E. Racine
Department of Medicine and Department of Social and Preventative Medicine, Université de Montréal, Montréal, QC, Canada

E. Racine
Division of Experimental Medicine and Department of Neurology and Neurosurgery, McGill University, 110 avenue des Pins Ouest, Montréal, QC H2W lR7, Canada

Introduction

Research on Brain-Computer Interfaces (BCIs) has grown considerably in the last decades [1]. Among many recent publications in this area, there was a successful attempt to create a so-called "live cockroach cyborg" steered by a human brain [2]. Another study reported the case of a patient with spinal cord injury who regained control over isolated finger movements with the help of a BCI [3]. The recent news that Elon Musk, the Chief Executive Officer of Tesla and SpaceX, is investing in the BCI-development company Neuralink further illustrates the high expectations this field has now raised [4]. Nevertheless, there remain many issues limiting the widespread and reliable use of BCIs. The translation of the technology from the laboratory to everyday life is just beginning [1, 5–7]. However, BCIs have already stimulated many ethical debates [8] and further ethical questions may arise, notably whether such technologies represent a threat to human autonomy [9].

The answer to this questions is complex, especially because existing concepts of human

Published online: 18 April 2018

autonomy are diverse. It is generally agreed that voluntariness is a key component of autonomy [10]. However, the general and idealized picture of an autonomous and voluntary subject, with a coherent and conscious self, with access to truth and moral rightness, fully rational and in control of her actions has been questioned and corrected from various quarters. Nietzsche, Freud, and Marx all questioned this idea [11–13], along with contributions from feminist relational autonomy theorists [14], who have offered less idealistic and individualized conceptions of autonomy. Most recently, scientific findings on psychological phenomena, such as biases and non-conscious cognitive processing have also provided additional avenues of critique and food for thought [15]. Despite conceptual challenges, the general idea of autonomy is highly valued in human life and in ethics. Accordingly, understanding whether and how technological advancements like BCIs pose a challenge to human autonomy is of crucial relevance to ethics as a whole, and in particular to ethics of technology [16].

In this paper, we aim to analyze whether BCIs positively or negatively impact human autonomy, which we consider a multi-component ability. Three *component-abilities* are inherent to many accounts of autonomy and will be discussed in terms of how they are affected by BCIs impact: (1) the ability to use information and knowledge to produce reasons; (2) the ability to ensure that intended actions are realized effectively (i.e., control); and (3) the ability to enact intentions within concrete relationships and contexts.

These three abilities are common and relevant to several theories about autonomy though we do not tie our analysis to any single specific theory. Furthermore, the analysis presented does not exhaust the ways component-abilities can be relevant for autonomy or the normative implications of the described component-abilities.

Prior to discussing further the three component-abilities and their relation to autonomy, we will briefly introduce BCIs and point to their features of interest given our focus, particularly on the mental strategies they require their users to apply. We then describe the three component-abilities and critically review how BCIs might affect them in positive or negative ways.

Overview of Brain-Computer-Interfaces

BCIs are commonly viewed as a subtype of neuroprosthesis that detect and process brain activity in order to direct external electrical devices. By providing an artificial output channel for the user, the user's activity is enabled without the involvement of peripheral nerves and muscles [6, 17–19]. It is important here to note that BCIs cannot "read thoughts." Rather, they associate specific patterns of brain activity to a specific output of the device.

Brain activity detection in BCIs can be either invasive or non-invasive [20]. The most commonly used non-invasive recording method is electroencephalography (EEG) [21, 22]. Invasive methods either place electrodes on the surface of the brain [23] or within the brain tissue [5, 22], and require the user to assume the risks associated with surgery. The long-term functioning of such devices is also not guaranteed [21]. Generally, the signals obtained by invasive methods tend to more precisely locate electrical brain activity, and detect activity in distinct or small regions of the brain, compared to non-invasive methods [17, 20]. New approaches of recording brain activity continue to emerge however, for example, by implanting stent electrodes into cortical veins [24].

Typical clinical output devices for BCIs can be grouped depending on their purpose. Medical-assistive devices which increase communication and/or motor control and can include a spelling program or cursor navigation on a screen, driving a wheelchair, moving a robotic arm, or other types of neuroprostheses or assistive devices, but they may also activate the patient's own muscles [3, 22]. Other uses for BCIs in medical contexts include neurorehabilitation for example after strokes or spinal cord injuries [5, 25] or other neurological conditions such as epilepsy [26]. BCIs have also been tested for the treatment of psychiatric conditions, such as attention deficit hyperactivity disorder or mood disorders [27, 28]. In addition to applications in the medical field, BCIs may also serve non-medical purposes such as gaming [29]. Brain-to-brain interfaces, which seek to provide a means of directly communicating information from one brain to another, are among the latest, most impressive and morally controversial projects [2, 30, 31].

It is essential to keep in mind that almost all of these systems and proposed uses of BCIs rely on closed loop

technology, meaning that the BCI user is presented with real-time feedback on the output. The feedback can be visual, auditory, tactile, vestibular, or proprioceptive, though it is most commonly visual (e.g., the user sees the movements of a cursor on a screen) [32]. There are also approaches which utilize a more direct or more invasive way of closing the loop, for example by stimulating the somatosensory cortex with electrodes implanted in the brain [33].[1] The potential seen with these new approaches has prompted new areas of research, for example in the treatment of depression or obsessive-compulsive disorder (OCD), by recording and modulating patients' brain activity directly [34]. In this case theoretically, these systems would modulate pathological brain activity and thereby reduce symptoms, without the user being necessarily conscious of it.

In order to assess the impact of BCI systems on autonomy, it is helpful to distinguish between three categories of BCIs: passive, reactive, and active. *Passive* BCIs rely on brain activity that is not voluntarily modulated [35], like mental workload [36], drowsiness, or affective states [37] by capturing changes in brain activities as input. This means that the user does not have to apply any mental effort to intentionally cause a certain outcome for the device. Passive BCIs could be used, for example, to avoid dangerous situations in a working environment caused by deficits in attention [38] or to improve gaming-experiences [39].[2]

In contrast, *reactive* BCIs require some degree of user activity, and are based on changes in brain activity (e.g., specific P300 spikes) that occur when users focus their attention on an external stimulus among a group of stimuli provided by the BCI. The stimuli can be auditory, somatosensory or, most typically, visual (e.g., letters on a screen that flicker in different frequencies) [17]. One great advantage of reactive BCIs is that they only require a minimum amount of learning since users must only focus their attention selectively. On the other hand, some of these systems might still require some functioning of peripheral nerves and muscles, such as when the user's gaze has to be focused on a certain object. Thus, such systems might not be suitable for patients with conditions like complete locked-in-syndrome (cLiS).

Finally, in the third case of *active* BCIs, users are required to deploy specific mental strategies to create a certain brain activity that will be recorded, processed, and identified as commands by the BCI system. One of the most commonly used strategies is that of motor imagery [17]. To intentionally produce changes in so-called sensorimotor rhythms (SMR), i.e., oscillations in brain activity recorded from somatosensory and motor areas, it is sufficient for a user to imagine body-movements (e.g., moving a limb) without actually moving the body at all. Since imagining different types of movements produces different changes in brain activity, these shifts in the imagining of an activity can be used to distinguish different commands [17]. Active BCIs may be preferable to other systems since they do neither rely on external stimulation nor on gaze control by the user [17, 42].

Following this description of BCI technology, we devote the next section to the concept of autonomy, to demonstrate how it is a multi-component ability.

Autonomy as a Multi-Component Ability

"Autonomy" means etymologically "self-rule", "self-government" or "self-determination" [43].[3] Autonomy implies – with significant differences for the specification of each of its aspects – the ability for an agent to act or live her life according to her own wishes or reasons without interference of strong external influences (mild

[1] In the following we will refer to the more direct/invasive feedback methods when discussing closed loop BCIs.

[2] While it makes sense conceptually to distinguish between passive and (re-)active BCIs, this distinction might not be clear in real-world applications [35]. Implications of passive BCIs become especially troubling in settings where signals like a P300 might be used to extract information from a user that she or he does not actively want to share, see [40, 41].

[3] The term 'self-determination' indicates less strong conditions because a rule or law is absent, contrary to 'autonomy'. A strong and normatively not neutral conception of autonomy as moral self-rule can be traced back to Kant's account of autonomy, where moral reasoning according to principles that allow for universalizing the maxims of actions is required [44]. We think that the distinction between 'self-determination' and 'autonomy' is reasonable in many discussions of bioethics and should be made clear more often. Nevertheless, we will use both terms synonymously here, since the distinction is not relevant for the specific aim of this paper. Further, we do not discuss here, how Kant's understanding of autonomy and of human dignity that must be ascribed to all beings capable of reason, can be reconciled with considering sub-abilities of autonomy. We presuppose that they can build two compatible perspectives on human autonomy.

or moderate influences being usually tolerated).[4] Autonomous thought and decision versus autonomous action [50] are sometimes sharply distinguished; yet they may appear interwoven in certain contexts [51], such that autonomy may be considered as an ability [52, 53]. Here, we focus our analysis on component-abilities relevant for autonomy, which are compatible with most accounts of autonomy. These abilities are: (1) the ability to use information and knowledge to produce reasons; (2) the ability to ensure that intended actions are realized effectively (control); and (3) the ability to enact intentions within concrete relationships and contexts. For the purposes of this paper, these component-abilities are presented ("Ability to use information and knowledge to produce reasons", "Ability to ensure that intended actions are realized effectively (control)" and "Ability to enact intentions within concrete relationships and contexts" sections) and discussed ("Critical analysis of autonomy in light of BCI research" section) in light of recent developments in BCI research. The aim of the study undertaken here is not to develop an entire new model of autonomy, or to exhaustively explore the normative implication for each of the component-abilities of autonomy. Nonetheless the study is an introductory attempt, to offer initial insights and outlooks how BCIs could support or undermine certain component-abilities of autonomy.

Ability to Use Information and Knowledge to Produce Reasons

Based on the different accounts and definitions of autonomy available, it seems that in order to consciously and deliberately develop (or change) one's own reasons to act, a person uses *information* or *knowledge* about a situation at hand, about the options from which to choose, about the way how to achieve certain aims, about her beliefs, emotions and desires, about her evaluation of the situation, and so on. Understanding and evaluating this information and knowledge behind an act presupposes a degree of self-trust and self-confidence [54–56].

According to Beauchamp and Childress [57], an *autonomous* person has to understand the facts surrounding an action, such as the consequences that may arise from a medical procedure or its omission. This caveat assumes that she evaluates the information with regard to her own actual and expected long-term appreciation of the options at hand, her desires and preferences, and her value system, which again presupposes an understanding of these. Having information about personal values and preferences, and about different situational or contextual aspects, is also highly relevant for coherentist or hierarchical approaches to autonomy. Critical reflection, evaluation, and endorsement of her reasons is an important element, which again presupposes information about actual and long-term emotional states, desires, beliefs, and values, etc., as well as about the situation [50, 58, 59]. The normatively not neutral conception of autonomy as moral self-rule in the work of Kant also requires knowledge [44]. Knowledge seems also to be a key issue if we consult relational or externalist accounts of autonomy. A person needs to know if she is a victim of self-deception, or if other autonomy-enabling and autonomy-inhibiting external circumstances are present [60, 61].

Ability to Ensure that Intended Actions are Realized Effectively (Control)

Related to the relevance of knowledge for autonomy, is the ability to ensure that intended actions are realized effectively, which presupposes some kind of control. If autonomy represents an ability to guide one's behavior

[4] *Volitions* usually refer to some 'inner activity', with the power of producing changes of the world, usually including bodily movements [45]. Acting in accordance with one's *own reasons* is a common presumption in current philosophical discussions about autonomy [46]. In the following, we will mostly refer to the term *reasons*, but we would like to avoid misunderstandings, which might result from the well-known controversy between philosophical traditions emphasizing the motivating force of reasons (e.g., Kant), or that of emotions or rather passions (e.g., Hume). Reasons can justify, motivate or explain an action; for many authors, reasons include the mental states of a pro-attitude and of a belief [47]. We do not use 'reasons' as a normatively laden term (as e.g., Kant) and do not want to neglect the role of emotions in motivating actions in the right way. The question what role emotions play and of how to control and cultivate the 'right', ethically appropriate emotions or affective states has been a key issue of discussions about autonomy throughout the history of philosophy, from Aristotle to Stoics, to Kant, etc. There is also a philosophical history, e.g., in the work of Husserl or Sartre, of conceptualizing emotions as intentional, as an important aspect of being in the world, and sometimes even as acts. We assume that emotions can influence non-consciously the attentional focus and decision-making, or the way an action is performed, and they can influence the modulation of behavioral dispositions [48, p. 58–63]. We won't discuss it here, but we would say that emotions can not only influence in a non-conscious way (e.g., desires), but also inform our critical reflection, our evaluative judgments in important ways [48, 49]. Therefore, emotions also can be a relevant source of autonomous actions in reason-responsive conceptions of autonomy.

by knowledge, or eventually reasons, then some ability to realize the intended action effectively is implied for this reason to produce the intended action. If a person is unable to transform her reasons to act into the intended action, she lacks a substantial capacity to act or to live her life in a self-determined, or autonomous way. Accordingly, it seems clear that an autonomous person has to exercise some kind of control over the process leading from intention to intended event in order to be able to realize the intended action effectively. This requirement is reflected in many accounts of autonomy[5] and for the sake of simplicity, we will call this condition 'control' here. Lack of control is usually interpreted as a sign of non-autonomous decision or action. For example, some people suffer from inner coercion due to mental disorders (e.g., obsessions) and akratic actions in everyday life are usually considered non-autonomous. For instance, imagine the case of eating a chocolate bar even though you believe that not eating the bar would be better for you, and not eating the bar is what you actually want.[6] It seems that if you eat the bar, you are not in control of what you want to do, or of what you determined yourself to do in light of your own reasons. As a result, doubts could be raised about your autonomy in the situation and the use of your ability to be self-govern. Of course, different capacities can be further necessary for control. Reflective self-monitoring, for example, can be required in many cases, which may presuppose the monitoring of potentially influencing emotions [48, 49].

Ability to Enact Intentions within Concrete Relationships and Contexts

A third aspect of autonomy, namely the ability to enact intentions within concrete relationships and contexts, is especially emphasized in relational or externalist accounts of autonomy [60, 61]. According to such conceptions, developing, realizing and maintaining autonomy always takes place in relation to objects, to other persons, and to the world. For example, the ability to carry out a given intention is contingent on the availability of the (human, physical) resources to enact intentions. The question is: which aspects of such relationships or circumstances enable or inhibit autonomy?[7] One requirement (e.g., in the conception of Beauchamp and Childress [57]) is the *absence of controlling influences* that would determine an action, namely coercion, problematic persuasion (in contrast to rational persuasion like being convinced by sound arguments), or manipulation. Besides the difficulties of establishing a precise threshold where excessive external influence begins, it seems obvious that strong manipulation or coercion cannot result in reasons considered to be *one's own*. The question can also be dealt with in a broader sense by distinguishing autonomy-enabling and autonomy-inhibiting external circumstances [60]. The way in which reasons are developed and the issue of possible self-deception also play a key role here [61].

Critical Analysis of Autonomy in Light of BCI Research

BCIs provide the user with a wide range of new options in the possibility to act and therefore it might seem evident that BCIs therefore improve the ability to live life according to the user's own reasons.[8] If, for example, lost motor function after a stroke or injury is restored by BCI systems, or patients with cLiS regain a way to communicate with their surroundings via BCIs, their ability to act according to their reasons are significantly improved [5, 65]. Despite such positive impacts of BCIs on autonomy, in relation to improved motor function and communication, we must also account for and analyze less self-evident, philosophical implications of BCI use for autonomy, both positive and negative.

[5] Competence is a necessary condition in the conception of Beauchamp and Childress e.g., [57], which in accordance with the classic belief-desire model by Hume, is a property that is present when a person transforms her reasons into actions without external circumstances getting in the way, granted that these reasons are based on that person's true beliefs. For Hume, it is clear that reasons have instrumental value and are the "slaves" of desires (or passions, in his terminology), which alone motivate our actions [62]. The endorsement of a person's own reasons and the causal efficacy of her own reasons is also highly relevant, e.g., for coherentist conceptions of autonomy [50, 58, 59].

[6] There is an extensive and old debate about whether akratic actions exist at all and if so, how we can conceptualize them, see e.g., [63, 64].

[7] If we consider persons only as the product of social or external relations, there is little place for the ideas of the internalist conceptions of autonomy. It seems uncontroversial to suggest that individuals are influenced by social and external circumstances and to search for the autonomy-enabling external conditions is important in our view. Yet we grant space for the subject to gain distance from such influences, we find moderate accounts that still allow also for internalist conceptions of autonomy.

[8] As mentioned in Footnote 4, we do not use 'reasons' as a normatively laden term and furthermore do not exclude emotions in motivating actions in the right way.

We will proceed by considering each of the component-abilities of autonomy that we have introduced above, and the respective impact of BCIs.

BCIs Impact on the Ability to Use Information and Knowledge to Produce Reasons

Does the use of BCIs produce knowledge that affects the ability of a person to carry out autonomous decisions and actions or to reflect critically on which reasons should guide decisions and actions? At a fundamental level, the knowledge of the availability of BCI use may result in adjusting goals to act according to this existing tool. In *active* and *reactive* BCIs, we further gain new insights into the interplay between intentionally caused brain activity and resulting actions. People usually do not focus on their brain activity and do not need to possess knowledge about the way they can change their brain activity to carry out an action.[9] Therefore, the knowledge about the interplay between one's own brain activity and its effects in the world via, for example, a prosthesis seems to be qualitatively novel. However, this information could also be considered knowledge about new means to achieve aims, not significantly different from knowing how to use one's hands to cut with a knife.

The knowledge someone may gain about their own brain activity through *passive* BCIs, which gives feedback about workload [36] or affective states [37], could lead to a subsequent development of reasons to act based on this knowledge. Such a scenario points towards a more relevant direction for autonomy. The information the person receives from a *passive* BCI might generate new reasons or a change of previous reasons to act. Imagine, for instance, the following case. A person has a higher-order volition (following hierarchical accounts of autonomy) that she wants to maintain her relationship with a friend and she therefore decides to write her regularly; simultaneously she receives the information through *passive* BCI that every time she tries to write an email to this person, her affective state changes from good to very bad – a change that we assume she would not realize without the BCI.[10] The information gleaned via BCI might lead to a critical reflection about the previous motivation to write to this person, her relationship, and her emotions towards this person. Consequently, this process could result in changing her previous reasons to act. Such reflective changes of evaluative judgements present a positive impact of BCIs on self-determination. In some cases, this could also be interpreted as avoiding akrasia, if we understand akrasia, as some authors have described, as failing to appreciate the force of powerful feelings or emotions, while evaluating your own beliefs and forming your own reasons to do X [66]. On the other hand, the higher quantity and differing quality of information gained via *passive* BCIs could also result in inhibiting autonomy. If the person from the letter-example is confused by the new information delivered via BCI and is not able to report or to handle her emotional state, the new information might also undermine her autonomy.

We should further consider the role and the information received and saved about the brain activity of the user, if we are evaluating the ability to use information and knowledge to produce reasons. Imagine a BCI system in which the computer saves and 'learns' about the brain activities of the user in different situations.[11] This function could permit easier handling or automatization in BCIs. A higher degree of automatization might be an improvement of self-determination from the perspective of *control* (we come back to this later), as automatization can allow an easier way to transform one's reasons into actions. Nevertheless, the saved and used information about the user's brain activity by the BCI could result in fewer computer-generated options for the user to choose from. Imagine an application of a *passive* BCI, which is combined with smartphone use, where the computer receives and saves information about the mental states of the user while using certain websites. After a while the BCI provides the user only with websites that might suite her current mental state. The computer might only provide the user with those algorithm-derived information and options to act that were based on brain activity under similar contexts. In short, this technology could not only leave people more constrained to their own past states and decisions, but also limit their development of habits, thinking patterns, and actions. It may also constrain the actions to those

[9] There are other activities where we focus on our brain activity (neurofeedback training e.g.) in order to gain better results for a certain action.

[10] BCI applications that allow for a better perception of emotions compared to unaided humans is speculative. Nevertheless, emotion perception via BCI already is an aim of current research [39].

[11] Such BCIs could be described for all types, active, reactive and passive and such intelligent systems are under development; see e.g., [67].

deemed reasonable or acceptable by the BCI. The information and options to act provided by the BCI, e.g., with the suggested websites example, might further be presented by the computer system in an emotionally attractive way for the user, again based on previous information about the user.[12] This might nudge behavior in a direction (re-) calculated by the BCI [71]. In some cases, the user may not even know that the options to act have been reduced or computer-directed, or that a command to act has already been given by the BCI system. The person might know, in theory, that BCIs can reduce the options for humans to act. She might even be aware of ways in which this could happen, e.g., by adjusting the light, turning off your computer in cases of tiredness. However, the user would not necessarily know when or in what exact cases such alterations or nudges occur.

By restricting, altering, or algorithmically directing users' knowledge of the range of options to act, BCIs can undermine and decrease the ability to choose and enact different actions. Moreover, if the computer system 'makes decisions' for the user based on algorithms, without informing the user or giving her the option to decide or override the computer-based decision, then the person's ability for autonomous action could be significantly compromised. Both of these possibilities might be at odds with one necessary condition of self-determination, namely the *absence of controlling influences* determining an action [57].[13] The information from which users develop their reasons would be a product of automated data and information processing, where the user has little insight or control, and her options to choose might be subject to algorithmic influences or manipulations. Profiling users, subsequently providing them only with selected information and choosing for them is, in some cases, inhibiting the development of their own reasons in an 'acceptable' way, in a manner that violates precepts of human autonomy.

The issue of whether, and to what degree, the user is aware of or has control over the development of her reasons to act in BCIs has a profound impact on her self-determination. As Goering et al. address, a limit for closed loop applications undermining the individual's autonomy would be cases of complete manipulation, where the entire behavior of a person would be subverted to the control of the machine [34]. However, it is still unclear at what point or to what extent, machine initiated decision-making may infringe upon the autonomy of the individual. An understanding of the balance between machine and human 'decision-making' needs further social and ethical consideration, such as informing the patient and allowing her to voice her preferences. Some patients may want to trade some amount of autonomy for other goods. However, it could be ethically problematic to constrain a person's autonomy based on the trade-off between different components of autonomy (e.g., diminished autonomy in terms of the ability to use information and knowledge, but enhanced autonomy in terms of ability to self-control) and potentially even between autonomy and other values or outcomes.[14]

We can conclude that the use of BCIs can produce knowledge that affects in positive and negative ways the ability of a person to carry out autonomous decisions and actions, or to reflect critically on which reasons should guide her decisions and actions. On the positive side, BCIs could establish new means to achieve aims. In *passive* BCIs more information about mental states could further allow for some re-evaluation of reasons to act, or for overcoming akratic actions. As stated however, information about the user's brain activity in BCIs could also undermine autonomy, if the machine self-reliantly influences the user's options to act. After having addressed some possible implications of BCIs for the ability to use information and knowledge to generate reasons, we will now turn to the possible implications of BCIs on the ability to realize intended actions effectively.

BCIs and the Ability to Ensure that Intended Actions Are Effectively Realized (Control)

After developing reasons for a given decision, a person has to transform these reasons into concrete actions in order to be considered to act autonomously. In the context of BCIs, different aspects of *control* are necessary to ensure that intended actions are effectively realized. The first step a person takes to

[12] There are strong similarities here to the recent discussions of nudging, see e.g., [68, 69] and to discussions of restrictions of choice due to (non-conscious) manipulation [70].

[13] Similar questions have recently been discussed for closed-loop scenarios in BCIs [34, 72]. Goering et al. have pointed out that the user's agency and autonomy can be undermined if she doubts her authorship of an action due to the way the machine operates [34].

[14] This is analogous to cases where DBS patients sometimes have to choose between symptom relief and neurological or psychological side effects [73].

make her reasons effective is to put on an EEG cap, for example. The second step requires that she (in both *active* and *reactive* BCIs) might purposefully modulate her brain activity (e.g., mental imagery or focused attention). Here, she seems to control her *executive command* to act consciously and willingly. This conscious control over the executive command could be in fact stronger than in paradigmatic, everyday actions because in many paradigmatic actions, people have no conscious control over the initiation of an action [74]. Per contra, in *passive* BCIs, the user does not have control over the executive command. Here, the resulting events are not actions, and are, therefore, not self-determined actions.

Active and Reactive BCIs

After receiving an executive command as a result of purposefully produced brain activity in the case of either *reactive* and *active* BCIs, the computer system seems to take over *guidance control* until the intended event in the world is achieved. In BCI applications, there are no bodily movements of the user, unlike in normal paradigmatic actions, without a BCI. Rather, the computer system guides a device. The person might be able to inhibit the 'commanding' brain activity or to produce a different brain activity, if she wants to veto the initiated process. However, once a command is given for the computer system, the person does not have the same control over action guidance as in other everyday actions. It can be difficult also in normal, everyday paradigmatic actions, like in sports, to have action guidance or to veto an action after the decision was made for a certain movement. Nevertheless, humans seem to be able to change the course of their actions up until very shortly before the act is carried out in such normal paradigmatic cases [75], whereas the time needed and the current difficulties in technological readiness in BCIs to change or veto an initiated action might result in a difference in action guidance.

Some have concluded from this difference that BCIs produce a lack of control and therefore a responsibility gap [76]. However, similar to other actions undertaken with technological tools, users do not need to possess guidance control throughout the whole process, in order to ascribe responsibility to them [77]. After initiating the executive command through the relevant brain activity, the user transfers guidance control to the BCI device, similar to cases of other semi-autonomous technologies or tools, like in motor vehicles.

How might such a difference in control in BCIs change self-determination? In the introduction to this paper, we mention that reflective self-monitoring during action guidance, like monitoring potentially influencing emotions, can be essential for guidance control and self-determination [48, 49]. For example, the impulses people receive through emotions between time 1 (t1) (executive command) and time 2 (t2) (event performed, for example wheelchair or prosthesis moved) can help to maintain appropriate control in action guidance. These affective reactions can point to situational factors between t1 and t2 (e.g., recognizing that the person whom the BCI user tried to hit with her prosthesis is not a thief, but a beloved person), which the person could not plan for beforehand, but which are important to act according to the reasons in t1. In everyday life, people can use such input to adjust action guidance accordingly, and oftentimes they might not even be consciously aware of the input and subsequent adjustments. In comparison, it is clearly imaginable that guidance control could be jeopardized by BCIs.

The control of action guidance with the help of adjustments, which result from situational input, requires BCI users to make a greater effort than in paradigmatic actions, if such adjustments are at all possible in current BCI use. After the situational input, (e.g., emotions are recognized by a user) the user has to consciously stop producing previous brain activity or start producing new brain activity in order to reorient action guidance. A more developed BCI technology might help to overcome such difficulties in controlling action guidance through an easier integration of situational input into action guidance.

The positive outlook on action guidance and BCIs for autonomy, would be in the case of transferred control in action guidance to the machine, which could help the person to withdraw from contrary influencing impulses during action guidance between t1 and t2. In paradigmatic actions, for example, emotions might arise between t1 and t2, which might (unconsciously) influence the performance of the event in t2, contrary to the initiating volition in t1. Such influences could be rendered mute in *active* and *reactive* BCIs for the event in t2. Transferring parts of guidance control to the machine in BCIs might therefore indicate a new way of bringing one's own reasons into actions effectively. The use of *active* and

reactive BCIs might therefore indicate a new way to overcome akrasia.[15]

As aforementioned, certain automatization processes in BCI use might provide also higher effectiveness in bringing reasons into actions. The machine might be problematic in terms of reducing the options to choose for the user (see above; ability to use information and knowledge to produce reasons), but from the perspective of control, it might allow the person to act in accordance with her own reasons more easily.

Passive BCIs

In *passive* BCIs, the opposite direction regarding situational input, for example due to emotions, seems to be plausible. It is not the neglect of the machine of human emotions in action guidance, but rather the 'knowledge' of the user's emotions via the machine that might help to enhance self-determination. If, for example, a person has formed the higher-order volition that she does not want to write emails to her students or boss while she is in a negative affective state, the BCI information about her current affective state might help her to succeed in her aim. After the BCI offers her information that her affective state has become negative (let us suppose that she does not recognize this usually), she could turn off the computer or e-mail program. Alternatively, for a more extreme case, imagine a person who strongly wants to stop her addictive behavior: she endorses this wish wholeheartedly and wants it to be effective in her actions. The same person knows that, in some affective states (of craving), she is unable to stop her addiction. If the *passive* BCI provides her with information about the beginning of such an affective state, this could help her to stay away from the addictive substance before it is too late. It could be argued, though, that this warning or risk-aversion BCI capability could be achieved in a far easier manner. For example, the person could avoid storing the addictive substance at home. Further, the person may also ignore the anti-addiction-help in both cases.

Nevertheless, the potential information about brain states provided by the *passive* BCI could be a qualitatively novel way to maintain control in action guidance. Such applications of *passive* BCIs might therefore help to improve autonomy. If affective BCIs someday are able to measure affective states exactly [39] the feedback about such states could help the person make her reasons effective in actions. To encapsulate this discussion in a nutshell, the autonomy-relevant ability to ensure that intended actions are effectively realized can be enhanced, but also undermined in BCI use, depending on the BCI type we address and on the aspect of control we focus our considerations on.

BCIs and the Ability to Enact Intentions within Concrete Relationships and Contexts

The next component of autonomy concerns the fact that the development and realization of one's own reasons always takes place in relation to the world, to other persons and material conditions. Many factors of the interaction with others can enhance or inhibit autonomy [52, 53]. In other words, people's relation to the world shapes their autonomy, and their relation to the world is, in turn, often shaped by technology. The fact that technology shapes this relation is sometimes referred to as 'mediation.' Technology mediates the relation to the world and other people but it also mediates subjectivity. It is important to note here that the world and the self are not fixed entities, so that technology simply connects the two, but the two are co-shaped in the technological mediation [78, 79]. The manner in which subjectivity is molded by technology is evident when we look at tracking or monitoring devices that shape the users' relation to their bodies and how people view their health and illness [80]. There are examples of cases in which recording and attending to recordable bio-data may lead users to neglect non-quantifiable signals and may lead to a more quantitative view on life [81].

Importantly, human behavior is shaped by technology in myriad ways and not all of these ways bear on autonomy. For example, the road influences the behavior of driving a car and the arrangement of the keyboard limits what letters you can type but we would not go so far as to call this an infringement on autonomy while engaging in those activities. However, there are cases where technologies influence us in ways that are not so clear-cut. For example, it is well known that certain social-media platforms algorithmically screen, limit,

[15] BCIs could then also be interpreted as an unusual technological option for 'technologies of the self', i.e. practices that are supposed to allow a person to gain a productive distance from her passions and desires, in order to augment the ability to carry out intended actions. Practices (usually not in combination with technologies like BCI) to control desires and passions in order to live a self-determined life have been discussed in philosophy since the ancient times, e.g., by Aristotle, the Stoics and in the last century by Foucault.

and pick the information and news that users receive, which may have more profound effects on core components of autonomy. Shifting our attention to concrete interactions with machines, one may wonder how interactions with machines impact our ability to enact intentions within concrete relationships and contexts. Philosophy of technology has already discussed relevant aspects of this question in some depth. For new technologies, it now seems inadequate to define the interaction only in terms of people *using* technology like BCIs [16]. Indeed, there are some new characteristics of human-machine interactions that are not well conveyed by the terminology of use, user, and usage. Some technologies (like invasive BCIs or DBS combined with some monitoring function) are literally inside the body of the user; they react intelligently to the user and the environment, act upon self-reliant decisions, and co-create the human environment [16].

Considering BCIs, it seems clear that this technology fosters a more autonomous life for many people. In most of the current applications, BCI helps individuals to (re)gain communication and motor abilities. The notion that technology enslaves humans, an idea sometimes put forth in public discourse, seems to lose its force given such advantages, particularly with medical applications. Blinded by so many advantages, it would be thus simple to underestimate the potential negative influences of BCI technology on relational aspects of autonomy. For example, if an intelligent BCI system takes over decision-making or guidance control, we must learn first how to evaluate the way the machine does this, before being able to understand their impact on the course of actions. In usual human-human interaction, there is no certainty about whether we have correctly interpreted another person's behavior, but we can rely on our lifelong experience, social codes and habits. If an intelligent machine in BCI use generates algorithmically-derived options to choose from, or if the machine 'decides' instead of us, we can expect more difficulties assessing the interaction, e.g., regarding its potential for manipulation. In social interaction with human beings, it may also be difficult to identify manipulative behavior, but at least there is usually the possibility of communication and social exchange in order to verify the causes behind what others say or do. In the interaction with a BCI using machine learning algorithms, there always remains a black box inside the device which cannot be accessed. For instance, there is no option to question a BCI directly, and ask it to explain why it made certain decisions or made certain outputs. Given that in BCI technology at least some of the decisions and choices are outsourced to algorithms, and because of the potential impact of BCI on autonomy, this technology is a case where an 'ethics of algorithms' [82] seems to be more than warranted.

Conclusion

When asked whether BCIs have a positive or negative impact on user autonomy, most people would likely claim that they have a positive impact, since BCIs can restore, enhance, or improve human abilities to ensure that intended actions are performed. The idea that BCIs would improve autonomy thus seems almost uncontroversial. In this paper however, we identified several possible implications of BCI with respect to autonomy, including potential negative outcomes. We focused on three component-abilities relevant for autonomy: (1) the ability to use information and knowledge to produce reasons; (2) the ability to ensure that intended actions are effectively realized (control); and (3) the ability to enact intentions within concrete relationships and contexts. Regarding the first two component-abilities, we showed that the knowledge of the BCI user about her brain activity could improve autonomy. This knowledge could be used to establish certain brain activities as means. Knowledge provided (in *passive*) BCIs of affective states in certain situations could further allow the user to re-evaluate her reasons to act. Feedback via BCI about current or impending affective states might also help, in some cases, to overcome akratic actions. Akratic actions could also be prevented in BCI use in another way: by transferring guidance control to the machine after the executive command has been made, this could override otherwise disturbing influences (e.g., from competing desires). An improvement of autonomy could also result from a higher grade of automatization due to decision-making or guidance control transferred to the machine. Nevertheless, exactly the higher control by the machine in the previous cases could also be interpreted as a reduction of human autonomy.

In contrast to potentially positive effects on autonomy, BCIs have potential for negative impacts on making our own reasons effective in actions. Compromises in human autonomy could arise from the lack of integration of situational factors during action guidance in BCIs, due to missing or different feedback procedures

or due to ignorance of situational feedback (e.g., ignorance of emotions). Furthermore, based on previous information about the user, the machine could influence the development of the user's reasons by altering the user's options to act self-reliantly. An attenuation or absence of options to choose from – without the explicit endorsement of the user – has the potential to impact self-determination since controlling influences are present. It appears that when considering autonomy as a multi-component ability, we have to reflect seriously on what it means to act autonomously in our interactions with new technologies and BCIs and to understand the trade-offs between different components of autonomy. Accordingly, BCI technologies increase the complexity of the evaluation of the aims and the degree of the control of the interacting 'partner,' or the degree to which machines influence human decision-making.

Acknowledgments Writing of this article was supported by a joint grant from the Federal Ministry of Education and Research (BMBF) in Germany, the Canadian Institutes of Health Research and the Fonds de recherche du Québec – Santé (European Research Projects on Ethical, Legal, and Social Aspects (ELSA) of Neurosciences) as well as a career award from the Fonds de recherche du Québec – Santé (ER). We would like to thank the anonymous reviewers for their in depth and constructive review of our paper and for very helpful comments, further we thank Stephanie Simpson, Rose Richards, Mary Clare O'Donnell and Christoph Bublitz for providing helpful comments on a previous version of this manuscript.

References

1. Shih, J.J., D.J. Krusienski, and J.R. Wolpaw. 2012. Brain-computer interfaces in medicine. *Mayo Clinic Proceedings* 87 (3): 268–279.
2. Li, G., and D. Zhang. 2017. Brain-computer interface controlling cyborg: A functional brain-to-brain interface between human and cockroach. In *Brain-computer interface research. A state-of-the-art summary 4*, ed. Christoph Guger, Gernot Müller-Putz, and Brendan Allison, 71–79. Cham: Springer International Publishing.
3. Bouton, C.E., A. Shaikhouni, N.V. Annetta, M.A. Bockbrader, D.A. Friedenberg, D.M. Nielson, G. Sharma, P.B. Sederberg, B.C. Glenn, and W.J. Mysiw. 2016. Restoring cortical control of functional movement in a human with quadriplegia. *Nature* 533 (7602): 247–250.
4. Winkler, R. 2017. Elon musk launches Neuralink to connect brains with computers. *Wall Street Journal*. https://www.wsj.com/articles/elon-musk-launches-neuralink-to-connect-brains-with-computers-1490642652. Accessed 29 Sept 2017.
5. Chaudhary, U., N. Birbaumer, and A. Ramos-Murguialday. 2016. Brain-computer interfaces for communication and rehabilitation. *Nature Reviews Neurology* 12 (9): 513–525.
6. Marchetti, M., and K. Priftis. 2015. Brain–computer interfaces in amyotrophic lateral sclerosis: A metanalysis. *Clinical Neurophysiology* 126 (6): 1255–1263.
7. Käthner, I., S. Halder, C. Hintermüller, A. Espinosa, C. Guger, F. Miralles, E. Vargiu, S. Dauwalder, X. Rafael-Palou, and M. Solà. 2017. A multifunctional brain-computer interface intended for home use: An evaluation with healthy participants and potential end users with dry and gel-based electrodes. *Frontiers in Neuroscience* 11: 286.
8. Brunner, C., N. Birbaumer, B. Blankertz, D. Guger, A. Kübler, D. Mattia, J. del R. Millán, F. Miralles, A. Nijholt, E. Opisso, N. Ramsey, P. Salomon, and G.R. Müller-Putz. 2015. BNCI horizon 2020: Towards a roadmap for the BCI community. *Brain-Computer Interfaces* 2 (1): 1–10.
9. De Mul, J., and B. van den Berg. 2011. Remote control: Human autonomy in the age of computer-mediated agency. In *Law, human agency, and autonomic computing: The philosophy of law meets the philosophy of technology*, ed. Mireille Hildebrandt and Antoinette Rouvroy, 46–64. London: Routledge.
10. Racine, E., and S. Rousseau-Lesage. 2017. The voluntary nature of decision-making in addiction: Static metaphysical views versus epistemologically dynamic views. *Bioethics* 31 (5): 349–359.
11. Kallinikos, J. 2011. Technology and accountability: On autonomic computing and human agency. In *Law, human agency, and autonomic computing: The philosophy of law meets the philosophy of technology*, ed. Mireille Hildebrandt and Antoinette Rouvroy, 161–179. London: Routledge.
12. Nietzsche, F.W. 1998. *On the genealogy of morality*. Cambridge: Cambridge University Press.
13. Foucault, M. 1982. The subject and power. *Critical Inquiry* 8 (4): 777–795.
14. Mackenzie, C., and N. Stoljar. 2000. *Relational autonomy: Feminist perspectives on automony, agency, and the social self*. Oxford: Oxford University Press.
15. Kahneman, D. 2011. *Thinking, fast and slow*. London: Penguin Books.
16. Verbeek, P.-P. 2011. Subject to technology on autonomic computing and human autonomy. In *Law, human agency, and autonomic computing: The philosophy of law meets the philosophy of technology*, ed. Mireille Hildebrandt and Antoinette Rouvroy, 27–46. London: Routledge.
17. Graimann, B., B. Allison, and G. Pfurtscheller. 2009. Brain–computer interfaces: A gentle introduction. In *Brain-computer interfaces*, ed. Bernhard Graimann, Gert Pfurtscheller, and Brendan Allison, 1–27. Heidelberg: Springer.
18. Daly, J.J., and J.R. Wolpaw. 2008. Brain–computer interfaces in neurological rehabilitation. *Lancet Neurology* 7 (11): 1032–1043.
19. Wolpaw, J.R., N. Birbaumer, D.J. McFarland, G. Pfurtscheller, and T.M. Vaughan. 2002. Brain–computer interfaces for communication and control. *Clinical Neurophysiology* 113 (6): 767–791.
20. Lebedev, M.A., and M.A. Nicolelis. 2006. Brain-machine interfaces: Past, present and future. *Trends in Neurosciences* 29 (9): 536–546.
21. Yuan, H., and B. He. 2014. Brain-computer interfaces using sensorimotor rhythms: Current state and future perspectives. *IEEE Transactions on Bio-Medical Engineering* 61 (5): 1425–1435.

22. Mak, J.N., and J.R. Wolpaw. 2009. Clinical applications of brain-computer interfaces: Current state and future prospects. *IEEE Reviews in Biomedical Engineering* 2: 187–199.
23. Schalk, G. 2010. Can electrocorticography (ECoG) support robust and powerful brain-computer interfaces? *Frontiers in Neuroengineering* 3: 9.
24. Oxley, T.J., N.L. Opie, S.E. John, G.S. Rind, S.M. Ronayne, T.L. Wheeler, J.W. Judy, A.J. McDonald, A. Dornom, and T.J. Lovell. 2016. Minimally invasive endovascular stent-electrode array for high-fidelity, chronic recordings of cortical neural activity. *Nature Biotechnology* 34: 320–327.
25. Salisbury, D.B., T.D. Parsons, K.R. Monden, Z. Trost, and S.J. Driver. 2016. Brain-computer interface for individuals after spinal cord injury. *Rehabilitation Psychology* 61 (4): 435–441.
26. Maksimenko, V.A., S. van Heukelum, V.V. Makarov, J. Kelderhuis, A. Lüttjohann, A.A. Koronovskii, A.E. Hramov, and G. van Luijtelaar. 2017. Absence seizure control by a brain computer interface. *Scientific Reports* 7: 2487.
27. Zafar, M.B., K.A. Shah, and H.A. Malik. 2017. Prospects of sustainable ADHD treatment through brain-computer interface systems. Paper presented at the 2017 Conference for *Innovations in Electrical Engineering and Computational Technologies (ICIEECT)* in Karachi, Pakistan.
28. McFarland, D.J., J. Daly, C. Boulay, and M.A. Parvaz. 2017. Therapeutic applications of BCI technologies. *Brain-Computer Interfaces* 4 (1–2): 37–52.
29. Ahn, M., M. Lee, J. Choi, and S.C. Jun. 2014. A review of brain-computer interface games and an opinion survey from researchers, developers and users. *Sensors* 14 (8): 14601–14633.
30. Grau, C., R. Ginhoux, A. Riera, T.L. Nguyen, H. Chauvat, M. Berg, J.L. Amengual, A. Pascual-Leone, and G. Ruffini. 2014. Conscious brain-to-brain communication in humans using non-invasive technologies. *PLoS One* 9 (8): e105225.
31. Rao, R.P.N., A. Stocco, M. Bryan, D. Sarma, T.M. Youngquist, J. Wu, and C.S. Prat. 2014. A direct brain-to-brain interface in humans. *PLoS One* 9 (11): e111332.
32. Alonso-Valerdi, L.M., R.A. Salido-Ruiz, and R.A. Ramirez-Mendoza. 2015. Motor imagery based brain–computer interfaces: An emerging technology to rehabilitate motor deficits. *Neuropsychologia* 79: 354–363.
33. Flesher, S.N., J.L. Collinger, S.T. Foldes, J.M. Weiss, J.E. Downey, E.C. Tyler-Kabara, S.J. Bensmaia, A.B. Schwartz, M.L. Boninger, and R.A. Gaunt. 2016. Intracortical microstimulation of human somatosensory cortex. *Science Translational Medicine* 8 (361): 361ra141.
34. Goering, S., E. Klein, D.D. Dougherty, and A.S. Widge. 2017. Staying in the loop: Relational agency and identity in next-generation DBS for psychiatry. *AJOB Neuroscience* 8 (2): 59–70.
35. Zander, T.O., and L.R. Krol. 2017. Team PhyPA: Brain-computer interfacing for everyday human-computer interaction. *Periodica Polytechnica Electrical Engineering and Computer Science* 61 (2): 209.
36. Gerjets, P., C. Walter, W. Rosenstiel, M. Bogdan, and T.O. Zander. 2014. Cognitive state monitoring and the design of adaptive instruction in digital environments: Lessons learned from cognitive workload assessment using a passive brain-computer interface approach. *Frontiers in Neuroscience* 8: 385.
37. Fan, J., J.W. Wade, A.P. Key, Z. Warren, and N. Sarkar. 2017. EEG-based affect and workload recognition in a virtual driving environment for ASD intervention. *IEEE Transactions on Biomedical Engineering* 99: 43–51. https://doi.org/10.1109/TBME.2017.2693157.
38. Martel, A., S. Dahne, and B. Blankertz. 2014. EEG predictors of covert vigilant attention. *Journal of Neural Engineering* 11 (3): 035009.
39. Mühl, C., B. Allison, A. Nijholt, and G. Chanel. 2014. A survey of affective brain computer interfaces: Principles, state-of-the-art, and challenges. *Brain-Computer Interfaces* 1 (2): 66–84.
40. Martinovic, I., D. Davies, M. Frank, D. Perito, T. Ros, and D. Song. 2012. On the feasibility of side-channel attacks with brain-computer interfaces. *Presented as part of the 21st USENIX Security Symposium,* USENIX Security 12: 143–158.
41. Bonaci, T., R. Calo, and H. J. Chizeck. 2014. App stores for the brain: Privacy & security in brain-computer interfaces. *Proceedings of the IEEE 2014 International Symposium on Ethics in Engineering, Science, and Technology*: 47.
42. Mousavi, M., A.S. Koerner, Q. Zhang, E. Noh, and V.R. de Sa. 2017. Improving motor imagery BCI with user response to feedback. *Brain-Computer Interfaces* 4 (1–2): 74–86.
43. Feinberg, Joel. 1986. *Harm to self*. New York, Oxford: Oxford University Press.
44. Kant, I. 1997 [1785]. Grundlegung zur Metaphysik der Sitten [Groundwork of the metaphysics of morals]. In: *Ak IV*, ed. Mary Gregor, 387–436. Cambridge: Cambridge University Press.
45. Wilson G. and S. Samuel. 2016. Action. In: *The Stanford encyclopedia of philosophy, winter 2016 edn*., ed Edward N. Zalta. https://plato.stanford.edu/archives/win2016/entries/action. Accessed 29 Sept 2017.
46. Betzler, M. 2009. Authenticity and self-governance. In *Emotions, ethics, and authenticity*, ed. Mikko Salmela and Verena Mayer, 51–67. Amsterdam/Philadelphia: John Benjamins Publishing Company.
47. Davidson, D. 1963. Actions, reasons, and causes. *Journal of Philosophy* 60 (23): 685–700.
48. Tappolet, Christine. 2016. *Emotions, values, and agency*. Oxford: Oxford University Press.
49. Jones, K. 2003. Emotion, weakness of will, and the normative conception of agency. In *Philosophy and the emotions*, ed. Anthony Hatzimoysis, 181–200. Cambridge: Cambridge University Press.
50. Dworkin, Gerald. 1988. *The theory and practice of autonomy*. Cambridge: Cambridge University Press.
51. Racine, E., D. Larivière-Bastien, E. Bell, A. Majnemer, and M. Shevell. 2013. Respect for autonomy in the healthcare context: Observations from a qualitative study of young adults with cerebral palsy. *Child: Care, Health and Development* 39 (6): 873–879.
52. Racine, E., and V. Dubljević. 2017. Behavioral and brain-based research on free moral agency: Threat or empowerment? In *Neuroethics: Anticipating the future*, ed. Judy Illes. Oxford: Oxford University Press.
53. Racine, E., and V. Dubljević. 2016. Porous or contextualized autonomy? Knowledge can empower autonomous moral agents. *The American Journal of Bioethics* 16 (2): 48–50.

54. Willett, C., E. Anderson, and D. Meyers. 2016. Feminist perspectives on the self. In: *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), ed. Edward N. Zalta, https://plato.stanford.edu/archives/win2016/entries/feminism-self/. Accessed 25 Feb 2018.
55. Govier, T. 1993. Self-trust, autonomy, and self-esteem. *Hypatia* 8 (1): 99–120.
56. McLeod, C. 2002. *Self-trust and reproductive autonomy*. Cambridge: MIT Press.
57. Beauchamp, T., and L. Childress. 2009. *Principles of biomedical ethics*. Oxford: Oxford University Press.
58. Frankfurt, H.G. 1971. Freedom of the will and the concept of a person. *Journal of Philosophy* 68 (1): 5–20.
59. Quante, M. 2011. In defence of personal autonomy. *Journal of Medical Ethics* 37: 597–600.
60. Oshana, M. 1998. Personal autonomy and society. *Journal of Social Philosophy* 29: 81–102.
61. Christman, J. 1991. Autonomy and personal history. *Canadian Journal of Philosophy* 21 (1): 1–24.
62. Hume, D.. 1975 [1739]. *A treatise of human nature*. Oxford: Clarendon Press.
63. Davidson, D. 1970. How is weakness of the will possible? In *Moral concepts*, ed. Joel Feinberg. Oxford: Oxford University Press.
64. Stroud, S. 2014. Weakness of will. In: *The stanford encyclopedia of philosophy, spring 2014 edn*., ed Edward N. Zalta. https://plato.stanford.edu/archives/spr2014/entries/weakness-will/. Accessed 29 Sept 2017.
65. Guger, C., R. Spataro, B.Z. Allison, A. Heilinger, R. Ortner, W. Cho, and V. La Bella. 2017. Complete locked-in and locked-in patients: Command following assessment and communication with vibro-tactile P300 and motor imagery brain-computer interface tools. *Frontiers in Neuroscience* 11: 251.
66. Arpaly, N. 2000. On acting rationally against one's best judgment. *Ethics* 110 (3): 488–513.
67. Wu, S.L., Y.T. Liu, T.Y. Hsieh, Y.Y. Lin, C.Y. Chen, C.H. Chuang, and C.T. Lin. 2017. Fuzzy integral with particle swarm optimization for a motor-imagery-based brain-computer interface. *IEEE Transactions on Fuzzy Systems* 25 (1): 21–28.
68. Bell, E., V. Dubljevic, and E. Racine. 2013. Nudging without ethical fudging: Clarifying physician obligations to avoid ethical compromise. *American Journal of Bioethics* 13 (6): 18–19.
69. Cohen, S. 2013. Nudging and informed consent. *American Journal of Bioethics* 13 (6): 3–11.
70. Kiesel, A., A. Wagener, W. Kunde, J. Hoffmann, A.J. Fallgatter, and C. Stöcker. 2006. Unconscious manipulation of free choice in humans. *Consciousness and Cognition* 15 (2): 397–408.
71. Ihde, D. 2011. Smart? Amsterdam urinals and autonomic computing. In *Law, human agency, and autonomic computing: The philosophy of law meets the philosophy of technology*, ed. Mireille Hildebrandt and Antoinette Rouvroy, 12–27. London: Routledge.
72. Kellmeyer, P., T. Cochrane, O. Mueller, C. Mitchell, T. Ball, J.J. Fins, and N. Biller-Andorno. 2016. The effects of closed-loop medical devices on the autonomy and accountability of persons and systems. *Cambridge Quarterly of Healthcare Ethics* 25 (4): 623–633.
73. Glannon, W. 2014. Prostheses for the will. *Frontiers in Systems Neuroscience* 8: 79.
74. Bargh, J.A., and T.L. Chartrand. 1999. The unbearable automaticity of being. *American Psychologist* 54 (7): 462–479.
75. Schultze-Kraft, M., D. Birman, M. Rusconi, C. Allefeld, K. Görgen, S. Dähne, B. Blankertz, and J.-D. Haynes. 2016. The point of no return in vetoing self-initiated movements. *Proceedings of the National Academy of Sciences* 113 (4): 1080–1085.
76. Tamburrini, G. 2009. Brain to computer communication: Ethical perspectives on interaction models. *Neuroethics* 2 (3): 137–149.
77. 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.
78. Verbeek, P.-P. 2014. Some misunderstandings about the moral significance of technology. In *The moral status of technical artefacts*, ed. Peter-Paul Verbeek and Peter Kroes, 75–88. Netherlands: Springer.
79. Verbeek, P.-P. 2008. Obstetric ultrasound and the technological mediation of morality: A postphenomenological analysis. *Human Studies* 31 (1): 11–26.
80. Kiran, A.H., and P.-P. Verbeek. 2010. Trusting our selves to technology. *Knowledge, Technology, and Policy* 23 (3–4): 409–427.
81. Van Den Eede, Y. 2015. Tracing the tracker: A postphenomenological inquiry into self-tracking technologies. In *Postphenomenological investigations: Essays on human technology relations*, ed. Robert Rosenberger and Peter-Paul Verbeek, 143–158. Lanham, MD: Lexington Books.
82. Mittelstadt, B. D., P. Allo, M. Taddeo, S. Wachter, and L. Floridi. 2016. The ethics of algorithms: Mapping the debate. Big Data & Society 3 (2): 2053951716679679, 205395171667967.

9.) Literaturverzeichnis

1. Schöne-Seifert, B. 1996. Medizinethik. In *Angewandte Ethik. Die Bereichsethiken und ihre theoretische Fundierung. Ein Handbuch*, ed. Julian Nida-Rümelin, 552-648. Stuttgart: Alfred Körner Verlag.
2. Beauchamp, T.L., und J.F. Childress. 2001. *Principles of biomedical ethics*. New York, Oxford: Oxford University Press.
3. Pömsl, J., und O. Friedrich. 2017. Why Enhancing Autonomy Is Not a Question of Improving Single Aspects of Reasoning Abilities through Neuroenhancement. *Neuroethics* 10(2): 243-254. doi:10.1007/s12152-016-9299-y.
4. Clausen, J., und H.J. Ehni. 2012. Enhancement. Einführung. In *Ethik in der Medizin. Ein Studienbuch*, ed. Urban Wiesing, 472-479. Stuttgart: Reclam.
5. Schleim, S., und B.B. Quednow. 2018. How realistic are the scientific assumptions of the neuroenhancement debate? Assessing the pharmacological optimism and neuroenhancement prevalence hypotheses. *Frontiers in Pharmacology* 9: 1-7. doi:10.3389/fphar.2018.00003.
6. Friedrich, O., Racine E., Steinert, S., Pömsl, J. und R. Jox. 2018. An Analysis of the Impact of Brain-Computer Interfaces on Autonomy. *Neuroethics*. doi:10.1007/s12152-018-9364-9.
7. Graimann, B., Allison, B., und G. Pfurtscheller. 2009. Brain–Computer Interfaces: A Gentle Introduction. In *Brain-Computer Interfaces*, ed. Bernhard Graimann, Gerd Pfurtscheller, und Brendan Allison, 1-27. Heidelberg: Springer.
8. Lotte, F., Nam, C.S. und A. Nijholt. 2018. Introduction: Evolution of Brain-Computer Interfaces. In *Brain-Computer Interfaces Handbook: Technological and Theoretical Advance*, ed. Nam S. Chang, Anton Nijholt, und Fabien Lotte, 1-11. Boca Raton: Taylor & Francis.
9. DeGrazia, D. 2014. Moral enhancement, freedom, and what we (should) value in moral behaviour. *Journal of Medical Ethics* 40(6): 361-368. doi:10.1136/medethics-2012-101157.
10. Crockett, M.J. 2014. Moral bioenhancement: a neuroscientific perspective. *Journal of Medical Ethics* 40(6): 370-371.
11. Schaefer, G.O., Kahane, G. und J. Savulescu. 2014. Autonomy and Enhancement. *Neuroethics* 7(2): 123-136.
12. Mühl, C., Allison, B., Nijholt, A. und G. Chanel. 2014. A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges. *Brain-Computer Interfaces* 1(2): 66-84. doi:10.1080/2326263X.2014.912881.
13. Turkheimer, E. 2016. Weak Genetic Explanation 20 Years Later: Reply to Plomin et al. (2016). *Perspectives on Psychological Science* 11(1): 24-28. doi:10.1177/1745691615617442.
14. Betzler, M. 2013. *Autonomie der Person*. Münster: Mentis.
15. Buss, S. und A. Westlund. 2016. Personal Autonomy. In *Stanford Encyclopedia of Philosophy (Spring 2018 Edition)*, ed. Edward N. Zalta. https://plato.stanford.edu/archives/spr2018/entries/personal-autonomy. Zuletzt aufgerufen 01. April 2019.