

La verdad en un realismo epistemológico naturalista

Resumen de los aspectos centrales de una tesis doctoral
Prof. Carlos Alberto Garay

La afirmación central de nuestra tesis es: la verdad es una relación entre procesos que ocurren en el sistema nervioso humano y procesos que ocurren en el mundo. El mundo existe independientemente de cualquier individuo, es decir, existió antes de que existiera algún ser humano y, probablemente, continúe existiendo cuando éste ya haya desaparecido.

Intentamos proveer una explicación de lo que es "ser verdadero" en contraposición con lo que es "tener por verdadero" o "saber de algo que es verdadero". Estos últimos puntos sólo se tratarán incidentalmente.

Un concepto clave de la teoría es el de representación. Una representación, en el presente contexto, es un estado propio del sistema nervioso que se reconoce como originado a partir de la interacción de factores filo y ontogenéticos. Es algo así como una impronta surgida del choque entre el impulso causal heredado y los impulsos causales recibidos por el individuo del entorno en el que se ha desarrollado.

La verdad se predica primariamente de las representaciones y sólo por derivación, muchas veces engañosa, de otra clase de ítem.

La verdad acerca de un segmento espacio-temporal del mundo es siempre parcial. Los seres humanos no tenemos, como cuestión de hecho, acceso a todos los aspectos del universo.

No todas las cosas de las que tradicionalmente se predica la verdad tienen el mismo tipo de correlato neurofisiológico. Las proposiciones que se usan en la lógica y en la teoría del conocimiento no son homogéneas en lo que respecta a su instanciación en el sistema nervioso. Es particularmente importante notar que algunas de ellas han de analizarse recurriendo a unidades estructurales de distinto nivel, dependiendo de éstas algunas propiedades importantes en teoría de la verdad, como por ejemplo, la duración y permanencia de los procesos, la categorización de objetos, la atribución de propiedades y relaciones, la teoría de la referencia lingüística, la accesibilidad epistémica y el grado en el que guardan la relación apropiada con el mundo.

La expresión lingüística de una verdad es uno de los últimos procesos que deben analizarse, pues esconden una complejidad muy superior a la supuesta por la mayoría de los filósofos. Así, nuestra teoría no es una teoría del "decir verdad" ni de "afirmar algo verdadero". Dicho de otra manera, existe un desplazamiento extensional entre las categorías de la psicología popular (folk) y las provenientes de la neurociencia.

Es falso que pueda hablarse de la verdad de las proposiciones (o de las afirmaciones, oraciones, creencias, etc.) sin tener en cuenta las diferencias que existen entre ellas, más allá de la división en analíticas y sintéticas o en a priori y a posteriori.

Para comprender lo que se atribuye a una proposición toda vez que se dice de ella que es verdadera, es absolutamente necesario hacer referencia a los procesos que le dieron origen, es decir, a los procesos que le dieron existencia.

Muchos de estos procesos son comunes a la mayoría de los seres humanos. Pero otros son particularmente idiosincráticos.

Los filósofos se han dejado engañar por la similitud entre los signos lingüísticos y creyeron que cuando dos individuos pensaban, creían o afirmaban una misma proposición p , el ser verdadera de p , debía ser lo mismo para ambos.

Muy difícilmente existan dos individuos que compartan una y la misma creencia. Tan difícil como que tengan las mismas huellas digitales, en razón de la diversidad biológica a la que estamos sujetos. Pero es probable que puedan compartir creencias parecidas en algún grado, de la misma manera que nuestras huellas digitales comparten algunos patrones comunes.

Las teorías del significado de las palabras y de las oraciones, y también las teorías del conocimiento basadas en ellas, han enfrentado situaciones paradójicas y difíciles por utilizar supuestos psicológicamente falsos, por ejemplo, todos los referidos a las relaciones de suposición e implicación lógicas. Este tipo de relaciones son una herramienta útil a la hora de sistematizar el conocimiento adquirido por otros medios y también a la hora de efectuar predicciones. Pero son sólo un medio auxiliar de otros procesos cognitivos más vastos y complejos.

Convención: toda vez que identifiquemos una representación mediante recursos lingüísticos, es decir, mediante oraciones o símbolos oracionales, deberá tenerse en cuenta que lo hacemos sólo por economía expositiva, y que nos estamos refiriendo a una instancia representativa particular y concreta. Por ejemplo, podremos decir de un individuo que él posee una representación de que el gato está sobre la alfombra. Esto no significa que cualquier otro individuo que piense o crea que el gato está sobre la alfombra ha de tener necesariamente la misma representación. Una representación es siempre un sistema dinámico complejo que interviene en diversos procesos cognitivos. Como tal, está determinado por un conjunto único de acontecimientos. Aunque el lector y yo podamos tener algo en común, pongamos por caso, el concepto "gato", los gatos vistos, oídos o tocados por Ud. y por mí han sido, probablemente, distintos. Incluso la representación de "gato" puede ser distinta en un mismo individuo cuando cree que el gato está sobre la alfombra que cuando cree que el gato es un mamífero.

El primer nivel de organización del sistema nervioso que llama nuestra atención es el de redes neuronales (RN). Se trata de un nivel intermedio entre los circuitos locales y los sub-sistemas del sistema nervioso. Su papel es central en todos los procesos cognitivos. Es modelizable mediante redes neuronales artificiales (RNA) las que, a su vez, son testables empíricamente. Una RN es un conjunto de neuronas, interconectadas entre sí en distintos grados, que cumplen un papel funcional conjunto. Las RNs no mantienen un patrón de conexiones estático sino que cambia según diferentes factores y en distintas escalas temporales. También cambian algunas de sus propiedades fisiológicas, es decir, diversos factores pueden modificar el funcionamiento general de la RN. Estas alteraciones afectan positiva o negativamente el desempeño funcional global de la RN.

Una RNA es un sistema de procesamiento de la información no algorítmico, no es digital y trabaja en paralelo. No se programa, se entrena. Consiste en varios procesadores simples, altamente interconectados, llamados nodos o también

neurodos, análogos a las células nerviosas humanas, aunque tienen importantes diferencias con éstas. Los neurodos están conectados por un gran número de lazos de diferente peso, a través de los cuales pasa la señal. Cada neurodo recibe muchas señales de otros neurodos o del mundo externo (por ejemplo, fotones captados por una matriz de fotorreceptores, o un patrón de señales presentado a la red por el diseñador). Aunque recibe muchas señales, sólo produce una señal de output, la cual puede dividirse y alcanzar a otros neurodos. En cada una de las divisiones la señal tiene la misma intensidad. Algunos neurodos arrojan su señal de salida fuera de la red, generando de esta manera patrones de respuesta o de control.

La señal de salida de cada neurodo depende de tres factores: a) las señales de entrada, b) el peso asignado a la conexión y c) una función numérica definida para cada neurodo.

Las señales de entrada se miden en magnitudes de intensidad. El peso es una cantidad que se multiplica por cada una de las señales de entrada separadamente y, sumando luego todos los productos, se obtiene la totalidad del input recibido. Por último, se procesa este número aplicándole una función que determinará, finalmente, la señal de salida del neurodo. Existen numerosas funciones que pueden cumplir este papel, pero básicamente pueden caer dentro de tres categorías: funciones lineales, de umbral y sigmoides. En las primeras, la salida fluctúa proporcionalmente a la entrada. En las de umbral, hay dos salidas fijas: se elige una u otra según la intensidad del input total recibido supere o no cierta magnitud. Si utilizamos funciones sigmoides, la salida variará de modo no lineal según varíe la entrada. En la ilustración suponemos que las esferas azules son los neurodos de entrada, las rojas son neurodos de procesamiento u ocultos, y las amarillas los de salida. En conjunto, constituyen una red neural artificial elemental. Podemos representar los valores de entrada mediante un vector. A título de ejemplo, permitiremos que los valores que puede tomar cada elemento del vector varíen de 1 a 10 en valores absolutos. Y, aunque no es para nada necesario, asignemos arbitrariamente un peso de 0,1, 0,2, y 0,3 a la primera, segunda y tercera conexión que recibe el neurodo rojo izquierdo, y 0,3, 0,4 y 0,5 a las conexiones que recibe el neurodo rojo derecho. También, arbitrariamente, asignemos de izquierda a derecha los siguientes pesos a las conexiones entre las esferas rojas y las amarillas: 0,5, 0,6, 0,7, 0,8, 0,9 y 0,01. Establezcamos, además, que las esferas rojas realizan una transformación lineal $y = x + 3$. Y ya tenemos los elementos necesarios para que a cada vector de entrada, le corresponda un vector de salida. Así, si la entrada de las esferas azules tiene una intensidad $\langle 3, 4, 2 \rangle$, comenzamos el proceso multiplicando cada elemento por el peso asignado. De esta manera obtenemos los valores 0,3, 0,8 y 0,6. Luego, sumándolos, tenemos que el total de la entrada de la esfera roja izquierda será igual a 1,7. Sometemos ahora este valor a la función de salida, obteniendo $y = 1,7 + 3$, o sea, 4,7. La esfera roja izquierda tiene una salida igual a 4,7. Aplicando el mismo procedimiento, llegamos a que la intensidad de salida del neurodo rojo izquierdo es de 6,5.

En las esferas amarillas se repite el proceso. Multiplicamos el peso de la primera conexión, 0,5, por 4,7, lo que da 2,35. El peso de la segunda conexión es 0,6. Por 6,5 es igual a 3,9. Sumadas dan 6,25, es decir, la salida total de la esfera amarilla izquierda. El patrón de salida, correspondiente a la entrada $[3, 4, 2]$, es $[6,25, 5,2, 4,295]$.

La siguiente tabla que muestra los pesos asignados a cada conexión, juntamente con la función de salida asignada a cada neurodo oculto de la capa intermedia constituyen la

representación interna del input dado. Vale decir, dado un input de intensidad [3, 4, 2], y dado el estado interno de la red, el resultado tendrá una intensidad (un grado de activación) de [6,25, 5,2, 4,295].

0,5 0,6 0,7 0,8 0,9 0,01 0,1 0,2 0,3 0,3 0,4 0,5

Existen 1000 vectores diferentes de entrada de tres elementos, como los de nuestro ejemplo, que pueden tomar valores de 1 a 10. Esto significa que nuestro modesto ejemplo de un sistema con ocho neuronas es capaz de representar internamente 1000 estímulos diferentes.

Un cerebro humano normal contiene aproximadamente 100 mil millones (10^{11}) de neuronas.

Uno puede entrenar una RNA para que realice determinadas tareas. Son especialmente aptas para el reconocimiento de patrones. Esto incluye: facilidad para distinguir propiedades en circunstancias inciertas, ver analogías, centrar la atención en regiones específicas del input externo, reconocer cualidades sensibles difícilmente identificables (como un aroma, o la voz de una persona). Hasta hay quienes les adjudican el poder gobernarse a sí mismas en un entorno social y moral con responsabilidad y en vistas de algún propósito.

En 1974, Paul Werbos, descubrió un procedimiento matemático que permitía ajustar automáticamente el patrón de salida de una RNA utilizando el entrenamiento mediante la exposición de la red a ejemplos. Dicho sencillamente, consiste en comparar la salida actual con el resultado deseado, y luego ir modificando sistemáticamente los pesos de las conexiones para acercarse al máximo a ese resultado. El ingeniero que diseña la red proporciona los ejemplos, y es el que determina cuándo la red ha alcanzado el resultado esperado. En este respecto, es comparable a un padre que educa a sus hijos. Él les dice qué es lo correcto tanto en el modo de pensar como en el de actuar.

Cuando el tejido nervioso involucrado en funciones psíquicas superiores, como la percepción, el pensamiento o el habla, se daña, la función no suele desaparecer por completo, sino que muestra un déficit parcial. El daño puede deberse a la muerte de algunas neuronas, por traumatismo o accidente vascular por ejemplo, o también a una alteración en las condiciones de funcionamiento. Hay factores neuroquímicos que facilitan o inhiben la comunicación entre neuronas: algunos favorecen el establecimiento o fortalecimiento de las conexiones, otros aumentan o disminuyen el flujo de la estimulación. Análogamente, se pueden producir daños en RNA con efectos similares a los observados en seres humanos.

Supongamos que el sistema nervioso humano utilice recursos similares a los establecidos para las RNA, quiero decir, especificables matemáticamente de la misma forma. Una representación verdadera es una colección de estados del sistema nervioso causada por una colección de estados del mundo que acontecen realmente, independientemente de que el individuo que posee la representación sea consciente de ella, e independientemente de que esté o no justificado de algún modo para tenerla, es decir, independientemente de la evaluación que, en algún momento dado, el individuo haga sobre ella. El portador de verdad puede simbolizarse por la matriz numérica de los pesos asignados a la red más las funciones de salida asignadas a las neuronas. El verificador, por su parte, es el estado del mundo que produjo los estímulos, es decir, el que proveyó la matriz energética pertinente que estimuló los receptores. El factor diferenciador entre representaciones verdaderas y representaciones falsas viene de la mano de lo que llamaremos "estímulos

pertinentes". Una ilusión es causada por un estado real del mundo tanto como lo es una representación verdadera. Sin embargo, podemos distinguirlas en cuanto las ilusiones reconocen como causas estados de cosas insuficientes o no captados suficientemente por nuestros receptores y también puede ocurrir que el proceso haya sido realizado por una red neural defectuosa. Una determinada combinación de representaciones en un sistema nervioso puede dar lugar a representaciones falsas. Por ejemplo, puedo creer que lo que veo allí es una oveja, cuando en realidad se trata de un perro. Esto puede ocurrir porque el animal se encuentra medio oculto y no me llega el estímulo necesario. Pero también puede ocurrir que éste sea suficiente, pero que otras creencias o estados mentales (esperanzas, temores, etc.) me fueren a preferir creer en la oveja. Mis redes neurales no estarían adecuadamente preparadas para captar el hecho. Mi representación de que allí hay una oveja sólo será verdadera cuando allí haya una oveja, y esa oveja, directa o indirectamente, haya causado mi creencia¹. La causación será directa cuando haya sido percibida por mí. En cualquier otro caso, será indirecta.

Cuando contamos con información insuficiente del entorno, tendemos a completarla sobre la base de nuestra experiencia pasada. En algunas ocasiones estos estímulos convergerán hacia una representación verdadera, en otras pueden provocar el efecto contrario. Es más, el concepto de "información suficiente" es relativo al estado de nuestro sistema nervioso en un momento dado: puede haber condiciones ambientales óptimas, pero si estoy borracho, o jamás había visto una oveja o animal alguno en toda mi vida, la información proveniente de mi entorno no será suficiente para generar una representación verdadera. Recordemos que una representación siempre es compleja y que no se puede aislar a partir, por ejemplo, del significado o contenido de una única oración.

Así, cuando dos personas creen que allí hay una oveja, sus creencias pueden ser parcialmente falsas (o verdaderas) en diferentes grados. Probablemente, la representación que posea un individuo que no sabe que las ovejas son comestibles sea más falsa que la de otro que las conozca mejor. Cuando más experimentamos respecto de un tipo de cosas, obtenemos una representación más fiel de ese tipo de cosas y, por lo tanto, más verdadera.

Una representación puede ser completamente falsa respecto de un estado del mundo o de un tipo de estados del mundo, como cuando alucinamos un helado de frutillas cuando tenemos delante una oveja en el campo. Pero nunca es completamente verdadera, porque la realidad siempre es más rica que cualquier representación que de ella podamos forjarnos.

El estar la oveja en el campo es una colección de sucesos en un espacio de sucesos. El creer que la oveja está en el campo es también una colección de sucesos en un espacio de sucesos, en particular, consiste en la codificación neural del estar la oveja en el campo. Cuando el estar la oveja en el campo es la causa de la creencia, la creencia es (más o menos) verdadera. Cuando la causa no es la apropiada, o cuando la codificación es defectuosa, la creencia es (más o menos) falsa.

Nota 1. Obsérvese que no puede darse el caso de que a) haya allí una oveja, b) yo crea que allí hay una oveja, y c) aún así, mi creencia no haya sido causada por la oveja, a menos que entendamos por "oveja" una oveja cualquiera y no eso que está allí.