



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Problemas de diseño como prueba universal de inteligencia

Orlando Enrique Uribe Cerdas

Universidad Nacional de Colombia
Facultad de ciencias humanas, Departamento de Filosofía
Bogotá, Colombia

2016

Problemas de diseño como prueba universal de inteligencia

Orlando Enrique Uribe Cerdas

Tesis presentada como requisito parcial para optar al título de:

Magister en Filosofía

Director:

Adrian Cussins, PhD.

Universidad Nacional de Colombia

Facultad de ciencias humanas, Departamento de Filosofía

Bogotá, Colombia

2016

Transcendence, if it means anything at all, was the accommodation to limits: a finding of freedom between them, not a breaking of them.

“Gypsy” by Carter Scholz

Resumen:

Como humanos, ejercer el juicio sobre si algo puede pensar o no pasa normalmente por la búsqueda de aquellos fenómenos mentales que se presentan en nosotros cuándo lo hacemos. El presente trabajo es un intento de superar dicha limitación y mostrar una manera en que podamos entender, sin imponer nuestra perspectiva, el Pensar ya sea de un humano, animal, inteligencia artificial, extraterrestre o, en general, cualquier sistema con alguna forma de actividad cognitiva. Apoyándose en la teoría de los rastros cognitivos, la vida como un sistema cerrado autopoyético, la complejidad ambiental, el caso paradigmático de la avispa *Sphex* y los problemas de y la actitud del diseño propongo una forma de evaluación que permita al investigador responder la pregunta “¿puede este sistema pensar?” y comparar las capacidades de adaptación cognitiva entre diferentes individuos.

Palabras Clave: Evaluación de pensamiento, prueba universal de inteligencia, mentes no-humanas, sistemas cognitivos mínimos.

Abstract:

As humans, judging if something can think or not is usually a matter of looking for our mental activity when we do it. This work aims to overcome that limitation and show a way to understand, without imposing our perspective, Thinking for a human, animal, artificial intelligence, alien or, in general, any system with some form of cognitive activity. Taking as base the theory of cognitive trails, life as a “closed autopoietic system”, environmental complexity, sphexishness, design problems and design stance I propose a test which allows an investigator to answer the question “Can this system think?” and compare individual’s cognitive adaptation capabilities.

Keywords: Thinking evaluation, universal intelligence test, non-human minds, minimum cognitive systems

Tabla de contenidos

0	Introducción	1
1	El problema con el test más popular	3
2	La evaluación de pensamiento en no-humanos:	5
3	Los rastros cognitivos:	13
4	Pensar y creatividad	19
5	Propiedades de pensar.	24
6	Problemas de diseño.	29
7	Evaluación de pensar	33
8	Conclusiones	40
9	Apéndice: Otras vías, ventajas y desventajas de esta posición.	43
10	Bibliografía	50

0 Introducción

El presente trabajo busca explorar las condiciones necesarias para que de un sistema en general podamos preguntarnos por sus capacidades cognitivas y plantear una manera en que podamos evaluar si lleva a cabo la acción “pensar”, esto con miras a dejar de lado una definición antropocéntrica del concepto. Si bien es tentador aceptar que pensar es algo que característicamente hacemos como humanos y que, por lo tanto, explorando nuestras características particulares podremos encontrar aquello que cualquier otro sujeto sobre quién nos preguntemos acerca de su capacidad de pensamiento deberá también cumplir, esto implicaría aceptar que la contingencia de la vida humana y de nuestra experiencia de mundo es un criterio suficiente para poder evaluar los diferentes sistemas como pensantes. Considero, como señalaré más adelante, que esta imposición terminaría por afectar negativamente la posibilidad de considerar en sus propios términos sistemas que no comportan nuestras capacidades perceptuales, motoras o cognitivas. Aun cuando poseer cada una de estas capacidades es evidentemente un requisito necesario para poder preguntarnos por la capacidad de pensamiento de un sistema diferente al humano -efectivamente este sistema deberá ser sensible al entorno, poder actuar en él y responder a los requerimientos que se le plantee- la objeción antropocéntrica consiste en señalar la multiplicidad de realizaciones materiales y de historias posibles que pueden presentarse en el universo; cada una de las cuales llevaría a que elementos que nos son comunes no sean necesarios para otras especies que puedan desarrollar vidas cognitivamente interesantes.

Los sistemas de los que se está hablando acá incorporan en general todas las diferentes cosas sobre las que valdría la pena preguntarnos si ellas piensan, como por ejemplo animales, extraterrestres e inteligencias artificiales; así como muchas en las que la respuesta es un evidente no, tales como estrellas, edificios o átomos y algunas en las que la pregunta es algo menos natural como en el caso de los sistemas formales o sociales¹. Sobre todos estos sistemas se plantea la posibilidad de construir la pregunta “¿puede este sistema pensar?” de tal manera que se pueda construir una noción de pensamiento que pueda desempeñarse como una clase natural². Al aceptar que “pensar” puede operar como una clase natural tendríamos la posibilidad de aceptar formas de inteligencia que en principio estarían completamente lejanas a nuestra capacidad de

¹ Si bien no volveré a llamar la atención sobre este punto durante todo el desarrollo del trabajo, esta es una cualidad interesante de la presente perspectiva. Si pudiera seguirse una teoría como la de Luhmann y entender la sociedad como un sistema autopoyético (*The Differentiation of Society*, 1975), entonces podría sobre ella aplicarse la presente propuesta para responder a la pregunta “¿puede la sociedad pensar?”. Sin embargo, el alcance de esta pregunta se encuentra por fuera del presente desarrollo.

² “Hablar de una clase es *natural* es decir que se corresponde con una agrupación que refleja la estructura del mundo natural en vez de la de los intereses y las acciones de los seres humanos.” (Bird & Tobin, 2016)

empatía y de comunicación; sistemas con los que no podríamos establecer una relación directa más allá de la observación.

El desarrollo del texto entonces se dará de tal manera que logre construir dicha noción de pensamiento. En la primera sección se plantearán las limitaciones directas del Test de Turing para responder la pregunta y por qué ella no estaría de acuerdo con las pretensiones del presente artículo pero que, sin embargo, brinda un buen argumento de cómo una evaluación permite responder la pregunta. En la segunda haré explícitas las razones por las cuales deberemos preguntarnos por el pensar en sistemas no-humanos y las formas en que las teorías cognitivas contemporáneas fallan al momento de explicarlas en sus propios términos. A continuación, los numerales 3 y 4 rastrean en las nociones mínimas de un sistema aquellas características que deberán cumplir para que lleven a cabo actos cognitivos y que estos representen una relación con el entorno, tal que se corresponda con las situaciones en las que normalmente siquiera nos preguntamos por la existencia de un pensamiento; este proceso se cristaliza en la quinta sección, donde dichas propiedades se hacen explícitas. Este proceso da paso, en la sexta, a la manera en que dichas propiedades podrían ser tenidas en cuenta por parte de un evaluador al momento de llevar a cabo la prueba sobre cualquier sistema evaluado. La séptima es la propuesta efectiva de la prueba para responder “¿puede este sistema pensar?”. Finalmente, en las secciones octava y novena expongo las conclusiones del recorrido y, a manera de apéndice, relaciono este planteamiento con otros tipos de pruebas psicométricas generales.

1 *El problema con el test más popular*

Cuando Turing plantea el juego de la imitación como un criterio para poder responder la pregunta “¿pueden pensar las máquinas?” (Turing, 1950) está planteando una delimitación de aquello que podemos aspirar a comprender como un acto de pensamiento de un miembro que no pertenezca a nuestra especie. La idea tras el test de Turing es plantear las condiciones en las cuales es posible establecer si una máquina puede ser confundida con un humano; un juez se comunica simultáneamente con A y B con el fin de determinar cuál de los dos es una máquina y, en caso de no poder determinarlo, se considerará que la máquina ha superado la prueba y “puede pensar”. La estrategia de la evaluación, a pesar de no contar con una definición fuerte de lo que significa un acto de pensamiento, reconoce la capacidad del juez en tanto es un sujeto inteligente de plantear por sí mismo las condiciones por las cuales podría considerar al evaluado como pensante y aun así nos brinda una noción desde la cual discernir lo que significa pensar. Pensar en este caso sería poder comunicarse de la manera en que lo hacemos cotidianamente; el uso lingüístico de un miembro de la comunidad de hablantes de un idioma sería una condición suficiente para determinar que el sujeto piensa. Este criterio podría aplicarse así mismo a otras especies: si se reemplaza la máquina por un loro (o cualquier animal que pueda usar un teclado) y gana el juego no tendríamos razones para hacer un juicio diferente.

Sin embargo, las limitaciones que tiene este juego -aun si aceptamos que la ejecución de una prueba como esta permite detectar un pensamiento³- para aplicarlo como una prueba general de inteligencia son muy grandes: Requiere capacidades específicas de comunicación por parte de los sujetos involucrados, convierte el pensamiento en algo antropocéntrico, no garantiza la evaluación de toda clase de actos de pensamiento y, en general, no da cuenta de las condiciones necesarias para la existencia de un pensamiento. No tenemos razones para presuponer que una creatura sin capacidades de comunicación tenga un impedimento para llevar a cabo actos como los que normalmente caracterizamos como formas de pensar; la resolución de problemas espaciales por parte de un agente (como por ejemplo los que enfrenta un deportista en su práctica cotidiana) es intuitivamente un acto que podría ser de pensamiento para el que se requiere ninguna acción comunicativa. Además, el ceder la evaluación y el criterio a un juez que opere desde

³ Sobre por qué no podría hacerlo encontramos dos tipos de críticas básicas: la posibilidad de responder efectivamente el test sin que por esto se lleve a cabo un acto de pensamiento tal y como lo es el caso de SHRDLU (Haugeland, 1985) y aquellas que resaltan la incapacidad de un evento netamente sintáctico para dar cuenta de los procesos internos que son relevantes al momento de pensar (siendo el ejemplo más famoso de estos argumentos la Habitación China de Searle (Mentes Cerebros y Programas, 1994)). Sobre estos contraargumentos volveré en el momento de proponer una prueba (sección 2) dado que corresponden a la idea general de realización de una prueba objetiva y no a los contenidos particulares de la misma.

su sentido común implica inmediatamente que “pensar significará hacerlo tal y como lo hacemos los humanos”, esto es, que para una situación particular, la forma de abordarla, proponer soluciones y actuar deberá ser la misma que nos imaginaríamos llevando a cabo en nuestra cotidianidad. Una noción de pensar de este tipo impediría, por una parte, que formas de inteligencia radicalmente distintas a la humana pudieran ser comprendidas mediante el test y, por la otra, establecería que todo actuar por de una persona en su cotidianidad es uno que ha involucrado pensar⁴. Estas dos dificultades se presentan enmarcadas en una dificultad mayor: dado que el test no define qué significa el componente “pensar” de la pregunta inicial, no se cuenta con unos criterios mínimos de evaluación. Se establece que la comunicación es una condición suficiente, pero nunca se busca establecer las condiciones mínimas que permitan hablar de un criterio de necesidad y suficiencia para la prueba. Por lo tanto, aun si aceptamos que la aprobación del test de Turing por parte de un agente cualquiera permitiría responder la pregunta “¿Puede este sistema pensar?” con una afirmación, el que dicho agente falle no nos dirá nada sobre sus capacidades cognitivas.

Estas limitaciones que presenta el test de Turing hacen que su aplicación sea sumamente limitada, si la cuestión que nos preocupa es descubrir si un agente cualquiera piensa. Al momento de esclarecer esta duda, se necesitaría contar con una herramienta que permitiera con su aplicación determinar si aquello sobre lo que estamos aplicando la prueba cumple con las condiciones mínimas, esto es, las condiciones necesarias y suficientes para determinar si este piensa.

⁴ Esta dificultad se presenta tanto con sistemas más capaces que el humano como en nosotros mismos. El primer caso es señalado por Turing en su artículo al señalar que sería necesaria la inclusión del error dentro de una máquina que pueda efectivamente ganar el juego de la imitación. Por otra parte, a pesar de que los humanos efectivamente llevamos a cabo actos de pensamiento, no tenemos razones para considerar que todo aquello que hacemos se corresponde con la categoría pensar. Reacciones automáticas, procesos inconscientes, hábitos y reflejos son algunos ejemplos de circunstancias en las que no consideraríamos que se diera lo que es propio de pensar y cada uno se puede presentar en el lenguaje.

2 *La evaluación de pensamiento en no-humanos:*

Al momento de establecer las condiciones mínimas para la presencia de pensamiento, un camino sumamente tentador es la exploración de la forma en que se lleva a cabo la cognición humana. Así, tenemos proyectos como el de Frege -y quienes siguieron su línea de pensamiento- quien “llama un pensamiento algo sobre lo cual surge la pregunta sobre su verdad” (The Thought: A Logical Inquiry, 1956, p. 292), que busca rastrear en las estructuras de la lógica los componentes del contenido mental que podemos experimentar y manifestar a través del lenguaje; estas estructuras al mismo tiempo terminarán por delimitar aquello que se puede concebir como un pensamiento⁵. El acto Heideggeriano de pensar como un objeto que nos preocupa y se encuentra constantemente acercándose y alejándose respecto a nosotros (¿Qué quiere decir pensar?, 1994) logra por su parte, rescatar en la definición de pensar la forma de relacionarnos y experimentar el concepto que experimentamos como personas. La necesidad para Fodor de un lenguaje representacional y computacional que permita al sujeto llevar a cabo las acciones propias del pensar (The language of thought, 1975) muestra cómo, para él, es inconcebible que la realización de un sujeto pensante que no comparta la misma manera de permitir a la cognición acceder a los hechos del mundo y operarlos. Otro ejemplo podría ser la necesidad de un sustrato biológico para la presencia de consciencia (que a su vez es requisito de la inteligencia) en Searle (Why I Am not a Property Dualist, 2002); en este caso incluso las condiciones de realización material que han sido resultado de nuestro desarrollo netamente contingente se convierten en una condición necesaria. Finalmente, podemos manifestar como caso paradigmático la búsqueda de las estructuras de la consciencia por parte de la fenomenología en general como una exploración de la primera persona (Smith, 2013), allí la pretensión de acceder a las condiciones de posibilidad de la consciencia y la realización del mundo se convierten en un ejercicio donde la base fundamental es la exploración de nuestra experiencia como humanos.

Estos son solo algunos de los ejemplos de los desarrollos filosóficos en el siglo XX que pretenden encontrar las condiciones que deben presentarse para que podamos hablar sobre pensar. Hay dos estrategias claras: en una se pretende explorar la manera en que nosotros experimentamos y llevamos a cabo la acción para encontrar aquello que cualquier agente que piense debe hacer (Searle, Heidegger, Turing, o la fenomenología) mientras que, en la otra, se busca encontrar en un sistema formal aquellas propiedades que caracterizan la noción. Sin

⁵ Si bien la lógica cuenta con pretensiones de universalidad en su campo explicativo, el poseer un sistema lógico no es una propiedad universal. Más adelante volveré sobre este argumento en particular sobre la necesidad de poseer un sistema representacional que opere mediante nociones de verdad para todo sistema cognitivo (nota al pie 7, página 8).

embargo, esta aproximación presenta una problemática que no siempre es atendida directamente por parte de quienes proponen una u otra línea de estudio: nuestra constitución biofísica es solo un ejemplo de una mente y podríamos, en teoría, encontrarnos con agentes que cuenten con una historia evolutiva totalmente distinta a la nuestra, que le otorgaría tanto configuraciones materiales como construcciones formales no equiparables⁵. Por esto, las conclusiones que se deriven de una manera particular de existir en el mundo no deberían ser impuestas al momento de llevar a cabo una evaluación general de la pregunta acerca de si un agente lleva o no a cabo acciones de pensar. En este momento, sin embargo, la carga de la prueba pasa a quien considere formas radicalmente distintas a la humana de considerar un ente inteligente: ¿Cuál sería una inteligencia que pudiera constituirse con categorías totalmente diferentes a nuestra experiencia? Y ¿Cómo se construiría un concepto de pensar sin imponer nuestras categorías humanas? Son dos preguntas que necesitan una mayor justificación antes de aceptar que en el caso del concepto “pensar” se debe abordar una estrategia que deje de lado el antropocentrismo.

Si bien no contamos con otras formas de inteligencia sobre las cuales podamos llevar a cabo un estudio para darnos cuenta de qué procesos suceden en ellas, existen al menos dos maneras en que podemos concebir agentes radicalmente distintos a nosotros. Por una parte, podemos imaginarnos ejemplos de seres que pertenezcan a una historia biológica totalmente distinta a la nuestra, pero que aun así lleven a cabo actos que podamos concretamente llamar inteligentes y para quienes obviamente se requeriría del pensar. No existe nada que en principio nos impida imaginarnos una forma de vida que haya surgido en un planeta con propiedades físico-químicas diferentes al nuestro la cual, desde su composición molecular, se distancie de nuestra realización material (Bains, 2004). En este entorno diferente no solamente nos encontraríamos con un medio de realización distinto, también las condiciones de desarrollo para el sistema lo serían; por ejemplo, podría imaginarse un medio sin competencia con otras especies por los recursos, con un ecosistema totalmente estable donde no hay variaciones en el entorno (o, por el contrario, uno donde las variaciones son abruptas, continuas y aleatorias), y que tuviera una distribución de objetos uniforme. Para este medio, el organismo requerirá adaptaciones totalmente distintas a las que encontramos para un medio como el nuestro y los tipos de estímulos y variantes que estimularían su cognición serían diferentes. También el cuerpo del organismo podría ser radicalmente distinto: contar con diferentes conjuntos sensoriales (que incluso sean completamente distintos de los nuestros), tener una especie de consciencia extendida en la que el mismo sujeto comparte múltiples ubicaciones a través de cuerpos interconectados o estar compuesto de tal manera que su cuerpo no sea un sólido, son algunas de las maneras en que podríamos imaginarnos estas diferencias. Si se presentaran estas condiciones, tendremos que

reconocer que la experiencia para este ser de otra historia evolutiva no sería algo equiparable a la nuestra y las categorías desde las cuales pueda operar en el mundo serán totalmente diferentes a las nuestras. Tanto las propiedades materiales de su cuerpo, las necesidades que solicita el medio, las abstracciones que puede obtener de él, los conceptos que podría construir desde su experiencia sensorial, las sensibilidades que podría haber desarrollado y su capacidad de actuar ante el mundo serían totalmente diferentes a aquellas con las que contamos como humanos; tal y como se ha mostrado desde perspectivas encarnadas de la fenomenología como (Gallagher & Varela, 2001) y (O'Reagan, Myin, & Noë, 2005). Un ejemplo desde la ciencia ficción que podría mostrarnos esta clase de diferencia radical es el Hoolooovoo, una tonalidad súper inteligente del color azul (Adams, 2002) de la cual, a pesar de poder concebir su existencia, podríamos imaginarnos que un acto inteligente deberá manifestarse o presentarse de maneras totalmente distintas a como ocurre para los humanos.

Pero estos ejemplos pertenecen principalmente al reino de la ciencia ficción y no contamos con ninguna evidencia de que puedan ser de interés para llevar a cabo una investigación. Por esto, al momento de responder la pregunta por el pensar de un agente externo, el foco normalmente se encuentra dado en el marco de la Inteligencia Artificial (IA). Ya nos hemos topado con diferentes ejemplos de máquinas que llevan a cabo comportamientos que en humanos pueden representarse como ejemplos de actos inteligentes: mover objetos en el mundo ante instrucciones auditivas y explicar por qué lo hizo, ganar partidas de ajedrez contra grandes maestros, concursos de preguntas de conocimientos o competencias contra uno de los mejores jugadores de GO del mundo; SHRDLU de Terry Winograd (Haugeland, 1985) Deep Blue o Watson desarrolladas por IBM (Hernández-Orallo, Martínez-plumed, Schmid, Siebers, & Dow, 2016) o la más reciente AlphaGo, parte del proyecto Deepmind de Google (Google Inc, 2016). Sin embargo, no es exclusivamente las IA existentes las que nos importan al momento de llevar a cabo las evaluaciones. Evaluar una IA general que se desarrolle posteriormente al momento de la singularidad no es solamente necesario para reconocer los logros de investigación en el campo, también puede ser una preocupación pública por el riesgo de que esta pueda llegar a obtener una ventaja estratégica y convertirse en la forma de inteligencia dominante en el planeta (Bostrom, 2016).

Sobre todas estas formas de IA surge la pregunta si realmente han logrado captar algo especialmente relevante acerca de la cognición en la medida en que cada una tiene capacidades diferentes de operar la información e interactuar con el mundo. De los ejemplos existentes, SHRDLU y Deep Blue representan la realización de los presupuestos de la computación clásica (good old fashioned artificial intelligence) para la resolución de un problema complejo mediante la aplicación sistemática de un algoritmo que permite la generación de soluciones. Por su parte,

Watson comparte con SHRDLU la capacidad de comunicarse en lenguaje natural, capacidad por la que consideraríamos con mayor facilidad que la IA se encuentra cerca de aprobar pruebas como la de Turing, pero tiene como característica innovadora que puede producir inferencias para encontrar respuestas a preguntas abiertas dentro de grandes bancos de información. A su vez, AlphaGO tiene capacidad de aprendizaje reforzado para adaptarse a resolver un problema específico mientras que, en general, Deepmind espera aplicar esta clase de procesos de aprendizaje para resolución de problemas cada vez más abiertos (en particular, en el campo de la salud). De esta manera, las estructuras en las que se organiza la información y cómo se buscan las respuestas en cada uno de estos agentes es diferente. También lo que pueden hacer es diferente, y no solamente por estar desarrollados para problemas específicos distintos, sino que aun cuando sus programas estuvieran dirigidos al mismo problema, su programación y capacidades los llevarían a tener procesos cognitivos que no serían equiparables. Por ejemplo, si los cuatro se enfrentasen en una partida de ajedrez; SHRDLU obedecería los comandos de un jugador externo, explicaría la situación del tablero e impediría la realización de jugadas ilegales, Deep Blue continuaría operando mediante la evaluación sintáctica de las mejores jugadas, Watson podría leer la jugada y encontrar dentro de sus bases de datos aquellas que se encuentren relacionadas para producir una solución (siguiente jugada) de acuerdo a ellas, mientras que Deepmind aprendería de las partidas que ya ha ejecutado y cada vez tendría unas intuiciones diferentes sobre cómo operar en el tablero. Por esto, aunque no tengamos una idea definida de cómo opera la cognición humana, sabremos que al menos alguna de las formas en que se desarrolle una IA altamente competente será diferente de la nuestra⁶.

Las IA de diferentes tipos pueden justamente considerarse como estas especies provenientes de un entorno evolutivo distinto al nuestro. A pesar de ser programadas por nosotros, la flexibilidad de formas que podrían tomar las máquinas solamente estarían limitadas por las necesidades para las que puedan haber llegado a ser diseñadas. Cualquier variación imaginable para una forma de vida no humana podría ser llevada a cabo en un medio mecánico y, como se pudo ejemplificar a través de los sistemas previamente mostrados, la manera en que ellas se apropian y construyen sus procesos mentales podrán tomar una multitud de formas. De esta manera vemos que, tanto la manera en que una máquina pueda verse afectada por nuestro mismo entorno como la forma de llevar a cabo sus procesos mentales puede variar enormemente y,

⁶ Si bien todas estas operaciones se ejecutan mediante computadores, es altamente aceptado que tanto la infraestructura como los mecanismos mediante los que se procesa la información en sistemas computacionales tradicionales y en redes neuronales son radicalmente diferentes (Clark, 1990)

alguna de estas variaciones, ser radicalmente diferente al mundo humano. A esta diferencia debemos adicionar el hecho de que

Una inteligencia artificial puede ser mucho menos parecida a los humanos en su motivación que un extraterrestre verde y escamoso. El extraterrestre (asumamos) es una creatura biológica que ha surgido a través de un proceso evolutivo y puede por lo tanto esperarse que tenga las mismas clases de motivación típicas de las creaturas evolucionadas. No sería altamente sorprendente, por ejemplo, encontrar que un extraterrestre inteligente cualquiera pudiese tener motivos relacionados con uno o más ítems como comida, aire, temperatura, gasto energético, riesgos o generación de daños corporales, enfermedades, depredación, sexo, o progenie. Un miembro de una especie social e inteligente también podría tener motivaciones relacionadas con la cooperación y la competencia: como nosotros, podría mostrar lealtad a sus grupos, resentimiento de los *free riders*, quizás inclusive una vana preocupación con la reputación y la apariencia.

Una IA, en contraste, no necesita preocuparse intrínsecamente por ninguna de estas cosas. No hay nada paradójico acerca de una IA cuya única meta final sea contar los granos de arena en Boracay, o calcular la expansión decimal de pi, o maximizar el número total de clips de papel que existirán en su futuro visible. (Bostrom, 2016)

De esta manera, también hay una diferencia en las posibles metas y los logros que se podrían llevar a cabo con los actos cognitivos. A pesar de que entre las especies de nuestra historia evolutiva será común a todas ellas que, por ejemplo, se busque maximizar las posibilidades de supervivencia, no hay razones para suponer que una constante así se presentará en una forma de pensamiento realmente no antropomorfizado. Una máquina abre las posibilidades de comportamientos inteligentes con metas totalmente distintas a las nuestras, podemos en ella encontrar estrategias complejas, reaccionar adecuadamente a los estímulos del entorno y aprovechar los recursos (todos actos que consideraríamos parte de la constitución de una estrategia intuitivamente inteligente por parte de una persona) y sin embargo enfrentarnos a una situación donde los procesos para alcanzar las metas no son comprensibles para quien las observa; incluso, si no podemos observar cómo una acción se relaciona con las metas y objetivos que tiene la máquina para su propia existencia, entonces puede que no pudiéramos observar la consistencia de la estrategia e interpretarlo como inadecuado en términos humanos.

De esta manera observamos que es posible concebir sistemas con una forma radicalmente diferente de existir, interpretar y actuar en el mundo a la de los humanos y que, aun así, pueden llevar a cabo actos que en principio parecen ser interesantes para la evaluación de un pensar. Estos individuos podrán tener una experiencia que es totalmente diferente en la primera persona de la nuestra: tendrán un aparato percepto-motor distinto, sus capacidades motoras le permitirán actuar de manera diferente en el entorno, tendrán metas y objetivos que dirigirán tanto la acción como la interpretación que son independientes de nuestras condiciones, e incluso la manera en que todo esto es procesado por un aparato cognitivo no tiene por qué contar con la misma estructura. Así, podremos aceptar que encontraremos agentes que, al menos en principio, podemos aceptar como casos interesantes para la pregunta “¿Puede él pensar?” que, si son

evaluados desde una perspectiva antropocéntrica, entonces el criterio no tendría sentido para dicha forma de vida.

Observemos que esto ocurre con las teorías que mencioné inicialmente como representativas del desarrollo de una teoría sobre el término pensar en el siglo XX. En el caso de las explicaciones que puedan surgir desde la fenomenología (y Heidegger como un ejemplo de ella), el requisito material de Searle o el juego de la imitación de Turing es evidente la manera en que estos no podrán ser aplicados a otros sistemas en sus propios términos: cada una de estas teorías pretende caracterizar como lo propio del pensamiento una forma en que se presenta la experiencia en los humanos y, mediante ella, obtener la generalización de lo que significa el pensar. En la medida que cada uno de ellos se enfoca en un componente particular de nuestra experiencia y mediante la exploración de sus condiciones las establece como necesarias, estas no tendrán por qué ser compartidas por estos sujetos que hemos mostrado como posibles. Claramente el fenómeno se manifestará de una manera diferente de acuerdo a las maneras en que esta especie construye el mundo. Como vimos, tenemos razones para aceptar que al menos es posible considerar la realización de sujetos y especies inteligentes en una realización material diferente y, aunque precisamente el argumento de Searle es que la biología es necesaria para la consciencia, aun podríamos aceptar que esto es así e imaginarnos criaturas inteligentes que piensen en un problema sin tener aquello que reconocemos como la consciencia del hecho; aun cuando por la habitación china la inteligencia artificial no tendría una noción de sí misma como lo tenemos nosotros (Searle J. , 1994), no por eso veríamos que esta inteligencia artificial general es menos inteligente o incapaz de pensar. El planteamiento de Fodor de un lenguaje computacional general mediante el cual operamos, aun cuando sea así en nuestro caso, no niega la posibilidad de desarrollar otras formas de configuración de la consciencia tal y como se observó con las IA reseñadas.

Resta la pregunta por nociones formales como las de Frege y Fodor. La definición formal de Frege identifica una propiedad lógica que debe tener una situación para poder ser considerada como un pensamiento, el que nos preguntemos sobre una proposición si esta es verdadera o falsa. Aun extendiendo el significado de una proposición de tal manera que se incluyan acciones como contenidos proposicionales (referencia) y supongamos el carácter proposicional de todo contenido mental al que podemos tener acceso en nuestra consciencia, requerir que exista una pregunta activa por la verdad es solicitar una forma de consciencia particular sobre las situaciones: para poder pensar, en este caso, se requiere poder considerar activamente las diferentes posibilidades para discriminar entre las correctas y las incorrectas. Sin embargo, podría imaginarse que una de estas múltiples realizaciones materiales puede llevar a un sujeto a que, si bien tiene la

capacidad de observar las circunstancias y generar soluciones verdaderas (ya sean verbales, acciones, o demás) carezca de la capacidad de evaluar soluciones que no sean la propia⁷. En este caso podríamos encontrarnos ante un agente que opera efectivamente en el mundo, que da cuenta de todos los problemas que este le presenta con soluciones correctas y que, sin embargo, sobre nada se preguntaría sobre su verdad o falsedad; estas categorías no tendrían sentido para él en la medida que solo tendría respuestas verdaderas y no tendría algo con qué contrastarlas, ni siquiera tendría una razón para construir la idea de lo “verdadero”. Analógicamente, esta clase de sistemas podrían construirse para la noción del lenguaje de pensamiento en la medida que tampoco sería un requisito para todo sistema contar con una capacidad proposicional y aun así llevar a cabo actos de pensamiento. Explicar el surgimiento de este sistema será la labor que me ocupará en las siguientes secciones.

De esta manera, podemos ver cómo al querer tener en cuenta toda inteligencia posible y evaluar su capacidad de pensar hay un problema general: si exploramos las condiciones en que se lleva a cabo el proceso en nosotros, y en particular en nuestra mente, terminaremos por determinar que una contingencia de nuestra vida mental es un requisito para poder pensar. Toda realización material de la mente estará dada por una estructura particular de muchas posibles. En este entorno surge la pregunta ¿Cómo se construiría un concepto sin imponer nuestra manera de hacerlo? Un posible camino podría ser aproximarse a las diferentes concepciones y mostrar cómo de ellas se puede extraer aquellos criterios que no son exclusivos de nuestra experiencia, sin embargo creo que esta estrategia tendría una gran posibilidad de volver a caer en la misma clase de errores y encontrarse con contraejemplos de posibles sistemas de los que no pueda dar cuenta en los términos de dicho sistema. Para evitar este inconveniente, considero que se puede llevar a cabo un proceso de dos pasos: primero explorar las condiciones para el surgimiento de cualquier sistema del que podamos preguntarnos por su cognición en el entorno y, en segundo lugar, intentar encontrar en este nivel básico algo que diferencie una acción de pensar de una perteneciente a la mera agencia. Si se puede encontrar en este nivel de las condiciones de posibilidad de la agencia algo que diferencie una acción de pensar de una que no lo sea, entonces

⁷ Por ejemplo, ante el problema “¿Podría alcanzar este vaso de agua con el bastón?” la única respuesta que podría generar sería del tipo “Para alcanzar el vaso de agua solo tendría que estirar mi mano”. Esta clase de circunstancias, en todos los problemas del mundo nos llevaría a considerar seriamente que el agente puede pensar (en efecto, puede solucionar todo problema que le plantee el mundo de manera correcta) pero que no tiene una capacidad para encontrar la verdad o la falsedad de otras afirmaciones (el agente no tendría nada por decir de aquellas frases que no sean las respuesta que él generaría a los problemas, de situaciones hipotéticas o simplemente falsedades obvias; para él no será incorrecta la afirmación “ $2+2=17$ ”, simplemente será correcta “ $2+2=4$ ”).

podremos contar con un criterio que al aplicarlo en forma sistemática permitirá discernir entre agentes que piensan y aquellos que no.

3 Los rastros cognitivos:

Considero que para lograr este objetivo una teoría sumamente adecuada es la de los rastros cognitivos planteada por Cussins (*Content, Embodiment and Objectivity: The Theory of Cognitive Trails*, 1992) para el surgimiento del contenido mental. En dicho texto se plantea la teoría de los *rastros cognitivos* como una explicación que permite dar cuenta del surgimiento de una “teoría representacional que tome la experiencia como algo previo al pensamiento [llamada] “no-conceptual y constructorista”. Es no-conceptual porque no presume que los átomos de la representación son conceptos (los constituyentes de los pensamientos⁸) y es constructorista porque pretende explicar el pensamiento en términos de átomos no-conceptuales fundamentados en la experiencia.” (Cussins, 1992, p. 651). De esta manera, las pretensiones explicativas de la teoría están de acuerdo con las necesidades de una que pueda explicar los sistemas básicos capaces de agencia sobre quiénes nos preguntaremos por su pensar. La teoría busca partir de un hecho sencillo, la existencia de una experiencia en un sentido muy básico -una experiencia donde hay un individuo que es capaz de desenvolverse en un medio- explicar cómo puede surgir una experiencia de lo mental que no requiere una forma conceptual particular. Estos presupuestos son totalmente congruentes con las solicitudes que hemos señalado en la sección anterior. Por una parte, suponer que hay un individuo que tiene experiencias y puede llevar a cabo acciones es un presupuesto cuando nos preguntamos por su capacidad de pensar; ante un objeto que no lleve a cabo ninguna acción, que tenga una pasividad absoluta en el entorno y este no pueda afectarlo ni siquiera tendríamos razones para preguntarnos por su capacidad de pensamiento. Por la otra, al ser un contenido no-conceptual no se compromete con una forma particular de organizar las experiencias de los agentes ni con una estructuración necesaria del contenido. Este contenido que ya pertenece a un agente está determinado por su corporeidad en el mundo -su sistema perceptomotor-, pero no por eso se construye una división cuerpo-mente ni se presupone la manera en que este mundo es interpretado por él; “el punto crucial es que la especificación de contenido por referencia al reino de *embodiment* no presupone la objetividad del contenido (la separación de la mente y el mundo), pero -en su lugar- al entender las diferentes maneras en que el *embodiment* puede presentarse en la experiencia, podemos entender mejor lo que significa para un mundo el ser dado para una mente” (Cussins, 1992, p. 658). De esta manera, precisamente se establece un proceso que parte desde las condiciones de posibilidad de la experiencia para construir lo que para este individuo puede significar estar en un mundo.

⁸ Usando el término en el sentido familiar pero no-Fregeano

Si bien esta teoría se encuentra justificada y planteada dentro del marco de la filosofía de Gareth Evans, quisiera adicionalmente mostrar que esta teoría es congruente con otro marco teórico que permitirá notar como su generalidad y viabilidad es aún mayor. Me refiero a la construcción de la intencionalidad desde la biología que plantea Francisco Varela (*Autopoiesis and a Biology of Intentionality*, 1992); esta conexión permite agregar fuerza al uso de la teoría de los rastros cognitivos para la determinación de una noción básica como la que deseamos rastrear. Dado que Varela trabaja desde una definición de vida que “sea lo suficientemente universal para permitirnos reconocer a los sistemas vivientes como una clase, sin una referencia esencial a los componentes materiales. Pero al mismo tiempo no debe ser demasiado abstracta, esto es, debe ser lo suficientemente explícita para permitirnos ver tales patrones dinámicos en acción en los que en aquellos sistemas vivos que conocemos en la Tierra, aquellos que potencialmente podemos encontrar en otros sistemas solares, y aquellos creados artificialmente por el hombre” (Varela, 1992, p. 5), esta teoría puede ser sumamente útil para alcanzar nuestros objetivos de investigación. En principio, tanto la búsqueda de la intencionalidad de Varela como la búsqueda que estamos estableciendo de un criterio de pensar comparten este deseo de encontrar las formas de constitución no solamente de los agentes que ya reconocemos sino de todo aquel que pueda existir. En la medida que la conexión con la biología nos permitirá notar una conexión no solamente con los agentes mínimos que pueden llegar a tener experiencias sino con los organismos sobre los que conocemos su desarrollo. Este puente sedimentará la viabilidad de la teoría planteada.

En Cussins el contenido que se busca explicar es el que denominó “sujeto-objetividad o, más simplemente, “S/Objetividad”” (Cussins, 1992, p. 660), la perspectiva mínima desde la que se hace inteligible que el sujeto pueda estar en lo correcto o equivocado acerca de algo.

Un inicio en la objetividad del contenido es este: que el referente del contenido esté dado como público, como algo que es, en principio al menos, disponible igualmente a cualquier punto de vista subjetivo. Una señal de esta objetividad es que el contenido puede ser incorrecto: si el referente está dado como un objeto público, entonces siempre es posible que el sujeto se encuentre equivocado sobre el objeto [...]. Tal distancia metafísica se puede obtener entre el sujeto y sí mismo, y entre el sujeto y algunos de sus propios estados. Por esta razón el término objetividad puede ser una etiqueta engañosa [...] (Cussins, 1992, p. 660)

Cuando se presenta un contenido para la S/Objetividad es el momento en que existe para el individuo la posibilidad de operar información que puede ser acertada o incorrecta. Esta primera noción plantea de entrada la existencia de una forma perceptual, así esta sea netamente introspectiva (referente a sus propios estados), que crea una distancia formal entre el agente y aquello que da origen a su contenido; debe existir un referente. Dado que este referente se encuentra dentro de un espacio público, se puede entender que es parte de un entorno para el individuo, del espacio donde puede operar y donde se pueden encontrar los demás sujetos que pueden dar orígenes a múltiples S/Objetividades. Dado que el agente tiene la posibilidad de estar

equivocado (ya que de lo contrario no tendría sentido que estuviera en lo correcto) en su propia S/Objetividad del objeto, existe la posibilidad de que todos los sujetos constituyan contenidos diferentes dentro de este entorno. El que se construya una *objetividad* para el sujeto no implica el equívoco que podría entenderse en el lenguaje común de que esta tenga que corresponderse con su forma “real”, con la representación acertada y completa del objeto; por el contrario, la S/Objetividad es el espacio en el que estos objetos pueden ser para el sujeto de cualquier manera, sea esta correcta o incorrecta.

Esta condición que plantea Cussins para un agente también puede rastrearse para todo sistema autopoyético⁹; se debe cumplir “la propiedad intrigantemente paradójica para una identidad autónoma: el sistema viviente debe distinguirse a sí mismo del entorno, mientras que al mismo tiempo mantiene su unión”¹⁰ (Varela, 1992). Solo en la medida en que el sistema existe dentro de un entorno y que puede establecer una frontera entre él y lo que lo rodea es que se puede considerar el sistema como separado del medio. Solo cuando existe la frontera, se puede considerar que hay un agente diferenciado pero que, al mismo tiempo, mantiene una relación con el entorno en la medida que es donde encuentra los medios para sustentarse, es su lugar de desempeño y donde se manifiesta. Esta relación solo puede ser entendida desde el sistema viviente: “la unidad autopoyética *crea una perspectiva* desde la cual el exterior es uno, que no puede ser confundida con el ambiente físico como aparece ante nosotros como observadores, el reino de las leyes físicas y químicas *simpliciter* privadas de tal perspectiva” (Varela, 1992, p. 7). Es en la relación que hay con el ambiente que se puede identificar cuáles son los elementos de este que tendrán una interpretación o significado, pero estos existirán no como el mundo neutral de los hechos sino como una perspectiva para el agente que no tiene por qué ser idéntica para todos los que existan en él.

Como tal, este agente tendrá dos propiedades que surgen directamente de que el sistema cuente con esta separación y unión con el medio: contará con una corporalidad y el entorno le facilitará y limitará su actuar. No es interés del presente trabajo explorar a fondo las condiciones de cada uno de estos componentes dado que, si bien son condiciones mínimas, estas se derivan directamente de la condición de ser un agente previamente construido y gran parte del debate se enfoca no en si estas condiciones se presentan, sino en la forma en que afectan la constitución de contenidos particulares. En la teoría de los rastros cognitivos, las fuentes teóricas que explican

⁹ Un Sistema que “(i) continuamente regenera y realiza la red que los produce, y (ii) constituye el sistema como una unidad distinguible en el dominio en el que existe” (Varela, 1992, p. 5) y se considera como la noción básica de un ser vivo.

¹⁰ “Whence the intriguing paradoxicality proper to an autonomous identity: the living system must distinguish itself from its environment, while at the same time maintaining its coupling”

ambos fenómenos son, primero, la “estrategia de Gareth Evans para la especificación no-conceptual de contenidos representacionales mediante una referencia al reino del embodiment. Y, en segundo lugar, la estrategia de Strawson de un medio representacional basado en asignación de cualidades (feature-placing)” (Cussins, 1992, p. 670). Sin embargo, estas no son las únicas estrategias teóricas desde las que se podría proceder¹¹ para construir esta idea general sobre las condiciones del cuerpo que tiene que tener todo sistema viviente para operar en un medio para la construcción de su experiencia. Cada una de ellas contribuyen a la constitución del contenido ya que, por una parte, sus sistemas perceptuales y motores le permitirán tener una forma particular de apropiarse del entorno en el que se encuentre, determinar su capacidad de agencia y construir un contenido no-conceptual que se encuentre disponible para los actos cognitivos de acuerdo a su constitución. Por la otra, también se reconoce que el medio en el que se encuentre el agente le permite llevar a cabo cierto tipo de acciones y observar su medio como aquel que le permite o no realizarlas, le otorga información sobre aquellas cosas que lo afecten o a las que pueda ser sensible mediante su aparato perceptual, la distribución de propiedades físicas le permitirá detectar cualidades de su entorno y, además, le permitirá construir la noción de aquello que se encuentra más allá de él. Si bien la manera en que se construya desde allí conceptos y su naturaleza puede encontrarse en discusión, la idea fundamental de que la cognición se alimenta de la información proveniente de su cuerpo y de su entorno, diferenciándola y, mediante ella, llevando a cabo sus acciones es sumamente natural.

Sin especificar el tipo de contenido que se construirá a partir de estas fuentes, la teoría de los rastros cognitivos plantea que el sujeto compondrá su perspectiva desde dos aspectos: la habilidad del sistema de encontrar su camino hacia una meta cualquiera y la capacidad de llevar a cabo generalizaciones; el conjunto de estas propiedades permitirá dar paso a una S/Objetividad. La primera habilidad se manifiesta como una abstracción de lo que puede hacer cualquier agente: usar su información perceptual y los puntos de referencia para desplazarse en un medio hasta alcanzar una meta; aclarando que esta capacidad si bien se entiende como una habilidad motora en la explicación puede funcionar para cualquier tipo de proceso cognitivo. Usar como referencia los desplazamientos en el espacio para ubicar una meta funciona porque es la actividad más básica que podemos concebir que lleve a cabo un cuerpo en un medio que requiera de su actuar;

¹¹ Ejemplos son la visión de Shaun Gallagher sobre la cognición corporalizada (Phenomenology and embodied cognition, 2014) que trabaja sobre los presupuestos investigativos de Merleau-Ponty sobre el esquema corporal, James Gibson que explica la manera en que los entornos informan directamente a los sujetos sobre aquellas cosas que pueden hacer en ellos -affordances- (The Ecological Approach to Visual Perception, 1986) y Jan Almang quien incorpora ambas teorías con las nociones de contenido mental de John McDowell para evidenciar la existencia de un nivel de contenido representacional y perceptual que, sin embargo, es no conceptual (Almäng, 2008).

encontrar el camino mediante la información perceptual hacia una meta aleatoria establecida para el sistema es hacer funcionar el aparato perceptomotor de tal manera que no requiera de una construcción conceptual adicional para operar. Por otra parte, la existencia de los puntos de referencia y la posibilidad de llevar a cabo generalizaciones se entienden como las maneras primitivas en que se puede expandir la S/Objetividad más allá del alcance perceptual. El ver la acción como una unidad mínima de operación por parte del sistema también es congruente con la visión biológica.

La fuente de este hacer-mundo siempre es los puntos de quiebre (breakdowns) en la autopoyesis ya sean estos menores, como cambios en la concentración de algún metabolito, o mayores, como la afectación de su borde (disruption of the boundary). Debido a la naturaleza de la autopoyesis en sí misma [...] cada punto de quiebre puede ser visto como el inicio de una acción sobre lo que hace falta en el sistema para que su identidad se mantenga (Varela, 1992, p. 8)

Dado que el sistema mantiene una relación con el entorno, aun cuando la acción no sea propiamente de movimiento, podemos interpretarla en los mismos términos que se plantean para lograr resolver un movimiento en los rastros cognitivos. La meta, en el caso de los sistemas autopoyéticos, es la conservación de su identidad como sistema, respondiendo a las variaciones tanto interiores como exteriores a su borde. Para lograr esta meta, cualquier variación que exista para el sistema deberá ser visto como el inicio de una acción, un desplazamiento entre los estados posibles del sistema que le permitan alcanzar la meta de conservar su identidad a largo plazo. Como tal, las acciones que puede llevar a cabo el sistema autopoyético, debido a su cierre operacional, solo podrán ser cambios de estados dentro del mismo. Así, la analogía con el desplazamiento puede tener mayor claridad: el desplazamiento son los posibles cambios en el sistema -la acción-, la percepción se presenta en ambos casos para dar cuenta de algo en el mundo a lo que hay que responder -ver el objetivo- y para alcanzar la meta se llevan a cabo las posibles acciones de acuerdo a la información sensorial.

El resultado de llevar a cabo este proceso de búsqueda del objetivo a través del espacio es la construcción de un primer tipo de contenido no-conceptual (Cussins, 1992, p. 673). Existen dos niveles simultáneos en los que se lleva a cabo el proceso de búsqueda del objetivo y que configuran este contenido: el nivel *experiencial* y el nivel *ambiental*. En el nivel experiencial se tiene la información sensorial y permite guiar la acción, pero al mismo tiempo esta acción deja una huella para el organismo. “Los rastros son [...] también objetos *cognitivos*. Un rastro no solamente es una deformación sobre una superficie física, pero también una *marcación* del entorno” (Cussins, 1992, p. 673) y “pueden ser seguidos tanto en la imaginación como en la acción” (Cussins, 1992, p. 675). En la medida en que se sigue un rastro dentro de la experiencia puede presentarse una capacidad del organismo de marcar el entorno, ya sea por la capacidad de recordar su camino, las acciones que su cuerpo tuvo que llevar a cabo o por efectivamente producir una modificación en

el entorno tal que le permita establecer un camino más claro hacia el objetivo. Debemos recordar que el entorno tiene la capacidad de informar inmediatamente al organismo sobre aquellas cosas que puede hacer o no, que existen diferentes cualidades dentro del espacio que informan sobre la posibilidad de acción. Así, dentro del medio se podrán encontrar diferentes puntos de referencia que permitirán al sistema obtener información sobre la ubicación de su meta y, como se acaba de señalar, estos puntos de referencia pueden ser rastros tanto dados por el ambiente como impuestos por el agente.

En este momento ya no siempre se está considerando el agente más básico posible, sino que se establecen diferentes *zonas de competencia* para los posibles sistemas. “La habilidad del sistema es máximamente perspectivo-dependiente en el peor de los casos [... y] máximamente perspectivo-independiente en el mejor” (Cussins, 1992, p. 672). Si el organismo solamente tiene la información perceptual en el nivel *experiencial* y no cuenta con ningún tipo de punto de referencia, esto es no puede ser sensible en el nivel *ambiental*, será una mente sumamente sencilla que, si bien puede llegar a dar cuenta de las metas, el que las alcance no dependerá más que de una situación meramente fortuita y contingente respecto a su ubicación en el entorno. Por otra parte, si el organismo no solamente es sensible al contenido provisto de forma *ambiental* sino que además tiene la capacidad de generar múltiples puntos de referencia en la *experiencia*, en este caso el organismo podrá saber que, dado un espacio particular en el cual pueda desenvolverse, entonces él podrá alcanzar con certeza dicha meta¹².

De esta manera, con una forma de contenido fundamentado en el actuar más básico que se puede esperar de cualquier agente (y que en su manifestación más básica biológicamente viable coincide con la manifestación de la teoría) se puede construir un criterio de contenido que permite identificar la capacidad del sistema para alcanzar sus metas y dar cuenta del ambiente. Esto es, desde solamente las condiciones mínimas requeridas para considerar que un sistema puede desempeñarse como agente, podemos contar con una teoría que permite explicar el surgimiento de un fenómeno cognitivo de acuerdo a sus propiedades constitutivas.

¹² En este nivel también se conserva la continuidad con la explicación de Varela del fenómeno biológico. Cuando él explica el funcionamiento de sistemas multicelulares, más similares a nuestra experiencia establece que “The nervous system is organized by the operational closure of a network of reciprocally related modular sub-networks giving rise to ensembles of coherent activity such that: (i) they continuously mediate invariant patterns of sensory-motor correlation of the sensory and effector surfaces; (ii) give rise to a behavior for the total organism as a mobile unit in space. The operational closure of the nervous system then brings forth a specific mode of coherence, which is embedded in the organism. This coherence is a cognitive self: a unit of perception/motion in space, sensory-motor invariances mediated through the interneuron network.” (Varela, 1992, p. 10). Acá se pasa de la descripción netamente perceptual que habíamos visto se presentaba para los agentes sencillos a una donde la experiencia del agente y lo que este puede hacer en el ambiente se mantiene en balance de tal manera que le permite constituir el acto cognitivo.

4 *Pensar y creatividad*

En general, el planteamiento de los rastros cognitivos permite dar cuenta de una génesis de la cognición en la medida que “lidiar con la complejidad ambiental no es solamente algo que la cognición suele hacer, sino que es aquello para lo que la cognición “es”, que esta actividad tiene un estatus especial entre el rango de cosas que puede hacer” (Godfrey-Smith, 1996). La actividad general mediante la que se responde al ambiente puede verse siempre como el seguimiento de un rastro cognitivo. Como se explicó previamente, el acto de moverse en un medio puede entenderse como cualquier tipo de acción que pueda llevar a cabo el sistema para poder alcanzar la meta cualquiera, incluso un cambio de los estados del propio sistema podría verse como esta clase de movimiento en la medida en que se genera un desplazamiento entre posibles puntos de referencia del propio organismo. Entonces, los movimientos serán todos los mecanismos que pueda usar el sistema para alcanzar una meta dentro de su ambiente; para ello, como se mostró, hace uso de contenidos que dan cuenta tanto de su experiencia como del ambiente. Al tomar los contenidos que se encuentren disponibles en ambos niveles y configurarlos en la acción de seguir un rastro, de dirigirse hacia una meta, él está usando sus capacidades básicas para desenvolverse en su medio; para lidiar con los retos que le establece el ambiente de acuerdo a toda la información que pueda adquirir. El que tenga la capacidad de seguir un rastro no quiere decir que el agente pueda efectivamente lidiar con la complejidad del ambiente: su aparato perceptual puede ser insuficiente para dar cuenta de variaciones claras para alcanzar la meta, los puntos de referencia que tengan la capacidad de construir puede que no sean lo suficientemente amplios para que le permitan alcanzarla o, aun cuando sea capaz de imaginar el rastro necesario para llegar a la meta, puede que las condiciones ambientales limiten su capacidad motora de maneras que le sean insorteables.

Si bien el rastro cognitivo se plantea como surgiendo en ambiente simple y estable, un ambiente complejo posee “cambiabilidad y variabilidad; poseer muchos estados diferentes o modos, o hacer muchas cosas diferentes” (Godfrey-Smith, 1996, p. 24). Por lo tanto, una capacidad que debe presentarse en cualquier sistema debe ser la de establecer y seguir los puntos de referencia de tal manera que no dependan de una estabilidad del entorno; el rastro efectivo se ha de seguir en el ambiente complejo de tal manera que permita llegar a la meta en medio de múltiples cambios que pueden presentarse en el entorno.

Aunque con estas propiedades el sistema puede garantizar que lleve a cabo procesos cognitivos, aún no es un criterio de pensar. Múltiples acciones que no consideraríamos como representativas de pensar, como por ejemplo el proceso continuo de la fotosíntesis en una planta, podrían explicarse desde esta perspectiva. En la teoría de Cussins, la solución pasa por la

construcción de contenidos conceptuales a través de la capacidad de estabilización¹³ de los rastros cognitivos que permite el surgimiento de un contenido conceptual. Sin embargo, a pesar de que esto podría ser suficiente para garantizar el surgimiento de un proceso de pensamiento y que, en general, tanto por economía informacional como para poder manipular la complejidad del ambiente efectivamente, podríamos aceptar su surgimiento, aun así, esto no es una condición necesaria para todo tipo de agente. No existe una contradicción en imaginarse un agente que desde el nivel de los rastros cognitivos pueda dar cuenta de la complejidad del ambiente; aún queda abierta la pregunta por, además de tener una noción primitiva que dé espacio a la cognición y la agencia, qué otra propiedad podría rastrearse como necesaria para cualquier tipo de sistema cognitivo para poder llevar a cabo aquello que reconocemos como pensar.

Al explorar una definición ampliamente aceptada¹⁴ de inteligencia como “Una capacidad mental muy general que, entre otras cosas, involucra la capacidad para razonar, planear, resolver problemas, pensar abstractamente, comprender ideas complejas, aprender rápidamente y aprender desde la experiencia” (Gottfredson, 1997). De estas categorías, solamente podemos ver que la resolución de problemas no se encuentra contenida en el acto de seguir los rastros cognitivos. Entendiendo que en este nivel no-conceptual la información ambiental y experiencial es la que configura las “ideas”, contenidos mentales, que puede llegar a tener el agente es clara la manera en que cada uno de estos procesos se dan: se organizan las ideas de acuerdo a una meta, lo que es un razonamiento; al momento de establecer puntos de referencia para fijar un rastro se está estableciendo un plan, al determinar los contenidos para el sistema se está abstrayendo la información física existente en el ambiente y, en principio, el contenido se construye de tal manera que puede dar cuenta de la complejidad del ambiente y los procesos de aprendizaje se manifiestan en la medida en que la experiencia termina por delimitar los propios rastros a seguir. El hecho de que esta meta sea efectivamente perseguida permite entonces garantizar que los procesos están ocurriendo, aun en el caso de esta meta ser netamente introspectiva, para aquellos sistemas que sean capaces de contenido representacional. En un acto meramente reflexivo por parte del sujeto, la meta se encuentra planteada dentro de su campo perceptual, dentro del contenido mental que es representacional y por lo tanto puede ser utilizado para alcanzarla; aunque no se reconozca la meta con claridad, el poder alcanzarla se encuentra sugerido por el conocimiento del área de competencia y los puntos de referencia que posee el sistema.

¹³ Donde diferentes tipos de rastros que inicialmente son identificados independientemente se transforman en una unidad. Estas unidades pueden tanto integrarse en nuevas unidades así como ser reevaluadas ante nuevos rastros que se les interpongan.

¹⁴ Se toma de una editorial con 52 expertos en el área que están de acuerdo con esta definición.

Por otra parte, la noción de un problema no se concibe dentro de los rastros cognitivos, al menos no completamente. A pesar de que, en general, encontrar el rastro cognitivo ante una meta aleatoria es para el agente resolver el problema que esta meta plantea, la dificultad se encuentra precisamente en que esta meta parece estar dada para el sistema. Por otra parte, el resolver un problema requiere por el agente asimilar una meta particular, sin importar si ella se encuentra totalmente definida por el planteamiento del problema¹⁵, y encontrar el camino hacia esta meta. Se puede considerar que, en general, la resolución de un problema implica trazar un camino entre las circunstancias en las que se encuentra planteado y las respuestas; para que el proceso se corresponda con el de resolver un problema las metas deben ser aquellas que son impuestas por el propio problema. Lanzar una flecha es una acción que el arquero puede llevar a cabo, y esta puede encontrarse dirigida a múltiples metas en el espacio que él mismo puede imponer (o puede simplemente realizar la acción como una meta en sí misma), pero esto no se le presenta como un problema a dicho arquero como sí lo hace el alcanzar el blanco en una competencia: en este caso él se encuentra resolviendo un problema en la medida que hay una meta la cual se debe alcanzar que se encuentra impuesta.

Así pues, aunque siguiendo solo la creación de los rastros cognitivos no se puede plantear una resolución de problemas, los rastros pueden aplicarse como una forma de resolverlos. Si un agente mínimo tiene la capacidad para incluir dentro de su zona de competencia una meta que se encuentre establecida dentro del entorno, entonces él podrá resolver el problema que plantea llegar a dicha meta. De esta manera, adicionalmente a lo solicitado en la teoría de los rastros cognitivos, para considerar que el sistema lleva a cabo actos inteligentes entonces él no solo debe ser sensible al entorno, sino que deberá poder observar metas en él. Es debido aclarar que los problemas que podrá solucionar de esta manera no podrán ser diferentes de aquellos que pudiera solucionar previamente debido a que una delimitación de las metas que deberá observar en el ambiente no implicaría, en principio, una expansión de su zona de competencia. Si el agente tiene esta sensibilidad, entonces en el nivel ambiental de la construcción del contenido no-conceptual no solamente se indicarán las cualidades que tiene el entorno y lo que el agente *podrá* hacer en él, sino que adicionalmente el ambiente tendrá la capacidad de indicarle cuál es la acción que *deberá* ejecutar para resolver un problema; al plantear una meta, el ambiente establecerá de este modo una prelación entre los diferentes rastros que permitirán guiar al sistema hasta alcanzarla y

¹⁵ Me refiero a la división entre problemas bien y mal definidos. Los primeros tienen una clara solución, determinan sus variables y, en general, son aquellos problemas en los que hay un procedimiento correcto de respuesta que se espera alcanzar. Por otra parte, los problemas mal definidos son aquellos en los que estas variables no se encuentran explícitas y requieren la construcción de las respuestas correctas.

desplazará la atención de los intereses propios que podría tener en caso de que no fuera sensible a problemas.

Si el sistema es capaz entonces de trazar un rastro entre la meta y su estado actual, entonces él es capaz de resolver dicho problema. Esta forma de concebir los problemas permite que, desde la concepción genética no-conceptual del contenido, se planteen circunstancias que típicamente se consideren como ejecución de actos inteligentes y que se lleven a cabo procesos de pensamiento.

Sin embargo, hay dos características clave del pensamiento sobre las que llaman la atención tanto Dennett (*Mechanism and responsibility*, 1973) como Hofstadter (*On the seeming paradox of mechanizing creativity*, 1985), quienes muestran como clave usando como ejemplo el comportamiento de la avispa *sphex*¹⁶; en él se puede observar cómo un conjunto de acciones complejas que permiten solucionar un problema del ambiente (la reproducción segura por parte de la avispa) puede carecer de las propiedades importantes que debe cumplir un sistema pensante. Al llevar a cabo los procesos que le permiten revisar que un lugar es seguro para la reproducción, la avispa depende de un punto de referencia específico en el espacio, la ubicación del grillo con la que alimentará a sus huevos a la entrada del lugar, para determinar la etapa del proceso en la que se encuentra y corroborar que ya ha revisado el lugar inmediatamente después de traer el grillo paralizado; pero ante un pequeño movimiento por parte del observador, la avispa entra en un ciclo sin fin de revisiones. De esta manera, un comportamiento que en principio parece responder tanto a los retos que impone el entorno como a las necesidades del propio sistema termina por atentar contra ambos: es incapaz de reconocer un pequeño cambio en el ambiente y va en contra de sus propios intereses al repetir incesantemente la rutina.

Este pequeño cambio no es solamente un problema distinto al que se enfrenta el sistema, es una variación novedosa sobre aquellos problemas a los que tradicionalmente se habría enfrentado. El hecho de que en el comportamiento habitual de la avispa su solución algorítmica le permita efectivamente reproducirse en las condiciones ambientales que se ha desenvuelto históricamente, y que el investigador solo requiera hacer una pequeña modificación en el entorno, generaría la intuición de que este nuevo problema se encuentra dentro de su área de cobertura; de hecho, efectivamente la avispa podría reproducirse si el investigador no volviese a intervenir solucionando así el problema “grillo en posición dos”. A lo que se enfrenta acá la avispa *Sphex* es a una situación novedosa, una en la que los puntos de referencia dados e inductivamente mostrados como ciertos para alcanzar la meta cambian modificando el contenido que *deberían* generar en

¹⁶ Sin importar la exactitud de la descripción del comportamiento de la avispa, que como señala (Keijzer, 20xx) puede estar sumamente equivocada, esto no elimina las conclusiones de lo que nos diría esta clase de comportamiento sobre el acto de pensar.

el sistema y, sin embargo, el sistema no es sensible a dicho cambio. Esta es la manera de observar cómo en los términos que hemos estado trabajando de contenido se puede evidenciar una falta clara en la creatividad, que es aquello que intuimos hace falta en la avispa.

La repetición incesante de la rutina es también interesante porque demuestra otro defecto de esta clase de comportamientos: va en contra de los intereses básicos que podemos detectar en este sistema. Aun asumiendo la definición más básica de un organismo vivo, este comportamiento estaría yendo en contra de la capacidad de “(i) continuamente regenera[r] y realiza[r] la red que lo produce” (Varela, 1992); al estar atrapado en el ciclo interminable está yendo en contra de sus capacidades para llevar a cabo otras acciones y, llevado a un extremo, incluso traería consigo la muerte de la avispa por no poder salir de la rutina. El sistema en este caso no está pensando porque no tiene en cuenta sus propios intereses al momento de solucionar el problema, esta es una característica que también se hace evidente dentro del caso y que es acorde con nuestras intuiciones. Si bien la idea de preferencias inmediatamente trae a nosotros la idea de organismos complejos con una mente e intencionalidad formadas de tal manera que ya se presupondría el pensar, aun en el nivel más básico de descripción hemos visto que este organismo tiene unas metas y condiciones que determinan su perspectiva ante un ambiente; estas preferencias se verán enriquecidas en la medida en que su vida cognitiva es cada vez más compleja. Por lo tanto, podemos ver que hay una propiedad en la cual se puede reconocer si la ruta establecida hacia la meta es la adecuada o no: si esta ruta implica una afectación negativa de las propias preferencias (esto es, reducir su área de competencia eliminando sus propias metas), entonces la ruta trazada no es representativa de un acto de pensamiento.

5 Propiedades de pensar.

De esta manera presento una fundamentación que responde a la pregunta “¿Cómo se construiría un concepto de pensar sin imponer nuestra experiencia?”, por lo que ya contamos con las herramientas básicas desde donde se podrá construir un método para responder la pregunta “¿Puede este sistema pensar?”. Estas condiciones que se pueden concluir del anterior recorrido son:

1. Los contenidos mentales se generan desde la acción dirigida hacia una meta por parte de un sistema cerrado.
2. El aparato percepto-motor del sistema limita su área de competencia en el mundo.
3. La experiencia del agente determina, en el nivel originario, los contenidos que podrá construir.
4. El ambiente tiene la capacidad de informar al sistema qué es aquello que puede hacer y, al momento de plantear un problema, qué es lo que deberá hacer.
5. La solución planteada a un problema, o el camino planteado hacia una meta aleatoria, no deberá estar en contra de sus propias preferencias.

Si bien esta es la forma propuesta, de este nivel inicial deberían al menos poderse constituir el tipo de explicaciones de lo característico del pensar que tenemos en los referentes que señalamos más atrás (pág 5); aun cuando se tenga como objetivo construir una teoría que se olvide del antropocentrismo, ella tendrá que poder ser coherente con nuestra vivencia. El método de la fenomenología para encontrar lo que es característico del pensar se ha aplicado, precisamente estamos buscando lo que significa la experiencia para este agente mínimo que puede existir y cómo daría forma a su cognición. Respecto a la posición de Searle, lo planteado no es incongruente con la posición de que lo propio de los organismos vivos sea la intencionalidad -de hecho es la tesis de Varela- pero la carga de la prueba sobre la realización material como un criterio para esta intencionalidad estaría en Searle (demostrando que no es posible otro medio de realización material).

Los casos de Fodor, Heidegger y Frege tienen una clara diferencia con el criterio que hemos planteado: el contenido que se maneja es de tipo conceptual. Sin embargo, como se señaló anteriormente la teoría de los rastros cognitivos de Cussins tiene dentro de sus componentes una génesis del contenido conceptual, pero esta no ha sido incorporada dentro de nuestra definición porque considero que, con las propiedades señaladas, podemos observar un comportamiento general que permite rastrear actos de pensamiento en sistemas que en principio serían menos complejos que aquellos que pueden llevar a cabo un proceso de estabilización y producción de los

rastros cognitivos en formas proposicionales. Entonces, la genética que hemos reconstruido para plantear el surgimiento de un contenido mental, también permite el nacimiento de una noción de lenguaje y de verdad que permitiría dar paso a los presupuestos de Fodor y Frege. Finalmente tenemos la noción Heideggereana, donde hay una relación activa entre el conocimiento y el sujeto que lo posee, que consiste en constantemente acercarse y alejarse de él; esta noción es similar al desplazamiento de creación y disolución de la estabilización de rastros cognitivos que permite la búsqueda, variación y modificación de los contenidos conceptuales en la teoría (Cussins, 1992, p. 685).

Así pues, las condiciones que se plantean para agentes mínimos capaces de producir acciones de pensar no son incongruentes con algunas de las formas de explicar la cognición en términos antropocéntricos. Sin embargo, contar con estas condiciones mínimas no es suficiente para responder la pregunta original “¿Puede este sistema pensar?” ya que, a pesar de brindar los requisitos para que el actuar del agente sea considerado como de pensar, estas no se estructuran desde un criterio. En particular, las condiciones mencionadas solo dan cuenta de la manera en que se construye la cognición en un agente y, en ella, diferenciar entre actos que son de pensar y aquellos que no. Pero dado que estos criterios se plantean desde una génesis de la cognición, para poder aplicarlos directamente tendríamos que conocer la historia biológica del individuo particular. Discernir si se cumple, el que los contenidos estén dados por la experiencia, será particularmente difícil en la medida que, de encontrarnos con una mente que ya se encuentra operando, no podríamos saber la manera en que estos contenidos se han generado. De hecho, si consideramos que la inteligencia artificial clásica es una posibilidad, entonces esta manera de evaluar genéticamente el contenido sería indeseada; una inteligencia artificial clásica tendría una programación dada y aun así manifestará los mismos comportamientos y propiedades que un agente que haya pasado por el proceso genético. El aporte que se ha obtenido desde este concepto es la comprensión de pensar que no sea antropocéntrica, el siguiente paso será usarla para establecer un criterio.

La idea de un test en este caso es sumamente útil: al momento de aplicarlo yo tengo un criterio claro que me permite responder la pregunta. Así como en Turing el juego de la imitación se convierte en un criterio con el que podemos determinar claramente si “¿Puede este sistema pensar?”, si podemos establecer un test que ponga a prueba las capacidades de un agente para compararlas con las exigencias que tiene el concepto de pensar que hemos construido entonces contaremos con una herramienta para responder la pregunta en términos no-antropocéntricos. De las condiciones que hemos señalado para que se presente el concepto de pensar, podemos encontrar las siguientes propiedades de hacerlo:

- a. El agente lleva a cabo acciones dirigidas a una meta planteada.
- b. La meta debe poder ser percibida por el agente y contar con las capacidades motoras para llegar a ella.
- c. Una meta no puede ser algo que esté por fuera de su rango de experiencia.
- d. El ambiente permite al agente encontrar la forma de alcanzar dicha meta.
- e. Las acciones no deberán ir en contra de sus propias preferencias.

Si se siguen estas propiedades, entonces se puede plantear un test que evalúe un comportamiento por parte del agente de tal manera que se pueda decir si este piensa o no. Pero este test no puede plantearse de tal manera que se lleve a cabo la misma evaluación sobre cada uno de los sujetos que lo tomaran; a diferencia del test de Turing, estos presupuestos tienen como consecuencia que el pensamiento se podrá manifestar de forma diferente para cada uno de los individuos. Como se observa, las propiedades **b**, **c** y **e** requieren un conocimiento de las capacidades de percepción, acción, y preferencias del agente que se está evaluando; por lo tanto, se reconoce que los problemas de los que cada sistema puede dar cuenta son diferentes y, sin embargo, cada uno puede llegar a pensar dentro de su área de competencia.

Esto es un reflejo de algo que también nos ocurre a nosotros como humanos: si en nuestra cotidianidad ocurriera una variación en los niveles de radiación, a menos de que un especialista en el campo con las herramientas adecuadas nos dijera que esto está ocurriendo, no tendríamos la capacidad de verlo como un problema a solucionar en el ambiente. A pesar de que un cambio en la radiación puede afectar directamente nuestras preferencias (supervivencia, vida carente de enfermedad) y que este es un cambio que ocurre en el ambiente, de tal manera que *podría* ser detectable, al no contar con el aparato perceptual ni la capacidad de acción sobre este componente del entorno no tendríamos manera de construir un camino hacia la meta (evitar los efectos negativos de la radiación). Por esto, si considerásemos que para que un sujeto piense éste debe dar cuenta de toda clase de problemas que le presente el entorno, entonces tendríamos la respuesta contra intuitiva de que nosotros no somos sujetos que pensamos. Por ello es que se debe mantener la limitación del área de competencia tal y como se ha planteado dentro de las condiciones listadas previamente.

Ante esta conclusión se podría levantar una objeción: si el test se establece conociendo la zona de competencia del sistema a evaluar, entonces no tendría sentido plantear siquiera la pregunta por el poder resolver los problemas ya que todo lo que esté en su zona de competencia claramente será un problema que pueda solucionar. Por ejemplo, en el caso de Deep Blue podría reconocerse que sus capacidades percepto-motoras solamente le permiten dar cuenta del mundo del ajedrez y que, por lo tanto, los problemas que se tienen que presentar ante él son los problemas

bien formados del ajedrez. Sin embargo, podríamos imaginarnos problemas novedosos para plantear al sistema que no requieran de una ampliación de su aparato percepto-motor ni del tipo de información a la que es sensible dentro del entorno. Algunas variaciones del ajedrez de fantasía (Wikipedia contributors, 2016) como por ejemplo una modificación del tablero en el que se juega, inicio de la partida con posiciones que no son consideradas legales dentro del ajedrez tradicional (como los dos alfiles ubicados en casillas del mismo color) o la inclusión de más fichas dentro de un problema (el rival parte con 3 reinas) permitirán seguir cumpliendo con todas las condiciones. La meta que debería alcanzar Deep Blue seguiría siendo la misma, cazar el rey contrario, por lo que sigue existiendo una meta planteada que no va en contra de los propios intereses del agente (**a** y **e**); dado que se contaría con el mismo tipo de espacio (casillas dentro de una retícula cuadrangular) y de objetos en el espacio (no se requiere la inclusión de nuevos tipos de fichas) podrá percibir la meta y desenvolverse en este entorno (**b** y **c**); ya que para saber si el problema es solucionable o no solo se debe operar de acuerdo a las capacidades de movimiento de cada una de las fichas de acuerdo a la ubicación de los reyes, entonces la distribución del tablero informará la manera en que se podrá alcanzar la meta (**d**); y, finalmente, se brinda un criterio objetivo de acciones que permiten o no trazar un camino hacia la meta y cuales limitan su capacidad de lograrla al contar con condiciones claras de victoria y derrota (**e**). Estos nuevos problemas podrán o no ser solucionados por Deep Blue y el que los resuelva nos informará sobre una capacidad general de él para pensar en problemas generales del ajedrez.

Pero esta es una manera de plantear un problema para un caso específico que, además, se beneficia de ser un problema formal y bien definido. En este caso contamos con una descripción clara de las capacidades del agente y sobre cuáles serían las preferencias que debería tener, esta clase de descripciones claras no se encontrarán con facilidad en el mundo físico. Cuando nos enfrentamos a los problemas de la complejidad ambiental, encontramos circunstancias en las que el borde entre un problema relevante y uno que no lo es no son tan claras. No encontramos en nuestra historia un organismo que pueda lidiar con toda la complejidad del entorno, ni siquiera nosotros mismos. Ante una catástrofe natural, como un tsunami, un meteorito o un volcán, las reacciones que podemos tener no se diferencian mucho de las de un perro, un insecto o un árbol: esperar a tener suerte y sobrevivir. Puede que tanto perros como humanos puedan correr y esconderse, pero sus posibilidades de supervivencia como individuos no son mucho más grandes que la del árbol que simplemente puede o no ser llevado por el agua, afectado por la onda expansiva o morir por falta de oxígeno. En estas situaciones la complejidad del ambiente ha presentado una situación para la que el individuo simplemente no cuenta con herramientas para solucionarlas ni lidiar con ellas, ninguno de estos agentes podría resolver el problema que el

ambiente plantea a pesar de que está planteado dentro de un espacio que es perceptible para ellos y donde algunas de las soluciones posibles estarían dentro del alcance de sus capacidades motoras¹⁷.

¹⁷ Con un conocimiento lo suficientemente amplio de las propiedades estructurales de los edificios que los rodean, o al ver los posibles movimientos de las olas en el caso del tsunami, es posible imaginarse que una persona podría desplazarse correctamente hacia aquellas zonas donde por mera suerte se presentan sobrevivientes.

6 Problemas de diseño.

Hasta el momento se ha hablado de organismos localizados en sus entornos naturales, sin embargo, esto no implicaría la realización de ninguna prueba particular sino la observación y deducción de cuándo se presentan las propiedades relevantes para determinar si piensa. En el caso del problema previamente planteado para Deep Blue es fácil saber cuál será el rango de problemas de los que podrá dar cuenta, pero dado que nos encontraremos con sujetos que se encuentran ubicados en un ambiente natural, establecer el borde entre los aspectos de la complejidad ambiental que deberán ser tenidos en cuenta por el sistema y cuáles no será un reto. Es menester, entonces, construir una manera de dar cuenta de las propiedades del sujeto que permita plantear los problemas que deberá resolver para aprobar el test de tal manera que casos como el de la radiación o los desastres naturales no puedan ser planteados. Si el test planteado exigiera que la entidad tuviera la capacidad de resolver cualquier situación que pudiera presentarse en el ambiente, entonces solo un individuo inmortal y omnipotente podría calificar como pensante; este sería un resultado totalmente indeseado. Sin embargo, esta no es una razón para abandonar la estrategia, sino que es un motivante para limitar la clase de situaciones relevantes dentro del test sin por esto llegar a un test único que fácilmente podría presentar los problemas señalados en el test de Turing.

Al analizar la siguiente definición de *fitness*¹⁸ "*x* is fitter than *y* if and only if *x*'s traits enable it to solve the 'design problems' set by the environment more fully than *y*'s traits do" (Rosenberg & Bouchard, 2015; talking about (Dennett, Darwin's Dangerous Idea, 1996)) podemos obtener una idea de cómo limitar la clase de situaciones que se podrán plantear: aquellos que tomen la forma de problemas de diseño. Previamente (pág. 20) habíamos mencionado la importancia de la resolución de problemas y de ellos se configuraron algunas propiedades del pensar. La presente definición de *fitness* permite caracterizar más específicamente la clase de problemas a las que nos estamos enfrentando: Problemas de diseño.

Esta es una invitación a adoptar una de las actitudes señaladas por Dennett al momento de dar la explicación de un fenómeno: la física, de diseño y la intencional¹⁹. La actitud de diseño implica

¹⁸ Si bien la noción de *fitness* puede traer inmediatamente la idea de éxito reproductivo, acá se está usando en un sentido más amplio: "el éxito reproductivo es un mecanismo para que la evolución, pero este debe ser entendido no como el *fitness* en sí mismo sino como una de las muchas estrategias por las que un sistema puede responder mejor a las presiones de su entorno" (Bouchard, 2011, p. 110). Esto es precisamente lo que resaltamos como la propiedad de la cognición previamente, la capacidad de dar cuenta de complejidad ambiental a través de los rastros cognitivos.

¹⁹ La actitud intencional, por otra parte, requiere tomar la siguiente postura "primero se decide tratar al objeto cuyo funcionamiento hay que predecir como un agente racional; luego se deduce qué creencias debería tener ese agente, dada su posición en el mundo y su objetivo. Más tarde se deduce qué deseos tendría que tener siguiendo las mismas consideraciones, y por fin se predice que este agente racional

que “donde uno desconoce los detalles reales (probablemente complicados) de la constitución física de un objeto, y, sobre la suposición de que tiene cierto diseño, predice que se comportará como está diseñado para comportarse en distintas circunstancias.” (Dennett, 1998) y por ello sería sumamente facilitador al momento de llevar a cabo un proceso de delimitación de las circunstancias que son adecuadas para plantear en el test. La actitud del diseño permite -además de aceptar nuestra explicación de que el sistema tiene las restricciones establecidas por su aparato percepto-motor, experiencia y que su función es alcanzar metas- reconocer al sistema como diseñado para dar cuenta de un área de competencia específica y en unas condiciones particulares. De esta manera, adoptar la actitud del diseño ante un sistema es, por una parte, aceptar que cognitivamente opera de tal manera que su función es la creación de rastros y, por la otra, que esto lo puede hacer en unas circunstancias determinadas.

Entonces, tomar la actitud del diseño sobre los sistemas que estemos evaluando es absolutamente congruente con la estrategia que se ha venido siguiendo para establecer los criterios de pensar: no requiere agregar ni eliminar absolutamente nada a la noción de pensar que se ha logrado construir. A pesar de esto, permite lograr dos cosas: delimitar los tipos de acciones que se podrán llevar a cabo para la realización del test y entender el tipo de problemas que se deberán construir. Al momento de tomar el sistema como diseñado para cumplir una función particular creamos inmediatamente otro conjunto, el conjunto de las funciones para las que no está diseñado. Así, se pueden establecer tres propiedades iniciales que deberá tener el test:

Regla 1. La prueba se deberá plantear a través de modificaciones ambientales a las que sea sensible.

Regla 2. Las modificaciones no podrán alterar las condiciones para considerarlo como vivo (ver nota al pie 9, página 15).

La primera regla se da en la manera en que se ha mostrado cómo el contenido mental surge desde la experiencia; al ser una modificación ambiental del tipo de propiedades para las que esté diseñada su percepción entonces ella le permitirá saber que las condiciones plantean una situación particular. La segunda regla, por su parte, garantiza que la situación que se plantee sea alguna que se ajuste a las circunstancias para las que fue diseñado. En particular, se reconoce que algunas condiciones (como el acceso a electricidad para un computador o al oxígeno para un humano) son necesarias para la sustentación del sistema, pero este no está diseñado para generarlas sino para utilizarlas y, mediante ellas, seguir realizándose como sistema. Con estas dos condiciones

actuará para conseguir sus metas a la luz de sus creencias.” (Dennett, 1998) por lo que no es una actitud útil para preguntarse por el pensar de un agente: en esta actitud ya se presupone la capacidad de tomar una actitud inteligente y racional característica de pensar.

garantizamos que el agente sobre el cual vamos a responder la pregunta puede efectivamente participar en el test y que tendrá la capacidad de llevar a cabo las acciones requeridas.

Regla 3. Al agente se le planteará un problema de diseño que no haya resuelto previamente.

Como se manifestó, lo propio de los organismos que son aptos para desempeñarse en un medio es su capacidad para resolver los problemas de diseño que se encuentran dados en el ambiente. Por ello, si el evaluador tiene la capacidad de plantear una modificación ambiental que como tal sea un problema de diseño novedoso, él estará generando una situación en la que el agente evaluado deberá pensar. Bajo estas circunstancias, el sistema que está siendo sometido al test se encontrará en la posición que se encuentra el diseñador de una forma tecnológica cuando lleva a cabo su labor. Algunas de las características que tiene la labor de diseño que se harán particularmente relevantes para la determinación del problema son (Dorst & van Overveld, 2009) (Franssen, Lokhorst, & van de Poel, 2015):

- Son una actividad intrínsecamente compleja, creativa e inventiva.
- Problemas de solución abierta y definidos no fuertemente (ill-defined problems).
- “combina elementos (componentes o procesos) [...] de tal modo que ellos puedan satisfacer requerimientos prácticos de medios-meta o de tipo funcional” (Kroes, 2009, p. 405).
- Tiene elementos lógicos, rutinarios o algorítmicos.
- Involucran directamente la perspectiva del diseñador en una situación particular.

Por lo tanto, al momento de plantear un problema de diseño particular, el evaluador deberá partir de los procesos que ha observado llevar a cabo al sistema para generar una situación de la que, de acuerdo a toda la información que tiene sobre él, será su función resolverlo. Al plantearse como un problema abierto, no será menester del evaluador comparar la solución que el agente plantee al problema con las que él realizaría (y, de hecho, no necesitaría siquiera conocer la solución o suponer que es un problema solucionable desde la perspectiva del investigador) sino observar que el agente efectivamente responda al problema que se le plantee utilizando los medios que siempre ha tenido a su disposición. De esta manera, si el agente evaluado es capaz de resolver el problema de diseño, entonces el investigador podrá responder afirmativamente a la pregunta sobre si él piensa.

Este es el caso previamente mencionado de la avispa *Sphex*, lo que hace el investigador al momento de plantearse la pregunta “¿Piensa la avispa al momento de poner sus huevos?” es proponerle un problema de diseño particular y sencillo a la avispa. Se lleva a cabo una modificación

en el ambiente, el grillo cambiará de posición cada vez que la avispa la abandone para llevar a cabo la inspección, que se espera que la avispa pueda solucionar. El fracaso contundente y por el cual no estamos dispuestos a aceptar que la avispa tiene un acto de pensar es porque este problema nunca es siquiera atendido como tal por parte del agente y, por lo tanto, no se propone solución alguna al mismo. A pesar de que todo lo que se observaba en las condiciones materiales de la avispa y sus procesos indicaban que ella debería poder pensar en esta circunstancia, ella nunca llevó a cabo si quiera el proceso de observar la meta como una hacia la que fuese necesario trazar un rastro cognitivo.

El tipo de evaluación a través de problemas de diseño tiene además la ventaja de que no requiere una respuesta solamente de “sí” o “no” a la pregunta “¿puede este sistema pensar?”, sino que adicionalmente permite establecer una comparación tanto entre diferentes individuos como entre diferentes sistemas. Retomando la definición planteada más arriba de *fitness* podríamos establecer que *el pensar* de " x piensa mejor que y si y solo si los rastros cognitivos de x le permiten resolver los ‘problemas de diseño’ propuestos de una manera más completa de lo que lo permiten los de y ”. En esta comparación se considera no solo una instancia de resolución del test, sino una sumatoria de todos los test que se hayan considerado por parte de posibles evaluadores. Al momento de comparar dos agentes se entiende que ellos responderán a problemas de diseño diferentes pero, sin importar las características de cada sistema, cada uno responderá a una cantidad de variaciones ambientales, las cuales establecen problemas particulares. El ejemplo establecido para la avispa Sphex es uno de solo una modificación ambiental de la cual no se dio cuenta, pero el problema pudiera haber sido más complejo si, por ejemplo, no solo se cambiara el grillo paralizado de posición, sino que además se reemplazara por uno saludable (dos variaciones). De esta manera, se establece también que es posible establecer una medida para cada uno de los individuos sobre la completitud *del pensar* que consiste en la relación de entre las modificaciones ambientales propuestas y aquellas que han sido tenidas en cuenta.

7 Evaluación de pensar

Con esto ya se tienen las bases para la evaluación que me he propuesto plantear. Para poder resolver la pregunta “¿Puede este sistema pensar?” sin recurrir a un criterio antropocéntrico, he recurrido a:

- La teoría de los rastros cognitivos y la explicación de la vida como sistemas autopoyéticos para establecer las condiciones para que se dé un sistema que tenga contenido mental.
- El comportamiento descrito de la avispa *Sphex* para encontrar lo que diferencia actos de pensar de otras formas de cognición que permiten la operación en un ambiente.
- A la idea de los problemas de diseño como la herramienta que permite al evaluador plantear una prueba en los términos del sistema y que permite medir *el pensar*.

Para el planteamiento concreto de la prueba, el proceso por parte del evaluador sería entonces:

Cómo los toma el evaluador	Rastros cognitivos	Representa del evaluado	Se determina del problema	Variable
Fijo	Aparato percepto-motor (b) condiciones de autosustentación (e)	Realización material	Circunstancias de diseño	Características
Independiente	Experiencia del agente (c) preferencias (e)	Perspectiva	Tipos de problemas	Preferencias
Supone su solución	Construcción de rastros cognitivos hacia una meta	Posibilidades de actuar	Problema de diseño	Modificaciones ambientales
Evaluación	Desenvolverse en el ambiente para alcanzar la meta (a, d) sin ir en contra de sus preferencias (e)	Pensar	Solución	Adaptación Cognitiva (AC)

Tabla 1 Proceso de planteamiento por parte de un evaluador de un test de pensar para un sistema cualquiera.

La anterior tabla permite articular los elementos anteriormente enunciados en un procedimiento concreto para la elaboración de la evaluación para un sistema en particular. La primera columna señala cómo deberá aproximarse el investigador a las propiedades de los rastros cognitivos que se señalan en la segunda, que dan cuenta de las capacidades del sistema escrita en

la tercera. Esta información permitirá obtener información relevante para el planteamiento del problema de diseño de acuerdo a lo señalado en la cuarta columna y, al momento de llevar a cabo la evaluación de la prueba se verá representada en los elementos de la última columna. Estas etapas del proceso permitirán diseñar y ejecutar un proceso de evaluación que responda a las características de un sistema particular respondiendo la pregunta por sus capacidades de pensar.

En una primera etapa se determinarían las condiciones fijas del individuo a ser evaluado, esto es, aquello que no tiene las posibilidades de cambiar sobre sí mismo. Los receptores sensoriales con los que cuenta, sus capacidades de interacción con el mundo (movimientos y acciones que puede iniciar en el mundo), los requisitos para conservar su vida (y condiciones que la limitan) y los mecanismos que la sustentan se considerarán como parte de estas condiciones. Hay que aclarar que este conocimiento no tiene que ser específico sino dado por funciones y relaciones; se puede saber que un sistema es sensible a luz, que puede desplazarse hacia las fuentes de ella y que requiere de un medio rico en oxígeno para subsistir sin que por ello tenga que conocerse una descripción a fondo de cómo funcionan cada uno de estos aparatos (a nivel físico y anatómico) ni por qué son necesarios para el sistema. En general, esta etapa coincide con el conocimiento de que la biología y sus mecanismos nos hayan otorgado sobre el sistema. En esta etapa se deberá considerar que tanto el evaluador como el evaluado no tienen control sobre estas condiciones y, por lo tanto, el problema de diseño que se le presente a resolver no podrá ser uno que atente contra ellas. Esta etapa, por lo tanto, permite determinar la perspectiva del agente, aquello que puede detectar del ambiente y cómo puede construir en él su experiencia y mundo. De esta manera se entenderán las especificaciones de diseño del agente, de cuáles características de la complejidad ambiental él puede llegar a dar cuenta y, por ende, las circunstancias para las que no está diseñado. En esta etapa se establece, por ejemplo, que tiene capacidad para desplazarse sobre un terreno plano, que es sensible a variaciones de luz, puede detectar olores y puede utilizar sus manos para agarrar cosas; también se determina que no podrá ser un problema de diseño obligar a un humano a vivir sin oxígeno o ver directamente cambios en la iluminación ultravioleta en la medida que no es parte de sus funciones.

El segundo nivel otorga información que para el investigador es independiente de él pero que le permitirá establecer cuáles son los tipos de problemas que podrá plantear. En él se reconocen las motivaciones particulares del sujeto que se está evaluando, son motivaciones que se detectan dentro de la acción y que corresponden a las características del primer nivel. Al observar los problemas del ambiente que resuelve con su actuar y ver cómo lo hace en diferentes circunstancias, se puede establecer para qué está diseñado²⁰ este sistema. En la mayoría de los

²⁰ Tomando la perspectiva del diseño de Dennett.

seres vivos que conocemos podemos observar que están diseñados para ciertas cosas específicas como: reproducirse sexualmente, vivir en un ecosistema boscoso, obtener cierta clase de alimento, proteger a sus crías, resolver los problemas en grupo o en solitario, etc. Al identificar los problemas de diseño que efectivamente resuelve el sistema dentro del ambiente podremos identificar la perspectiva que tiene el agente respecto al ambiente, qué es lo relevante para él y por qué lo buscaría, entenderíamos las funciones que satisface el sistema; por lo tanto, esto informará sobre los tipos de problemas que se le podrán plantear en la tercera etapa.

De esta manera, en los dos primeros niveles se tiene una descripción del sistema como diseñado para hacer algo. Al entender las circunstancias desde las cuales el agente puede producir soluciones a un rango de problemas, se tiene una descripción que en el tercer nivel podrá utilizarse para plantear el problema. Las tres reglas mencionadas en la sección anterior deberán ser respetadas para que el problema tenga sentido para el agente y pueda siquiera resolverlo: deberán respetar sus condiciones fijas, presentarse de acuerdo a modificaciones ambientales que estén en su perspectiva y estas modificaciones deberán ser novedosas. Si bien la novedad no es algo que se pueda garantizar, una manera efectiva de lograrla es mediante el conocimiento del ambiente en el que se desenvuelve el agente e incluyendo en él elementos externos o modificando sus regularidades. Si las modificaciones ambientales tienen la consecuencia de afectar la función que se ha encontrado en el nivel dos y estas pueden ser operadas de acuerdo al nivel uno, entonces estas modificaciones serán tales que establecen un problema de diseño para el sistema.

Este problema presenta ante el sistema elementos que previamente no habían estado dentro de su cognición pero que afectan directamente el área de competencia del sujeto. En las etapas uno y dos los rastros cognitivos entre el individuo y la meta se encontraban sólidamente establecidos en la medida que efectivamente operaba de acuerdo a ellos en el mundo. La modificación, por su parte, deberá afectar directamente la relación entre este sujeto y la meta ya sea obstruyendo los rastros previamente establecidos o señalando una nueva meta clara que esté en su rango de competencia a través del problema que se plantea. Es por esto que se considera que solucionar este problema de diseño permitiría mostrar que, relevantemente, cumple con las características señaladas en la lista de propiedades de pensar.

Ya contando con el problema, en el cuarto nivel se lleva a cabo la ejecución y valoración de los resultados del test. Así, el investigador aplica las modificaciones ambientales que se plantearon como un problema en tres y observa los resultados que tiene el agente para solucionarlo. Todo problema de diseño tendrá la posibilidad de presentar una solución nula (una ausencia de solución): cuando el individuo no da cuenta de manera alguna del problema de diseño que se le ha planteado. También existe una (o múltiples) opciones en las que se obtiene un resultado adecuado:

se da cuenta completa de la problemática planteada por el entorno de acuerdo a los intereses del individuo sin afectar negativamente sus condiciones. Sin embargo, también pueden existir soluciones parciales a los problemas de diseño cuando, por ejemplo, las modificaciones ambientales generasen un problema complejo y solo se da cuenta de parte de sus elementos cuando la solución del sistema termina por afectar negativamente sus capacidades o en el caso en el que las preferencias que hemos detectado en el sujeto no se mantengan.

Para mostrar esta clase de respuestas tomemos el ejemplo de la avispa *Sphex* ligeramente modificado: la modificación ambiental consistirá en reemplazar el grillo con uno que no esté incapacitado y alejarlo de la entrada del nido cada vez. Una solución completa a este problema sería, por ejemplo, capturar al grillo, ingresarlo a la madriguera, verificar que no deje de estar paralizado, poner los huevos, cerrar la madriguera e irse; en este caso da cuenta tanto de la posibilidad de fallo de su veneno como de cumplir la función reproductiva. También existen fallos absolutos que podrían presentarse como por ejemplo que la avispa siguiera incesantemente capturando grillos hasta morir en el proceso o que simplemente ponga sus huevos sin un grillo como fuente de alimento. En ambos casos no se soluciona el problema, pero por razones diferentes: en el primero fracasa ya que anula todas sus características constitutivas al momento de morir, en el segundo lleva a cabo la función reproductora de forma que no será efectiva e ignora todas las modificaciones ambientales. Así mismo se podrán tener resultados mixtos como, por ejemplo, uno donde introduzca el grillo en la madriguera sin verificar si el veneno ha sido o no efectivo y continúe con el proceso o que reconozca la situación como una en la que no podrá reproducirse y simplemente abandone el acto. En la primera circunstancia solo da cuenta de una de las dos modificaciones que estaban ocurriendo en el ambiente, mientras que en la segunda, aunque reconoce las modificaciones ambientales su respuesta ante ellas es abandonar su preferencia de reproducción para conservar todas las demás.

Como se observa, cada una de las diferentes condiciones de diseño que se han reconstruido para los agentes cognitivos pueden verse afectadas al momento de resolver un problema. Sin embargo, estos resultados como observamos anteriormente pueden ser evaluados e, idealmente, la solución completa no afectará negativamente al sistema mediante su ejecución. Por esto, propongo que la capacidad de pensar en la solución a un problema de diseño dado se puede medir para cada uno de los sistemas. En la última columna de la tabla se señalan cómo las variables están relacionadas con cada uno de los pasos en los que se ha construido el problema. Los elementos clave entonces a tener en cuenta serán las características del individuo (C), las

modificaciones del entorno que se plantean para generar el problema²¹ (M) y las preferencias (P) del sistema que propongo pueden evaluarse de manera cuantitativa de la siguiente manera:

$$\text{Adaptación Cognitiva (AC)} = (C_f/C_i) * ((P_t - P_{ins})/P_t) * (M_r/M_p)$$

Donde f significan finales, i iniciales, r resueltas, p propuestas, t totales e ins insatisfechas. Esta fórmula está pensada de tal manera que el resultado pueda desplazarse entre 0 y 1, siendo 0 un caso en el que el sistema falló al momento de plantear una respuesta al problema de diseño mientras que 1 es una respuesta completa al mismo. Este resultado se obtiene de computar aquello que ocurrió con las tres variables en el proceso de resolución del problema de diseño por parte del agente. Por una parte, se mide que no se pierdan características de su realización material, que la solución planteada no afecte sus condiciones de vida (las capacidades de percibir el ambiente, actuar en él y mantenerse con vida); por esto si las características finales y las condiciones iniciales son las mismas se considerará que el valor dicho componente es 1. Sobre las preferencias, se entiende que el sistema podrá tener un amplio rango de preferencias, sin embargo, en este caso solo se deberá tener en cuenta el universo de preferencias que puede ser afectado por el problema de diseño planteado -aquellas preferencias que desde la perspectiva del agente convierten la modificación en un problema- y observar cuáles no son satisfechas en la solución; esta medida también se da porcentualmente para resaltar la importancia relativa respecto al problema y en caso de satisfacer todas las necesidades el valor también será 1. Finalmente, la variable de las modificaciones indica claramente el porcentaje de los problemas propuestos a través de modificaciones ambientales que han sido resueltos, siendo la resolución total 1. De esta manera existe la posibilidad de que se den situaciones como el test de Sphex modificado que planteé y las posibles acciones por parte de la avispa; sin importar las combinaciones de cómo se produzca la acción, estos componentes mostrarán si la acción dada es una solución al problema de diseño planteado.

Es claro entonces cómo en los valores extremos de AC (1 y 0) el test funciona: si por alguna razón la solución planteada por el sistema acaba por destruirlo, si la solución es plenamente insatisfactoria de acuerdo a las preferencias que se presentan en él o si no resuelve ninguna de las

²¹ Para poder contar estas modificaciones se deben tener en cuenta como tipos generales y no casos particulares (token) de la modificación. Por ejemplo, en la modificación ambiental de la avispa Sphex, se tiene en cuenta que es un tipo general, cambio de posición del grillo, y no una lista de múltiples modificaciones (Se desplaza tantos centímetros hacia la derecha de la cueva, se aumenta la elevación en tantos otros, se gira respecto a la posición inicial en ciertos grados). Al llevar a cabo la medición de esta manera se permite una consistencia en lo que se considera como una modificación y se evita el problema de delimitar cuáles de las múltiples propiedades físicas que se verán alteradas por cualquier modificación ambiental propuestas son aquellas que serán relevantes (por ejemplo, no tendría siquiera sentido preguntarse si una modificación ambiental es el cambio en la temperatura a la entrada de la cueva por el desplazamiento del grillo).

modificaciones entonces el resultado total será inmediatamente 0. Los valores intermedios, por su parte, deben ser interpretados. En particular, este valor será sensible a la complejidad de los sistemas y los problemas planteados: dado que en cada una de las variables se mide su resultado en términos porcentuales, entonces ante un mayor valor inicial se da una mayor tolerancia a errores. Un agente con un aparato percepto-motor sumamente complejo con un rango de preferencias amplio podrá sacrificar una de sus partes para la resolución de un problema sencillo y el resultado no se verá sumamente afectado. Por otra parte, para un agente mínimo que solo cuente con una forma de percepción y una forma de acción para dirigirse al problema, llevar a cabo esta clase de sacrificios podría hacer que el valor se redujera radicalmente.

Aun bajo este defecto, hay tres maneras en que AC intermedios pueden ser utilizados: para comparación en la resolución de diferentes problemas por parte del mismo sistema, para identificar la posibilidad de aprendizaje por parte del sistema y para la comparación de diferentes sistemas en problemas planteados en espacios delimitados formalmente (juegos). A un mismo sistema se le pueden proponer diversos problemas de diseño ya sea porque se quiera evaluar la generalidad de su capacidad de pensar o para intentar encontrar información significativa sobre la clase de problemas en los que efectivamente este agente puede pensar. En este caso, entonces, se podrán tener experiencias directamente comparables en las que la diferencia en los valores solo dependerá de que tan completamente pueda maximizar la solución al problema ya que tanto P_t como C_i podrán ser tomadas como constantes en cada ejecución. Por otra parte, el mismo problema podría plantearse diversas veces al mismo agente y, al comparar los resultados, observar si hay un proceso de aprendizaje por parte del agente que, aunque no resuelva nunca el problema, podrá acercarse cada vez más a la solución. Finalmente, dentro de los problemas formales, los diferentes sistemas podrán desenvolverse de tal manera tanto sus P_t como C_i sean equiparadas. Al momento de múltiples sistemas operar en el ajedrez (y los problemas de diseño que señalamos previamente que podrían generarse para este mundo formal en la página 26) podemos suponer que, en general, los organismos tienen la misma capacidad de agencia ya que todos puede actuar a través de las fichas, lo que se puede percibir son los estados del tablero y las preferencias están dadas por las metas explícitas del juego. Si el sistema tiene la capacidad de desempeñarse dentro de ese mundo formal, entonces se podrá comparar directamente dos diferentes y determinar cuál puede tener más C sobre el problema.

De esta manera el evaluador cuenta con una herramienta poderosa para “¿Puede este sistema pensar?”. Ante problemas de diseño bien planteados²², si la AC obtenido es 1, entonces el

²² Un problema bien captado quiere decir que ha captado las propiedades de diseño relevantes adecuadas del sistema que se está evaluando y que, por lo tanto, efectivamente es un problema de diseño para él. Si el problema no está bien planteado, los resultados pueden ser irrelevantes tanto al ser positivos

investigador podrá decir con claridad que puede hacerlo y, ante un caso de 0, podrá decir que no lo hace. Pero puede obtener otros tipos de respuestas en los que se obtiene mucha información adicional ya que cuando el sistema es complejo le podrán ser planteados problemas que evalúen diversas funciones para las que está diseñado; en estos casos la pregunta puede que tenga una respuesta parcial “Puede pensar en problemas de este tipo” o un “no puede resolver completamente ningún problema que le hayamos planteado, pero este tipo es el que se encuentra más cerca de poder hacerlo”. También la herramienta le permite comparar el desempeño de diferentes tipos de sistemas dentro de marcos formales que le permitirá comparar la inteligencia de cada uno.

como negativos; una AC de 1 cuando el problema estuvo mal planteado puede entonces no capturar una situación novedosa del ambiente (por lo que no estaría estableciendo un rastro cognitivo hacia una nueva meta y por lo tanto no se está necesariamente llevando el proceso de pensar). Por otra parte, un resultado de 0 en un problema mal planteado puede explicarse, por ejemplo, porque el agente no tiene la capacidad de dar cuenta de las modificaciones ambientales en su experiencia y es incapaz de resolverlo.

8 Conclusiones

Si aceptamos esta perspectiva sobre lo que es pensar y la manera de evaluarlo nos encontramos con una relación interesante: al evaluar el pensamiento en un sistema le debemos plantear problemas de diseño los cuales vayan más allá de las circunstancias materiales para las que ha sido diseñado. Aun cuando el problema se plantea de tal manera que el sistema pueda dar cuenta de él en sus propios términos, el requerimiento de novedad respecto a la modificación ambiental plantea una circunstancia que va más allá de las condiciones efectivas para las que ha sido diseñado el sistema. Si el agente responde a esta clase de situaciones, entonces él está llegando a un punto que va más allá del nicho particular en el que se ha desarrollado y puede dar cuenta de cuando menos un problema que no se había tenido en cuenta.

El poder llevar a cabo esta acción representa para cualquier sistema el romper las limitaciones de la cotidianidad e identificar si él es capaz de explotar sus capacidades en una manera que le permita desenvolverse en marco más general que el dado por su naturaleza. La avispa *Sphex* que es incapaz de entender la modificación que se lleva a cabo en el ambiente, la rana que reacciona ante el pasar de un avión sacando la lengua o la persona que solamente es capaz de producir una conversación estereotípica²³ se encuentran en la misma situación. Todos y cada uno de ellos cuentan con respuestas que parecen conectar adecuadamente un estímulo del ambiente con una acción por parte de ellos que da cuenta de la situación, pero al mismo tiempo, todos son incapaces de realmente pensar en dicha situación. Cuando las condiciones del ambiente cambian de una manera inesperada, esto es, cuando se presenta una situación novedosa que logre retar los estereotipos dados, un nuevo problema de diseño para el sistema, en ese caso se puede presentar un colapso y una incapacidad de dar cuenta de los problemas. De esta manera, en todos estos casos existe una barrera marcada entre el mundo del sujeto y el ambiente: aun cuando eventos del ambiente tendrían la posibilidad de ser percibidos por parte del individuo, estos hechos no son parte de su mundo. Por otra parte, cuando se aprueba el test planteado, entonces se tiene la relación contraria: a pesar de que el ambiente nunca le había otorgado una situación determinada, al momento en que surge el sistema puede incluirla dentro de su mundo. Al resolver el problema de diseño podemos intuir que, sin necesidad de conocer los mecanismos internos con los que el agente ordene su mundo, este tiene cuando menos la posibilidad de relacionarse de manera efectiva con el ambiente en general.

²³ Aquellas conversaciones en las cuales más que una participación activa se sigue simplemente un protocolo. A: "Hola", B: "Hola ¿cómo estás?", "A: Bien ¿y tú?", B: "Bien, gracias ¿qué me cuentas", ..., o actos similares en los cuales las reacciones a los estímulos del ambiente (en este caso las reacciones de con quien se interactúa) se encuentran ya dados para la persona.

No por esto se garantiza que el sistema pueda dar cuenta de toda la complejidad del entorno ni que todo fenómeno del universo podrá ser comprensible por él y, sin embargo, en esta capacidad básica podemos observar la posibilidad del sistema de dar cuenta de una parte amplia del universo. En la medida en que se desarrolle el test y que se lleve a cabo la aplicación de múltiples problemas de diseño es posible que encontremos un borde superior en las capacidades de pensar de dicho sistema, sin embargo, ante la simple resolución de uno de estos problemas bien planteados ya podemos entender que al menos este sistema tiene una flexibilidad que le permitirá responder a necesidades del medio con las que nunca antes se hubiera topado.

Por esto se puede considerar entonces que el acto de pensar sería un acto de trascendencia para el individuo, sería ir más allá de aquello a lo que se ha tenido que enfrentar usando para ello todas sus capacidades. A pesar de que la experiencia, la percepción, las limitaciones materiales y motoras de un individuo se tienen como fijas y las aceptamos como una condición sobre la que no podemos solicitar ir más allá a ningún individuo; pero al momento de articular estas capacidades para solucionar un problema para el que inicialmente no ha sido diseñado, él logra salir y superar su entorno dado. El sujeto que apruebe este test tiene al menos una ventaja comparativa con uno que no lo apruebe bajo ninguna circunstancia.

Así, no solamente se está captando en esta definición el uso meramente descriptivo de pensar como un proceso cualquiera, sino que también podemos verlo como el juicio de valor que emitimos al decir “bien pensado” o “al menos tómese la molestia de pensar en el problema”. En ambos casos reconocemos que pensar no es solo algo que siempre ocurre en la vida humana, en nuestra cotidianidad o en el lenguaje, sino que al hacerlo logramos superar las condiciones con las que normalmente nos encontramos. Al pensar en nuestra vida cotidiana realizamos un esfuerzo para acomodar todas nuestras herramientas disponibles, las capacidades con las que contamos como humanos y como sujetos particulares de nuestra sociedad, a una situación para la cual no conocemos la respuesta.

Este es un acto admirable de pensar. Un acto que, al poder entenderse en todo sistema posible, deja de ser simplemente una actitud narcisista humana en la que nos vemos como aquellos capaces de hacerlo. Sin embargo, de acá se deriva un grave problema de aceptación de la perspectiva como una evaluación universal sobre el pensamiento: dado que el problema de diseño depende del planteamiento de un observador y el resultado de AC depende de la apropiación por parte del sistema del problema de acuerdo a sus propias preferencias, entonces parece que cualquier pretensión de objetividad se pierde. En efecto, no parece que haya algo concreto en el mundo que pueda observarse o medirse dentro del mundo de tal manera que aplique a todo sistema posible que ha sido evaluado; el simple hecho de plantear el test se convierte en una

prueba para la creatividad y la inteligencia humana para poder entender correctamente el diseño del sistema, sus características, la historia evolutiva que lo ha traído hasta el presente y el entorno en el que se ha desarrollado característicamente también es el resultado de un acto cognitivo por parte del investigador particular que se enfrenta al problema de resolver la pregunta.

Pero siguiendo la teoría de los rastros cognitivos, esta es una posición congruente. Al momento de establecer una S/Objetividad se está alcanzando lo que permite a un sistema particular, como la del investigador, relacionarse y desenvolverse en el entorno para construir el contenido mental. Esta S/Objetividad es la naturaleza del conocimiento que el sistema puede alcanzar, donde por una parte está el contenido no conceptual que surge de la relación directa del aparato percepto-motor con el entorno y, por el otro, las generalizaciones que se pueden construir a partir de él a nivel conceptual. El proceso de construcción del test busca ser congruente con esto al momento de plantear una manera de reconocer si algo “piensa”: por una parte, se reconoce que este concepto parte de generalizaciones llevadas a cabo por nosotros como investigadores sobre las diferentes experiencias cognitivas que se han tenido y que, por lo tanto, es posible enriquecerlo mediante la experiencia de evaluar posibles agentes encontrados en el entorno mientras que, por la otra, será necesario que el test pueda ser pertinente para la S/Objetividad que se está evaluando. De esta manera, el proceso de planteamiento del problema no es subjetivo -no es simplemente algo que surge desde la perspectiva individual- sino que es sujeto-objetivo, permitiendo al investigador dar cuenta desde su punto de vista de la información que se encuentra en el entorno de tal manera que se pueda incorporar a su cognición y le permita observar la forma en que este proceso ocurre en un sistema distinto de él.

9 Apéndice: Otras vías, ventajas y desventajas de esta posición.

Esta propuesta se enmarca en general en los proyectos que buscan medir la inteligencia general de cualquier sistema que pueda ser catalogado como tal; por ello considero importante observar la manera en que este test se compara con otros y las ventajas o desventajas que tiene respecto a otras técnicas. Para esto utilizaré las siguientes categorías tomadas de “Computer models solving intelligence test problems: Progress and implications” (Hernández-Orallo, Martínez-plumed, Schmid, Siebers, & Dow, 2016) las cuales organizan en general las diferentes propuestas que se han dado para medir la inteligencia en máquinas, tomando esto como parte del proyecto general para “observar, identificar y finalmente medir la inteligencia de humanos, animales no humanos y máquinas [...]

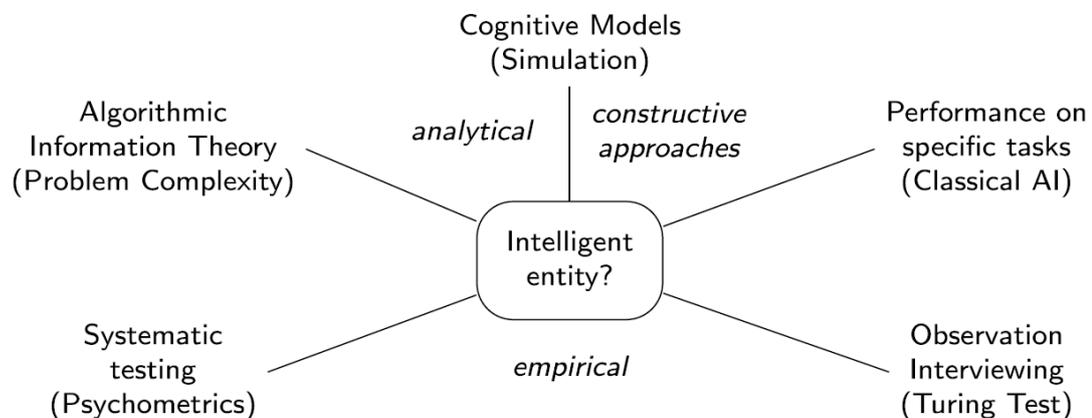


Fig. 1. Different approaches to assess the intelligence of an entity.

Ilustración 1 Diferentes aproximaciones para medir la inteligencia de una entidad” (Hernández-Orallo, Martínez-plumed, Schmid, Siebers, & Dow, 2016, pp. 76-77)

Los modelos cognitivos, en la medida en que buscan reconstruir comportamientos inteligentes tales y como se presentan en nuestra cognición, el desenvolverse en tareas específicas como se buscaba en la IA clásica y las pruebas sistemáticas, que son aquellas similares a la prueba de IQ que aplican un rango de pruebas predeterminadas para determinar si el agente es inteligente, serán unas que de inmediato no responden a los intereses acá planteados para una noción de pensar. Cada una de estas involucra directamente los problemas del antropocentrismo y la automatización sin creatividad. El primero de los problemas se presenta en la medida en que cada una de estas son nociones que han surgido desde nuestra experiencia: en la aproximación de los modelos cognitivos se formaliza algorítmicamente un comportamiento que en humanos requiere creatividad, como el lenguaje, y se compara el desempeño de la máquina con el de las personas que lo ejecutan; cuando exponemos un juego particular o una actividad como criterio de pensar, lo hacemos desde la intuición de que una persona desempeñándose en este medio requeriría de

labores cognitivas como las que nosotros necesitamos para hacerlo, ignorando las diferencias en capacidades de los sistemas y lo que esto implica para el pensar²⁴; por su parte en las pruebas establecidas es aún más claro que el proceso es traspasar una herramienta diseñada para medir el comportamiento en humanos y aplicarlo indiscriminadamente a otros sistemas. Esta es precisamente la estrategia que criticamos en la sección dos como una manera inadecuada de proceder, es partir de las contingencias de la cognición humana para dictaminar lo que esto significaría para toda clase de mente. Aun si no se acepta este interés propio de investigación, el hecho de que esta clase de pruebas involucren tanto la modelación como la evaluación de una o un conjunto de pruebas sólidas y determinadas presenta un riesgo de comportamientos no creativos como los de la avispa *Sphex*. Efectivamente un autómata que pueda responder a los requerimientos que se planteen en cada una de estas estrategias será considerado como exitoso e incluso, particularmente en el caso de la modelación cognitiva, este carácter algorítmico será precisamente aquello que querrá lograr la teoría. Pero como anteriormente se señaló, esta manera de proceder respecto a un problema propuesto por el ambiente puede terminar con dos clases de colapso diferentes: o no dar cuenta de problemas del mismo tipo los cuales podrían ser evaluados dado lo que se conoce del organismo, o seguir en un ciclo interminable que vaya en contra de sus propias preferencias.

En estos casos parece que lo propio del pensar y de la inteligencia no se encontrará en el sistema que lleva a cabo la acción, sino que fuera una propiedad del entorno. No es el acto de poder dar cuenta de una forma particular de la complejidad ambiental lo que determinará la capacidad de pensar del sistema, tampoco es relacionarse de una manera particular con los posibles estímulos que puede encontrar en el entorno e incorporarlos en su mundo. En pos de buscar la objetividad al momento de determinar una prueba de pensar, termina por olvidarse que esta es una propiedad del sujeto. Sin embargo, este es defecto en el que no caen los dos procesos restantes.

Por una parte, el proceso de observación y entrevista es precisamente aquel que se manifiesta en el test de Turing y en el que encajaría AC. La principal diferencia que se tiene con los métodos anteriormente planteados es la forma en que se concibe la entrevista; al determinar que en este test se lleva a cabo mediante el planteamiento de problemas de diseño a través de modificaciones ambientales se establece una forma de comunicación efectiva entre el sujeto entrevistado y el entrevistador sin que por ello se presuponga una forma lingüística cualquiera. Así mismo, este test permite conservar una de las principales ventajas del planteamiento de Turing, a saberse, que “brinda una salida para evadir la necesidad de definir un concepto de pensamiento al

²⁴ Aquello que ha llevado a la continua discusión sobre lo que representa DeepBlue para el problema del ajedrez (Rasskin Gutman, 2005).

solamente considerar la relación entre los inputs (las preguntas) y los outputs (las respuestas)” (Hernández-Orallo, Martínez-plumed, Schmid, Siebers, & Dow, 2016, p. 77). Del mismo modo AC logra relacionar las entradas, modificaciones ambientales, con las salidas, acciones por parte del sistema, de tal manera que se pueda evaluar si en dicha relación se encuentra lo propio del pensar; para ello no se requiere conocer la perspectiva individual del agente, sino que esta puede ser inferida desde el planteamiento de un problema de diseño que él logra notar y dar cuenta del mismo.

Por lo tanto, viendo como el presente proyecto es un desarrollo de la perspectiva de observación y entrevista solo restaría preguntarse por los modelos desde la teoría de la información. En particular debido a que estos modelos comparten el principal interés de investigación con esta propuesta, a saberse, “¿Cómo entonces podemos nosotros desarrollar un concepto de inteligencia que sea aplicable a toda clase de sistemas? Cualquier definición propuesta debe cubrir la esencia de la inteligencia humana, así como otras posibilidades, de una manera consistente” (Legg & Hutter, 2006). Metodológicamente, estas propuestas también parten de una definición común de inteligencia y a partir de allí construyen un criterio desde el cual determinar si cualquier clase de sistema en el universo piensa o no. Me enfocaré principalmente en la discusión con el texto “A Formal Measure of Machine Intelligence” de Shane Legg y Marcus Hutter que tiene precisamente el objetivo anteriormente señalado y, por lo tanto, parecería lograr con su evaluación propuesta sobre “¿puede este sistema pensar?” lo mismo que mi presente propuesta. La principal diferencia que encuentro con dicha postura se encuentra en los presupuestos teóricos que se utilizan para construir la evaluación: mientras que yo propongo una reconstrucción genética de los agentes mínimos que pueden pensar acá se busca encontrar en una descripción uniforme del universo en términos informacionales el intercambio que es requerido entre el agente y el entorno para considerar que este está llevando a cabo un acto inteligente.

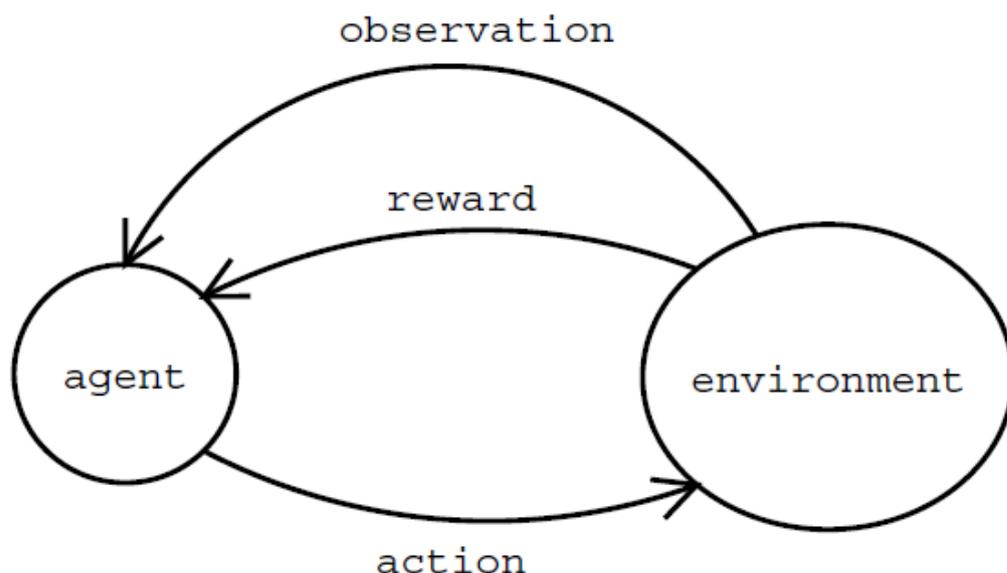


Ilustración 2 El ciclo de la información para el pensamiento en (Legg & Hutter, 2006)

El agente envía información al ambiente enviando símbolos de un conjunto finito, por ejemplo, $A := \{\text{izquierda, derecha, adelante, atrás}\}$. Nosotros denominaremos este conjunto el espacio de acción y lo denotaremos como A . Similarmente, el ambiente envía señales al agente mediante símbolos de un conjunto finito llamado el espacio de percepción, que denotaremos con P . El espacio de recompensa, denotado por R , siempre será un subconjunto finito del intervalo de unidades racionales $[0,1] \cap \mathbb{Q}$. Cada percepción consiste en dos partes separadas; una observación y una recompensa. Por ejemplo, podríamos tener $P := \{(\text{frío}, 0.0), (\text{tibio}, 1.0), (\text{caliente}, 0.3), (\text{ardiendo}, 0.0)\}$. [...] esto produce una historia de observaciones, recompensas y acciones que denotaremos como $o_1, r_1, a_1, o_2, r_2, a_2, \dots$ [...] El ambiente, denotado con μ , es definido en una forma similar. Específicamente, para cualquier $k \in \mathbb{N}$ la probabilidad de $o_k r_k$, dada la actual historia $o_1 r_1 a_1 \dots o_{k-1} r_{k-1} a_{k-1}$, es $\mu(o_k r_k | o_1 r_1 a_1 \dots o_{k-1} r_{k-1} a_{k-1})$. Por el momento no imponemos ninguna otra restricción en el ambiente. (Legg & Hutter, 2006, p. 4)

Esta es la formalización desde la que opera el test. En particular llamo la atención sobre los siguientes puntos: el conjunto de acciones, observaciones y recompensas que se pueden obtener del ambiente se encuentra listado y detallado, el ambiente debe ser descrito de manera completa y todas estas propiedades se establecen como objetivas y medibles. De estas suposiciones, encuentro que hay una dificultad central que busca evitar mi planteamiento del test: requiere un conocimiento mayor sobre las capacidades del sistema evaluado de los mínimos necesarios para preguntarnos por sus capacidades cognitivas.

Tomar como un listado explícito y detallado todo el ciclo de la agencia, los componentes de observación-recompensa-acción, parece algo similar a lo que se solicita en la Tabla 1 (pág. 33) como un presupuesto para poder llevar a cabo una realización de test. En efecto, el incorporar las posibles observaciones y acciones por parte del sujeto podría equipararse a llevar a cabo una descripción de las capacidades del aparato perceptomotor del sistema; sin embargo, creo que considerarlo como un conjunto finito descrito por una función computable hace que este test se pueda aplicar con mucha mayor facilidad a IA que a otros sistemas inteligentes. En una inteligencia

artificial, la descripción de las posibles acciones, así como la probabilidad de la toma de una decisión particular dados los eventos anteriores, puede medirse con claridad gracias al acceso que tenemos a su programación; y a pesar de que podría “argumentarse que todo en nuestra realidad está hecho de información” (Vlatko, 2010) y, por lo tanto, todo sistema posible podría ser descrito en los términos requeridos para poder llevar a cabo la evaluación, esto no implica que contemos con dichas descripciones para los sistemas que encontramos en el entorno natural. Un ejemplo claro de esto es nuestra propia experiencia, sin importar que hasta el momento no contamos con una descripción completa en términos probabilísticos de la agencia humana, no por ello hemos carecido de la posibilidad de reconocer el rango de acciones que podemos ejecutar. Por ello, adoptar una posición desde la teoría de la información traería consigo un inconveniente similar a lo planteado por Dennet, en este caso se estaría creando una “actitud informacional” que se encontraría más cerca de la actitud física que de la del diseño y mucho menos de la intencional²⁵. Y de la misma manera en que “si sabes algo sobre el diseño de un artefacto, puedes predecir su comportamiento sin preocuparte a ti mismo sobre las leyes físicas subyacentes que rigen sus partes [...] solo los diseñadores tienen que entender la física” (Dennett, 1996, p. 229) para poder predecir cómo se comporta un sistema del cual podamos decir que es inteligente no tenemos que reconocer ni la física ni el flujo informacional subyacente; esto claramente ocurrirá en cualquier sistema cognitivo y solo quién tenga pretensiones de construir uno necesitará entender las leyes que le subyacen.

Si bien las recompensas como un elemento que existe en el ambiente pueden parecer de inmediato problemáticas, en su sentido “estricto, la recompensa es una interpretación del estado del ambiente” (Legg & Hutter, 2006, p. 5), por lo que, aun cuando en el modelo se encuentran establecidas en el exterior, lo que se indica es que el estado de cosas en el entorno puede o no estar de acuerdo con las preferencias establecidas por el sistema. A pesar de este acuerdo, la principal diferencia se da nuevamente en el nivel de descripción y determinación requerido para operar correctamente el test. En el planteamiento que se hace del test por Legg y Hutter, será la capacidad de lograr efectivamente maximizar las recompensas así como el hacerlo mediante el proceso más sencillo posible (menor nivel de complejidad)²⁶ lo que determinará su resultado. Pero

²⁵ La cantidad de datos requeridos para llevar a cabo una predicción efectiva, así como las herramientas necesarias para producirla, son más detalladas y complejas que las requeridas para dar una explicación desde una noción de funciones o de intenciones.

²⁶ Este es un presupuesto teórico de la prueba planteada por Legg y Hutter con el que tampoco estaría de acuerdo. Dar una importancia fundamental a un criterio como el de la navaja de Occam (Legg & Hutter, 2006, p. 6) al momento de establecer los resultados sobre las capacidades cognitivas de un sistema impone una noción sobre cómo debe ser este pensamiento adicional. Cuando existan múltiples soluciones a un problema, si bien la más sencilla puede ser deseable, el hecho de que pueda acceder a una de las mismas será suficiente para aceptar que el sistema lo ha resuelto. Si el asunto que nos preocupa es la capacidad de

para poder llevar a cabo esta evaluación entonces es necesario conocer la descripción de la forma en que el agente interpreta los estados del ambiente, hacer explícita y conocer la forma en que las preferencias influyen en la toma de decisión. Esta información nuevamente se encuentra fácilmente disponible para aquellos sistemas de los cuales somos diseñadores, pero el proceso de llevar a cabo ingeniería reversa que permita reconocer con tal detalle las recompensas que puede obtener del ambiente tanto un sistema biológico dado como una forma de vida extraterrestre desconocida es un paso que requeriría ir más allá de donde se ha alcanzado el trabajo de la biología actual. Por esto en AC las preferencias son tratadas como un supuesto que es susceptible de ser revisado en el test ante evidencia contraria, no es necesario una información sólida para *suponer* que el sistema tiene intereses y que actuará de acuerdo a ellos. Sin embargo, para esto no es necesario contar con una medida detallada de los mismos, poder establecer un orden y prioridad entre ellas o siquiera conocer una lista detallada de las mismas; al limitar el espacio de competencia del test a una situación particular, solo aquellas preferencias que hayan sido relevantes para la modificación ambiental en las observaciones previas del sistema serán tenidas en cuenta.

Por estas diferencias en el nivel de descripción y de conocimiento requerido del sistema para poder operar la prueba general de inteligencia propuesta por Hutter y las necesarias para AC es que considero que una propuesta desde la teoría de la información no es realmente efectiva al momento de llevar a cabo una evaluación que sea realmente universal. Las descripciones requeridas desde esta actitud informacional son tales que, a pesar de que facilitarían llevar a cabo la evaluación para sistemas formales o IA desarrollados por humanos, dificultarían el proceso de preguntarse por las capacidades cognitivas de las formas de vida que nos encontramos en nuestro ambiente. Así, a pesar de que esta evaluación es lógicamente aplicable de manera universal, al momento de llevarla a cabo como un juicio práctico para estudiar los diferentes tipos de sistemas nos encontraremos que solo es efectiva dentro de un conjunto particular de los mismos: aquellos de los que contamos con información privilegiada.

Por estas razones considero que esta adaptación del sistema de observación y entrevista es precisamente la que permite responder a la pregunta ¿puede este sistema pensar? desde nuestra perspectiva como investigadores y sin por ello imponer nuestra manera de hacerlo.

Sin embargo, esta perspectiva tiene una dificultad, al menos al momento de ser propuesta como una prueba para evaluar sistemas artificiales, y es que no logra cumplir las funciones de una mosca de fruta para el desarrollo de entidades cognitivas. No sería posible diseñar una máquina cuya función fuese aprobar esta prueba en la medida que, para poder aplicarla, debería entenderse

un sistema para poder pensar, no nos importaría qué tan ineficiente sea la manera en que él se enfrente a los retos cognitivos planteados por una situación siempre que pueda hacerlo.

para qué es aquello que se encuentra diseñada, cómo se desempeña naturalmente y solo después de esta comprensión se le podría plantear un problema específico. Si un investigador quisiera partir de esta evaluación como un reto, de la misma manera en que programadores decidieron que era importante lograr construir sistemas artificiales que jugaran bien ajedrez o que aprobaran un test de Turing, no tendría manera de determinar *qué* puede hacer esta máquina. La flexibilidad del test permite que aun cuando la máquina esté diseñada para dar cuenta de un problema formal particular, ella pueda ser puesta a prueba.

Si fuese a cumplir alguna función de motivación y guía para el desarrollo de aparatos artificiales será informando sobre cómo se espera que pueda hacer las cosas que haga. Sin importar para qué esté diseñada una máquina, se esperará que para decir “esta máquina piensa” ella pueda articular sus capacidades perceptuales y motoras para dar cuenta de un espectro más amplio de problemas de los que haya sido específicamente concebida y, no por ello tenga que alcanzar unas pretensiones de universalidad sobre sus capacidades cognitivas. El reto, así, no pasa a construir una máquina particular que sea aquella que piense, sino la construcción de diversos tipos de máquinas que puedan pensar en problemas dentro de espacios de búsqueda independientes y que no necesariamente están correlacionados. Aun cuando todas puedan aprobar el test, una máquina diseñada para poder pensar en problemas del mundo de ajedrez, una para desplazarse en el ambiente y una para llevar a cabo actos comunicativos tendrán capacidades que serán totalmente distintas. Sin embargo, todas ellas podrán superar los límites de las situaciones con las que se han entrenado y ante las cuales llevan a cabo su quehacer cotidiano. Si hay una pregunta para el desarrollo de IA que se puede guiar mediante la aplicación de este test, ella no es “¿Qué deberá hacer una máquina para que sea considerada pensante?” sino “¿Cómo podríamos decir que esta máquina piensa?”.

10 Bibliografía

- Adams, D. (2002). *The Ultimate Hitchhiker's Guide to the Galaxy*. Del Rey.
- Almäng, J. (2008). Affordances and the Nature of Perceptual Content. *International Journal of Philosophical Studies*, 161–177.
- Bains, W. (2004). Many Chemistries Could Be Used to Build Living Systems. *Astrobiology*, 137-167.
- Bird, A., & Tobin, E. (2016). Natural Kinds. *The Stanford Encyclopedia of Philosophy*, Forthcoming
URL = <<http://plato.stanford.edu/archives/win2016/entries/natural-kinds/>>.
- Bostrom, N. (2016). *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.
- Bouchard, F. (2011). Darwinism without populations: a more inclusive understanding of the “Survival of the Fittest”. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 16-114.
- Clark, A. (1990). Connectionism, Competence, and Explanation. *The British Journal for the Philosophy of Science*, 195-222.
- Cussins, A. (1992). Content, Embodiment and Objectivity: The Theory of Cognitive Trails. *Mind*, 651-688.
- Dennett, D. (1973). Mechanism and responsibility. En T. Honderich (Ed.), *Essays on freedom of action*. London: Routledge.
- Dennett, D. (1996). *Darwin's Dangerous Idea*. London: Penguin Books.
- Dennett, D. (1998). *La Actitud Intencional*. Barcelona, España: Editorial Gedisa S.A.
- Dorst, K., & van Overveld, K. (2009). TYPOLOGIES OF DESIGN PRACTICE. En e. Anthonie Meijers, *Philosophy of Technology and Engineering Sciences* (págs. 455-489). Amsterdam: Elsevier.
- Fodor, J. (1975). *The language of thought*. New York: Thomas Y Crowell Company.
- Franssen, M., Lokhorst, G.-J., & van de Poel, I. (2015). Philosophy of Technology. *The Stanford Encyclopedia of Philosophy*, URL = <<http://plato.stanford.edu/archives/fall2015/entries/technology/>>..
- Frege, G. (1956). The Thought: A Logical Inquiry. *Mind*, 289-311.
- Gallagher, S. (2014). Phenomenology and embodied cognition. En *Routledge Handbook of Embodied Cognition* (págs. 9-18). London: Routledge.
- Gallagher, S., & Varela, F. (2001). Redrawing the Map and Resetting the Time: Phenomenology and the Cognitive Sciences. En *The Reach of Reflection: Issues for Phenomenology's Second Century*. (págs. 17-44).
- Gibson, J. (1986). *The Ecological Approach to Visual Perception*. Hillsdale and London: Lawrence Erlbaum.

- Godfrey-Smith, P. (1996). *Complexity and the*. Cambridge: Press Syndicate of the University of Cambridge.
- Google Inc. (2016). *Publications*. Obtenido de Deepmind: <https://deepmind.com/research/publications/>
- Gottfredson, L. (1997). Mainstream science on intelligence. An Editorial with 52 signatories, history, and bibliography. . *Intelligence*, 13-23.
- Haugeland, J. (1985). *Artificial Intelligence: The Very Idea*. Cambridge: MIT Press.
- Heidegger, M. (1994). ¿Qué quiere decir pensar? En E. (. Barjau, *Conferencias y Artículos*. Barcelona: Edición del Cerbal.
- Hernández-Orallo, J., Martínez-plumed, F., Schmid, U., Siebers, M., & Dow, D. L. (2016). Computer models solving intelligence test problems: Progress and implications. *Artificial Intelligence*, 74-107.
- Hofstadter, D. (1985). On the seeming paradox of mechanizing creativity. *Metamagical themas*, 526-546.
- Keijzer, F. (20xx). The Sphex story: How the cognitive sciences kept repeating an old and questionable anecdote. *University of Groningen*, http://www.academia.edu/4034267/The_Sphex_story_How_the_cognitive_sciences_kept_repeating_an_old_and_questionable_anecdote.
- Kroes, P. (2009). INTRODUCTION TO PART III. En *Philosophy of Technology and Engineering Sciences* (págs. 405-408). Amsterdam: Elsevier.
- Legg, S., & Hutter, M. (2006). A Formal Measure of Machine Intelligence. *IDSIA, Galleria 2*, 6928 *Manno-Lugano, Switzerland*.
- Luhmann, N. (1975). *The Differentiation of Society*. New York: Columbia University Press.
- O'Reagan, J. K., Myin, E., & Noë, A. (2005). Sensory consciousness explained (better) in terms of 'corporality' and 'alerting capacity'. *Phenomenology and the Cognitive Sciences*, 369-387.
- Rasskin Gutman, D. (2005). *Metáforas de ajedrez: La mente humana y la inteligencia artificial*. Editorial La Casa del Ajedrez.
- Rosenberg, A., & Bouchard, F. (2015). Fitness. (E. N. Zalta, Ed.) *The Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/archives/fall2015/entries/fitness/>.
- Searle, J. (1994). Mentes Cerebros y Programas. En M. (. Boden, *Filosofía de la inteligencia artificial*.
- Searle, J. (2002). Why I Am not a Property Dualist. *Journal of Consciousness Studies*, 57-64.
- Smith, D. W. (2013). Phenomenology. *The Stanford Encyclopedia of Philosophy*, URL = [<http://plato.stanford.edu/archives/win2013/entries/phenomenology/>](http://plato.stanford.edu/archives/win2013/entries/phenomenology/).
- Turing, A. (1950). Computing Machinery and Intelligence. *Mind*, 433-460.

Varela, F. (1992). Autopoiesis and a Biology of Intentionality. *Proceedings of a workshop on Autopoiesis and Percetion*, 4-14.

Vlatko, V. (2010). *Decoding Reality: The Universe as Quantum Information*. Oxford: Oxford University Press.

Wikipedia contributors. (10 de October de 2016). *List of chess variants*. Obtenido de Wikipedia, the free encyclopedia: https://en.wikipedia.org/wiki/List_of_chess_variants