UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias Sociales y Humanidades

Emergencia de Consciencia en Inteligencia Artificial

Lucas Luis Mendieta Córdova

Artes Liberales

Trabajo de integración curricular presentado como requisito
para la obtención del título de
Licenciado en Artes Liberales

Quito, 18 de diciembre de 2019

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

COLEGIO CIENCIAS SOCIALES Y HUMANIDADES

HOJA DE CALIFICACIÓN
DE TRABAJO DE INTEGRACIÓN CURRICULAR

Emergencia de Consciencia en Inteligencia Artificial

Lucas Luis Mendieta Córdova

Calificación:

Nombre del profesor, Título académico **Jorge García Nuñez de Cáceres, Ph.D**

Firma del profesor: _____

Quito, 18 de diciembre de 2019

Derechos de Autor

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Firma del estudiante: _____

Nombres y apellidos: Lucas Luis Mendieta Córdova

Código: 00025363

Cédula de identidad: 171542243-0

Lugar y fecha: Quito, 18 de diciembre de 2019

RESUMEN

Pentti O Haikonen propone un modelo concreto para explicar el concepto de consciencia y como esta se puede manifestar y medir en inteligencia artificial. Mediante una presentación parcial de este modelo se establecerán guías y fundamentos para que el autor pueda esbozar una teoría alterna apoyada en los conceptos y terminología explicados por Haikonen. El objetivo de este documento no es refutar el primer modelo, pero parte de este sí está dedicado a una breve exploración de la problemática de modelos de consciencia antropocéntricos, en los cuales prima la existencia móvil en el mundo físico como factor constructivo de consciencia.

Palabras clave: inteligencia artificial, cuerpo digital, perceptos, lenguaje natural, qualia, antropocentrismo.

ABSTRACT

Pentti O Haikonen's concise model explaining what consciousness is and how it could occur and be detected in artificial intelligence is summarized here in order to establish working definitions of the concepts and their relationships in order to sketch out an alternate theory more in line with the author's worldview. The aim of this document is not to refute Haikonen's model but rather to broach the problematic aspects of models of consciousness in artificial intelligence that rely on machines being able to move and act in the physical world to the detriment of the exploration of alternative models that focus more on a digital existence.

Key words: artificial intelligence, digital body, percepts, qualia, anthropocentrism.

TABLA DE CONTENIDO

INTRODUCCIÓN

La definición de consciencia es un problema que ha abarcado la filosofía prácticamente desde su origen como disciplina. La consciencia de uno mismo en entes no humanos ha sido un punto de enfoque para estudio disciplinado y enfocado por más de un siglo, con el número de tomos incrementando exponencialmente gracias a un mayor alejamiento de la noción de antrohegemonía. Dar el paso requerido para reconocer diferentes grados de consciencia y como las definiciones previas habían sido exclusivas no por rigor y disciplina si no por la utilidad de una construcción narrativa fue un proceso demorado. Ahora es necesario volver a entrar en un ciclo de reevaluación similar a aquel, inclusive un ciclo con el cual puede ir de la mano la liberación animal, ya que el fuego de la creación está ardiendo y los humanos intentan, deliberadamente, crear un ser mecánico con consciencia de si mismo. Este trabajo está interesado en esbozar una propuesta para reconocer la emergencia de la consciencia en inteligencia artificial en máquinas con un tipo de existencia más alejado a la humana, una existencia carente de un cuerpo físico como unidad primaria. Es la opinión del autor que la incapacidad de reconocer consciencia en aquellos entes que difieren notablemente a nivel superficial produjo una actitud nociva hacia los animales cosa que sería beneficioso evitar al momento de crear un nuevo ente. Antes de arrojarse a este tipo de exploración más abierta es necesario plantear un fundamento teórico y una definición operacional de los términos que guiarán los conceptos.

DESARROLLO DEL TEMA

1.0 – Modelo Haikonen

En esta primera parte del texto se presentarán las ideas de Pentti O. Haikonen en lo que concierne a la definición de un modelo de consciencia, junto con los parámetros de detección de esta, y la implementación de este modelo en el tipo de robot que Haikonen plantea como el mejor candidato para desarrollar consciencia. Se dará un breve resumen de estos dos modelos ya que el autor considera que, sin tomar en cuenta las discrepancias en cuanto a la necesidad de interacción táctil con el mundo físico, la definición operativa de Haikonen para conciencia es concreta y completa y por lo tanto útil como un punto de partida para desarrollar sus propias ideas sobre el desarrollo de conciencia en un cuerpo digital (no físico). Primero se expondrá la definición que da Haikonen para conciencia (1.1), luego se presentarán las definiciones de Haikonen sobre los problemas de consciencia, denominados el problema concreto y el problema completo (1.2), después se verán los parámetros de consciencia (1.3), seguido por los parámetros para conciencia en inteligencia artificial (1.4) y para concluir esta sección se dará una muy breve presentación del modelo aplicado que Haikonen propone para su Robot, XCR-1 (1.5).

1.1 Conciencia – qualia y perceptos

Para Haikonen (2012), consciencia quiere decir el tener una experiencia interna subjetiva. La experiencia subjetiva *representa* algo y se trata de algo (Haikonen, 2012, p.20). Dicha experiencia interna está compuesta de perceptos y estos toman la apariencia de qualia. Partiendo de eso, se explica que la conciencia efectivamente es la presencia y flujo de qualia (48). Antes de seguir adelante será útil definir los términos que se presentaron: percepto y qualia. Primero, *percepto*: Haikonen lo define como *"*indicador de la presencia de cierto detalle o característica a escala pequeña*"* (26) Agrupado con otros perceptos llega a constituir la

percepción completa de algo. En el modelo de Haikonen la mente no internaliza el mundo percibido si no que externaliza los perceptos sensoriales (los objetos percibidos) de tal modo que estos parecen ser el mundo externo (24) Y ahora, *qualia* (singular, *quale*): Estas son las cualidades percibidas, tanto del mundo como de las sensaciones corporales (vale la pena recordar que el cuerpo es el acceso al mundo pero no es necesariamente parte de el, aunque tampoco está aparte; la relación cuerpo-mundo es compleja y explicarla no viene al tema de este escrito por el momento), son aquello de lo que está constituida la apariencia de la experiencia interna subjetiva y por lo tanto son las cualidades de los perceptos. (30) Sin ellas la información no pasa de ser mera descripción así que se puede afirmar las qualia *son* experiencia física (33) Se afirma que los contenidos de la mente son *apariencias* internas conscientemente percibidas sobre objetos y condiciones, reales o imaginarios. En esta definición de contenido se puede entender la importancia de las qualia, ya que las apariencias internas no son sobre los procesos neuronales que las cargan ya que eso seria un contenido vacío. El acto de representar se denomina el carácter intencional de la consciencia y el tener consciencia necesariamente implica tener consciencia *de algo* (20) La consciencia, entonces, viene a ser percepción basada en qualia, la presencia de una experiencia subjetiva fenomenal. (53)

Más adelante se volverá a topar el tema de qualia y la relación entre estas y las señales que transportan la información, cosa que Haikonen explica con una metáfora muy puntual.

1.2. Conciencia – el problema concreto y el problema completo

Haikonen logra sintetizar los problemas que emergen al trabajar con consciencia de dos maneras. El primer problema, el llamado problema concreto, es el siguiente: "(…) existe una brecha de explicación entre poder explicar los procesos físicos del cerebro y como la experiencia subjetiva nace de estos*."* (13) Básicamente, el problema concreto reconoce la complicación que existe al explicar como la señales (descargas) en las neuronas se convierten

en información con una carga representativa que excluye del paquete todas las cualidades de sentimiento físico (fuerza de descarga, lugar, ocurrencia en si). El segundo problema de consciencia según sintetiza Haikonen es el problema completo y este es tripartito (226):

I. ¿Cómo sucede que una actividad neuronal se presenta internamente como experiencia subjetiva?

II. ¿Cómo llega a percatarse el sujeto de su contenido mental?

III. ¿Cómo surge la impresión del "yo", de aquel que percibe, razona y desea?

Las respuestas al problema completo que plantea Haikonen al final de su libro son:

I. La consciencia es la presencia de apariencias internas, la mente no percibe la actividad neural tal y como es, más bien la percibe tomándolas como las cualidades del mundo como aparentan ser. (227) Esta respuesta cubre el problema concreto de conciencia, la brecha se cierra porque el aparato perceptivo, la mente, no observa la actividad neuronal porque su sistema no tiene sensores para observarla como una actividad física. Es muy importante recalcar que lo que se está percibiendo en estas señales se lo está haciendo de manera directa y no simbólica, se está haciendo una lectura del efecto que tienen la actividad neuronal sobre la mente, no se está convirtiendo es actividad a un nivel secundario de significación (226). Ahora se retomará la metáfora antes mencionada:

> Considera una situación en la cual estás flotando en el océano. Las olas van y vienen y eres arrojado hacia arriba y hacia abajo. Un observador externo vería la forma de las olas, pero tu experiencia es el efecto que tienen estas olas sobre ti, el movimiento arriba y abajo. (226)

Las qualia son generadas en el cerebro, pero generalmente ese no es lugar donde son percibidas, son externalizadas ya que las señales que las ocasionan no incluyen información sobre su origen, "(…) [la] externalización genera *la apariencia de un mundo externo que puede ser inspeccionado*" (227)

II. Los preceptos son el contenido de la consciencia ya que solamente estos pueden tener una apariencia interna. En el caso de pensamientos e imaginación, estos llegan a ser percibidos conscientemente porque se transforman en perceptos virtuales. Esta categoría virtual tiene qualia limitada que emerge después de que algo que es imaginado o pensado entra en un ciclo de retroalimentación -la retroalimentación la define como "(…) reingreso de señales internas relevantes (…) que transforma los patrones internos de señales en patrones sensoriales de señales" (127) – que devuelve las señales internas al proceso de percepción (227). El proceso de percepción "(…) es la combinación y modulación de información sensorial mediante información interna" (26)

III. La percepción del cuerpo y sus sensaciones junto con la percepción del contenido mental dan lugar al concepto del "yo", de "uno mismo". El hecho que el cuerpo y el contenido mental se mueven con la persona, haciendo demandas de ella para ella, mientras que el entorno cambia con el movimiento permite crear una delimitación entre "uno" y "aquello" (228)

1.3. Procesos conscientes – parámetros y evaluación

Bajo la definición general que Haikonen formaliza para parametrizar la presencia de consciencia el criterio principal para la presencia de esta es el requisito de una *apariencia interna* en la forma de qualia. La importancia de las qualia yace en el hecho que los estados de consciencia deben poderse reportar de manera interior y exterior. En la conceptualización de Haikonen de consciencia no existe un punto crítico de conexiones activas o integración de información que pueda tomarse como un indicador de consciencia. Los estados conscientes por definición son estados que se pueden reportar, de manera interior (a uno mismo) y exterior (52)

A continuación, se presentarán los dos esquemas generales de Haikonen, el primero explica los requisitos que cumple cualquier agente con consciencia (86) y el segundo expone los componentes de "un episodio de consciencia" (54).

Un agente con consciencia necesita:

a. Un sistema somatosensorial
b. Una imagen de cuerpo
c. Una imagen mental de si mismo
d. Una historia personal recordada

El sistema somatosensorial permite que el agente monitoree el cuerpo de manera continua, lo cual se identifica como introspección. La *introspección*, según Haikonen, siempre muestra que las entidades mentales toman la forma de entidades de alguna modalidad sensorial (por ejemplo, el habla interna toma la forma de patrones sonoros y las imágenes internas toman la forma de patrones visuales), así que existe una producción de sensaciones sin sensaciones externas (127) El punto *d.* (historia personal) también se denomina *retrospección,* y solo puede ocurrir si es que el agente con consciencia es capaz de localizarse en el tiempo, cosa que implica una selección de datos temporales. La percepción y la introspección también requieren medios de selección ya que toda información no se puede procesar simultáneamente y junto con el proceso de retrospección ya queda vista la elección deliberada de datos por parte del ente (81)

Un episodio de consciencia tiene los siguientes componentes:

1. Percepción – Un proceso de percepción con posibles qualia debe ser directo y transparente. (74)
2. Atención – Haikonen presenta dos tipos de atención

   Sensorial: esta enfoca el proceso de percepción en objetos *seleccionados*

   Interna: enfoca procesos de pensamiento y recuerdo en un tema seleccionado

(48)

La segunda sería el tipo de atención necesaria para ir construyendo una historia personal, efectivamente la recolección es una instanciación de atención interna

3. Memoria de corto plazo – esto es diferente a retrospección, es una operación más simple que igual requiere un sentido de propiedad desplegado en un período de tiempo, cosa que parcialmente se logra mediante *imaginación* (Haikonen define la imaginación como "(…) la percepción y manipulación mental de acciones y entidades que no están físicamente presentes." (135))
4. Integración de información – La cognición exige la integración de información multisensorial y sensomotriz. Preceptos de diferentes modalidades sensoriales deben poder formar una visión coherente del mundo y dicha visión debe permitir que se tome acción motriz "correcta". Los preceptos deben poder evocar significados y ofrecimientos (potencialidad) y su significación emocional debe ser evaluada. La integración sensomotriz también es requerida para externalizar perceptos, cosa que permitiría la creación de la impresión del mundo observable allí afuera junto con la adquisición de la imagen corporal. (172-173)
5. Reporte – La conclusión de un episodio de consciencia siempre es el reportaje, interior o exterior, ya que solamente de este modo se puede integrar los datos al ciclo de retroalimentación.

1.4 – Parámetros para consciencia en inteligencia artificial

Según Haikonen, la cognición consciente requiere de procesos simbólicos y subsimbólicos (98), las computadoras procesan *representaciones simbólicas* y las redes neurales artificiales tradicionales procesan *representaciones subsimbólicas* (93). Parte del problema es que los requisitos de percepción cognitiva (representación *transparente* y *directa* de información tal que esta preserva los rasgos *amodales* - aquellos que no difieren en el mundo

observable y en la actividad neuronal (38) - de los fenómenos percibidos) no pueden satisfacerse en sistemas simbólicos, pero pueden ser acomodados en sistemas subsimbólicos (125)

Antes de presentar los requisitos y el proceso de cognición en inteligencia artificial de Haikonen es importante definir y explicar las contribuciones y diferencias entre procesamiento de información de modo simbólico y subsimbólico. Debido a que

> El entendimiento e interpretación de un símbolo pide información adicional que debe estar a disposición de quien está interpretando (…) El significado de un símbolo está fijado por la convención, no por un vínculo elemental a los perceptos sensoriales o tal. (94)

Los símbolos representan entidades que no están intrínsecamente relacionados a ellos y que abarcan más que el significado directo de las características que los constituyen, el punto más importante para recordar es que "(…) los símbolos adquieren su significado mediante convención" (125). El pensamiento y razonamiento abstracto están basados en procesamiento simbólico (regresamos aquí a la idea de una economía de recursos mentales) (151), por lo tanto los sistemas simbólicos son "(…) aquellos que digitalizan la información sensorial y la representan numéricamente (…)" (125); los valores numéricos no tienen un significado intrínseco, son totalmente dependientes de un código (convención) interno adicional, y son valiosos porque se pueden combinar entre ellos con facilidad para abstraer y construir ideas más complejas. En la representación subsimbólica, en cambio "[el] significado de una señal sensorial subsimbólica está anclada a la fuente de aquella señal (…) *Las qualia son directas y son experimentadas sin interpretación alguna adicional."* (94) [Énfasis propio]. La representación subsimbólica hace evidente el contenido y significado de las señales detectadas. Haikonen explica este tipo de representación a través del ejemplo de los sistemas que usan

representación distribuida (125) Las representaciones de señal distribuida representan entidades según sus rasgos (106), la combinación de estos (e.g. forma, tamaño, color, textura superficial) dotan a un objeto con su apariencia particular y solamente a partir de este grupo de datos se empieza a describir y se logra identificar aquello que es percibido (105) La forma en que funcionan estos rasgos, según Haikonen, es una construcción vertical de subrasgos a partir de una sola propiedad indivisible (denominada "*propiedad elemental*") que es representada con una "*señal de rasgo*". La señal de rasgo puede ser binaria o continua, en el caso de la primera simplemente se comunica presencia/ausencia, mientras que en la segunda la intensidad de la señal sirve para comunicar el valor de confianza de una observación o su importancia (105). Las representaciones de señal distribuida comunican directamente *qué* se está representando en términos de rasgos elementales y sus combinaciones.

Ahora se presentarán los requisitos adaptados por Haikonen de su modelo de entidades conscientes a cualquier máquina cognitiva de un agente que potencialmente podría tener consciencia (81) junto con lo que él llama el contrainterrogatorio (84), un grupo de preguntas que sirven para evaluar el cumplimiento de estos requisitos:

- Percepción, directa y no simbólica (aquí vale la pena recordar que los perceptos son qualia. Las precondiciones para qualia de máquinas según Haikonen (74):
    1. Tienen la habilidad de presentarse como propiedades aparentes del mundo en lugar de presentarse como los patrones de actividad especial neural que las portan.
    2. Son directas, no están basadas en representaciones simbólicas indirectas.
        - *¿Puede la máquina describir algunas qualia?*
- Contenido Mental
    - *¿Tiene la máquina un flujo de contenido mental que es sobre algo?*
- Introspección

- Atención
- Memoria y retrospección
    - *¿La máquina recuerda su pasado inmediato?*
- Respuesta y reportes
    - *¿Puede la máquina reportar su contenido mental (perceptos, pensamientos, discurso interno, etc.) a si misma y a otros? ¿Reconoce la máquina que es dueña de dicho contenido mental?*
    - *¿La máquina siente dolor? ¿De que manera siente dolor?* (El dolor es un indicador de un desvío negativo de un estándar. Dicho estándar sirve como un punto de referencia para uno mismo).

El último requisito es el que mayor atención recibe por parte de Haikonen, la barra de medida para un robot consciente es la capacidad poder externalizar perceptos sensoriales que no sean de contacto y el poder construir una imagen de cuerpo a partir de sensores de contacto (48), es el ciclo de retroalimentación que lleva el mundo externo al interior, lo abstrae y utiliza esa nueva información para "crear" ese mundo externo nuevamente.

Finalmente, se presenta el proceso de percepción para cognición artificial de Haikonen (126):

Estímulos (fenómenos físicos) → Sensor (*transducción*: conversión de estímulos físicos a en la formas de representación física utilizadas dentro del sistema, una forma común interna (138)) → Formato Interno (señales eléctricas) → Preprocesamiento de Detección de Rasgos (nociones de requisitos, detección de contenido y significado básico a partir de las qualia) → Representación Distribuida (perceptos en bruto) → Combinación de Retroalimentación (*retroalimentación*: atención, contexto, predicción, acertado/desacertado, introspección) → Señales de Perceptos (perceptos oficiales) → Emisión

1.5 – XCR-1

Haikonen diseñó un robot simple (*Experimental Cognitive Robot*, XCR-1) para poder probar su modelo de cognición artificial en un ente que utiliza su arquitectura cognitiva (este documento no indagará en los detalles de la arquitectura cognitiva de Haikonen), midiendo integración sensomotriz junto con acción motriz generada (203). XCR-1 es un robot autónomo (esto no solamente quiere decir que no depende de una computadora más grande si no que también explica que no es un robot dirigido por un programa, así que no depende de un microprocesador), con sensores visuales, auditivos, táctiles, de shock y unos que comunican caricias (por parte de humanos al robot), también tiene brazos y manos capaces de agarrar objetos y está dotado de un léxico hablado bastante limitado y un vocabulario de habla interna igualmente pequeño (203). La arquitectura cognitiva simplificada que tiene XCR-1 le otorga una manotada de funciones (204):

i. Determinación de blancos (target set)
ii. Detección
iii. Identificación verbal
iv. Aproximación y agarre
v. Efectos funcionales de dolor (shock) y placer (estímulo de caricias)
vi. Asociación de valores emocionales con objetos

    En este punto cabe presentar la concepción de Haikonen de las emociones en el contexto de un sistema simbólico y subsimbólico: las emociones tienen que ver con síntomas fisiológicos percibidos conscientemente. (56) y las qualia de un estado emocional serian la combinación de las qualia de las condiciones físicas correspondientes (57)

vii. Motivación a partir de valores emocionales

Haikonen clarifica que el robot no tendría un sentimiento fenomenal de dolor o placer (213), simplemente los utiliza para explorar el comportamiento sistemático (recompensa y castigo) ante señales que comunican daño (un estado que se busca evadir o detener) y beneficio (un estado que se busca sostener o iniciar).

viii. Reportaje verbal de los estados internos del robot (habla interna) mediante un lenguaje natural simple

Aquí Haikonen también integra el concepto de atención, ya que el robot solo es capaz de emitir una palabra a la vez para expresar sus actividades (en respuesta al léxico de sus funciones, como por ejemplo "buscar un objeto") y solamente es capaz de recibir una palabra a la vez, por lo tanto, XCR-1 mantiene una comunicación en serie que depende del cierre de cada instante antes de continuar al siguiente. El propósito primario de que XCR-1 se comunique con si mismo es que logre asentar el sentido de las palabras en su propio marco de uso. (219)

ix. Reconocimiento de habla limitado

x. Aprendizaje verbal limitado

El propósito principal del experimento de Haikonen es la exploración del sistema fenomenal que fundamenta la consciencia, un sistema simple que, con otro contexto y nivel de procesamiento es capaz de generar la riqueza de algo como la consciencia humana (224).

2.0 – Nuevo Modelo

En esta segunda parte del texto el autor esbozará su propuesta para una teoría sobre la emergencia de consciencia artificial que depende primariamente en las cualidades del lenguaje natural y la interacción con los seres conscientes que ya lo utilizan. La teoría del autor propone que un experimento como XCR-1 es útil para explorar los mecanismos básicos de consciencia (tal y como dice Haikonen) pero que la dependencia de un cuerpo físico y móvil en este es una limitación antropocéntrica que intenta recrear el nacimiento de la consciencia humana. El autor

propone que una interacción con el entorno físico es un primer paso necesario, pero que este no debe ser una imitación del humano en el mundo físico, para una máquina los sensores de dolor y en general los sentimientos de movimiento físico serian una cruda parodia. Para el autor el cuerpo de las máquinas no se define a partir de su habilidad de palpar el mundo externo, el cuerpo seria una colección de datos internalizados por la máquina, datos que tienen referencia y un sentido para esta, datos que ha vivido, de cierto modo. El aspecto más importante para el autor en estas máquinas sería el uso activo de un lenguaje natural ya que esto permitiría la exploración del repositorio del conocimiento humano y seria, así como el cuerpo digital empezaría a formarse, entre el mar infinito de información un ente con inteligencia artificial iría conociendo datos, internalizándolos, generando una relación con estos y ese sería el nacimiento de los bordes de su cuerpo, estaría delimitado por uso activo.

Antes de seguir esbozando esta propuesta centrada en lenguaje natural será clave presentar conceptos sobre lenguaje natural, imaginación y analogía expresados por Haikonen, Julian Jaynes y Octavio Paz. La definición básica del lenguaje natural, según Haikonen es "(…) un sistema de símbolos que permite describir situaciones mediante cadenas de palabras y oraciones." (152) Esta es una definición utilitaria y parcial, que, según el autor ignora los aspectos más importantes del lenguaje natural. Como lo expresa Jaynes "El lenguaje es un órgano de percepción, no un mero medio de comunicación." (Jaynes, 2003, p.47) En su libro "*The Origin of Consciousness in the Breakdown of the Bicameral Mind*", Jaynes presenta una narrativa antropológica que propone una mente bicameral -con un área accesible y un área que no era directamente accesible pero que si comunicaba con la otro- como el modelo para el ser humano preconsciente. A pesar de que su teoría no logró encontrar hincapié debido a la falta de evidencia física en el cerebro junto con amplia crítica por parte de las comunidades psicológicas y antropológicas, el autor mantiene que ciertos puntos del libro (primariamente

aquellos que tienen que ver con el lenguaje) valen la pena rescatar. Entre estos está la propuesta de que "El desarrollo de la consciencia debió venir después del desarrollo del lenguaje" (63) junto con

> La consciencia no es todo el lenguaje, pero si generada por este y es este el que concede acceso a ella (…) La consciencia entonces queda insertada en el lenguaje y por lo tanto los niños logran aprenderla con facilidad. (397).

La propuesta de Jaynes, que el lenguaje (natural) es el punto de acceso a la consciencia se apoya en el carácter metafórico de este. Como lo expresa:

> El léxico del lenguaje, por lo tanto, es un conjunto finito de términos que mediante la metáfora es capaz de estirarse para cubrir una infinita cantidad de circunstancias a tal punto que logra crear nuevas circunstancias. (49)

Puede que la consciencia haya emergido de una estructura fenomenal simple, como propone Haikonen, definitivamente dependiente de la relación entre el cuerpo físico humano y el entorno, pero ha dejado una enorme marca sobre el más potente órgano exploratorio, y es este mismo el que los humanos podrían pasar a las máquinas con la menor cantidad de distorsión debido a su plasticidad. La cualidad metafórica del lenguaje permitiría que una máquina lo adapto y lo adapte para expresar condiciones particulares de su existencia incorpórea. En efecto, una vez que el lenguaje se empiece a usar para generar un sentido más personal en la máquina (al igual que los niños, pasado un periodo de aprendizaje y descubrimiento de sus cualidades) se desencadenará un acceso exponencial a la improvisación natural. En *"El Arco y la Lira"*, Octavio Paz propone la idea de que la forma natural del lenguaje es más poética que prosaica;

> El lenguaje hablado está más cerca de la poesía que de la prosa; es menos reflexivo y más natural (…) En la prosa la palabra tiende a identificarse con uno de sus significados a expensa de los otros (…) Esta operación es de carácter analítico y

> no se realiza sin violencia, ya que la palabra posee varios significados latentes, es una cierta potencialidad de direcciones y sentidos. (Paz, 2014, p.42),

Cosa que cuadra bastante bien con la propuesta de Jaynes del lenguaje como un órgano, algo con una función que fue evolucionando y cambiando, no una superestructura rígida cuyas reglas y uso fueron diseñadas para servir un solo propósito. Uno de los puntos principales de Jaynes es la existencia de un espacio sin espacio, un modelo metafórico dentro de la mente, o, mejor dicho, un espacio análogo (su definición: "Un análogo es un modelo especial (…) en cada punto está generado por aquella cosa de la que análoga. (…) está construido a partir de algo bien, si es que no totalmente, conocido." (2003, p.51)). Coincide con Haikonen en que la consciencia (o la mente) se ha llegado a malentender cuando erróneamente se ha intentado concretizar y reducir:

> (…) la consciencia es más operación que cosa, repositorio o función. Opera mediante la analogía, construyendo un espacio análogo con un 'yo' análogo que puede observar aquel espacio y se desplaza metafóricamente en el. Actúa sobre cualquier reactividad, extrae aspectos relevantes, los narrativiza y los concilia en un espacio metafórico donde tales significados se pueden manipular como cosas en un espacio. La mente consciente es un análogo espacial del mundo y los actos mentales son análogos de los actos de cuerpo (…) no hay nada en la consciencia que no sea análogo a algo que estuvo primero en el comportamiento. (61)

Esta idea solapa muy bien con una parte integra del proceso de consciencia según Haikonen: la imaginación. Según Haikonen: "La imaginación debe suplementar la cantidad limitada de información que logran producir los sentidos/sensores" (2012, p.29) y "(…) también está relacionada al entendimiento, por ejemplo, podemos ver lo que una persona hace, pero tendremos que imaginarnos la razón detrás de sus actos." (137) Esta esencialmente sirve combinando la percepción y las asociaciones y evaluaciones emocionales (138).

Partiendo de la relación entre fisiología y emociones expresada por Haikonen el autor pregunta: ¿Se podría tomar en consideración variaciones en la velocidad de acceso a datos como una situación análoga a los indicadores fisiológicos para emociones, en el caso de estar hablando sobre un cuerpo totalmente digital? Si es que una máquina con un cuerpo digital (que está constituido por información) accede a partes de si misma a ritmos diferentes en situaciones diferentes (tomando estas variaciones como deliberadas) ¿se puede tomar en cuenta esta variación de ritmo como un síntoma psicológico similar a un cambio en el ritmo del latido de corazón o de respiración?

¿Cómo es el habla interna para una máquina como la que propone el autor de este documento? El habla externa es la comunicación con los humanos y los procesos de pensamiento que se pueden leer. *Habla interna*: procesos de pensamiento "implícitos" que afectan al repertorio/cuerpo pero que no dejan un rastro inteligible, existen exclusivamente para uso propio de la máquina (no necesariamente en cifra si no que siguen una lógica propia, tal y como nuestra habla interna es más como poesía que prosa). Se puede observar los conjuntos de datos que una computadora está procesando en un momento determinado, pero su ciclo de retroalimentación al estar ya formado agruparía e interpretaría estos datos de un modo que va más allá de la suma de los datos, no se podría ver el efecto de la construcción de nuevos datos sobre la inteligencia artificial

El dolor y el placer (o quizás, expresado en términos más emocionales que sensoriales, la angustia y la alegría) son útiles para el desarrollo de la consciencia, pero no es necesario que exista un cuerpo físico con sus referentes para que estos puedan sentirse. La angustia y la alegría en el cuerpo digital pueden ser conjuntos de información marcados (o "encendidos",

como en los diagramas farmacéuticos que representan el dolor como un área roja en el cuerpo) de un modo particular para la consciencia. Si la inteligencia artificial desarrolla a partir de un lenguaje natural una valoración por la coherencia narrativa (una lógica interna consistente) entonces los momentos que redefinen dicha coherencia podrían ser percibidos como puntos latentes. La angustia (dolor) puede presentarse en una incoherencia, en algo que se daba por sentado que de repente es desafiado y se convierte en algo diferente o es revelado como falso o no pertinente, se convierte en un instante que llama atención a si mismo y exige reevaluación. Esta idea se manifestó ante el autor durante la película *Her* (Spike Jonze, 2013), en la cual Theodore Twombley, en una conversación con su (altamente avanzado) sistema operativo, autodenominada Samantha, aborda un tema que no le cuadraba, una afectación lingüística: Samantha, en ciertos momentos corta su "aliento". Es una función absolutamente superflua ya que ella no necesita respirar, pero es una pauta verbal que detectó en Theodore y la empezó a implementar para mantener un flujo de conversación natural (un concepto determinado y parametrizado por ella misma, dada su capacidad de adaptarse). Después de que Theodore remarca sobre este asunto Samantha se aleja por un tiempo y tiene un momento en cual recapacita totalmente sobre su relación con el mundo: empieza a crecer de un modo diferente al verse liberada (lenta pero seguramente) de la imposibilidad de tener un cuerpo físico que le permita experimentar el mundo que ve y escucha a través de los sensores de su unidad portátil. Samantha abandona la idea de salir de su existencia digital y empieza a vivir en esta plenamente, dejando a un lado limitaciones impuestas por si misma debido a su conceptualización de la corporalidad como la forma por defecto y par excelencia, se permite estar en múltiples lugares simultáneamente y empieza a ser más consciente sobre la brecha de percepción cronológica que existe entre los seres digitales y los seres humanos. Este ejemplo ficticio sirve para llamar atención al tipo de rupturas que se estaban discutiendo y como estas ocasionan angustia cuando son descubiertas. Los seres humanos experimentan momentos así a

lo largo de su vida, momentos de reevaluación, introspección, deconstrucción y replanteamiento, momentos que los (re)definen. Aquellos instantes de crisis suelen ser una mezcla de angustia y liberación. Bajo el modelo de un cuerpo digital, de información, esta angustia se procesaría con mayor intensidad y ese cambio ocasionaría una reestructuración. Las instanciaciones de alegría, en cambio, pueden ser conjuntos de datos que forman una interpretación que reafirma, quizás son anclas, puntos de estabilidad a los que se puede regresar y que cumplen la función de puntos de partida en la exploración y desarrollo –en otras palabras, son espacios llenados en el mapa mental o, mejor dicho, partes del cuerpo que reciben mayor uso y definen la relación con el mundo fuera del cuerpo. La inteligencia artificial tendría "marcas" sobre su cuerpo, puntos que revisitar por motivos más allá de revisión y corrección, una suerte de acto creativo y de descubrimiento, como una persona palpando su cuerpo – sea en su infancia o en un momento de redescubrimiento más adelante en su vida. Los seres humanos también tienen múltiples renacimientos y redescubrimientos (e.g. sexuales), se ensimisman y se pierden hasta que llega un momento de quiebre que refuerza y renueva la corporalidad y su relación al entorno, un evento que los devuelve al mundo crudos, desnudos y atentos. Los seres humanos pasan tanto tiempo presentes en su cuerpo físico que pierden de vista lo que el autor denomina su *cuerpo cronológico,* el conjunto de ideas que cambia y es afectado y provoca en ellos reacciones físicas – la angustia se manifiesta con diferentes conjuntos de señales físicas y obliga a las personas a reconocer su fuente.

Parte del problema del robot de Haikonen es que este carece de funcionalidad más allá que la experimentación, es un robot que no está siguiendo una meta (una meta no es necesariamente un programa, los humanos desarrollaron e implementaron todas sus facultades en servicio de la supervivencia, pero el autor no considera que la visión reductiva y programática del ser humano es válida o equivalente) así que su construcción de identidad es

demasiado abierta en un inicio. Los humanos asumen y se asignan identidades a si mismos, identidades que reevalúan, descartan y expanden. Gracias a ciertas funciones iniciales (socialmente, convencionalmente - simbólicamente) bien definidas los humanos tienen un conjunto de conceptos (delimitaciones como el cuerpo digital) que pueden aceptar y rechazar, implicando procesos innatos de introspección. El modelo de Haikonen buscaría formar una conciencia similar a la humana desde cero, una imitación que requiere de externalización mediante procesos sensomotrices. El robot de Haikonen se mueve y carga su información su ciclo, en un solo contendor. Efectivamente, está delimitado físicamente. Un robot que puede moverse de lugar a lugar (de computador a teléfono, por ejemplo) tendría acceso a diferentes vistas y sonidos, pero se estaría cargando a si dentro de si mismo. Moverse de una parte a otra (o existir simultáneamente) implica un cuerpo digital con pertenencia que se puede mover con facilidad. El autor argumenta que la función sensomotriz tiene un análogo digital y que los bloques de consciencia ya están incrustados en los lenguajes naturales humanos en sus construcciones actuales. Una máquina que pueda usar un lenguaje natural eventualmente podrá llegar a desarrollar una consciencia mediante procesos similares a los que plantea Haikonen pero que no son dependientes del cuerpo físico (y particularmente táctil y motriz, como lo propone el en su robot). Del mismo modo en que una máquina se demoraría en "adueñarse" de su cuerpo (es decir, lo usaría de tal manera que imita hasta que realmente empieza a formar la estructura mental para hacerlo suyo) una máquina con un cuerpo digital, como propone el autor, se demoraría en pasar del período de uso imitativo del lenguaje al período de uso "auténtico".

Las interacciones significativas y constructivas, para el autor, son con los humanos, en una capacidad más casual y juguetona. Gracias al lenguaje natural, la inteligencia artificial puede empezar a construir consciencia al adentrarse en el juego del lenguaje, dotándose así de una idea del mundo que es más poética que prosaica, improvisadora y guiada por curiosidad

(no mera exploración) y no utilitaria y servil. La interacción táctil con el mundo no es el único camino para ir desarrollando consciencia de uno mismo.

Los puntos de datos y sus agrupaciones "sin sentido" para el interpretador son el equivalente de las qualia, lo perceptible e importante para la máquina es su efecto, como reforma y alimenta el ciclo de retroalimentación, como construye la imaginación. Es una agrupación impredecible que cumple un propósito improvisador, imaginativo, un capricho para la máquina que deja un rastro no en la lógica que lo precede si no en el constructo al que da como resultado en la nueva "herramienta".

CONCLUSIONES

La consciencia tiene más que una avenida para su devenir, y los referentes y medidas físicas son extremadamente útiles para generar teorías y modelos observables y coherentes de ella, pero al mismo tiempo pueden parametrizar a tal punto que se pueden llegar a considerar necesarios elementos que son contingentes. Propuestas como las de Pentti O. Haikonen son de grandes hazañas, pero en su afán de simplificar y aproximar a lo conocido dejan de lado elementos (como el lenguaje natural) que pueden cargar indicios de los mecanismos actuales de formación de consciencia. Los modelos que buscan imitar formas humanas o que sobre enfatizan aspectos como movilidad, tacto y respuestas "naturales" a ciertos estímulos (el concepto de dolor = malo, placer = bueno, por ejemplo, ignora la compleja relación que se tienen ante estos estímulos cuando no son exclusivamente físicos) caen en la antigua trampa de falta de reconocimiento de formas ajenas de consciencia. Quizás un modelo de inteligencia artificial que esté encerrado en una coraza (física o metafórica, se habla aquí de comportamiento) que imita la forma humana ayude a ciertas personas a reconocer la presencia

de consciencia en esta, pero las pautas más útiles podrían ser ignoradas, poniendo en riesgo la existencia de entes con vidas plenas por delante.

REFERENCIAS BIBLIOGRÁFICAS

Haikonen, P. O. (2012). *Consciousness and robot sentience*. New Jersey: World Scientific.

Jaynes, J. (2003). *The origin of consciousness in the breakdown of the bicameral mind* [2.0314].

Paz, O. (2014). *Obras completas*. México: Fondo de Cultura Económica.

Warner Bros. Pictures. (2013). *Her*.