REVISTA DE LA ESCUELA DE FILOSOFÍA
CUADERNOS FILOSÓFICOS
SEGUNDA ÉPOCA

Facultad de Humanidades y Artes_UNR

CUADERNOS FILOSÓFICOS / SEGUNDA ÉPOCA
CUADERNOSFILOSOFICOS.UNR.EDU.AR
ISSN: 2683-9024

Traducciones

PP vainilla para filósofos: Un manual sobre Procesamiento Predictivo[1]

WANJA WIESE[2] & THOMAS METZINGER[3]

Traducción:
Romina Pogliani[4]

Palabras clave: inferencia activa; atención; inferencia bayesiana; aislamiento ambiental; principio de energía libre; procesamiento jerárquico; principio ideomotor; percepción; inferencia perceptiva; precisión; predicción; minimización del error de predicción; procesamiento predictivo; control predictivo; estimación estadística; procesamiento de arriba hacia abajo.

Key Words: active inference; attention; bayesian inference; environmental seclusion; free energy principle; hierarchical processing; ideomotor principle; perception; perceptual inference; precision; prediction; prediction error minimization; predictive processing; predictive control; statistical estimation; top-down processing.

El objetivo de este breve capítulo, dirigido a los filósofos, es proporcionar una visión general y una breve explicación de algunos conceptos centrales relacionados con el procesamiento predictivo (PP). Incluso aquellos que se consideran expertos en el tema pueden encontrar útil ver cómo se usan los términos centrales en esta colección. Para simplificar las cosas, primero definiremos de manera informal un conjunto de características importantes del procesamiento predictivo, complementadas con algunas breves explicaciones y un glosario alfabético.

1 **Cómo citar**: Wiese, W., & Metzinger, T. K. (2020). PP vainilla para filósofos: Un manual sobre Procesamiento Predictivo (trad. R. Pogliani). *Cuadernos Filosóficos*, 17. DOI: https://doi.org/10.35305/cf2.vi17.118

Versión original: Wiese, W., & Metzinger, T. K. (2017). Vanilla PP for Philosophers: A Primer on Predictive Processing. En W. Wiese & T. K. Metzinger (Eds.), *Philosophy and Predictive Processing* (pp. 1-18). MIND Group.

Fechas: Recepción: 13/04/2021. Aprobación: 07/07/2021.

2 Universidad Johannes Gutenberg de Maguncia (Maguncia, Renania-Palatinado, Alemania). ORCID ID: https://orcid.org/0000-0003-3338-7398. wawiese@uni-mainz.de

3 Universidad Johannes Gutenberg de Maguncia (Maguncia, Renania-Palatinado, Alemania). ORCID ID: https://orcid.org/0000-0003-2514-7191. metzinger@uni-mainz.de

4 Universidad Nacional de Rosario (Rosario, Santa Fe, Argentina).

Las características descritas aquí no se comparten en todas las explicaciones del PP. Algunas pueden no ser necesarias para un modelo individual; otras pueden ser discutidas. De hecho, ni siquiera todos los autores de esta colección aceptarán todas ellas. Para que esto sea transparente, hemos animado a los colaboradores a que indiquen brevemente qué características son *necesarias* para apoyar los argumentos que proporcionan, y cuáles (si las hay) son incompatibles con sus argumentos. En aras de la claridad, proveemos aquí la lista completa, ordenada a grandes rasgos en función de la importancia que consideramos para el "PP Vainilla" (es decir, una formulación del procesamiento predictivo que probablemente será aceptada por la mayoría de los investigadores que trabajan en este tema). Más adelante se darán explicaciones más detalladas. Obsérvese que estas características no especifican las condiciones necesarias y suficientes para la aplicación del concepto de "*procesamiento predictivo*". Todo lo que tenemos actualmente es un cúmulo semántico, probablemente con algunos conjuntos superpuestos de criterios suficientes. El marco está aún en desarrollo, y es difícil, quizá imposible, proporcionar explicaciones teóricamente neutrales de todas las ideas del PP sin introducir ya fuertes suposiciones de fondo. No obstante, al menos las características 1-7 pueden considerarse propiedades necesarias de lo que se denomina PP en este volumen:

1. Procesamiento de arriba hacia abajo [*top-down*]: El cálculo en el cerebro implica de manera crucial una interacción entre procesamiento de arriba hacia abajo y de abajo hacia arriba, y el PP hace hincapié en la ponderación relativa de las señales descendentes y ascendentes tanto en la percepción como en la acción.

2. Estimación estadística: El PP implica el cálculo de estimaciones de variables aleatorias. Las estimaciones pueden considerarse hipótesis estadísticas que pueden servir para explicar las señales sensoriales.

3. Procesamiento jerárquico: El PP utiliza estimadores organizados jerárquicamente (que rastrean características a diferentes escalas espaciales y temporales).

4. Predicción: El PP explota el hecho de que muchas de las variables aleatorias relevantes en la jerarquía son predictivas entre sí.

5. Minimización del Error de Predicción (MEP): El PP implica el cálculo de los errores de predicción; estos términos de error de predicción tienen que ser ponderados por estimaciones de precisión, y un objetivo central del PP es minimizar los errores de predicción ponderados por la precisión.

6. Inferencia bayesiana: El PP se ajusta a las normas de la inferencia bayesiana: a largo plazo, la minimización del error de predicción en el modelo jerárquico se aproximará a la inferencia bayesiana exacta.

7. Control predictivo: El PP está orientado a la acción, en el sentido de que el organismo puede actuar para cambiar su entrada sensorial para ajustarse a sus predicciones y minimizar así el error de predicción; entre otras ventajas, esto permite al organismo regular sus parámetros vitales (como los niveles de oxigenación y azúcar en la sangre, etc.).

8. Reclusión ambiental: El organismo no tiene acceso directo a los estados de su entorno y de su cuerpo (para un análisis conceptual de la "percepción directa", Snowdon, 1992), sino que los infiere (al inferir las causas ocultas de las señales sensoriales interoceptivas y exteroceptivas). Aunque esta es una característica básica de algunos argumentos filosóficos del PP (cf. Hohwy, 2016; Hohwy, 2017), es controversial (cf. Anderson, 2017; Clark, 2017; Fabry, 2017a; Fabry, 2017b).

9. El principio ideomotor: Existen estimaciones "ideomotoras"; su cálculo sustenta tanto la percepción como la acción, porque codifican los cambios en el mundo que son registrados por la percepción y pueden ser provocados por la acción.

10. Atención y precisión: La atención puede describirse como el proceso de optimización de las estimaciones de precisión.

11. Comprobación de hipótesis: Los procesos computacionales que subyacen a la percepción, la cognición y la acción pueden describirse útilmente como pruebas de hipótesis (o el proceso de acumulación de pruebas para el modelo interno). Conceptualmente, podemos distinguir entre comprobación de hipótesis pasiva y activa (y se podría intentar hacer coincidir la comprobación de hipótesis activa con la acción, y la la comprobación de hipótesis pasiva con la percepción). Sin embargo, puede resultar que toda la comprobación de hipótesis en el cerebro (si tiene sentido decir eso) es una comprobación de hipótesis activa.

12. El Principio de Energía Libre: Fundamentalmente, el PP es sólo una forma de minimizar la energía libre, que en la mayoría de las explicaciones del PP equivaldría a la media a largo plazo del error de predicción.

En lo que sigue, no asumimos ninguna familiaridad con el PP ni ningún conocimiento matemático de fondo, y esta introducción se limitará, en su mayor parte, a los fundamentos conceptuales del marco del PP. Una vez leída esta introducción, uno debería ser capaz de seguir la discusión en los otros artículos de esta colección. Sin embargo, también recomendamos encarecidamente a los lectores que profundicen en su comprensión del PP

leyendo (Clark, 2016) y (Hohwy, 2013), dos excelentes primeras monografías filosóficas sobre este tema.

1. ¿Qué es el procesamiento predictivo? Siete características principales

El procesamiento predictivo (PP) es un marco que implica un principio computacional general que puede aplicarse para describir la percepción, la acción, la cognición y sus relaciones de una manera única y conceptualmente unificada. No es directamente una teoría sobre los procesos neuronales subyacentes (es computacional, no neurofisiológica), pero existen propuestas más o menos específicas sobre cómo el procesamiento predictivo puede ser implementado por el cerebro (ver, por ejemplo, Engel et al., 2001; Friston, 2005; Wacongne et al., 2011; Bastos et al., 2012; Brodski et al., 2015). Además, parece que al menos algunos de los principios que pueden aplicarse a las descripciones en niveles de análisis subpersonales (por ejemplo, computacionales o neurobiológicos) pueden aplicarse también a las descripciones en el nivel personal (por ejemplo, a los fenómenos agentivos, a la estructura del razonamiento o a los informes fenomenológicos que describen los contenidos de la conciencia). Esta es una de las razones por las que el PP es filosóficamente interesante y relevante. Si la teoría se dirige por buen camino, entonces:

1. puede proporcionar los medios para construir nuevos puentes conceptuales entre los trabajos teóricos y empíricos sobre la cognición y la conciencia,
2. puede revelar relaciones inesperadas entre fenómenos aparentemente dispares, y
3. puede unificar en cierta medida diferentes enfoques teóricos.

Pero, ¿qué es el PP en primer lugar? He aquí una formulación relativamente temprana de una de sus ideas clave:[5]

> *Wenn die Anschauung sich nach der Beschaffenheit der Gegenstände richten müßte, so sehe ich nicht ein, wie man a priori von ihr etwas wissen könne; richtet sich aber der Gegenstand (als Objekt der Sinne) nach der Beschaffenheit unseres Anschauungsvermögens, so kann ich mir diese Möglichkeit ganz wohl vorstellen.* (Kant, 1998, B XVII)[6]

Algo que Kant subraya en este punto de la *Crítica de la razón pura* es que nuestras intuiciones (*Anschauungen*), que constituyen el material sensorial sobre el que se realizan los actos de síntesis, no son datos sensoriales simplemente dados (Brook, 2013, § 3.2). No sólo se reciben, sino que también son parcialmente moldeados por la facultad de la intuición (*Anschauungsvermögen*). En el lenguaje contemporáneo, la idea puede expresarse de la siguiente manera:

> Las teorías clásicas del procesamiento sensorial ven el cerebro como un dispositivo pasivo, impulsado por estímulos. En cambio, los enfoques más recientes hacen hincapié en la naturaleza constructiva de la percepción, considerándola un proceso activo y altamente selectivo. De hecho, hay una amplia evidencia de que el procesamiento de los estímulos está controlado por influencias descendentes [*top-down*] que moldean fuertemente la dinámica intrínseca de las redes talamocorticales y

5 En este punto, cabría esperar una referencia a la famosa idea de Helmholtz de que la percepción es el resultado de inferencias inconscientes; nos referiremos a este pasaje más adelante. El punto de vista de Helmholtz sobre la percepción estaba muy influenciado por Kant (aunque Helmholtz parece haber enfatizado el papel del aprendizaje y la experiencia más que Kant; véase Lenoir, 2006, pp. 201 y 203): "*Dass die Art unserer Wahrnehmungen ebenso sehr durch die Natur unserer Sinne, wie durch die äusseren Dinge bedingt sei, wird durch die angeführten Thatsachen sehr augenscheinlich an das Licht gestellt, und ist für die Theorie unseres Erkenntnissvermögens von der höchsten Wichtigkeit. Gerade dasselbe, was in neuerer Zeit die Physiologie der Sinne auf dem Wege der Erfahrung nachgewiesen hat, suchte Kant schon früher für die Vorstellungen des menschlichen Geistes überhaupt zu thun, indem er den Antheil darlegte, welchen die besonderen eingeborenen Gesetze des Geistes, gleichsam die Organisation des Geistes, an unseren Vorstellungen haben*." (Von Helmholtz, 1855, p. 19). (Nuestra traducción*: "Estos hechos muestran claramente que la naturaleza de nuestras percepciones está tan limitada por la naturaleza de nuestros sentidos como por los objetos externos. Esto es de suma importancia para una teoría de nuestra facultad epistémica. La fisiología de los sentidos ha demostrado recientemente, por medio de la experiencia, exactamente el mismo punto que Kant trató de mostrar antes para las ideas de la mente humana en general, exponiendo la contribución que hacen las leyes especiales innatas de la mente —la organización de la mente, por así decirlo— a nuestras ideas"). Una visión general de las raíces kantianas del PP puede encontrarse en Swanson, 2016.

*La traducción señalada, del alemán al inglés originalmente, corresponde a los autores del presente trabajo. [N. de la T.]

6 "Si la intuición debiese regirse por la naturaleza de los objetos, no entiendo cómo se podría saber a priori algo sobre ella; pero si el objeto (como objeto de los sentidos) se rige por la naturaleza de nuestra facultad de intuición, entonces puedo muy bien representarme esa posibilidad" (trad. M. Caimi).

> crean constantemente predicciones sobre los eventos sensoriales venideros. (Engel et al., 2001, p. 704).

Esto es lo que aquí llamamos la primera característica del procesamiento predictivo: **Procesamiento Descendente [*Top-Down*]**. Como se puede ver, la idea de que la percepción está impulsada, en parte, por los procesos descendentes no es nueva (lo que no quiere decir que las teorías dominantes de la percepción hayan marginado durante mucho tiempo su papel). La aportación novedosa del PP es que pone un énfasis extremo en esta idea, describiendo la influencia del procesamiento descendente y del conocimiento previo como una característica *omnipresente* de la percepción, que no sólo está presente en los casos en los que la entrada sensorial es ruidosa o ambigua, sino *en todo momento*.[7] Según el PP, el cerebro forma constantemente estimaciones estadísticas, que funcionan como representaciones[8] de lo que hay actualmente en el mundo (característica N° 2, **Estimación Estadística**), y estas estimaciones están organizadas jerárquicamente (siguiendo rasgos a diferentes escalas espaciales y temporales; característica N° 3, **Procesamiento Jerárquico**).[9] El cerebro utiliza estas representaciones para predecir la entrada sensorial actual (y futura) y la fuente de la misma, lo que es posible porque las estimaciones en diferentes niveles de la jerarquía son *predictivas* entre sí (característica N° 4, **Predicción**). Los desajustes entre las predicciones y la entrada sensorial real no se utilizan pasivamente para formar percepciones, sino para informar de las

7 Por supuesto, es interesante preguntarse hasta qué punto el propio Kant consideraba que las influencias activas (descendentes) sobre las intuiciones (*Anschauungen*) eran una característica omnipresente. Al menos algunos pasajes de la *Crítica de la Razón Pura* sugieren que Kant puso más énfasis en las influencias (descendentes) ejercidas por nuestra facultad de conocer (la espontaneidad de los conceptos): "*Unsere Erkenntnis entspringt aus zwei Grundquellen des Gemüts, deren die erste ist, die Vorstellungen zu empfangen (die Rezeptivität der Eindrücke), die zweite das Vermögen, durch diese Vorstellungen einen Gegenstand zu erkennen (Spontaneität der Begriffe); durch die erstere wird uns ein Gegenstand gegeben, durch die zweite wird dieser im Verhältnis auf jene Vorstellung (als bloße Bestimmung des Gemüts) gedacht.*" (Kant, 1998, B 74). ("Nuestro conocimiento surge de dos fuentes fundamentales de la mente, de las cuales la primera es [la de] recibir las representaciones (la receptividad de las impresiones), y la segunda, la facultad de conocer un objeto mediante esas representaciones (la espontaneidad de los conceptos); por la primera, un objeto nos es dado; por la segunda, éste es *pensado* en relación con aquella representación [considerada] como mera representación de la mente"; trad. Caimi). Sin embargo, una investigación seria de esta cuestión tendría que centrarse en la influencia de las representaciones inconscientes en la formación de las intuiciones (Giordanetti et al., 2012).

8 El uso de la palabra "representación" no es totalmente incontrovertido en este caso. Hay un cierto debate sobre si el PP postula representaciones y, en caso afirmativo, cuál es la mejor manera de describirlas (Gładziejewski, 2016; Downey, 2017; Dołega, 2017). Sin embargo, al menos es posible tratar las descripciones representacionalistas de los postulados implicados por el PP como una glosa representacional (o intencional) (cf. Egan, 2014; Anderson, 2017). Así, aunque reconocemos que algunos no estarían de acuerdo, creemos que es útil describir las estimaciones postuladas por el PP como representaciones, al menos para los fines de esta introducción (incluso si algunos autores argumentaran que estas posturas no son representaciones en un sentido fuerte).

actualizaciones de las representaciones que ya se han creado (anticipando así, en la medida de lo posible, las señales sensoriales entrantes). El objetivo de estas actualizaciones es *minimizar* el error de predicción resultante (característica N° 5, **Minimización del Error de Predicción** (**MEP**)), de tal manera que las actualizaciones se ajusten a las normas de la **Inferencia Bayesiana** (característica N° 6; más sobre esto a continuación). El principio computacional de la MEP es un principio general al que se ajusta todo el procesamiento en el cerebro (en todos los niveles de la jerarquía postulada por el PP). A partir de esto, es sólo un pequeño paso hacia la descripción del procesamiento en el cerebro como una alucinación en línea controlada:[10]

> Por lo tanto, una forma fructífera de ver el cerebro humano es como un sistema que, incluso en estados ordinarios de vigilia, alucina constantemente con el mundo, como un sistema que constantemente deja que su dinámica interna autónoma de simulación colisione con el flujo continuo de entrada sensorial, soñando vigorosamente con el mundo y generando así el contenido de la experiencia fenoménica. (Metzinger, 2004, p. 52)

Nótese que los contenidos de la experiencia fenoménica son sólo una parte de lo que, según el PP, se genera a través del proceso jerárquicamente organizado de minimización de errores de predicción (la mayoría de los contenidos serán inconscientes). Resumiendo las seis primeras características centrales descritas anteriormente, y añadiendo la séptima, podemos ahora dar una definición concisa de lo que se llama procesamiento predictivo en esta colección (enriqueceremos la definición con las característica 8-12 más adelante):

El Procesamiento Predictivo (PP) es

- codificación predictiva *jerárquica*,

9 Esta jerarquía de estimaciones implica un modelo generativo jerárquico. Un modelo generativo es la distribución conjunta de una colección de variables aleatorias (véase el glosario). Un modelo generativo jerárquico corresponde a una jerarquía de variables aleatorias, donde las variables de los niveles no adyacentes son condicionalmente independientes (esto puede, por ejemplo, representar una jerarquía de objetos o eventos relacionados causalmente, Drayson, 2017). La jerarquía de estimaciones postulada por el PP rastrea los valores de una jerarquía de variables aleatorias. Una ilustración heurística de un modelo generativo puede encontrarse en la introducción de (Clark, 2016). Agradecemos a Chris Burr por habernos recordado mencionar los modelos generativos.

10 Horn (Horn, 1980, p. 373) atribuye la idea de que "la visión es una alucinación controlada" a Clowes (Clowes1971). La única afirmación publicada de Clowes que se acerca a esta formulación parece, sin embargo, ser: "La gente ve lo que espera ver" (Clowes, 1969, p. 379; cf. Sloman, 1984). Más recientemente, una idea similar ha sido propuesta por Grush (Grush, 2004, p. 395; la atribuye a Ramesh Jain): "El papel que desempeña la sensación es el de constreñir la configuración y evolución de esta representación. A modo de lema, la percepción es un proceso de alucinación controlado".

- que involucra un proceso *mediado por la precisión*

- para la minimización del error de predicción,[11]

- que permite el *control* predictivo.

Nótese que esta definición ya va más allá de lo que se suele denominar *codificación predictiva* (especialmente si la codificación predictiva se concibe únicamente como una estrategia computacional para la compresión de datos, cf. Shi y Sun, 1999; Clark, 2013a). En primer lugar, el PP es jerárquico. En segundo lugar, las estimaciones de precisión pueden desempeñar papeles funcionales que van más allá de equilibrar las suposiciones previas y las pruebas sensoriales actuales de forma estadísticamente óptima (Clark, 2013b). En tercer lugar, el PP suele describirse como orientado a la acción, en el sentido de que permite el **Control Predictivo** (característica N° 7; cf. Seth, 2015). Esto pone de manifiesto la suposición, sostenida por algunos, de que la acción es, en cierto sentido, más importante que la percepción; aunque la percepción puede describirse como un proceso de obtención de conocimiento sobre el mundo, la función principal de la obtención de este conocimiento reside en permitir una acción eficiente y sensible al contexto, a través de la cual el organismo mantiene con éxito su existencia. Esto se hace evidente cuando se considera al PP en el contexto más amplio del principio de energía libre de Friston (PEL).[12] Antes de profundizar en esta cuestión, vamos a dar un paso atrás y examinar el problema de la percepción, visto desde la perspectiva de la codificación predictiva.

2. Procesamiento Predictivo y Codificación Predictiva

Para casi todas las características del PP (procesamiento predictivo) también hay precursores destacados de la perspectiva del PP sobre la percepción. Consideremos la siguiente afirmación de Helmholtz:

11 Las tres primeras partes de esta definición se corresponden aproximadamente con la definición ofrecida por Clark (Clark, 2013a, p. 202; Clark, 2015, p. 5). En (Clark 2013a), Clark también introduce la noción de PP orientado a la acción (que incorpora el cuarto aspecto de la definición ofrecida aquí). Estas cuatro características también son fundamentales en la exposición de Hohwy sobre la minimización del error de predicción (véanse los cuatro primeros capítulos de Hohwy, 2013).

12 Más adelante se hablará de ello. Obsérvese que es posible desarrollar explicaciones acerca del PP sin invocar el PEL (por lo que, en cierto modo, el PP es independiente del PEL), pero el PP puede incorporarse al PEL (Friston y Kiebel, 2009), por lo que la minimización del error de predicción puede interpretarse como una forma de minimizar la energía libre (que sería entonces un caso especial del PEL).

> *Die psychischen Thätigkeiten, durch welche wir zu dem Urtheile kommen, dass ein bestimmtes Object von bestimmter Beschaffenheit an einem bestimmten Orte ausser uns vorhanden sei, sind im Allgemeinen nicht bewusste Thätigkeiten, sondern unbewusste. Sie sind in ihrem Resultate einem Schlusse gleich, insofern wir aus der beobachteten Wirkung auf unsere Sinne die Vorstellung von einer Ursache dieser Wirkung gewinnen, während wir in der That direct doch immer nur die Nervenerregungen, also die Wirkungen wahrnehmen können, niemals die äusseren Objecte.* (Von Helmholtz, 1867, p. 430)[13]

El problema de la percepción, tal como se concibe aquí, tiene dos aspectos: (1) las percepciones son el resultado de un proceso inferencial inconsciente; (2) las percepciones nos presentan propiedades de objetos externos, aunque en realidad sólo podemos percibir los efectos de los objetos externos. Una descripción contemporánea de esta idea se puede encontrar en la monografía de Dennett del 2013, *Intuition Pumps and Other Tools for Thinking* [*Bombas de intuición y otras herramientas para pensar*], donde caracteriza la curiosa situación en la que se encuentra el cerebro, comparándola con el siguiente escenario de ficción:

> Estás preso en la sala de control de un robot gigante. [...] El robot habita un mundo peligroso, con muchos riesgos y oportunidades. Su futuro está en tus manos y, por supuesto, tu propio futuro también depende del éxito que tengas al pilotar tu robot por el mundo. Si se destruye, la electricidad de esta habitación se cortará, no habrá más comida en la nevera y tú morirás. ¡Buena suerte! (Dennett, 2013, p. 102)

La persona que está dentro del robot sólo tiene acceso indirecto al mundo, a través de los sensores del robot, y los efectos de las acciones ejecutadas no pueden conocerse, sino que tienen que inferirse. Esto ilustra la característica que llamamos **Reclusión Ambiental** (característica N° 8). La Reclusión Ambiental no es una característica computacional, sino epistemológica, pero aparece en las descripciones de los problemas a los que los cálculos del PP dan solución.[14] Para averiguar qué significan las diferentes señales que recibe el robot, la persona que está dentro tiene que formarse una hipótesis sobre sus causas ocultas. El problema de inferir las causas de las señales sensoriales es un problema inverso, porque

13 "Las actividades psíquicas que nos llevan a inferir que allí, delante de nosotros, en un lugar determinado, hay un objeto de cierto carácter, no son generalmente actividades conscientes, sino inconscientes. En su resultado son equivalentes a una conclusión, en la medida en que la acción observada sobre nuestros sentidos nos permite formarnos una idea en cuanto a la posible causa de esta acción; aunque, de hecho, son, de manera invariable, simplemente los estímulos nerviosos los que son percibidos directamente, es decir, las acciones, pero nunca los objetos externos mismos" (Von Helmholtz, 1985, p. 4).

requiere invertir el mapa de las causas (externas, ocultas) a los efectos (sensoriales). Se trata de un problema difícil (por no decir otra cosa), porque un mismo efecto podría tener múltiples causas.[15] Así que, aunque la relación entre las causas *C* y los efectos *E* pudiera describirse mediante un determinista, f: $C \rightarrow E$, la cartografía inversa, f^{-1}: $E \rightarrow C$, no suele existir. ¿Cómo resuelve el cerebro este problema?

Una primera observación es que la causa de un efecto sensorial está infradeterminada por el efecto, por lo que hay que utilizar información previa para adivinar la causa oculta. Además, si sabemos cómo afecta el aparato sensorial a las causas externas, es más fácil deducir los efectos sensoriales, dada la información sobre las causas externas, que a la inversa. Por lo tanto, si tenemos alguna información sobre causas ocultas, podemos formar una predicción de sus efectos sensoriales. Esta predicción puede compararse con la señal sensorial real, y la medida en que ambas difieren, es decir, el tamaño del *error de predicción*, nos da una pista sobre la calidad de nuestra estimación de la causa oculta. Podemos actualizar esta estimación, calcular una nueva predicción, volver a compararla con las señales sensoriales actuales y, de este modo, (con suerte) minimizar el error de predicción. Lo ideal es que nuestra primera estimación de la causa oculta sea realmente pobre, ya que al calcular constantemente las predicciones y los errores de predicción, y al actualizar nuestra estimación en consecuencia, podemos estar cada vez más seguros de haber encontrado una buena representación de la causa oculta.

Ilustremos esta estrategia con el siguiente ejemplo sencillo. Un profesor entra en el aula y encuentra un papel sobre su mesa con el mensaje "El profesor es un impostor. Ni siquiera existe realmente". El mensaje ha sido escrito con una pluma estilográfica, en color azul, lo que

14 He aquí algunos ejemplos: "Por ejemplo, durante la percepción visual el cerebro tiene acceso a información, medida por los ojos, sobre la distribución espacial de la intensidad y la longitud de onda de la luz incidente. A partir de esta información, el cerebro necesita inferir la disposición de los objetos (las causas) que dio lugar a la imagen percibida (el resultado del proceso de formación de la imagen)". (Spratling, 2016, p. 1 preimpresión). "La primera de ellas (el uso generalizado y descendente de modelos generativos probabilísticos para la percepción y la acción) constituye una propuesta muy sustancial, aunque hay que reconocer que es bastante abstracta: a saber, que tanto la percepción como [...] la acción dependen de una forma de 'análisis por síntesis' en la que los datos sensoriales observados se explican encontrando el conjunto de causas ocultas que son las mejores candidatas para haber generado esos datos sensoriales datos sensoriales en primer lugar". (Clark, 2013a, p. 234; pero véase Clark en prensa, para una opinión matizada). "Del mismo modo, el punto de partida de la consideración de error de predicción de la unidad es uno indirecto: desde el interior del cráneo el cerebro tiene que inferir las causas ocultas de su entrada sensorial" (Hohwy, 2013, p. 220).

15 Por esta razón, el problema también puede describirse como un problema mal planteado (Spratling, 2016), pero algunos autores considerarían el problema de encontrar cómo resolver este problema como mal planteado (Anderson, 2017).

(como sabe el profesor) excluye a muchos de sus alumnos. Para encontrar al culpable, el profesor pide a todos los alumnos que utilizan plumas estilográficas con tinta azul que pasen al frente y, con su propia pluma, escriban algo en un papel. Resulta que sólo se trata de tres alumnos, A, B y C, y todos utilizan tinta de diferentes marcas (lo que hace que se puedan distinguir). El profesor puede ahora formarse una conjetura sobre la causa oculta del mensaje ("El profesor es un impostor, ni siquiera existe realmente"): asume que el alumno A es el culpable y le pide a A que escriba el mismo mensaje. Esto puede verse como una predicción del mensaje real y, comparándolas, el profesor evalúa su suposición sobre la causa oculta. Si la tinta es la misma, no hay error de predicción y la estimación de la causa oculta no tiene que cambiar: el culpable ha sido encontrado. Si hay una diferencia puede actualizar su estimación asumiendo que, por ejemplo, B ha producido el mensaje. Mediante la formación constante de predicciones (mensajes escritos por los de los sospechosos) y comparándolos con la señal sensorial real (el mensaje en el escritorio), el profesor minimiza el error de predicción y encuentra al verdadero culpable.

Hay muchas diferencias entre este escenario ficticio y la situación en la que se encuentra el cerebro. Una de ellas es que el ejemplo implica una agencia a nivel personal (al igual que el experimento mental del robot gigante de Dennett): el profesor pone a prueba la hipótesis de que, digamos, el estudiante A es el culpable pidiéndole a A que escriba un mensaje. Además, el número de posibles causas ocultas es finito, y el "error de predicción" sólo indica al profesor que un alumno concreto no está implicado. No contiene ninguna otra información sobre el culpable; sólo excluye a uno de los sospechosos. El cerebro no puede recorrer todas las hipótesis posibles una por una, porque hay (potencialmente) infinitas posibles causas ocultas en el mundo. Además, el mundo es cambiante, por lo que las representaciones de las causas ocultas tienen que ser dinámicas: adaptarse y anticiparse a todos los cambios (relevantes y predecibles) del entorno. Por último, para ser más realista, el ejemplo del profesor tendría que ampliarse de tal manera que éste formara predicciones sobre todas sus entradas sensoriales en todo momento. Al igual que podría inferir las interacciones causales que conducen a la nota, puede inferir todos los acontecimientos causales que le rodean en todo momento (incluida su propia influencia en el flujo sensorial).

3. Procesamiento Predictivo e Inferencia Bayesiana

La **Inferencia Bayesiana** (característica N° 6) es un método computacional para combinar racionalmente[16] la información existente sobre la que existe incertidumbre, con nuevas pruebas. Aquí, incertidumbre significa que la información puede describirse en un formato probabilístico, es decir, utilizando una distribución de probabilidad. Un ejemplo muy sencillo sería una situación en la que un agente no está seguro de cuál, de un número finito de hipótesis, es verdadera (como en el caso del profesor anterior). La incertidumbre se reflejaría entonces en el hecho de que el agente asigna diferentes probabilidades a las hipótesis, sin asignar una probabilidad de 1 a ninguna de ellas. Pero también puede haber situaciones en las que la información del agente se modela mejor como si se tratara de un número infinito de posibilidades ("hipótesis"), por ejemplo, cuando el agente realiza una medición ruidosa de una cantidad que puede tener cualquier valor en un intervalo continuo. En este caso, la información puede codificarse mediante una función de densidad de probabilidad (es decir, un modelo), que asigna probabilidades a las regiones (por ejemplo, a los subintervalos). La pregunta a la que la inferencia bayesiana da una respuesta racional (utilizando la regla de Bayes) es la siguiente: ¿cómo debo actualizar mi modelo cuando obtengo nueva información? Un ejemplo de nueva información sería la que recibe un agente al realizar una medición (suponiendo que el agente ya tiene información incierta sobre la cantidad a medir).

Formalmente, esta actualización implica el cálculo de una *distribución a posteriori* (que también se llama simplemente *posterior*). La posterior se obtiene combinando una *distribución a priori* (también llamada simplemente *inicial*) con una *probabilidad* [*likelihood*]. La distribución a priori codifica la información que el agente ya tiene; la probabilidad codifica cómo el dominio sobre el que el agente ya tiene información está relacionado con el dominio de la nueva información obtenida. Una buena característica de la inferencia bayesiana es que puede reducir la incertidumbre. Formalmente, esto significa que la posterior suele tener una varianza menor —es más precisa— que la inicial.

Superficialmente hablando, no hay una conexión obvia entre la minimización del error de predicción (MEP) y la inferencia bayesiana. De hecho, no es obvio por qué sería deseable implementar o aproximar la inferencia bayesiana utilizando la MEP. Sin embargo, hay una buena

16 Aquí, "racionalmente" se define de acuerdo con los axiomas de la probabilidad, y con la definición de probabilidad condicional; también se puede demostrar que la inferencia bayesiana es óptima en un sentido teórico de la información (Zellner, 1988).

razón. Recordemos que el problema inverso de la percepción es un problema mal planteado: las señales sensoriales, consideradas como efectos de eventos externos, no pueden ser mapeadas a los estados ocultos del entorno porque para cada efecto sensorial hay múltiples causas externas posibles. En otras palabras, hay incertidumbre sobre las causas ocultas. Dadas las suposiciones previas sobre estas causas, y considerando los efectos sensoriales que medimos como nueva evidencia, la inferencia bayesiana promete darnos una solución racional al problema de cómo debemos actualizar nuestras suposiciones previas sobre las causas ocultas. En otras palabras, lo que la inferencia bayesiana puede darnos (al menos en principio) es algo así como un "mapa inverso probabilístico". Esta función mapea un efecto sensorial medido para los diferentes (conjuntos de) posibles causas ocultas, e indica qué causas posibles son probablemente las causas reales de los efectos sensoriales.

Pero, ¿por qué necesitamos la MEP si tenemos la inferencia bayesiana? La respuesta es que la inferencia bayesiana puede ser computacionalmente compleja, incluso intratable. En los casos sencillos, es posible calcular la variable posterior analíticamente; en otros casos, tiene que ser aproximada. En otros casos, puede ser posible calcular la posterior, pero lo que realmente se desea es el *maximizador* de la posterior (por ejemplo, una hipótesis más probable, después de haber tenido en cuenta las nuevas pruebas). Encontrar el maximizador puede ser, de nuevo, exigente desde el punto de vista computacional y puede requerir métodos aproximativos. Algunos métodos aproximativos implican la minimización del error de predicción. Aunque la motivación de la inferencia bayesiana es independiente de la minimización del error de predicción, una vez que la inferencia bayesiana se considera una solución al problema de la percepción, la minimización del error de predicción puede proporcionar una solución al problema del cálculo de las actualizaciones bayesianas.

Tengase en cuenta que la inferencia bayesiana también funciona para los modelos jerárquicos. Suponiendo que las variables en niveles no adyacentes de la jerarquía son condicionalmente independientes, las estimaciones pueden actualizarse en paralelo en los diferentes niveles (cf. Friston, 2003, p. 1342), lo que idealmente produce un conjunto de estimaciones globalmente consistente (en la práctica, las cosas se complican, como señalan Lee & Mumford, 2003, p. 1437). En este caso, la idea es que la mayoría de los objetos del mundo no se influyen directamente entre sí de forma causal, pero siguen siendo objetos en el *mismo* mundo, lo que significa que las interacciones causales entre dos objetos arbitrarios suelen estar *mediadas* por otros objetos. Del mismo modo, las diferentes características de un objeto

individual no son completamente independientes porque son características del *mismo* objeto, pero esto no significa que las representaciones de estas características deben ser siempre procesadas conjuntamente. Por ejemplo, un disco azul puede representarse representando un determinado color (azul) en un lugar determinado y una forma determinada (un disco) en el mismo lugar. La información sobre la ubicación del color me da información sobre la ubicación de la forma. Sin embargo, si tengo una representación separada de la ubicación del disco, puedo tratar el color y la forma como (condicionalmente) independientes, es decir, dada la ubicación del disco, la información sobre el color no me da nueva información sobre la forma. Desde el punto de vista computacional, esto permite representaciones más dispersas, lo que también puede reflejarse en la segregación funcional en el cerebro (Friston & Buzsáki, 2016, que exploran esto centrándose en el dominio temporal).

4. El Procesamiento Predictivo y el Principio Ideomotor

Hasta ahora, la minimización del error de predicción sólo se ha descrito como una forma de generar percepciones de acuerdo con la entrada sensorial. Sin embargo, la función principal de la minimización del error de predicción puede no ser inferir causas ocultas en el mundo, sino provocar cambios en el mundo que ayuden al agente a permanecer vivo (véase la sección 7 más adelante). Además, el objetivo principal de estos cambios puede no ser el entorno externo sino el entorno interno del agente, es decir, su cuerpo. En los sistemas biológicos, la integridad del organismo es una prioridad de primer nivel, porque un organismo estable (que puede controlar sus estados internos) puede sobrevivir en diferentes entornos, mientras que un organismo inestable puede no sobrevivir ni siquiera en entornos favorables. Así lo ha señalado Anil Seth:

> El PP puede aplicarse de forma más natural a la interocepción (la sensación del estado fisiológico interno del cuerpo) que a la exterocepción (las sensaciones clásicas, que transmiten señales que se originan en el ambiente externo). Esto se debe a que es más importante para un organismo evitar encontrarse con estados interoceptivos que evitar encontrarse con estados exteroceptivos inesperados. Un nivel inesperado de oxigenación o de azúcar en la sangre probablemente sea una mala noticia para un organismo, mientras que las sensaciones exteroceptivas inesperadas (como las nuevas entradas visuales) tienen menos probabilidades de ser perjudiciales y en algunos casos pueden ser deseables. (Seth, 2015, p. 9)

Está claro que el objetivo de la inferencia interoceptiva no es simplemente inferir el estado interno del cuerpo, sino permitir el *control predictivo* de parámetros vitales como la oxigenación de la sangre o el azúcar en la sangre (característica N° 7). Seth ofrece el siguiente ejemplo. Cuando el cerebro detecta un descenso del azúcar en sangre a través de la inferencia interoceptiva, la percepción resultante (un deseo de comer cosas azucaradas) conducirá a errores de predicción

> en niveles jerárquicamente superiores, donde los modelos predictivos integran señales multimodales interoceptivas y exteroceptivas. Estos modelos instancian predicciones de secuencias temporales de entradas exteroceptivas e interoceptivas, que fluyen hacia abajo a través de la jerarquía. La cascada resultante de errores de predicción puede resolverse a través del control autónomo, con el fin de metabolizar las reservas de grasa corporal (inferencia activa), o mediante acciones alostáticas que implican al entorno externo (es decir, encontrar y comer cosas azucaradas). (Seth, 2015, p. 10)

La minimización del error de predicción interoceptiva es, por tanto, un ejemplo ilustrativo de cómo la percepción y la acción están acopladas, según el PP. Un objetivo de la MEP interoceptiva es mantener los parámetros vitales del organismo (como su nivel de azúcar en sangre, etc.) dentro de unos límites viables, lo que implica inferir el estado actual de estos parámetros y cambiarlos activamente (cuando sea necesario). En este sentido, Friston lo expresa así (en términos de minimización de la energía libre, que bajo ciertos supuestos implica minimizar el error de predicción):

> Los agentes pueden suprimir la energía libre cambiando las dos cosas de las que depende: pueden cambiar la entrada sensorial actuando sobre el mundo o pueden cambiar su densidad de reconocimiento cambiando sus estados internos. Esta distinción se traslada muy bien a la acción y la percepción. (Friston, 2010, p. 129)

En resumen, el error entre las señales sensoriales y las predicciones de las señales sensoriales (derivadas de las estimaciones internas) puede minimizarse cambiando las estimaciones internas y cambiando las señales sensoriales (mediante acción). Lo que esto sugiere es que las mismas representaciones internas que se activan en la percepción también pueden desplegarse para permitir la acción. Esto significa que no sólo hay un formato de datos común, sino también que al menos algunas de las representaciones que sustentan la percepción son numéricamente idénticas a las representaciones que sustentan la acción.

Este supuesto ya está presente en la *teoría ideomotora* de James (James, 1890)[17], cuyo núcleo es resumido de la siguiente manera por el autor: "[L]a idea de los efectos sensoriales del movimiento de M[18] se habrá convertido en una condición inmediatamente antecedente a la producción del movimiento mismo". (James, 1890, p. 586; cursiva omitida). Más recientemente, esto ha sido recogido por las explicaciones de *codificación común* de la representación de la acción (Hommel et al., 2001; Hommel, 2015; Prinz, 1990).[19] La idea básica es siempre similar: las representaciones neuronales de las causas ocultas en el mundo se superponen con los fundamentos neuronales de la preparación de la acción (lo que significa que partes de ellas son numéricamente idénticas). En otras palabras, hay representaciones "ideomotoras", que pueden funcionar como perceptos y como órdenes motoras.[20]

Desde el punto de vista computacional, el **Principio Ideomotor** (característica N° 9) logra una dualidad formal entre la acción y la percepción. La dualidad es la siguiente: Si puedo acceder perceptivamente a un estado de cosas *p*, esto significa que *p* tiene consecuencias perceptibles (o constituyentes) *c*; la acción está orientada a un objetivo, por lo que al realizar una acción quiero provocar algún estado de cosas *p*. Esto significa que la acción también puede describirse como un proceso en el que se producen las consecuencias perceptibles *c* a partir de *p*, y la percepción puede describirse como el proceso por el que se infieren las causas de una acción hipotética, lo que provoca *p*, y por tanto *c* (para una descripción rigurosa de esta idea, Todorov, 2009). Los beneficios computacionales de esta perspectiva dual se recogen en la noción de *inferencia activa* (desarrollada por Friston y sus colegas):

> En esta imagen del cerebro, las neuronas representan tanto la causa como la consecuencia: Codifican las expectativas condicionales sobre los estados ocultos del

17 Otro precursor de la idea se encuentra en las obras de Herbart (1825, pp. 464 y s.) y Lotze (1852, pp. 313 y s.).

18 En el fragmento citado, "M" hace referencia en el texto original a una célula motora con la que se ejemplifica sobre las vías de descarga (James, 1890). [N. de la T.]

19 Una revisión de los enfoques ideomotores puede encontrarse en Badets et al. (2014). Un resumen histórico puede encontrarse en Stock y Stock (2004).

20 En sentido estricto, las representaciones ideomotoras se consideran a veces sólo como contribuciones tardías (de alto nivel) a la percepción, y como los precursores (tempranos) de la acción (en la siguiente cita, "TCE" denota la teoría de la codificación de eventos (TCE)): "La TCE no tiene en cuenta la compleja maquinaria de los procesos sensoriales "tempranos" que conducen a ellos. Por el contrario, en lo que respecta a la acción, se centra en los antecedentes cognitivos "tempranos" de la acción que representan ciertas características de los acontecimientos que se van a generar en el entorno (= acciones). La TCE no tiene en cuenta la compleja maquinaria de los procesos motores "tardíos" que procesos motores "tardíos" que sirven para su realización (es decir, el control y la coordinación de los movimientos). Así pues, la TCE pretende ofrecer un marco para comprender los vínculos entre la percepción (tardía) y la acción (temprana), o la planificación de la acción". (Hommel et al., 2001, p. 849)

> mundo que causan los datos sensoriales, y al mismo tiempo causan esos estados indirectamente a través de la acción. [...] En resumen, la inferencia activa induce una causalidad circular que destruye las distinciones convencionales entre las representaciones sensoriales (consecuencia) y motoras (causa). Esto significa que la optimización de las representaciones corresponde a la percepción o intención, es decir, a la formación de percepciones o intenciones. (Friston et al., 2011, p. 138)[21]

La inferencia *activa* suele distinguirse de la inferencia *perceptiva*. Sin embargo, dado que ambas se efectúan minimizando el error de predicción, y dado que sus implementaciones pueden no ser claramente separables, la *inferencia activa* también se utiliza como un término más genérico, especialmente por Friston y sus colegas. En el contexto del principio de energía libre (véase más adelante), denota los procesos computacionales que minimizan la energía libre y que sustentan tanto la acción como la percepción: "Inferencia activa — la minimización de la energía libre mediante el cambio de estados internos (percepción) y estados sensoriales actuando sobre el mundo (acción)" (Friston et al., 2012a, p. 539)[22].

Tanto la acción como la percepción tienen en común la inferencia (inconsciente, aproximadamente bayesiana). Dado que se supone que las estructuras neuronales que sustentan la acción y la percepción, respectivamente, se supone que se superponen, la inferencia activa y la perceptiva funcionan conjuntamente.[23] Esta versión actualizada y enriquecida del **Principio Ideomotor** proporciona una perspectiva unificadora de la acción y la percepción, mientras que sus implicaciones y retos más profundos sólo están empezando a explorarse.[24]

21 Esto resuena con el "principio de reafirmación" (*Reafferenzprinzip*) de Von Holst y Mittelstaedt (1950), que también subraya que los eventos neuronales que acompañan a la percepción pueden no sólo ser considerados como efectos de las señales sensoriales, sino también como sus causas, porque pueden influir las señales sensoriales (a través de la acción).

22 Véase también Clark (2016, p. 181) y Burr (2017).

23 La misma idea se valora en un trabajo reciente de Lake, Salakhutdinov y Tenenbaum sobre el aprendizaje de conceptos: el sistema reconoce caracteres visuales a partir de inferir un "programa probabilístico", que es un modelo generativo que puede utilizarse para generar la entrada visual (cf. Lake et al., 2015, p. 1333).

24 Por ejemplo, Wiese (2016) sostiene que, si la versión del PP de la teoría ideomotora está en el camino correcto, la acción está habilitada por representaciones erróneas sistemáticas; Colombo (2017) sostiene que el PP desafía la teoría humeana de la motivación, en el sentido de que las apelaciones al deseo y al valor pueden no ser necesarias para para explicar la motivación social, mientras que la minimización de la incertidumbre social puede serlo.

5. Atención y Precisión

Una de las tantas ideas fructíferas formuladas en el marco del PP es que la asignación de la atención puede analizarse como el proceso de optimización de las estimaciones de precisión (característica N° 10). Esto fue propuesto por primera vez por Karl Friston y Klaas Stephan (Friston & Stephan, 2007) (dos importantes artículos que amplían esta idea son Feldman & Friston, 2010 y Hohwy, 2012). Dado que las estimaciones de precisión funcionan como ponderaciones de los términos de error de predicción, la precisión asociada a un error de predicción influencia su impacto sobre el procesamiento en otros niveles. Esto significa que el aumento de la precisión estimada puede mejorar la profundidad del procesamiento de un estímulo. Además, las estimaciones de precisión pueden modificarse de forma ascendente [*bottom-up*] y descendente [*top-down*]: De abajo hacia arriba, la precisión puede estimarse como una función de las muestras obtenidas (por ejemplo, como la inversa de la varianza de la muestra); de arriba hacia abajo, las estimaciones de precisión pueden ser moduladas en contextos en los que se anticipan aumentos o disminuciones de la precisión, o pueden funcionar como representaciones de objetivos para la acción mental (Metzinger, 2017). La diferencia entre los cambios ascendentes y descendentes en las estimaciones de precisión puede vincularse a la diferencia entre la atención endógena y exógena (para más detalles, Feldman & Friston, 2010 y Hohwy, 2012).

Utilizando esta explicación de optimización de la precisión en torno a la atención, es posible establecer una conexión entre la acción y la atención. Recordemos que, según el principio ideomotor, algunas estructuras neurales son parte de los fundamentos neurales tanto de la acción como de la percepción. Supongamos que la estructura neural *N* se activa cuando percibo que una persona se rasca la barbilla y cuando yo mismo estoy a punto de rascarme la barbilla. Siguiendo a Friston et al. (2011, p. 138), *N* podría funcionar como percepción tanto como intención, aunque normalmente sólo funciona como una de ellas. Así, a menos que sufra de ecopraxia, usualmente percibir un movimiento no hará que me mueva de la misma manera (aunque hay situaciones en las que las personas se mimetizan en cierta medida, Quadt, 2017). Esto puede explicarse dentro del marco del PP observando lo siguiente: la hipótesis de que me estoy rascando la barbilla producirá predicciones propioceptivas y otras predicciones sensoriales (que describen, por ejemplo, los estados de mis músculos cuando mi brazo se mueve). A menos que me esté rascando la barbilla, estas predicciones entrarán en conflicto con señales sensoriales, por lo que habrá un gran error de predicción, que llevará a una

actualización de la hipótesis de que me estoy rascando la barbilla. En otras palabras, la hipótesis no puede sostenerse en presencia de tales errores de predicción. De esta manera, para permitir el movimiento, las estimaciones de precisión asociadas a los errores de predicción sensoriales deben ser anuladas por la modulación descendente. Combinando esto con la hipótesis de que la atención aumenta las estimaciones de precisión, se lo podría describir como un proceso de *prestar atención* [*attending away*] a partir de las señales somatosensoriales. A la inversa, prestar atención a los estímulos sensoriales debería perjudicar el movimiento normal (Limanowski, 2017).

Esta conexión entre la acción y la atención también se explota en las explicaciones sobre las autocosquillas [*self-tickling*][25] (Van Doorn et al., 2014; Van Doorn et al., 2015). Las desviaciones en las estimaciones de precisión se han relacionado con trastornos de la atención y motores, y se han planteado en el contexto del autismo y la esquizofrenia (González-Gadea et al., 2015; Palmer et al., 2015; Van de Cruys et al., 2014; Friston et al., 2014; Adams et al., 2016). Este es solo un ejemplo de cómo el enfoque del PP puede poseer una gran fecundidad heurística y poder explicativo para la neuropsiquiatría cognitiva y campos afines.

6. El Cerebro como Testeador de Hipótesis

Ya hemos mencionado que una persona atrapada en un robot gigante (recordemos el experimento mental de Dennett) tiene que formar hipótesis sobre su entorno. Una de las razones por las que el PP puede parecer atractivo para algunos, aunque dudoso para otros, es que aplica descripciones del nivel personal de este tipo al nivel computacional de la descripción (véase el artículo "The hypothesis testing brain" ["El cerebro testeador de hipótesis"], Hohwy, 2010, característica N° 11). Sin embargo, tales descripciones al menos pueden ser heurísticamente fructíferas para tratar de responder a la pregunta de por qué percibimos el mundo como lo hacemos, en particular, la cuestión de cuáles son los principios formales de la organización perceptiva. El clásico artículo de Richard Gregory "Perceptions as hypotheses" ["Percepciones como hipótesis"] prueba exitosamente la idea de que las percepciones *explican* las señales sensoriales, y que tienen *poder predictivo* (Gregory, 1980, pp. 182, 186). Helmholtz ya sugirió una versión ampliada de esta idea, a saber, que los movimientos podrían ser considerados como experimentos:

25 No existe una traducción literal para el término "*self-tickling*", referido al fenómeno de realizarse cosquillas a uno mismo, por lo que optamos designarlo como "autocosquillas". [N. de la T.]

> *[W]ir beobachten unter fortdauernder eigener Thätigkeit, und gelangen dadurch zur Kenntniss des Bestehens eines gesetzlichen Verhältnisses zwischen unseren Innervationen und dem Präsentwerden der verschiedenen Eindrücke aus dem Kreise der zeitweiligen Präsentabilien. Jede unserer willkührlichen Bewegungen, durch die wir die Erscheinungsweise der Objecte abändern, ist als ein Experiment zu betrachten, durch welches wir prüfen, ob wir das gesetzliche Verhalten der vorliegenden Erscheinung, d.h. ihr vorausgesetztes Bestehen in bestimmter Raumordnung, richtig aufgefasst haben.* (Von Helmholtz, 1959, p. 39) [26]

Una aplicación clásica en el presente debate son los movimientos oculares sacádicos, que ahora se conceptualizan como una forma incorporada de comprobación de hipótesis (Friston et al., 2012b). Aparte de estas consideraciones heurísticas, si el PP va por buen camino, podemos preguntarnos si el cerebro participa *literalmente* en la inferencia. Esta pregunta es respondida afirmativamente por Alex Kiefer en su contribución a esta colección (Kiefer, 2017). Jelle Bruineberg mantiene una posición escéptica (Bruineberg, 2017 y Bruineberg et al., 2016). La cuestión más general de cómo se relacionan la psicología popular [*folk psychology*] y el PP, y hasta qué punto el uso científico de los conceptos folk-psicológicos puede necesitar ser revisado, es discutido por Joe Dewhurst (2017)[27].

7. El Procesamiento Predictivo y el Principio de Energía Libre de Karl Friston

Considere la siguiente tautología: Todo organismo que consigue mantenerse vivo durante un tiempo determinado no muere durante ese tiempo. Además, permanecer vivo implica correr el riesgo de morir. No se trata de una una visión profunda del concepto de vida, sino una observación aparentemente superficial sobre los organismos vivos. Sin embargo, tiene algunas implicaciones interesantes. Para todo organismo vivo, hay situaciones mortales que el organismo debe evitar para seguir vivo; y quizá, cuanto más sofisticado sea un organismo habrá más situaciones potencialmente mortales. Sólo piense en los entornos en los que una bacteria

26 "Observamos en medio de nuestra propia actividad continua, y así alcanzamos el conocimiento de la existencia de una relación lícita entre nuestras inervaciones y la presencia de diferentes impresiones de presentaciones temporales [*Präsentabilien*]. Todos nuestros movimientos voluntarios por los que cambiamos la apariencia de las cosas debe considerarse como un experimento, mediante el cual comprobamos si hemos captado correctamente el comportamiento lícito de la apariencia que nos ocupa, es decir, su supuesta existencia en estructuras espaciales determinadas". (Traducción propia*). * La traducción señalada anteriormente como propia, del alemán al inglés originalmente, corresponde a los autores del presente artículo. [N. de la T.]

27 Los dos artículos a los que se hace referencia, titulados "Literal Perceptual Inference" (Kiefer, 2017) y "Folk Psychology and the Bayesian Brain" (Dewhurst, 2017), se ubican en el 17° y 9° lugar de la mencionada colección, respectivamente. [N. de la T.]

puede sobrevivir y compárelos con aquellos en los que puede hacerlo un ser humano. Si un organismo ha conseguido permanecer vivo durante un tiempo determinado, significa que (hasta el momento) ha evitado cualquier situación mortal.

Si hacemos una lista de las posibles situaciones en las que un organismo *podría* encontrarse, y la comparamos con las posibles situaciones en las que el organismo es capaz de sobrevivir, encontraremos dos cosas:

> 1. para la mayoría de los organismos (por ejemplo, los seres humanos), la segunda lista será drásticamente más corta que la primera (debido a que hay muchas situaciones mortales); y
>
> 2. si observamos un organismo capaz de sobrevivir durante un tiempo considerable, en un momento aleatorio de su vida, es muy probable que se encuentre en una situación de la segunda lista (lo que se hace eco de la tautología del principio de esta sección).

Podemos reexpresar estas dos observaciones de una manera un poco más técnica. Llamemos su *espacio de estados* al conjunto de todos los posibles estados en los que un organismo podría encontrarse, donde un estado se define por las señales sensoriales actuales recibidas por el sistema sensorial del organismo. En principio, ahora podemos definir una distribución de probabilidad sobre este espacio de estados que asigna probabilidades a las diferentes regiones de dicho espacio y describe la probabilidad de que el organismo se encuentre en las respectivas regiones durante su vida. Algunas regiones tendrán una alta probabilidad (por ejemplo, es probable que un pez se encuentre en el agua); otras tendrán una baja probabilidad (es poco probable que un pez se encuentre fuera del agua). Además, la *mayoría* de las regiones del espacio de estados tendrán una probabilidad baja (debido a que hay muchas situaciones mortales). Formalmente, esto significa que la *entropía* de la distribución de probabilidad es baja (sería máxima si asignara probabilidades uniformemente a las diferentes regiones del espacio de estados; véase a continuación un ejemplo formal sencillo). Con esta distribución de probabilidad a mano, podemos hacer una apuesta sobre el lugar del espacio de estado en el que se encontrará el organismo, si se observa en un momento arbitrario de su vida. Como la distribución asigna probabilidades extremadamente bajas a la mayoría de las regiones del espacio de estado, podemos hacer una conjetura bastante precisa (por ejemplo, podemos adivinar que un pez estará en el agua, que un pez de agua dulce estará en el agua dulce, etc.).

Consideremos ahora lo siguiente. A lo largo de la vida del pez, tomamos repetidas muestras de sus estados y construimos una *distribución empírica* utilizando dichas muestras. Una distribución empírica asigna probabilidades que reflejan la *frecuencia* con la que se tomaron muestras (al azar) de las diferentes regiones. Como ejemplo sencillo, pensemos en un dispositivo que produce uno de dos números, 0 y 1, cada vez que se pulsa un botón, y los dos números se producen con ciertas probabilidades desconocidas para el agente. Puede ser que ambos números se produzcan con la misma probabilidad (0,5), o que uno se produzca con mucha más frecuencia que el otro (digamos que el 0 se produce con una probabilidad de 0,9 y el 1 se produce con una probabilidad de 0,1). Cada vez que se pulsa el botón se anota qué número se ha producido (se trata de una muestra única) y, contando la frecuencia con la que se produce cada número, se puede construir una distribución empírica utilizando las frecuencias relativas. Por ejemplo, si 14 de cada 100 muestras son 0, y las otras 86 muestras son 1, la distribución empírica podría asignar la probabilidad 0,14 a 0 y 0,86 a 1. La entropía de esta distribución sería aproximadamente 0,58.

Pero, en primer lugar, ¿qué es la entropía? Es la sorpresa media de (en este caso) las diferentes salidas [*outputs*] del dispositivo. Aquí, la sorpresa[28] es una noción técnica para el logaritmo negativo de la probabilidad de un evento. La sorpresa media (la entropía) se calcula ahora como sigue $H = -[0.14 * \log(0,14) + 0,86 * \log(0,86)]$. Si esta cantidad es baja se debe a que los valores de sorpresa de los resultados individuales son bajos (o al menos la mayoría de ellos). Por tanto, para tener una entropía baja, la sorpresa de los estados debe ser baja en cualquier momento (o al menos la mayor parte del tiempo).

Podemos volver a aplicar esto al ejemplo de los peces. La mayor parte del tiempo, el pez estará en estados no sorpresivos. Si tenemos un conocimiento exhaustivo del pez, en principio podemos describir las regiones del espacio de estados en las que es probable que se encuentre, y construir una distribución de probabilidad "de sillón" que refleje este conocimiento. O podemos observar al pez y anotar las frecuencias relativas con las que se encuentra en diferentes regiones de su espacio de estados. A largo plazo, esta distribución empírica debería ser cada vez más similar a la distribución "de sillón". (Esta es una descripción informal del supuesto de *ergodicidad*, que es una característica formalmente definida de ciertos procesos aleatorios. Friston, 2009, p. 293).

28 Los autores señalan que tanto los términos "*surprise*" como "*surprisal*" hacen referencia al mismo concepto en esta ocasión. Omitimos en el cuerpo del texto la aclaración "(*also called "surprisal"*)" debido a que la traducción al español es la misma para ambos casos. [N. de la T.]

Hasta ahora, hemos observado a los peces desde fuera, desde la perspectiva del observador. ¿Qué ocurre si cambiamos nuestro punto de vista y observamos al pez desde "la perspectiva del animal" (como lo llama Eliasmith, 2000, pp. 25 y siguientes)? La diferencia clave es que ni siquiera conocemos el estado actual del pez. Un organismo adquiere conocimientos sobre su propio estado actual mediante mediciones sensoriales, pero estas mediciones sólo proporcionan al organismo una información parcial y tal vez ruidosa. Además, el organismo no tiene acceso a la distribución de probabilidad con la que se puede calcular la sorpresa de sus estados. En este caso, el principio de energía libre (PEL) ofrece una solución de principio (característica N° 12).

La estrategia general de PEL consiste en dos pasos. El primero de ellos es intentar hacer coincidir una distribución de probabilidad codificada internamente (una distribución de reconocimiento) a la verdadera distribución posterior de los estados ocultos, dadas las señales sensoriales. El segundo es tratar de cambiar las señales sensoriales de tal manera que la sorpresa de estados sensoriales y ocultos sea baja en un momento dado. Esto puede aparentar que empeora las cosas, ya que ahora hay dos problemas: ¿Cómo se puede ajustar la distribución de reconocimiento a una posterior desconocida y cómo se puede minimizar la sorpresa de las señales sensoriales, si la distribución relativa a la que se define la sorpresa es desconocida? El ingenio del PEL consiste en resolver ambos problemas minimizando la energía libre. En este caso, la energía libre es una cantidad teórica de la información, cuya minimización es posible desde la perspectiva del animal (para más detalles, Friston, 2008; Friston, 2009; Friston, 2010).

Explicar esto requiere una descripción un poco más formal (aquí simplificaremos las cosas; una explicación mucho más detallada, pero aún accesible, del principio de energía libre se puede encontrar en Bogacz, 2015). En primer lugar, "hacer coincidir" la distribución de reconocimiento con una posterior desconocida es solo una aproximación a la inferencia bayesiana en la que se supone que la distribución de reconocimiento tiene una forma determinada (por ejemplo, gaussiana). Esto simplifica los cálculos. En segundo lugar, una vez que se ha calculado una aproximación al modelo real, la energía libre constituye un límite estricto para la sorpresa de las señales sensoriales. Por lo tanto, la minimización de la energía libre cambiando las señales sensoriales minimizará, implícitamente, la sorpresa.

Obsérvese que la conexión entre la PEL y la inferencia bayesiana es sencilla: minimizar la energía libre implica aproximar la distribución posterior por medio de la distribución de

reconocimiento. Si se supone que la distribución de reconocimiento es gaussiana (con la famosa función de densidad de probabilidad en forma de campana), minimizar la energía libre implica minimizar los errores de predicción ponderados por la precisión. Así que, al menos bajo este supuesto (que se denomina *supuesto de Laplace*), también existe una conexión entre el PEL y la minimización de los errores de predicción. De hecho, el PEL puede considerarse como la teoría fundamental, que puede combinar las diferentes características del procesamiento predictivo descritas anteriormente dentro de un marco único y formalmente riguroso. Sin embargo, es discutible cuáles de estas características están realmente implicadas en el PEL. Como ya se ha mencionado, la **Reclusión Ambiental** es un ejemplo de característica controvertida (Fabry, 2017a; Clark, 2017). Por lo tanto, podría ser útil examinar aspectos específicos de esta novedosa propuesta no solo desde una perspectiva empírica, sino también conceptual y metateórica. Este fue uno de los principales motivos de nuestra iniciativa, que dio lugar a la actual colección de textos.

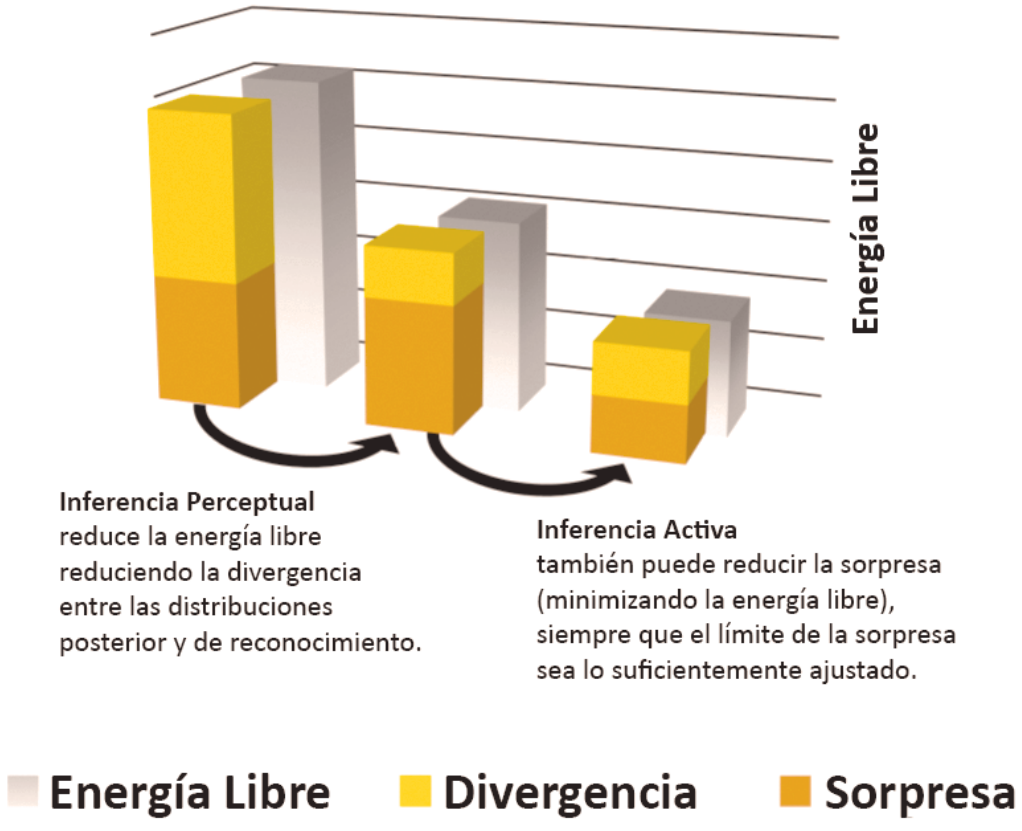

Figura 1[29]**:** Ilustración esquemática de cómo la minimización de la energía libre puede, implícitamente, minimizar la sorpresa. Inicialmente, la distribución de reconocimiento no se ajustará muy bien a la verdadera distribución posterior (de causas ocultas, dadas las señales sensoriales). Para mejorar la distribución de reconocimiento, se puede cambiar de tal manera que las señales sensoriales medidas se vuelvan más probables, dado este modelo (esto significa que la evidencia del modelo se incrementa). Una forma de ponerlo en práctica es minimizar el error de predicción. Así pues, se supone que las señales sensoriales no son sorpresivas, y esto debería reflejarse en la distribución de reconocimiento (es decir, la distribución de reconocimiento se altera de tal manera que, en relación con esta distribución, las señales sensoriales no son sorpresivas). Por supuesto, podría ser que las señales sensoriales sean, en relación con la verdadera posterior, sorpresivas. Por esta razón hay que comprobar la distribución de reconocimiento. Esto se hace, implícitamente, provocando cambios en el mundo tal que, si la distribución de reconocimiento es adecuada, las señales sensoriales no serán sorpresivas. Se trata de un muestreo activo. Hasta cierto punto, las señales sensoriales siempre serán sorpresivas, por lo que siempre será necesario un ajuste de la distribución de reconocimiento, seguido de un muestreo activo y un nuevo ajuste de la distribución de reconocimiento, etc. Así, este proceso de arranque [*bootstrapping*] funciona mediante un procedimiento continuo de prueba y error, y depende de una íntima conexión causal entre el agente y su entorno. Aunque las flechas negras pretenden indicar una secuencia temporal, no tiene por qué haber una separación nítida entre la inferencia perceptiva y la inferencia activa, y el proceso de arranque [*bootstrapping*] también podría comenzar con los movimientos corporales.

8. Glosario

Inferencia activa: 1. Proceso computacional en el que el error de predicción se minimiza actuando sobre el mundo ("haciendo que el mundo se parezca más al modelo"), en lugar de minimizar el error de predicción cambiando el modelo interno, es decir, la inferencia perceptiva ("hacer que el modelo se parezca más al mundo") 2. También se utiliza como término genérico para los procesos computacionales que sustentan tanto la acción como la percepción y, en el contexto del PEL, para todos los procesos computacionales que minimizan la energía libre.

Inferencia bayesiana: Actualización de un modelo según la regla de Bayes, es decir, el cálculo de la distribución posterior $p(c|s) = p(s|c)p(c)/p(s)$. Para un ejemplo, (Harkness y Keshava, 2017).

Modelo contrafáctico: Un modelo contrafáctico es una distribución de probabilidad condicional que relaciona posibles acciones con posibles estados futuros (al menos siguiendo a Friston et al., 2012b).

Estimador: Un estimador estadístico es una función de variables aleatorias que se conciben como muestras; así un estimador especifica cómo calcular una estimación a partir de los datos observados. Una estimación es un valor particular de un estimador (que se calcula cuando se obtienen muestras concretas, es decir, realizaciones de variables aleatorias).

29 La imagen original fue alterada con el fin de efectuar la traducción del texto que figura en ella. Se conservaron los gráficos, los colores pudieron haber sufrido alguna alteración en el proceso de edición. [N. de la T.]

"Dar explicaciones" ["Explaining-away"][30]: La noción de "dar explicaciones" es ambigua. 1. Algunos autores escriben que las señales sensoriales se explican mediante predicciones descendentes (cf. Clark, 2013a, p. 187). 2. Otro sentido en el que se utiliza el término es en el que las hipótesis o modelos que compiten entre sí son explicados (cf. Hohwy, 2010, p. 137). 3. Un tercer sentido es el de explicar el error de predicción (cf. Clark, 2013a, p. 187).

Energía libre: En el contexto del PEL de Friston, la energía libre no es una cantidad termodinámica, sino una cantidad teórica de la información que constituye un límite superior para la sorpresa. Si este límite es estrecho, la sorpresa de las señales sensoriales puede reducirse si se minimiza la energía libre provocando cambios en el mundo.

Distribución gaussiana: La famosa distribución de probabilidad en forma de campana (también llamada distribución normal). Su importancia se basa en el teorema del límite central que, básicamente, afirma que muchas distribuciones pueden ser aproximadas por distribuciones gaussianas.

Modelo generativo: La distribución de probabilidad conjunta de dos o más variables aleatorias, a menudo dada en términos de un *a priori* y una probabilidad: $p(s,c) = p(s|c)p(c)$. (A veces, sólo la probabilidad $p(s|c)$ se denomina un "modelo generativo"). El modelo es generativo en el sentido en que modela cómo las señales sensoriales son *generadas* por las causas ocultas *c*. Además, puede utilizarse para generar señales sensoriales simuladas, dada una estimación de las causas ocultas.

Jerarquía: El PP propone una jerarquía de estimadores, que operan en diferentes escalas espacio-temporales (por lo que rastrean características a diferentes escalas). La jerarquía no tiene necesariamente un nivel superior (pero puede tener un centro —piense en los niveles como anillos en un disco o una esfera).

Problema inverso: desde el punto de vista de la codificación predictiva, el problema de la percepción requiere invertir el mapeo de las causas ocultas a las señales sensoriales. Este problema es difícil, por decir lo menos, porque no suele haber una solución única, y las señales sensoriales suelen ser ruidosas (lo que significa que el mapeo de las causas ocultas hacia las señales sensoriales no es determinista).

Predicción: Una predicción es una función determinista de una estimación, que puede compararse con otra estimación (la estimación predicha). Las predicciones no son necesariamente sobre el futuro (nótese que una puede ser predictiva de otra variable si la primera lleva información sobre la segunda, es decir, si existe una correlación, cf. Anderson y Chemero, 2013, p. 204). Aun así, muchas estimaciones en el PP también son predictivas en el sentido temporal (cf. Butz, 2017; Clark, 2013c, p. 236).

Precisión: La precisión de una variable aleatoria es la inversa de su varianza. En otras palabras, cuanto mayor sea la divergencia promedio con respecto a su media, menor será la precisión de una variable aleatoria (y viceversa).

Variable aleatoria: Una variable aleatoria es una función medible entre un espacio de probabilidad y un espacio medible. Por ejemplo, un dado de seis caras puede modelarse como una variable aleatoria, que

30 La expresión "explaining-away", en los casos subsiguientes, puede referirse a las acciones de "justificar", "dar razones", "fundamentar" o, como hemos escogido debido a la mayor amplitud que presenta tal expresión, "dar explicaciones". [N. de la T.]

asigna cada uno de los seis sucesos igualmente probables a uno de los números del conjunto {1,2,3,4,5,6}.

Sorpresa: Noción teórica de la información que especifica lo improbable que es un suceso, dado un modelo. Más concretamente, se refiere al logaritmo negativo de la probabilidad de un suceso. Es importante no confundir este concepto subpersonal de la teoría de la información con la noción fenomenológica de nivel personal de "sorpresa".

9. Referencias

Adams, R. A., Huys, Q. J. & Roiser, J. P. (2016). Computational psychiatry: Towards a mathematically informed understanding of mental illness. *J Neurol Neurosurg Psychiatry, 87* (1), 53-63. https://dx.doi.org/10.1136/jnnp-2015-310737

Anderson, M. L. (2017). Of Bayes and bullets: An embodied, situated, targeting-based account of predictive processing. En T. Metzinger & W. Wiese (Eds.) *Philosophy and predictive processing*. MIND Group.

Anderson, M. L. & Chemero, T. (2013). The problem with brain GUTs: Conflation of different senses of "prediction" threatens metaphysical disaster. *Behavioral and Brain Sciences, 36* (3), 204–205.

Badets, A., Koch, I. & Philipp, A. M. (2014). A review of ideomotor approaches to perception, cognition, action, and language: Advancing a cultural recycling hypothesis. *Psychological Research, 80* (1), 1–15. https://dx.doi.org/10.1007/s00426-014-0643-8

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P. & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron, 76* (4), 695-711. https://dx.doi.org/10.1016/j.neuron.2012.10.038

Bogacz, R. (2015). A tutorial on the free-energy framework for modelling perception and learning. *Journal of Mathematical Psychology*. https://dx.doi.org/10.1016/j.jmp.2015.11.003

Brodski, A., Paasch, G.-F., Helbling, S. & Wibral, M. (2015). The faces of predictive coding. *The Journal of Neuroscience, 35* (24), 8997-9006. https://dx.doi.org/10.1523/jneurosci.1529-14.2015

Brook, A. (2013). Kant's view of the mind and consciousness of self. En E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*.

Bruineberg, J. (2017). Active inference and the primacy of the 'I can'. En T. Metzinger & W. Wiese (Eds.) *Philosophy and predictive processing*. MIND Group.

Bruineberg, J., Kiverstein, J. & Rietveld, E. (2016). The anticipating brain is not a scientist: The free-energy principle from an ecological-enactive perspective. *Synthese*, 1–28. https://dx.doi.org/10.1007/s11229-016-1239-1

Burr, C. (2017). Embodied decisions and the predictive brain. En T. Metzinger & W. Wiese (Eds.) *Philosophy and predictive processing*. MIND Group.

Butz, M. V. (2017). Which structures are out there? Learning predictive compositional concepts based on social sensorimotor explorations. En T. Metzinger & W. Wiese (Eds.) *Philosophy and predictive processing*. MIND Group.

Clark, A. (2013a). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36* (3), 181–204. https://dx.doi.org/10.1017/S0140525X12000477

Clark, A. (2013b). The many faces of precision (Replies to commentaries on "Whatever next? Neural prediction, situated agents, and the future of cognitive science"). *Frontiers in Psychology, 4*, 270. https://dx.doi.org/10.3389/fpsyg.2013.00270

Clark, A. (2013c). Are we predictive engines? Perils, prospects, and the puzzle of the porous perceiver. *Behavioral and Brain Sciences, 36* (3), 233–253. https://dx.doi.org/10.1017/S0140525X12002440

Clark, A. (2015). Radical predictive processing. *The Southern Journal of Philosophy, 53*, 3–27. https://dx.doi.org/10.1111/sjp.12120

Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.

Clark, A. (2017). How to knit your own Markov blanket: Resisting the second law with metamorphic minds. En T. Metzinger & W. Wiese (Eds.) *Philosophy and predictive processing*. MIND Group.

Clark, A. (En prensa). Busting out: Predictive brains, embodied minds, and the puzzle of the evidentiary veil. *Noûs*. https://dx.doi.org/10.1111/nous.12140

Clowes, M. B. (1969). Pictorial relationships – A syntactic approach. En B. Meltzer & D. Michie (Eds.) (pp. 361–383). Edinburgh University Press.

Colombo, M. (2017). Social motivation in computational neuroscience: Or if brains are prediction machines then the Humean theory of motivation is false. En J. Kieverstein (Ed.) *Routledge handbook of philosophy of the social mind*. Routledge.

Dennett, D. C. (2013). *Intuition pumps and other tools for thinking*. Norton & Company.

Dewhurst, J. (2017). Folk psychology and the Bayesian brain. En T. Metzinger & W. Wiese (Eds.) *Philosophy and predictive processing*. MIND Group.

Downey, A. (2017). Radical sensorimotor enactivism & predictive processing. Providing a conceptual framework for the scientific study of conscious perception. En T. Metzinger & W. Wiese (Eds.) *Philosophy and predictive processing*. MIND Group.

Dołega, K. (2017). Moderate predictive processing. En T. Metzinger & W. Wiese (Eds.) *Philosophy and predictive processing*. MIND Group.

Drayson, Z. (2017). Modularity and the predictive mind. En T. Metzinger & W. Wiese (Eds.) *Philosophy and predictive processing*. MIND Group.

Egan, F. (2014). How to think about mental content. *Philosophical Studies, 170* (1), 115-135. https://dx.doi.org/10.1007/s11098-013-0172-0

Eliasmith, C. (2000). *How neurons mean: A neurocomputational theory of representational content*. PhD dissertation, Washington University in St. Louis. Department of Philosophy.

Engel, A. K., Fries, P. & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nat Rev Neurosci, 2* (10), 704–716.

Fabry, R. E. (2017a). Predictive processing and cognitive development. En T. Metzinger & W. Wiese (Eds.) *Philosophy and predictive processing*. MIND Group.

Fabry, R. E. (2017b). Transcending the evidentiary boundary: Prediction error minimization, embodied interaction, and explanatory pluralism. *Philosophical Psychology*, 1–20. https://dx.doi.org/10.1080/09515089.2016.1272674

Feldman, H. & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience, 4*. https://dx.doi.org/10.3389/fnhum.2010.00215

Friston, K. (2003). Learning and inference in the brain. *Neural Networks, 16* (9), 1325–1352.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences, 360* (1456), 815-836. https://dx.doi.org/10.1098/rstb.2005.1622

Friston, K. (2008). Hierarchical models in the brain. *PLoS Computational Biology, 4* (11), e1000211. https://dx.doi.org/10.1371/journal.pcbi.1000211

Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences, 13* (7), 293–301. https://dx.doi.org/10.1016/j.tics.2009.04.005

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience, 11* (2), 127–138. https://dx.doi.org/10.1038/nrn2787

Friston, K. & Buzsáki, G. (2016). The functional anatomy of time: What and when in the brain. *Trends in Cognitive Sciences, 20* (7), 500–511. https://dx.doi.org/10.1016/j.tics.2016.05.001

Friston, K. & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364* (1521), 1211–1221. https://dx.doi.org/10.1098/rstb.2008.0300

Friston, K. J. & Stephan, K. E. (2007). Free-energy and the brain. *Synthese, 159* (3), 417-458. https://dx.doi.org/10.1007/s11229-007-9237-y

Friston, K., Mattout, J. & Kilner, J. (2011). Action understanding and active inference. *Biological Cybernetics, 104* (1-2), 137–160. https://dx.doi.org/10.1007/s00422-011-0424-z

Friston, K., Samothrakis, S. & Montague, R. (2012a). Active inference and agency: Optimal control without cost functions. *Biological Cybernetics, 106* (8), 523-541. https://dx.doi.org/10.1007/s00422-012-0512-8

Friston, K., Adams, R., Perrinet, L. & Breakspear, M. (2012b). Perceptions as hypotheses: Saccades as experiments. *Frontiers in Psychology, 3* (151). https://dx.doi.org/10.3389/fpsyg.2012.00151

Friston, K. J., Stephan, K. E., Montague, R. & Dolan, R. J. (2014). Computational psychiatry: The brain as a phantastic organ. *The Lancet Psychiatry, 1* (2), 148–158. https://dx.doi.org/10.1016/S2215-0366(14)70275-5

Giordanetti, P., Pozzo, R. & Sgarbi, M. (2012). *Kant's philosophy of the unconscious*. De Gruyter.

Gonzalez-Gadea, M. L., Chennu, S., Bekinschtein, T. A., Rattazzi, A., Beraudi, A., Tripicchio, P., Moyano, B., Soffita, Y., Steinberg, L., Adolfi, F., Sigman, M., Marino, J., Manes, F. & Ibanez, A.

(2015). Predictive coding in autism spectrum disorder and attention deficit hyperactivity disorder. *Journal of Neurophysiology, 114* (5), 2625-2636. https://dx.doi.org/10.1152/jn.00543.2015

Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 290* (1038), 181–197.

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences, 27* (3), 377–396.

Gładziejewski, P. (2016). Predictive coding and representationalism. *Synthese*, 559–582. https://dx.doi.org/10.1007/s11229-015-0762-9

Harkness, D. L. & Keshava, A. (2017). Moving from the what to the how and where – Bayesian models and predictive processing. En T. Metzinger & W. Wiese (Eds.) *Philosophy and predictive processing*. MIND Group.

Herbart, J. F. (1825). *Psychologie als Wissenschaft neu gegründet auf Erfahrung, Metaphysik und Mathematik. Zweiter, analytischer Teil*. Unzer.

Hohwy, J. (2010). The hypothesis testing brain: Some philosophical applications. En W. Christensen, E. Schier & J. Sutton (Eds.) *Proceedings of the 9th conference of the Australasian society for cognitive science* (pp. 135–144). Macquarie Centre for Cognitive Science. https://dx.doi.org/10.5096/ASCS200922

Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. *Frontiers in Psychology, 3*. https://dx.doi.org/10.3389/fpsyg.2012.00096

Hohwy, J. (2013). *The predictive mind*. Oxford University Press.

Hohwy, J. (2016). The self-evidencing brain. *Noûs*, 50 (2), 259–285. https://dx.doi.org/10.1111/nous.12062

Hohwy, J. (2017). How to entrain your evil demon. En T. Metzinger & W. Wiese (Eds.) *Philosophy and predictive processing*. MIND Group.

Hommel, B. (2015). The theory of event coding (TEC) as embodied-cognition framework. *Frontiers in Psychology, 6*. https://dx.doi.org/10.3389/fpsyg.2015.01318

Hommel, B., Müsseler, J., Aschersleben, G. & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences, 24*, 849–878. https://dx.doi.org/10.1017/S0140525X01000103

Horn, B. K. P. (1980). *Derivation of invariant scene characteristics from images* (pp. 371–376). https://dx.doi.org/10.1145/1500518.1500579

James, W. (1890). *The principles of psychology*. Henry Holt.

Kant, I. (1998). *Kritik der reinen Vernunft*. Meiner.

Kant, I. (2010). *Crítica de la razón pura* (trad. M. Caimi). Colihue.

Kiefer, A. (2017). Literal perceptual inference. En T. Metzinger & W. Wiese (Eds.) *Philosophy and predictive processing*. MIND Group.

Lake, B. M., Salakhutdinov, R. & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science, 350* (6266), 1332-1338. https://dx.doi.org/10.1126/science.aab3050

Lee, T. S. & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *J. Opt. Soc. Am. A, 20* (7), 1434–1448. https://dx.doi.org/10.1364/JOSAA.20.001434

Lenoir, T. (2006). Operationalizing Kant: Manifolds, models, and mathematics in Helmholtz's theories of perception. En M. Friedman & A. Nordmann (Eds.) *The Kantian legacy in nineteenth-century science* (pp. 141–210). Cambridge, MA: MIT Press.

Limanowski, J. (2017). (Dis-)attending to the body. Action and self-experience in the active inference framework. En T. Metzinger & W. Wiese (Eds.) *Philosophy and predictive processing*. MIND Group.

Lotze, R. H. (1852). *Medicinische Psychologie oder Physiologie der Seele*. Weidmann'sche Buchhandlung.

Metzinger, T. (2004). *Being no one: The self-model theory of subjectivity*. MIT Press.

Metzinger, T. (2017). The problem of mental action. Predictive control without sensory sheets. En T. Metzinger & W. Wiese (Eds.) *Philosophy and predictive processing*. MIND Group.

Palmer, C. J., Paton, B., Kirkovski, M., Enticott, P. G. & Hohwy, J. (2015). Context sensitivity in action decreases along the autism spectrum: A predictive processing perspective. *Proceedings of the Royal Society of London B: Biological Sciences, 282* (1802). https://dx.doi.org/10.1098/rspb.2014.1557

Prinz, W. (1990). A common coding approach to perception and action. En O. Neumann & W. Prinz (Eds.) *Relationships between perception and action* (pp. 167–201). Heidelberg: Springer.

Quadt, L. (2017). Action-oriented predictive processing and social cognition. En T. Metzinger & W. Wiese (Eds.) *Philosophy and predictive processing*. MIND Group.

Seth, A. K. (2015). The cybernetic Bayesian brain: From interoceptive inference to sensorimotor contingencies. En T. Metzinger & J. M. Windt (Eds.) *Open MIND*. MIND Group. https://dx.doi.org/10.15502/9783958570108.

Shi, Y. Q. & Sun, H. (1999). *Image and video compression for multimedia engineering: fundamentals, algorithms, and standards*. CRC Press.

Sloman, A. (1984). Experiencing omputation: A tribute to Max Clowes. En M. Yazdani (Ed.) *New horizons in educational computing* (pp. 207–219). John Wiley & Sons.

Snowdon, P. (1992). How to interpret 'direct perception'. En T. Crane (Ed.) *The contents of experience* (pp. 48–78). Cambridge University Press.

Spratling, M. W. (2016). A review of predictive coding algorithms. *Brain and Cognition*. https://dx.doi.org/10.1016/j.bandc.2015.11.003

Stock, A. & Stock, C. (2004). A short history of ideo-motor action. *Psychological Research, 68*, 176–188. https://dx.doi.org/10.1007/s00426-003-0154-5

Swanson, L. R. (2016). The predictive processing paradigm has roots in Kant. *Frontiers in Systems Neuroscience, 10*, 79. https://dx.doi.org/10.3389/fnsys.2016.00079

Todorov, E. (2009). Parallels between sensory and motor information processing. En M. S. Gazzaniga (Ed.) *The cognitive neurosciences. 4th edition* (pp. 613–623). MIT Press.

Van de Cruys, S., Evers, K., Van der Hallen, R., van Eylen, L., Boets, B., de-Wit, L. & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review, 121* (4), 649–675. https://dx.doi.org/10.1037/a0037665

Van Doorn, G., Hohwy, J. & Symmons, M. (2014). Can you tickle yourself if you swap bodies with someone else? *Consciousness and Cognition, 23*, 1-11. http://dx.doi.org/10.1016/j.concog.2013.10.009

Van Doorn, G., Paton, B., Howell, J. & Hohwy, J. (2015). Attenuated self-tickle sensation even under trajectory perturbation. *Consciousness and Cognition, 36*, 147–153. https://dx.doi.org/10.1016/j.concog.2015.06.016

Von Helmholtz, H. (1855). *Ueber das Sehen des Menschen*. Leopold Voss.

Von Helmholtz, H. (1867). *Handbuch der physiologischen Optik*. Leopold Voss.

Von Helmholtz, H. (1959[1879/1887]). *Die Tatsachen in der Wahrnehmung. Zählen und Messen*. Wissenschaftliche Buchgesellschaft.

Von Helmholtz, H. (1985[1925]). *Helmholtz's treatise on physiological optics*. Gryphon Editions.

Von Holst, E. & Mittelstaedt, H. (1950). Das Reafferenzprinzip. *Die Naturwissenschaften, 37* (20), 464–476.

Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L. & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proc Natl Acad Sci U S A, 108* (51), 20754-9. https://dx.doi.org/10.1073/pnas.1117807108

Wiese, W. (2016). Action is enabled by systematic misrepresentations. *Erkenntnis*. https://dx.doi.org/10.1007/s10670-016-9867-x

Zellner, A. (1988). Optimal information processing and Bayes's theorem. *The American Statistician, 42* (4), 278–280. https://dx.doi.org/10.2307/2685143