

EL ROL DE LA AGENCIA EN LA IDENTIDAD HÍBRIDA DE LOS PACIENTES NEUROTECNOLÓGICOS

THE ROLE OF AGENCY IN THE HYBRID IDENTITY OF NEUROTECHNOLOGICAL PATIENTS

Jorge Ignacio Fuentes¹

Pontificia Universidad Católica de Chile
Centro de Bioética UC, Facultad de Medicina
jfuenm@uc.cl
ORCID: 0009-0000-1414-2315

Abel Wajnerman Paz²

Pontificia Universidad Católica de Chile
Instituto de Éticas Aplicadas
abel.wajnerman@uc.cl
ORCID: 0000-0003-1551-527X

Recibido: 24-06-2024 • Aceptado: 14-10-2024

¹ Profesor asistente del Centro de Bioética de la PUC. Profesor de Matemáticas y Física (Universidad de Chile), magíster en Filosofía de las Ciencias (USACH) y doctor en Filosofía (Universidad Alberto Hurtado). Sus líneas de investigación incluyen la neuroética, la ética aristotélica y la filosofía de las neurociencias cognitivas. Indaga sobre asuntos ontológicos, epistemológicos y éticos sobre la computación neuronal y la artificial. Se encuentra investigando las implicaciones éticas de la interacción del cerebro humano y los implantes cerebrales.

² Profesor asistente de Neuroética en el Instituto de Éticas Aplicadas de la PUC. Doctorado en Filosofía en la UBA (2014) y becario posdoctoral Conicet (2015-2017) y Fondecyt (2018-2021). Sus principales áreas de interés son la filosofía de la neurociencia cognitiva y la neuroética. Su investigación se centra en la relación ontológica y epistémica entre los fenómenos cognitivos y los mecanismos neuronales, desde su dimensión ética. La escritura de este artículo es en el marco del proyecto Fondecyt iniciación 11220327, “La filosofía de la neurociencia como marco conceptual para los neuroderechos”.

RESUMEN

Recientemente, en la literatura sobre neuroética se ha analizado la integración entre neurotecnologías y ser humano desde la teoría de la mente extendida. La discusión se ha centrado en buscar dimensiones para evaluar la integración de los dispositivos en las funciones cognitivas de sus usuarios. Este enfoque ha tenido un relativo éxito en conceptualizar la paridad ética de los dispositivos respecto del cuerpo orgánico. Sin embargo, esta integración presenta al menos tres desafíos que no han sido abordados y requieren un nuevo marco conceptual. El primero, concierne a la distinción entre paciencia y agencia moral. El marco actual aborda solo el primero de estos aspectos. Los otros dos se desprenden en parte de este. El segundo, se relaciona con la distinción entre las dimensiones fenoménicas de los mencionados aspectos, esto es, el sentido de pertenencia corporal y el sentido de agencia. El foco en la paciencia moral descuida aspectos fenoménicos fundamentales de la integración con la neurotecnología. Finalmente, la posibilidad de que el dispositivo actúe coercitivamente sobre nuestros centros cognitivos constituye una forma de integración problemática que no puede abordarse adecuadamente desde los enfoques actuales. Proponemos que estos tres desafíos pueden ser estudiados explorando un nuevo abordaje sobre la identidad personal que enfatiza dichas dimensiones.

PALABRAS CLAVE: neuroética, neurotecnología, interfaces cerebro-computadora, mente extendida, identidad personal.

ABSTRACT

Recently, the integration between neurotechnologies and human beings has been analyzed using the extended mind theory in neuroethics literature. The discussion has been focused on finding dimensions to assess the integration of the devices in their users' cognitive functions. This approach has been relatively successful in elucidating the ethical parity of the devices regarding the organic body. However, the mentioned integration poses at least three additional challenges that have not been addressed in the literature and require a new conceptual framework. The first relates to the distinction between moral agency and patiency. The current framework addresses only the latter. The other two are in part entailed by this. The second one is the distinction between a sense of body ownership and a sense of agency. The focus on moral patiency disregards fundamental phenomenal dimensions of integration with neurotechnology. Finally, the possibility that the device acts coercively over our cognitive systems is a problematic form of integration that cannot be addressed adequately within current frameworks. We argue that these three challenges can be better faced by exploring a new approach to personal identity that highlights the ethical dimensions mentioned.

KEYWORDS: neuroethics, neurotechnology, brain-computer interfaces, extended mind, personal identity.

1. Introducción

RI Las neurotecnologías pueden definirse como dispositivos que modulan y/o detectan directamente la actividad del cerebro³. Podemos primero distinguir entre las tecnologías orientadas a modificar la actividad cerebral, como las tecnologías de estimulación cerebral de corriente directa, de aquellas empleadas para detectar las señales cerebrales, como el electroencefalograma o los dispositivos de resonancia magnética funcional, y dispositivos que, detectando las señales cerebrales, permiten al sujeto controlar por medio de estas señales objetos externos, como un brazo robótico. También encontramos neurotecnologías que combinan estas dos funcionalidades. A su vez, podemos distinguir entre neurotecnologías invasivas y no invasivas. Se denomina neurotecnologías invasivas a aquellas que requieren un implante en contacto con el cerebro insertado durante un proceso quirúrgico, mientras que las no invasivas no requieren un procedimiento de este tipo. Por último, podemos distinguir entre neurotecnologías médicas y no médicas. Muy frecuentemente el uso de estas tecnologías se da en el ámbito clínico para tratar a pacientes con afecciones complejas (por ejemplo, esclerosis lateral amiotrófica). Pero también se están desarrollando y comercializando neurotecnologías para usos cotidianos o para monitorear el desempeño de estudiantes o trabajadores. Un tipo neurotecnología médica invasiva ampliamente discutido son los dispositivos de estimulación cerebral profunda (en adelante, DBS por su sigla en inglés [Deep Brain Stimulation]), usados principalmente para disminuir los síntomas de la enfermedad de Parkinson. También han sido ampliamente discutidas en la literatura neuroética las interfaces cerebro-computadora (en adelante, BCI por su sigla en inglés [Brain-Computer Interface]). principalmente usadas para establecer una vía de comunicación con el mundo externo en personas con síndrome de enclaustramiento, o permitir la ejecución de comandos en un dispositivo externo (ej. robot de asistencia, cursor de un computador, prótesis, etc.), para personas con una significativa, o casi total, pérdida de movimiento voluntario de sus músculos, utilizando las ondas cerebrales. Los dispositivos de esta segunda clase para interpretar las señales cerebrales, requieren un proceso de entrenamiento del paciente y ajuste de los algoritmos. Estos dos tipos de neurotecnologías no son excluyentes. Por ejemplo, una segunda generación de DBS, llamada DBS de bucle cerrado (closed-loop DBS o adaptive DBS) se puede considerar como un tipo de BCI, en tanto modifica los parámetros de estimulación cerebral a partir del registro e interpretación algorítmica de patrones de actividad neuronal asociados a la condición que se está tratando por medio del dispositivo. Si bien hay diferentes tipos de neurotecnologías que podrían conllevar desafíos

³ 'Directamente' tiene que entenderse de manera relativa, en oposición a las tecnologías que permiten inferir estados de la mente o el cerebro de manera más indirecta por medio de datos conductuales.

similares, en adelante, cuando hablemos de neurotecnologías, nos referiremos particularmente a BCIs invasivas⁴.

En la interacción entre ser humano y neurotecnología se genera un sistema acoplado mente-máquina en el que el procesamiento neurocognitivo del usuario es detectado, modulado, o ambas cosas a la vez, por el dispositivo. En la literatura filosófica reciente, este acoplamiento ha sido caracterizado en términos del dispositivo formando parte de la mente de sus usuarios (por ejemplo, Heersmink 2015; Heinrichs, 2021) de acuerdo con un marco basado en la teoría de la mente extendida (Clark y Chalmers, 1998). Este enfoque enfatiza el daño que se puede provocar al usuario al manipular, dañar o hackear el dispositivo. Así, el debate ha girado en torno al estatus ético y legal de los dispositivos neurotecnológicos. Específicamente, quienes utilizan el marco de la mente extendida en el contexto de la discusión en neuroética han abordado este estatus desde la noción de paridad ética, esto es, la idea de que cuando existe un fuerte acoplamiento, deberíamos considerar un daño al dispositivo como un daño al núcleo de la persona, es decir, a su integridad mental, y no a su propiedad.

Aun cuando consideramos que este análisis llega a conclusiones correctas respecto de la magnitud del daño que se puede causar a un usuario por un ataque al dispositivo con el que se encuentra acoplado, sostenemos que esta manera de aproximarse al problema presenta dificultades al enfrentarse a tres desafíos adicionales vinculados con la agencialidad del sistema acoplado. Antes de referirnos a cada una en particular, es conveniente indagar en la importancia de dicha agencialidad.

El concepto de persona tiene desde sus orígenes una importante carga forense, es decir, se vincula esencialmente con la posibilidad de atribuir acciones a una unidad sustancial que actúa como autora de ellas. De este modo, John Locke (1694/1959, 2.27.26) define a una persona como una unidad a la que pueden atribuirse las acciones y sus respectivos méritos y culpas. Por ello, si queremos hablar de una extensión del usuario, aun cuando es fundamental conceptualizar el carácter del daño que pueden causar terceros al sistema acoplado mente-máquina, es también importante considerar este sistema en tanto que autor de acciones intencionales, es decir, como agente (Anscombe, 1957; Davidson, 1963, 1971; Schlosser, 2019).

⁴ El lector podrá preguntarse hasta qué punto hablamos aquí de una recuperación terapéutica y hasta qué punto es lícito añadir capacidades no-naturales al ser humano. El debate sobre la permisibilidad del aumento tecnológico de las capacidades, tanto físicas como psicológicas, la distinción precisa entre terapia y mejora, y sobre el transhumanismo, quedan más allá del alcance del presente estudio (para una revisión relativamente reciente de dicho debate véase Hofmann [2017]). Así, aquí no tratamos de dar una valoración normativa sobre cuándo realizar implantes, sino sobre cuál es el estatus moral de la persona ya implantada.

La primera limitación concierne a la distinción entre paciencia y agencia moral (Regan, 1985; Shevlin, 2021). En el presente artículo, consideraremos una noción mínima de paciencia: la idea de que un paciente moral es todo aquel que es sujeto de consideración moral. Por ejemplo, podríamos considerar como sujetos de consideración moral a todas las entidades que sean capaz de sufrimiento, dolor, placer y/o felicidad, y que son tales que nuestras acciones pueden influir sobre estos estados. Adicionalmente, consideraremos que todo agente es sujeto de consideración moral. Así, todo agente moral es paciente moral, pero no a la inversa. Podríamos considerar que ciertos animales no humanos no pueden ser considerados dignos de castigo o recompensa, que no son agentes responsables, pero que sin embargo su bienestar tiene que ser tenido en cuenta cuando analizamos las consecuencias éticas de nuestras acciones. Los enfoques neuroéticos⁵ de mente extendida, enfocados en la paridad ética, y en el posible daño al sistema acoplado, se centran en la paciencia moral de este. Pero, en importantes casos este sistema es, como un todo, autor de acciones intencionales.

La segunda limitación concierne al sentido de pertenencia corporal que tiene una correlación directa con el sentido de agencia (Gallagher y Zahavi, 2021). El sentido de agencia es lo que nos hace sentir que una acción es provocada por nosotros mismos. Los enfoques neuroéticos de mente extendida sobre el acoplamiento no tienen en cuenta esta importante dimensión que plausiblemente podríamos aplicar a los dispositivos acoplados.

Finalmente, existe la posibilidad de que el dispositivo actúe de manera coercitiva sobre los mecanismos neurocognitivos de toma de decisiones del usuario. En este caso, resulta razonable pensar que la agencia sería sustituida por el dispositivo en vez de ser mejorada. En dicho caso, el sistema acoplado no constituiría un agente y, por tanto, no podríamos hablar de una persona extendida, sino de una persona anulada.

Recientemente, Gilbert, Ienca y Cook (2023) han propuesto entender la integración en términos de una simbiosis humano-máquina. Dicho artículo analiza un caso en el que una persona que tuvo un implante exitoso de una BCI, cuya función era detectar una futura crisis de epilepsia y notificarle para que tomara medidas adecuadas. Al cabo de un tiempo, este dispositivo fue explantado contra su voluntad debido a problemas económicos del fabricante. Estos autores proponen

⁵ El calificativo 'neuroético' hace referencia a que estamos hablando específicamente del modo en cómo el marco extendido es entendido por los autores que buscan extraer sus implicaciones éticas. Podría ser que el marco mismo –tal como se entiende en filosofía de la mente– sea consistente con una perspectiva que enfatice las dimensiones éticas vinculadas a la agencia. Agradecemos a uno de los evaluadores por señalar este punto. En adelante cuando hablemos de mente extendida nos referiremos a su uso en el contexto de la neuroética.

que una explantación de un dispositivo integrado exitosamente en el procesamiento cognitivo de una persona sin su consentimiento y sin una razón médica de peso atacaría los (neuro)derechos fundamentales a la continuidad psicológica y a la integridad mental del nuevo agente simbiótico formado por el cuerpo orgánico y el dispositivo. Coincidimos en este punto, pero agregamos que la vulneración de esos derechos se da incluso si la explantación fuese hecha por una buena razón médica. Que la intervención busque restaurar la salud del paciente no excluye que tenga un impacto negativo sobre aspectos de su mente y su cuerpo que resultan constitutivos de su identidad. Además, que la intervención sea en contra de su voluntad –con buenas razones o no– vulnera su integridad psicológica, entendida como una forma de autonomía que el sujeto debe tener sobre sus procesos mentales (Wajnerman Paz *et al.*, 2024)⁶.

El enfoque de estos autores se centra en el daño causado en el usuario producto de la explantación. Sin embargo, de un modo muy original e interesante, introducen la expresión ‘persona fusionada’ y se enfocan en las nuevas capacidades agenciales habilitadas por la implantación y luego perdidas en la extracción. Propondremos que la noción de persona fusionada debe ser rigurosamente analizada en el marco del debate sobre las teorías de la identidad que abordan la llamada ‘pregunta por la reidentificación’. Asimismo, sostenemos que ninguna de las dos teorías dominantes en este debate, el marco de la continuidad psicológica y el marco de la continuidad biológica, abordan adecuadamente esta tarea, pues se necesita un enfoque híbrido diferente de los propuestos recientemente.

El artículo se desplegará del siguiente modo. En la segunda sección describiremos el modo en que el sistema acoplado es pensado por el marco neuroético de la mente extendida y por el marco de la simbiosis humano-máquina. En la tercera sección profundizaremos sobre las tres limitaciones antes descritas referentes a la neuroética de la mente extendida. En la cuarta sección introduciremos elementos del debate acerca de las teorías de la identidad que responden a la pregunta por la reidentificación. Finalmente, en la quinta sección mostraremos, de manera preliminar, qué pueden ofrecer las teorías de la identidad a la hora de pensar la noción de persona fusionada en el contexto de la integración entre el ser humano y neurotecnología.

6 Agradecemos a uno de los evaluadores por esta observación.

2. Marcos actuales

La integración entre el ser humano y la neurotecnología ha sido abordada desde dos puntos de vista complementarios en la literatura neuroética reciente. Ambos tienen como foco el respeto por la persona y la protección del sistema acoplado como un todo. Ya sea como mente extendida o como persona fusionada (o simbiote), estos marcos conceptuales resaltan que la integración tiene el carácter de un lazo íntimo en que se constituye un todo en que el procesamiento cognitivo es coejercido. Sin embargo, este segundo marco tiene la ventaja de considerar no solo aspectos de la paciencia moral, como en el caso del primero, sino que también las nuevas capacidades agenciales del sistema como un todo. Aun así, persiste la necesidad de dar una mayor precisión filosófica a la noción de persona fusionada.

2.1. Integración como mente extendida

El marco de mente extendida (Clark y Chalmers, 1998) se fundamenta sobre la idea de que en muchas ocasiones nuestras funciones cognitivas emplean elementos del entorno durante su ejecución. Por ejemplo, cuando realizamos una multiplicación de muchos dígitos, utilizamos lápiz y papel para desplegar un algoritmo que analiza dicha operación en multiplicaciones de dígitos individuales registrando cada uno de los resultados para su posterior adición. Sin el lápiz y el papel, la memoria quedaría sobrecargada y difícilmente podríamos concretar dicho cálculo. Ellos constituyen, en algún sentido, un dispositivo de memoria de trabajo extendida durante el desarrollo de dicha tarea.

Ahora bien, un criterio demasiado permisivo podría integrar en nuestro procesamiento cognitivo y, con ello, en nuestra mente excesivos elementos del entorno. Así, distinguir qué forma parte de nuestra mente y qué no requiere establecer criterios precisos de inclusión y exclusión. Clark y Chalmers proponen cuatro criterios para evaluar si un ítem es parte de una mente extendida. El primero es la constancia en el uso. El segundo, que la información representada en el ítem esté disponible directamente y sin dificultad. El tercero, que la información obtenida del ítem sea automáticamente aprobada como fiable. Finalmente, que la información contenida en el ítem haya sido conscientemente aprobada en el pasado y que esté ahí como consecuencia de dicha aprobación. Estos criterios han sido subsecuentemente debatidos y elaborados.

En neuroética algunos autores, entre los que podemos destacar a Heermink (2015) y de Heinrichs (2021) han intentado extraer consecuencias éticas de la

extensión de la mente. De acuerdo con Heinrichs, un dispositivo que no forma parte del cuerpo orgánico podría de todos modos ser un ‘par ético’ de una parte del cuerpo o de la mente si es que satisface los criterios para ser una extensión. Esta paridad ética es entendida como un cierto estado en el que deberíamos considerar un daño al ítem como un daño a su mente, entendido como uno de los núcleos de la persona, protegido por el derecho a la integridad mental (Wajnerman Paz *et al.*, 2024), y no a su propiedad. Dado que en la discusión neuroética el foco está puesto principalmente en las neurotecnologías, en adelante usaremos la expresión dispositivo en reemplazo de ítem. Esta última locución resultaba más adecuada dentro de los ejemplos de Clark y Chalmers que incluyen elementos como cuadernos y mapas.

Heermink (2015) desarrolla los cuatro criterios descritos en ocho dimensiones de evaluación. Así, para que un dispositivo forme parte de una mente extendida, debería ranquear alto en todas ellas. Estas dimensiones son 1) flujo de información, 2) consistencia, 3) durabilidad, 4) confiabilidad, 5) transparencia procedural, 6) transparencia informacional, 7) individualización y 8) transformación.

El flujo de información alude a la cantidad de información intercambiada entre el dispositivo y el organismo. La consistencia se refiere a la capacidad de un dispositivo de funcionar adecuadamente en contextos variados. La durabilidad concierne a la capacidad del dispositivo para ser usado en múltiples ocasiones. La confiabilidad está determinada por si tenemos una actitud positiva respecto del valor de verdad de la información que el dispositivo nos provee. La transparencia procedural refiere a que el dispositivo no requiera un gran esfuerzo para su uso y no requiera atención consciente. La transparencia informacional alude a que la información provista por el dispositivo sea fácilmente interpretable y comprensible. La individualización es la conexión con las necesidades y deseos específicos del usuario. Finalmente, la transformación consiste en la capacidad de aumentar las capacidades representacionales y cognitivas del cerebro.

Heinrichs sostiene que, aunque estas dimensiones son una herramienta evaluativa valorable, se postulan teniendo en mente un tipo de tecnología específica que no se ajusta al alcance del debate actual en neuroética. Así, un alto ranqueo en las dimensiones de flujo de información, transparencia informacional y la confiabilidad requiere que los dispositivos en cuestión porten información representacional que debe ser transada con el usuario a nivel personal. De acuerdo con Heinrichs (2021), este requisito amenazaría la unidad teórica de la neuroética, pues incluiría dentro de la mente extendida a dispositivos irrelevantes en su desarrollo –tales como los teléfonos inteligentes– y dejaría fuera dispositivos altamente debatidos, como los de DBS. Nos referiremos brevemente a las razones que sostienen esta última afirmación.

Los dispositivos de DBS estimulan pequeñas áreas de tejido subcortical con el objetivo de modularlas y restablecer su normal funcionamiento. Su interacción de grano grueso parece actuar allí sobre la dinámica de los neurotransmisores (Alosaimi *et al.*, 2022). De este modo, los dispositivos no actúan transfiriendo información a nivel personal y, por ello, las tres dimensiones descritas en el párrafo anterior ni siquiera pueden aplicarse a esta clase de dispositivos.

Consecuentemente, Heinrichs sostiene que la dimensión de flujo de información es problemática, porque presupone un tipo de tecnología computacional y representacional y un tipo específico de transacción informacional. Podemos añadir que esta dimensión no considera siquiera la posibilidad de una transacción informacional a nivel de procesamiento neurocognitivo subpersonal que afecte las representaciones neuronales (Thomson y Piccinini, 2018; Rusanen *et al.*, 2021). Más aún, añade Heinrichs, los dispositivos neurotecnológicos usados con fines terapéuticos tendrían una evaluación negativa de acuerdo con la dimensión de transformación, porque restauran las capacidades cognitivas y no añaden nuevas capacidades.

Heinrichs propone que 1) eliminemos la dimensión de flujo informacional y transparencia informacional; y 2) amplíemos la definición de la confiabilidad para incluir la que concierne al aporte del dispositivo a la ejecución de las funciones cognitivas.

Sin embargo, ambas aproximaciones presentan problemas para pensar la integración con las BCIs. Primero, están basadas en una noción de computación y representación que excluye los procesos subpersonales. Segundo, evalúan casi exclusivamente la dimensión de flujo desde el dispositivo al usuario y no en la dirección inversa. Como vimos, en el caso de las BCIs, el acoplamiento plausiblemente depende de la capacidad del dispositivo de detectar e interpretar las señales producidas por el cerebro. Más importante, consideran solo las consecuencias de la integración en términos de una paridad en paciencia moral, lo que limita las posibilidades de pensar acerca de las acciones en que participa el sistema acoplado como un todo. Las tres limitaciones que analizaremos en la tercera sección se derivan de este último aspecto.

2.2. Integración como simbiosis humano-máquina

Recientemente, Gilbert, Ienca y Cook⁷ han propuesto entender la integración en términos de una simbiosis humano-máquina que genera una persona fusionada. Su propuesta deriva del análisis de un caso y las consecuencias éticas –y jurídicas– que supone. Refirámonos en primer lugar al caso.

La paciente R participó en las pruebas experimentales tempranas de un dispositivo BCI implantado quirúrgicamente por la empresa Neurovista. El objetivo del dispositivo era alertar a pacientes epilépticos cuando están prontos a sufrir una crisis. Para ello, utilizaba un algoritmo ajustado de manera personalizada basado en una inteligencia artificial encargada de la predicción. El ajuste personalizado se realizaba mediante el entrenamiento de la red neuronal de dicha inteligencia artificial. El dispositivo funcionó de manera muy precisa, en el caso de la paciente R afectando su sentido de identidad y aumentando sus capacidades agenciales. Lamentablemente, el dispositivo resultó tener un bajo éxito predictivo en el resto de los pacientes. Finalmente, la empresa entró en estado de quiebra y el dispositivo de la paciente R fue explantado contra su voluntad y sin mediar una razón médica. Los autores recabaron relatos de la paciente a lo largo de seis años.

Entre los relatos de la paciente R que los autores reportan quisiéramos destacar los siguientes. Antes de la explantación, es decir, cuando la paciente hacía uso efectivo del dispositivo es posible observar una sensación tanto de unidad: “Nos enlazaron quirúrgicamente, y nos unimos al instante. Con la ayuda de la ciencia y los técnicos, ambos nos convertimos en uno”; como de mayor capacidad agencial: “Podía conducir, quedar con gente, era capaz de tomar mejores decisiones”. Después de la explantación, la paciente reporta asimismo tanto pérdidas en su identidad –“Diría que perdí, hasta cierto punto, el sentido de mi propia identidad”– como en sus capacidades agenciales –“Nunca he vuelto a sentirme tan segura y protegida, ni soy la mujer feliz, extrovertida y confiada que era”–.

Conviene explorar la noción de simbiosis subyacente al análisis de los autores, quienes la entienden como la ayuda mutua en las metas de cada uno. El dispositivo tiene teleología derivada de haber sido desarrollado para una tarea específica. Así, una ayuda al dispositivo es conceptualizada como una contribución para que ejerza de mejor manera su tarea. Así, mientras el dispositivo ayuda a la paciente R a adquirir la capacidad de realizar una mayor variedad de acciones –al usarlo no debe preocuparse de las consecuencias de una crisis en un momento imprevisto–,

⁷ Lamentablemente, estos autores no toman en consideración el debate descrito en la sección 2.1. sobre mente extendida y paridad ética. El cruce entre ambas propuestas nos parece relevante para pensar el problema de la integración entre lo humano y la neurotecnología.

el dispositivo se beneficia y que optimiza la función para la que fue diseñado (sobre la relación entre teleología y eficiencia, véase Fuentes 2023), mediante su ajuste sobre la base de los datos que obtiene durante su uso.

Ahora bien, una diferencia crucial entre el enfoque de mente extendida (véase sección 2.1) y el presente enfoque es que este último no parece atribuir un estatus ético al dispositivo acoplado a una persona, sino que lo que se busca proteger son los cambios que sufre la psicología de la persona misma mediante la implantación y explantación de él. Para ello, Gilbert, Ienca y Cook (2023) manifiestan adherencia a la perspectiva de continuidad psicológica respecto de la identidad personal. Por ello, cuando el dispositivo es explantado contra la voluntad de la paciente y sin una razón médica, sino que económica, los autores sugieren que habría una vulneración del (neuro)derecho a la continuidad psicológica. La persona fusionada es, de este modo, entendida como una persona que cambia su modo de interactuar con el mundo en el contraste presencia versus ausencia del implante.

Para finalizar esta sección, argumentaremos muy brevemente por qué creemos que las dos aproximaciones descritas hasta ahora no parecen suficientes. La principal razón es que consideramos necesario integrar las cualidades normativas positivas de ambas teorías en un único marco. Por un lado, nos parece que el enfoque de mente extendida tiene a su favor considerar el dispositivo un coejecutor de las funciones cognitivas en conjunto con los mecanismos neurocognitivos (Piccinini, 2020) del usuario. Por otro lado, el enfoque simbiótico toma en cuenta el hecho de que la identidad personal se ve afectada por la implantación o explantación de estos dispositivos: esta afectación está influida por la agencialidad mediada por estos dispositivos. En la sección 5 propondremos que estas virtudes solo pueden ser integradas mediante una aproximación particular a la identidad personal que supere el hiato entre continuidad psicológica y continuidad biológica.

3. Tres limitaciones del marco de mente extendida

El marco de mente extendida en la discusión neuroética posee una riqueza conceptual mayor que el de simbiosis debido a que su discusión ha sido más extendida en el tiempo, y han contribuido diversos autores con distintas miradas (por ejemplo, Anderson, 2008; Heerminsk, 2015; Heinrichs, 2021; Levy, 2007). Debemos recordar que el enfoque de simbiosis es sugerido por un artículo que estudia un caso específico y no posee una formulación independiente. Por ello, estudiar en detalle las limitaciones del primero resulta más factible, aun cuando al final de la sección anterior se ha dejado ver, porque consideramos la necesidad de una aproximación que vaya más allá de aquella de la simbiosis.

Las limitaciones que consideraremos en esta sección tienen en común el hecho de que el marco de mente extendida tal como es aplicado en neuroética no profundiza en las capacidades agenciales del sistema extendido, sino solo en la protección de las personas a través de la protección de los dispositivos que forman parte de su mente.

3.1. Paciencia y agencia moral

Desde la ética que estudia nuestra relación con los animales no humanos, se ha planteado la necesidad de distinguir entre ser causante de acciones morales en tanto que sujeto racional competente, y ser sujeto de consideración moral (Regan, 1985). Incluso cuando no seríamos capaces de culpar a un animal por un comportamiento que cause algún tipo de daño, dado que este se mueve solo por instinto, existe consenso respecto de que ciertos tipos de acciones constituyen maltrato animal. De este modo, paciente moral es definido como aquel que es sujeto (o eventualmente objeto) de consideración moral, teniendo como paradigma a los animales no humanos, y agente moral es el que realiza acciones sujetas a evaluación moral, esto es, acciones intencionales (Anscombe, 1957; Davidson, 1963, 1971), teniendo como paradigma al ser humano. Naturalmente, todo agente moral es paciente moral pero no a la inversa.

La noción de paridad ética, levantada en la aplicación del marco de mente extendida en neuroética, resalta el aspecto de la paciencia moral del dispositivo. Este debe ser objeto (no sujeto) de consideración moral, dado que forma parte de la mente de una persona usuaria. Además, el dispositivo no es cualquier paciente moral, pues un ataque a este (destrucción, hackeo, toma de control, etc.) es un ataque a una persona. Esto porque, el dispositivo está, *ex hypothesi*, a la par con las partes del cuerpo orgánico del usuario. Consideramos que esta aproximación es adecuada, dado el tipo especial de daño que constituiría un ataque a estos dispositivos cuando la relación con el mundo de una persona es mediada por estos dispositivos.

Sin embargo, una persona no es solo un paciente moral, sino que además es sujeto de acciones intencionales (véase la relación entre atribución de acciones y la persona como unidad en la cuarta sección). Por ello, ser parte de la mente de la persona no tiene solo consecuencias en términos de protección a la entidad participante, sino que debe estudiarse el modo en que se integra en su agencia, habilitando capacidades agenciales no ejecutables previamente, y el tipo de responsabilidad que emana de las acciones generadas por un sistema acoplado.

Estimamos que este énfasis en la paciencia moral podría estar facilitado por la manera en la que el marco se adaptó para ser aplicado a neurotecnologías como

las discutidas. Los ejemplos clásicos de Clark y Chalmers incluían casos donde los agentes interactúan de manera fluida y automática, pero frecuentemente a nivel personal, con dispositivos o ítems del entorno (como un cuaderno o una calculadora) sobre los que (potencialmente) podrían ejercer cierto control reflexivo o epistémico. En este tipo de escenarios es más factible pensar cómo los ítems extienden la agencia de los sujetos. En cambio, el intercambio informacional subpersonal y bidireccional con la neurotecnología parece dejar a los sujetos en una situación de mayor pasividad respecto del dispositivo. De hecho, la relativa autonomía de los dispositivos que tienen componentes computacionales, como las BCIs, involucran frecuentemente algún tipo de brecha de responsabilidad (Santoni de Sio y Mecacci, 2021). Si, por ejemplo, el algoritmo que permite que una persona parapléjica manipule un brazo robótico tiene que tomar decisiones respecto de cómo interpreta y responde a los comandos de esta, puede haber dificultades en asignar responsabilidad cuando ocurre un accidente que resulta de los movimientos del brazo. Sin embargo, consideramos que incluso en estos casos hay situaciones donde es razonable atribuir agencia (y responsabilidad) al sujeto, para lo cual necesitamos algún criterio que nos permita determinar cómo estos dispositivos pueden extender su agencia.

3.2. Sentido de pertenencia corporal y sentido de agencia

El sentido de pertenencia corporal (*body ownership*) se refiere al modo particular en que experimentamos el propio cuerpo como algo que me pertenece (Gallagher, 2000). Por su parte, el sentido de agencia (*sense of agency*) se refiere a la experiencia de tener el control sobre las propias acciones. A primera vista parece que podemos disociar analíticamente ambos tipos de experiencia, por ejemplo, en situaciones donde el cuerpo es sentido como propio, aunque sea movido de manera pasiva por otro agente. Sin embargo, hay autores que argumentan que hay una íntima relación entre ambos en la medida en que el control que experimentamos sobre el cuerpo es uno de los factores constitutivos de un sentido de pertenencia corporal pleno (Tsakiris y Haggard, 2005). En contraste, otras propuestas afirman que estos sentidos dependen de mecanismos completamente diferentes (Tsakiris, 2015).

Independiente de qué postura adoptemos sobre la relación entre estos dos sentidos, lo cierto es que se trata de dos maneras, al menos conceptualmente, diferentes en las que nos relacionamos con nuestro cuerpo, que podrían ser constitutivas de la paciencia y la agencia moral. Según algunos enfoques, el sentido de pertenencia corporal determina fenoménicamente el límite de aquello que me constituye como paciente moral. El cuerpo sentido, experimentado o vivido es caracterizado por algunos autores como aquello que determina el alcance de la protección de la

persona frente a intervenciones externas no consensuadas. Por ejemplo, Buller (2013) sostiene que el sentido moral de persona está estrechamente relacionado con la noción de persona como sujeto de experiencias y esto, a su vez, con los aspectos sensoriales y somatosensoriales del cuerpo. Por su parte, el sentido de la agencia podría considerarse como un aspecto constitutivo de nuestra capacidad como agentes morales. Como han señalado Zuk y Lazaro-Muñoz (2021), la concepción experiencial de la autonomía y la agencia afirma que la experiencia consciente de la regulación de la motivación y la acción es necesaria para la agencia autónoma.

Si aceptamos esta distinción entre aspectos fenomenológicos de la paciencia y la agencia moral, entonces un enfoque que solo da cuenta de la manera en la que los dispositivos se podrían integrar al cuerpo y la mente por medio de una experiencia de pertenencia, de “ser uno” o estar fusionado con el dispositivo, dejarían de lado el tipo de experiencia que permite que el dispositivo constituya parte constitutiva de nosotros en tanto agentes.

3.3. Estimulación coercitiva

Si nuestras intenciones, al momento de realizar una acción, son modificadas de algún modo por un dispositivo externo, sin mediar nuestro asentimiento, entonces las acciones que ejecutamos no pueden entenderse como emanando de nuestras intenciones y, por tanto, no pueden manifestar nuestra agencia. Si esto sucede con una frecuencia regular o en eventos importantes, entonces hemos perdido la capacidad de atribuirnos nuestras propias acciones, un aspecto clave de nuestra cualidad de persona. Así, nuestra propia condición de persona podría verse amenazada⁸. De ser así, difícilmente podríamos hablar de una mente (o de una persona) extendida.

Este tipo de consideraciones han sido analizadas en conexión con la permisividad ética de ciertas formas de estimulación cerebral profunda, específicamente los llamados DBS de bucle cerrado (*closed-loop DBS*) que regulan automáticamente los parámetros de estimulación cerebral sobre la base de la interpretación de las señales cerebrales que tienen la función de modular. Específicamente, Wajnerman Paz (2023) argumenta que hay ciertas aplicaciones de este tipo de DBS que podrían anular la agencia y la autonomía de los pacientes. Se trata de aplicaciones pensadas para sustituir (en lugar de apoyar) capacidades cognitivas que plausiblemente son necesarias para la agencia, como el control cognitivo. El contraste se da, por

⁸ No en cuanto a nuestro ser sujeto de derechos humanos o nuestra posesión de dignidad humana, sino en tanto a la actualización de las capacidades racionales que nos permiten decidir nuestros propios cursos de acción.

ejemplo, en aplicaciones que estimulan el cerebro para restaurar su capacidad para la autorregulación emocional o aplicaciones que se encargan de modular las emociones del usuario sin ningún tipo de control de su parte. Por supuesto, un análisis de este tipo debe sopesar tanto la necesidad terapéutica como la reducción de agencialidad caso a caso.

4. Las teorías de la identidad personal como respuesta a la pregunta por la reidentificación, ¿qué tipo de aproximación se necesita?

Si la mente extendida por un dispositivo es la mente de una persona, o si la simbiosis humano-máquina genera una persona fusionada, entonces, este sistema acoplado debe ser capaz de contener las propiedades que actualiza la clase de las personas. Así, este sistema debe constituir una unidad racional y prudencial, capaz de actuar sobre la base de sus propias inferencias prácticas, a la vez que se constituye como sujeto de protección normativa, tanto él como sus partes físicas propias (en nuestro caso, cuerpo orgánico más dispositivo neurotecnológico). Según hemos expuesto en la sección 2, los marcos actuales capturan este último aspecto. El marco de simbiosis tiene la ventaja adicional de que reflexiona además sobre las nuevas capacidades⁹ agenciales de la persona fusionada. Sin embargo, este marco tampoco provee una aproximación completa a la responsabilidad agencial del simbiote (sección 3.1), al sentido de agencia que pudiera o no conferir una prótesis neurotecnológica (sección 2.2), ni a si el sistema genera mermas en la capacidad de actuar intencionalmente (sección 3.3).

El motivo de esto último, es que el marco de simbiosis está construido sobre una teoría de continuidad psicológica de la reidentificación personal, que es impotente a la hora de decidir si algo físico forma parte de una persona. En la presente sección nos referiremos a los enfoques de continuidad psicológica, de continuidad biológica e híbridos sobre la reidentificación personal. Mostraremos que es necesario un nuevo enfoque y plantearemos qué características debería tener.

Primero, ¿en qué consiste una teoría de la identidad personal como respuesta a la pregunta por la reidentificación? Schechtman (1990) identifica que hay dos problemas distintos cuando queremos investigar la identidad personal. Por un lado, el

⁹ Nuevas capacidades respecto del estado previo a la cirugía. En ciertos casos terapéuticos, cuando la persona adquiere una discapacidad debido a un evento posterior al nacimiento, como un accidente, estas capacidades no son nuevas del todo, sino una restauración de capacidades que se tenían previamente (revisar también la nota 4 del presente artículo).

problema de la caracterización refiere a los aspectos narrativos que pueden utilizarse como respuestas a preguntas que pidan una justificación o explicación de nuestras acciones desde nuestra visión valórica. Tener una identidad narrativa es poseer una disposición a responder estas preguntas desde una narrativa estructurada y teleológica que manifieste nuestro esquema normativo en el largo plazo. Más específicamente, nuestra autonarrativa es una estructura interpretativa de nuestra experiencia y el mundo, un recorte de la misma que resalta ciertos rasgos de nuestro carácter, historia, relaciones, etc., que consideramos centrales o constitutivos de nuestra identidad.

Las teorías de la identidad narrativa han desempeñado un papel importante en los debates en neuroética sobre cómo la identidad puede ser afectada positiva o negativamente por la neurotecnología. Autoras como Mackenzie y Walker (2015) o Goering *et. al.* (2017) han argumentado que los pacientes que poseen implantes neurotecnológicos incorporan estos dispositivos como parte de la narrativa que define su identidad. Estas narrativas no son arbitrarias, sino que pueden estar sujetas a exigentes requisitos para ser consideradas válidas o constituyentes de la identidad de las personas. Por ejemplo, los enfoques relacionales de la identidad personal sostienen que la narrativa de una persona constituye su identidad cuando se encuentra en una posición de equilibrio con las narrativas que otras personas construyen sobre ella (Baylis, 2013). Si bien consideramos que este tipo de propuestas constituyen un aporte valioso y necesario para pensar la identidad en el contexto de la neuroética, creemos que la discusión de los casos que hemos descrito requieren de un criterio que no sea dependiente de la interpretación del paciente y de sus interlocutores, un criterio que se aplique transversalmente a diferentes personas que con distintas experiencias e interpretaciones respecto del rol que una neurotecnología juega en la constitución de su identidad¹⁰.

Por otro lado, el problema de reidentificación es el problema metafísico, con consecuencias normativas, acerca de qué nos hace ser la misma persona que fue antes y será después en cuanto unidad sustancial. Por los motivos mencionados, nos interesa aquí centrarnos en este tipo de propuestas.

Ahora bien, el ser responsables de nuestras propias acciones es un aspecto fundante de nuestro ser personas. Por ello, desde sus orígenes modernos, el término persona tiene una fuerte carga forense. John Locke explica del siguiente modo:

¹⁰ Cabe mencionar que entre los enfoques narrativos encontramos autoras como Schechtman (2010) que sostienen que las narrativas son válidas cuando cumplen con un criterio de adecuación a la realidad. Este tipo de exigencia sí se alinea más con el tipo de criterio que estamos buscando. Sin embargo, consideramos que cuando tengamos que especificar qué significa la realidad de una fusión o extensión de la mente, lo que obtendremos es simplemente un criterio no narrativo de la identidad.

Dondequiera que un hombre encuentra lo que él llama él mismo, allí, pienso, otro puede decir que es la misma persona. Es un término forense, que se apropia de las acciones y de su mérito; y, por tanto, pertenece tan solo a los agentes inteligentes, capaces de una ley, de felicidad y de miseria. (1694/1959, 2.27.26; trad. propia)

Indagaremos ahora acerca de los requisitos que se han propuesto para reidentificar una unidad de este tipo.

De entre las variantes de la teoría de la continuidad psicológica de la reidentificación personal, a veces llamadas neolockeanas ya que fue Locke el primero en sugerir una teoría de este tipo, la más célebre es la de Parfit:

X en el momento t_1 es la misma persona que Y en el momento t_2 si, y solo si, X es únicamente continuo psicológicamente con Y, donde la continuidad psicológica consiste en cadenas que se superponen de conectividad psicológica fuerte, a su vez esta última consiste en un número significativo de conexiones psicológicas directas tales como memorias, intenciones, creencias/metas/deseos, y similitud de carácter. (1984; trad. propia)

Esta teoría tiene la ventaja, de acuerdo con sus defensores, de funcionar bien en ciertos casos que el sentido común nos llevaría a hablar de una reidentificación personal. Primero, se enfatiza que usualmente consideraríamos que cuando se protege a una persona se le protege del sufrimiento, de condiciones que mermen la posibilidad de desarrollar su proyecto de vida, su libertad de pensamiento, entre otras cosas. Todas ellas tienen en común el estar construidas sobre propiedades psicológicas y no del cuerpo orgánico de la persona.

En segundo lugar, en el experimento mental del cerebro en la cubeta, un cerebro es alimentado y estimulado con falsos perceptos de modo que tenga la vivencia de estar en un mundo natural. Este cerebro difícilmente pueda ser considerado humano ya que carece de la estructura de sistemas orgánicos de un cuerpo humano. Por ello, pareciera ser que la teoría de continuidad psicológica tendría ventaja para analizar este caso sobre las teorías biológicas que identifican a la persona con un animal humano. Este cerebro en funcionamiento posee todas las propiedades psicológicas relevantes para poder planificar una vida en el mundo simulado y es capaz de sentir sufrimiento. Por ello, se considera como una persona no-humana.

En tercer lugar, tenemos el experimento mental en que las neuronas son reemplazadas gradualmente, una a una, por unidades computacionales equivalentes pero inorgánicas. Al final del proceso tendríamos un ser con capacidades psicológicas idénticas a las nuestras, aunque no posea unidades de procesamiento cognitivo humanas, por lo que no sería otro tipo de persona no-humana.

Por último, consideramos que una persona que ha sufrido muerte cerebral ha dejado de existir, aun cuando la tecnología médica permite mantener vivo al animal humano que somos por un tiempo.

En contraste, las teorías de la continuidad biológica plantean que las personas son esencialmente animales humanos (Olson, 1999, 1997). Consecuentemente, esta teoría es a veces referida como animalismo. Así, las propiedades psicológicas no serían relevantes (Olson, 1999) para definir la continuidad de una persona sino la persistencia de su organismo. En consecuencia, esta teoría rechaza la posibilidad de personas no humanas, ya que se equipara a la persona con el animal que pertenece a la especie *homo sapiens*. Dos casos soportan esta teoría. Primero, consideramos que las personas en estado vegetativo siguen siendo ellas mismas, aun cuando no puedan manifestar sus capacidades psicológicas. Segundo, consideramos que una vez nosotros mismos fuimos fetos en el vientre de nuestras madres, aunque careciéramos de estados psicológicos conscientes (Olson, 1997).

Finalmente, las variantes de la teoría híbrida disyuntiva (Langford, 2014; Madden, 2016; Noonan, 2021) se plantean como una salida a la necesidad de incorporar todos estos casos que se consideran exitosos para cada teoría bajo un marco único. Son híbridas en tanto que plantean que la continuidad de las capacidades psicológicas y biológicas son importante para la persistencia personal. Son disyuntivas en tanto plantean que una persona puede subsistir por continuidad psicológica y por continuidad biológica.

En la siguiente sección, mostraremos las razones por las que consideramos que ninguno de estos tres enfoques resuelve el problema planteado en este artículo, y qué tipo de enfoque debe buscarse en cambio.

5. ¿Por qué es necesario avanzar hacia una nueva teoría de la reidentificación? y ¿qué características debería tener?

Según hemos visto hasta aquí, la noción de persona, en su aspecto sustancial, es decir, tal y como es capturada por las teorías de la identidad entendida como reidentificación, tiene una importante carga forense y prudencial (Shoemaker, 2021). Por ello, son una base interesante para pensar en las capacidades agenciales de los sistemas acoplados entre humano y neurotecnología. De esto son conscientes Gilbert, Ienca y Cook (2023), cuando plantean la noción de persona fusionada. Sin embargo, su marco está construido sobre una variante de la teoría de continuidad psicológica. Argumentaremos que ni esta, ni las otras dos explicadas en la sección anterior, son adecuadas para pensar completamente todas las dimensiones asociadas al problema de la integración humano-neurotecnología. Así, sostendremos la necesidad

de una nueva aproximación al problema de la reidentificación. Mostraremos los *desiderata* para una teoría de este tipo y qué virtudes puede ofrecer para pensar la integración.

Empecemos por analizar la teoría de continuidad psicológica. Aun cuando puede dar cuenta de la importancia de los cambios psicológicos discontinuos provocados por los eventos de implantación y explantación, como correctamente concluyen Gilbert, Ienca y Cook (2023), no tiene las herramientas suficientes para ofrecer condiciones estructurales que permitan decidir si un dispositivo es parte integral de la identidad de la persona fusionada. La única condición ofrecida parece ser que su explantación genere una discontinuidad psicológica importante. De ser este el único criterio, entonces cualquier herramienta de asistencia para personas con discapacidad que tengan un efecto psicológico significativo en ellas (por ejemplo, una silla de ruedas que le da a la persona con discapacidad motriz un sentimiento de autonomía y agencia) y cuya ayuda pudiera ser removida podría resultar parte de la identidad de una persona fusionada.

Por supuesto, como sostenemos más adelante, la posibilidad de que un dispositivo que usualmente consideramos que no es parte de la persona (desde la mente extendida) o que no implementa capacidades de la persona (en el caso de la fusión) en realidad sí lo haga, no es algo que podamos descartar *a priori*. Por ejemplo, desde una perspectiva enactiva¹¹ (e.g., O’regan *et al.* (2005) para percepción y Hutto (2005) para una versión radical) tal vez no habría nada que impida considerar a una la silla de ruedas como parte de la mente del usuario, ya que las disposiciones psicomotrices del sujeto son parte constituyente de su mente, su cuerpo es el elemento que determina sus disposiciones psicomotrices, y la silla de ruedas estaría acoplada a ambos al habilitar sistemática y confiablemente tales disposiciones¹². Sin embargo, creemos que este tipo de enfoques deben complementarse con otros criterios para no correr el riesgo de una extensión o fusión trivial del sujeto que podría ser problemática desde una perspectiva ética y legal.

Por ejemplo, consideremos a una persona que está en un hospital conectada a un dispositivo que le permite comunicar sus intenciones y mover un brazo robótico para alimentarse y realizar diferentes tareas cotidianas. Supongamos que el dispositivo cumple con los requisitos para constituir una extensión de la mente del paciente y para considerarlo como fusionado con él. Asumamos adicionalmente que el dispositivo es propiedad del hospital y que depende de su funcionamiento de una red de electricidad suficientemente potente, como

¹¹ El enactivismo es la visión de que la cognición emerge de o está constituida por la actividad sensoriomotora. Esto es, la cognición es generada y especificada a través de la operación de procesos sensoriomotores que atraviesan el cerebro, el cuerpo y el mundo (Ward, Silverman y Villalobos 2017).

¹² Agradecemos a uno de los evaluadores por esta observación.

la que posee el establecimiento. Considerar que el dispositivo es parte de la persona (o que está fusionado con la misma) en este caso es problemático por diferentes razones, de las que podemos mencionar dos. En primer lugar, parece que el enfoque implicaría que la red de electricidad del hospital (sin la cual el dispositivo no funcionará) es parte del individuo o está fusionada con el mismo, lo que es problemático porque de la misma red podrían depender las funciones cognitivas y vitales de otras personas (Bublitz, 2022). Esto se complejiza por un segundo problema: en escenarios de escasez de recursos, los médicos tienen que priorizar el cuidado de ciertos pacientes por encima de otros, y podría darse el caso de que la red del hospital tenga un límite respecto de cuántos dispositivos pueden estar funcionando simultáneamente. Si la red fuese parte de un individuo o estuviese fusionada con él, tomar la decisión de no alimentar el dispositivo con el que está conectado sería equivalente a privarlo de un órgano o de una parte de su persona, lo que resulta absurdo (Bublitz, 2022). En la misma línea, incluso si hubiese posibilidad de restringir teóricamente aquello que forma parte del paciente al dispositivo mismo, esto sería problemático por el mismo motivo. El hecho de que se considera normalmente como propiedad del hospital y no como parte del individuo es lo que legitima que en los mismos escenarios de escasez de recursos los médicos puedan determinar sobre la base de criterios consensuados priorizar su uso por parte de ciertos pacientes por encima de otros (escenarios de triage). Esto, por supuesto, no quiere decir que en ciertas ocasiones los dispositivos, habiendo sido propiedad de empresas o instituciones, puedan devenir parte de una persona o fusionarse con ella. Solo es necesario señalar que necesitamos un criterio más robusto o más restrictivo que la mera contribución a sus capacidades psicológicas y/o biológicas, incluso más restrictivo que el proporcionado por los enfoques neuroéticos de la mente extendida.

¿Qué sucede en el caso de las teorías biológicas? Como vimos, estas no permiten que elementos no orgánicos puedan integrarse en una persona. Por ello, el problema ni siquiera puede plantearse en este caso. La cuestión es entonces que el enfoque psicológico ofrece un criterio demasiado permisivo (no hay un criterio claro para restringir el alcance de la fusión), mientras que el enfoque biológico resulta demasiado restrictivo (elimina *a priori* cualquier posibilidad de fusión). Las teorías híbridas disyuntivas parecen abrazar, al menos en la superficie, el hecho de que las personas están constituidas por cuerpos vivientes que poseen capacidades racionales. Sin embargo, por su naturaleza disyuntiva estas propiedades se presentan como una simple suma (lógica) que nos ocurre como especie particular. No se exploran las capacidades mentales en tanto que constituidas por un cuerpo en funcionamiento. Esto es, se presenta la identidad como determinada por diferentes factores que no guardan relación entre sí. Más allá de no ofrecer una teoría unificada de la identidad personal, lo que consideramos más problemático de estos enfoques es que no ofrecen un criterio

para restringir adecuadamente las extensiones de la mente y evitar el problema de la trivialización de la extensión y de la fusión.

La filosofía de las neurociencias cognitivas, especialmente el nuevo mecanicismo¹³, proporciona un marco interesante para entender por qué las propiedades funcionales y las biológicas podrían estar íntimamente conectadas y, de esta manera, superar el disyuntivismo en los enfoques híbridos. Según los enfoques mecanicistas, las propiedades estructurales de los mecanismos neuronales restringen el rango posible de funciones que un sistema puede implementar, mientras que las propiedades funcionales de un mecanismo (aun siendo múltiplemente realizables), constriñen el rango de estructuras que pueden implementarlas (e.g., Boone y Piccinini, 2016). Asimismo, el enfoque mecanicista señala que una comprensión adecuada de las capacidades psicológicas (esto es, una comprensión que pueda explicarlas en toda su complejidad, por ejemplo, cómo es modulada por diferentes contextos de desempeño) requiere entender su conexión con sus componentes neuronales constituyentes, las actividades de estos componentes y el modo en que están relacionados. Un abordaje puramente funcional, desligado de los mecanismos neuronales subyacentes, podrá, a lo sumo, darnos un primer acercamiento esquemático a dicha capacidad, pero no una comprensión de su funcionamiento efectivo (Craver, 2007).

Extendiendo esta idea a las teorías de la identidad discutidas, se podría decir que incluso si aceptamos que las capacidades psicológicas son componentes fundamentales de la identidad personal, esto no implica que podamos desentendernos completamente de los mecanismos (biológicos o no) que permiten implementar estas capacidades. En la medida en que las capacidades humanas están constituidas por ciertos aspectos de los mecanismos neuronales subyacentes, la pregunta de si un dispositivo logra restituir una capacidad psicológica genuina de la persona no puede responderse solo evaluando sus propiedades funcionales, sin atender a las características estructurales subyacentes del mismo. Por ejemplo, incluso los requisitos cognitivos más específicos proporcionados por los enfoques de mente extendida en términos de las características del flujo de información entre dispositivo y sujeto podrían restringirse más adecuadamente exigiendo que los dispositivos tengan ciertas propiedades estructurales propias de los mecanismos neuronales.

De este modo, una teoría de la reidentificación que permita pensar en el acoplamiento entre humano y máquina debe cumplir (al menos) con los siguientes requisitos:

¹³ Por ejemplo, Bechtel, 2009; Craver, 2007; Francken, Slors y Craver, 2022; Fuentes, 2024; Glennan, 2017; Piccinini, 2020; Wajnerman Paz, 2017.

1. Debe ofrecer condiciones estructurales para evaluar si una unidad externa efectivamente implementa aquellos procesos cognitivos de una persona que son esenciales para su identidad.
2. El dispositivo debe exhibir las propiedades estructurales que hacen posible la emergencia de un genuino sentido de agencia (y no meramente un sentido de pertenencia corporal) en las interacciones con el entorno que estén mediadas por él¹⁴.
3. La actividad del dispositivo no debe ser directamente coercitiva sobre los mecanismos que constituyen nuestra capacidad de toma de decisiones.
4. No se debe descartar *a priori* qué elementos externos puedan formar parte de la identidad de una persona, pero debe ofrecer una restricción razonable que tenga en consideración los intereses de otros agentes y pacientes morales

Una teoría así, contaría con las siguientes virtudes para pensar el problema en cuestión:

- a) Permitiría abordar las tres dificultades presentadas en la sección 3.
- b) Se integraría de modo más orgánico en la ciencia que explica las capacidades mentales, es decir, la neurociencia cognitiva. Superando la oposición entre estructura y función psicológica.
- c) Otorgaría la importancia debida al hecho de que una persona es esencialmente un agente racional.

Para finalizar quisiéramos añadir lo siguiente. Incluso cuando consideramos que estas tecnologías se pueden integrar, bajo ciertas condiciones, en el procesamiento cognitivo y la identidad de la persona, no se integran del todo en sus funciones biológicas. Por ejemplo, estos dispositivos no cumplen con los requisitos básicos de autopoiesis propios de los sistemas cerrados vivos (cf. Villalobos y Razeto Barry, 2020), por lo que pueden considerarse parte de la mente, pero no del cuerpo orgánico vivo de la persona. Esto provee un argumento posterior para mostrar la

¹⁴ Esto se sigue de la tesis mecanicista de que las capacidades y estados psicológicos no son independientes de los mecanismos subyacentes, sino que están constituidos por, al menos, algunas (tal vez abstractas) de sus propiedades estructurales. De hecho, si las propiedades necesarias para producir el tipo de fenomenología o experiencia consciente resulta en última instancia de propiedades demasiado concretas (propiedades fisiológicas y químicas) de los mecanismos neuronales, entonces, cabe la posibilidad de que esta no pueda ser implementada por un dispositivo artificial.

impotencia de la teoría de la continuidad biológica para pensar la integración.

6. Conclusiones

Hemos revisado críticamente el modo en que se ha pensado la integración entre el ser humano y la neurotecnología en los debates neuroéticos. Específicamente, analizamos dos aproximaciones existentes en la literatura especializada, una de ellas, la de mente extendida aplicada al debate neuroético, con una tradición de discusión de casi dos décadas y otra, mucho más reciente, en que toman relevancia las capacidades agenciales habilitadas por la neurotecnología, esta es, la de simbiosis humano-máquina.

Primero, nos enfocamos en los puntos positivos de cada una de ellas. En el caso del marco de mente extendida, concordamos en la protección especial, en sus dimensiones ética y legal, que requiere un aparato neurotecnológico funcional que se integra en nuestro procesamiento neurocognitivo. Sin embargo, consideramos que el argumento de la paridad ética enfatiza solo el aspecto de la paciencia moral del sistema acoplado humano-máquina. Por ello, pierde de vista la importancia del hecho de que un dispositivo de este tipo habilita capacidades agenciales que no existían en el sujeto previo a la cirugía de implante del dispositivo neurotecnológico. Segundo, hemos visto que este marco no pone suficiente énfasis en el sentido de agencia, que es un aspecto fenomenológico fundamental del agente extendido o fusionado. Además, no considera la posibilidad de que un dispositivo pueda generar alguna estimulación coercitiva, en cuyo caso difícilmente podamos hablar de la mente de una persona extendida, dado que algunas capacidades agenciales clave podrían ser capturadas por el dispositivo.

En el caso del marco de simbiosis humano-máquina, hemos valorado positivamente dos aspectos de ella. Primero, el protagonismo que toman las capacidades agenciales habilitadas después de la cirugía de implante. Segundo, la noción de persona fusionada, que centra el debate en torno a que el simbiote humano-máquina es una persona como un todo. Sin embargo, reconocemos que la teoría psicológica de la identidad personal que funda esta propuesta es insuficiente, porque considera al dispositivo como parte del todo tan solo en cuanto su continuidad psicológica es drásticamente afectada por la cirugía de implantación o por la de explantación. Sin embargo, argumentamos que esto ocurre con cualquier dispositivo de asistencia para una persona que lo necesite, incluso una silla de ruedas, de la que difícilmente podamos decir que forma parte de la mente de su usuario.

En consecuencia, hemos propuesto las bases y *desideratas* que debería tener una teoría de la identidad personal como reidentificación que pueda ayudar a conceptualizar la integración entre ser humano y neurotecnología de mejor manera.

Sobre todo, en contraste con los enfoques psicológicos, debe ofrecer condiciones estructurales, esto es, en términos físicos, para evaluar si una entidad externa se integra al procesamiento neurocognitivo del usuario y, en contraste con los enfoques biológicos, no debe descartar *a priori* que elementos que no forman parte de su sistema nervioso puedan formar parte de su identidad.

Agradecemos a los dos revisores anónimos de la Revista de Filosofía de la Universidad de Chile por sus valiosos comentarios que nos ayudaron a mejorar sustantivamente el contenido de este artículo.

Referencias

- Alosaimi, F., Boonstra, J. T., Tan, S., Temel, Y., y Jahanshahi, A. (2022). The role of neurotransmitter systems in mediating deep brain stimulation effects in Parkinson's disease. *Frontiers in Neuroscience*, 16, 998932. <https://doi.org/10.3389/fnins.2022.998932>
- Anderson, J. (2008). Neuro-Prosthetics, the Extended Mind, and Respect for Persons with Disability. En M. Düwell, C. Rehmann-Sutter, y D. Mieth (Eds.), *The Contingent Nature of Life* (vol. 39, pp. 259-274). Springer. https://doi.org/10.1007/978-1-4020-6764-8_22
- Anscombe, G. E. M. (1957). *Intention*. Basil Blackwell.
- Baylis, F. (2013). "I Am Who I Am": On the Perceived Threats to Personal Identity from Deep Brain Stimulation. *Neuroethics*, 6(3), 513-526. <https://doi.org/10.1007/s12152-011-9137-1>
- Bechtel, W. (2009). Looking down, around, and up: Mechanistic explanation in psychology. *Philosophical Psychology*, 22(5), 543-564. <https://doi.org/10.1080/09515080903238948>
- Boone, W., y Piccinini, G. (2016). The cognitive neuroscience revolution. *Synthese*, 193(5), 1509-1534. <https://doi.org/10.1007/s11229-015-0783-4>
- Bublitz, C. (2022). The body of law: Boundaries, extensions, and the human right to physical integrity in the biotechnical age. *Journal of Law and the Biosciences*, 9(2). <https://doi.org/10.1093/jlb/ljac032>
- Buller, T. (2013). Neurotechnology, Invasiveness and the Extended Mind. *Neuroethics*, 6(3), 593-605. <https://doi.org/10.1007/s12152-011-9133-5>
- Clark, A., y Chalmers, D. (1998). The Extended Mind. *Analysis*, 58(1), 7-19. <https://doi.org/10.1093/analys/58.1.7>

- Craver, C.F. (2007). *Explaining the Brain*. Oxford UP. <https://doi.org/10.1093/acprof:oso/9780199299317.001.0001>
- Davidson, D. (1963). Actions, Reasons, and Causes. *The Journal of Philosophy*, 60(23), 685. <https://doi.org/10.2307/2023177>
- _____. (1971). I. Agency. En R. W. Binkley, R. N. Bronaugh, y A. Marras (Eds.), *Agent, Action, and Reason* (pp. 1–37). University of Toronto Press. <https://doi.org/10.3138/9781442656963-002>
- Francken, J. C., Slors, M., y Craver, C. F. (2022). Cognitive ontology and the search for neural mechanisms: Three foundational problems. *Synthese*, 200(5), 378.
- Fuentes, J. I. (2023). Efficient mechanisms. *Philosophical Psychology*, 0(0), 1-24.
- _____. (2024). Computational systems as higher-order mechanisms. *Synthese*, 203(2), 55.
- Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences*, 4(1), 14-21.
- Gallagher, S., y Zahavi, D. (2021). *The phenomenological mind*. Routledge, Taylor & Francis Group.
- Gilbert, F., Ienca, M., y Cook, M. (2023). How I became myself after merging with a computer: Does human-machine symbiosis raise human rights issues? *Brain Stimulation*, 16(3), 783-789.
- Glennan, S. (2017). *The new mechanical philosophy*. Oxford UP.
- Goering, S., Klein, E., Dougherty, D. D., y Widge, A. S. (2017). Staying in the Loop: Relational Agency and Identity in Next-Generation DBS for Psychiatry. *AJOB Neuroscience*, 8(2), 59-70.
- Heersmink, R. (2015). Dimensions of integration in embedded and extended cognitive systems. *Phenomenology and the Cognitive Sciences*, 14(3), 577-598.
- Heinrichs, J. H. (2021). Neuroethics, Cognitive Technologies and the Extended Mind Perspective. *Neuroethics*, 14(1), 59-72.
- Hofmann, B. (2017). Limits to human enhancement: Nature, disease, therapy or betterment? *BMC Medical Ethics*, 18(1), 56.
- Hutto, D. D. (2005). Knowing What? Radical Versus Conservative Enactivism. *Phenomenology and the Cognitive Sciences*, 4(4), 389-405.
- Langford, S. (2014). On What We are and How We Persist. *Pacific Philosophical Quarterly*, 95(3), 356-371.
- Levy, N. (2007). *Neuroethics*. Cambridge UP.
- Locke, J. (1694/1959). *An essay concerning human understanding*, vol. 1. Dover Publications.
- Mackenzie, C., y Walker, M. (2015). Neurotechnologies, Personal Identity, and the Ethics of Authenticity. En J. Clausen y N. Levy (Eds.), *Handbook of Neuroethics* (pp. 373-392). Springer.
- Madden, R. (2016). Human persistence. *Philosophers' Imprint*, 16(17), 1-18.

- Noonan, H. W. (2021). Personal Identity and the Hybrid View: A Middle Way. *Metaphysica*, 22(2), 263-283.
- Olson, E. T. (1999). *The human animal: Personal identity without psychology*. Oxford UP.
- _____. (1997). Was I Ever a Fetus? *Philosophy and Phenomenological Research*, 57(1), 95.
- O'regan, J. K., Myin, E., y NOë, A. (2005). Sensory consciousness explained (better) in terms of 'corporality' and 'alerting capacity.' *Phenomenology and the Cognitive Sciences*, 4(4), 369-387.
- Parfit, D. (1984). *Reasons and persons*. Clarendon Press.
- Piccinini, G. (2020). *Neurocognitive Mechanisms: Explaining Biological Cognition*. Oxford UP.
- Regan, T. (1985). *The case for animal rights*. University of California Press.
- Rusanen, A. M., Lappi, O., Kuokkanen, J., y Pekkanen, J. (2021). Action control, forward models and expected rewards: Representations in reinforcement learning. *Synthese*, 199(5-6), 14017-14033.
- Santoni de Sio, F., y Mecacci, G. (2021). Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them. *Philosophy & Technology*, 34(4), 1057-1084.
- Schechtman, M. (2010). Philosophical reflections on narrative and deep brain stimulation. *The Journal of Clinical Ethics*, 21(2), 133-139.
- _____. (1990). Personhood and Personal Identity. *The Journal of Philosophy*, 87(2), 71-92. <https://doi.org/10.2307/2026882>
- Schlosser, M. (2019). "Agency". Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*.
- Shevlin, H. (2021). How Could We Know When a Robot was a Moral Patient? *Cambridge Quarterly of Healthcare Ethics*, 30(3), 459-471.
- Shoemaker, D. (2021). "Personal Identity and Ethics". Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*.
- Thomson, E., y Piccinini, G. (2018). Neural Representations Observed. *Minds and Machines*, 28(1), 191-235.
- Tsakiris, M. (2015). The relationship between agency and body ownership. Additive or independent? En P. Haggard y B. Eitam (Eds.), *The Sense of Agency*. Oxford UP, 235-249.
- Tsakiris, M., y Haggard, P. (2005). The Rubber Hand Illusion Revisited: Visuotactile Integration and Self-Attribution. *Journal of Experimental Psychology: Human Perception and Performance*, 31(1), 80-91.
- Villalobos, M., y Razeto Barry, P. (2020). Are living beings extended autopoietic systems? An embodied reply. *Adaptive Behavior*, 28(1), 3-13.
- Wajnerman Paz, A. (2017). Pluralistic Mechanism. *THEORIA*, 32(2), 161-175. <https://doi.org/10.1387/theoria.16877>

- _____. (2023). Identifying Relational Applications of Deep Brain Stimulation for Treatment Resistant Depression. *Review of Philosophy and Psychology*, 15, 499-521.
- Wajnerman Paz, A., Aboitiz, F., Álamos, F., y Ramos Vergara, P. (2024). A healthcare approach to mental integrity. *Journal of Medical Ethics*, 50(10), 664-669.
- Ward, D., Silverman, D., y Villalobos, M. (2017). Introduction: The Varieties of Enactivism. *Topoi*, 36(3), 365-375. <https://doi.org/10.1007/s11245-017-9484-6>
- Zuk, P., y Lázaro-Muñoz, G. (2021). DBS and Autonomy: Clarifying the Role of Theoretical Neuroethics. *Neuroethics*, 14(S1), 83-93.