

## El problema de la relevancia bajo el reflector

### The frame problem under the spotlight

Magalí La Rocca  
Universidad de Buenos Aires  
*laroccamagali@gmail.com*

#### Resumo

El “problema de la relevancia” en ciencias cognitivas es acerca de cómo la mente logra determinar lo que es relevante para realizar una acción o resolver un problema. Fodor (1983; 1987; 2000) afirma que no es posible solucionar este problema ya que no se puede entender cómo funciona la cognición en lo que concierne a la fijación de creencias y la toma de decisiones. En este trabajo mi tesis es que la teoría del espacio de trabajo global (Baars, 1988; 1997; Dehaene, 2015; Dehaene y Naccache, 2001) es una propuesta prometedora para resolver el problema de la relevancia, respecto con dos aspectos importantes: el aspecto computacional y el aspecto epistemológico. Sin embargo, para que ello suceda es necesario aclarar que se entiende por relevancia en estos procesos cognitivos.

#### Palavras-chave

Problema de la relevancia; Teoría del espacio de trabajo global; Conciencia.

#### Abstract

The frame problem for cognitive science is determining what information is relevant in order to perform a cognitive task. Fodor (1983; 1987; 2000) claims that the frame problem is unsolvable because it is impossible to understand how belief fixation and decision-making work. The aim of this paper is to show that the Global Workspace Theory (Baars, 1988; 1997; Dehaene, 2015; Dehaene and Naccache, 2001) is a promising approach in order to solve the frame problem, at least regarding two important aspects of the problem, namely: its computational aspect and its epistemological aspect. However, to evaluate this claim it is necessary to shed light on the notion of relevance for cognitive processes.

#### Keywords

Frame Problem; Global Workspace Theory; Consciousness.

### 1. Introducción

La “primera ley de la inexistencia de la ciencia cognitiva de Fodor” (Fodor, 1983) plantea que cuánto más isotrópico es un proceso cognitivo, menos se lo entiende. Por isotropía, Fodor entiende “el principio según el cual cualquier hecho puede resultar relevante (o irrelevante) para la confirmación de cualquier otro” (Fodor, 1983, p. 109). Esto quiere decir que cualquier información, en principio, puede ser relevante para cualquier tarea. Pero, ¿cómo logra la mente determinar qué es lo relevante? Esto es lo que se va a conocer como el problema del marco o problema de la relevancia. En este trabajo distingo entre el problema del marco, presentado por McCarthy y Hayes (1969) en el ámbito de la inteligencia artificial y el problema de la relevancia presentado posteriormente dentro del ámbito de las ciencias cognitivas.

Si bien dentro de la inteligencia artificial se han propuesto varias soluciones al problema del marco, Fodor (1983, 1987, 2000) sostiene que no es posible resolver este problema. Mi tesis es que la teoría del espacio de trabajo global (Baars, 1988, 1997; Dehaene, 2015; Dehaene y Naccache, 2001) es una propuesta prometedora para resolver el problema de la relevancia, respecto con dos aspectos importantes: el aspecto computacional y el aspecto epistemológico.

Sin embargo, para que ello suceda es necesario aclarar que se entiende por relevancia en los procesos cognitivos.

En la sección 2 presento el problema del marco y su posterior reinterpretación, i.e., el problema de la relevancia. En la sección 3, delinearé los principales aspectos de la teoría de espacio de trabajo global tal como la presenta Baars (1988, 1997). Este modelo permitiría resolver, en principio, el problema de la relevancia, lo cual se presenta en la sección 4. Por último, en la sección 5 resumo las conclusiones obtenidas en este trabajo.

## 2. Qué es el problema de la relevancia: origen e interpretaciones

Originalmente el “problema del marco” surge en el ámbito de la inteligencia artificial y consiste en la dificultad de especificar una fórmula lógica que describa las consecuencias de una acción sin tener que especificar otras fórmulas que describan, a su vez, que esa acción no cambia la mayoría de las propiedades de un objeto o del entorno. Imaginemos un robot que debe realizar la tarea de mover una caja de un punto X a un punto Y. El programa debe estar codificado de manera tal que no sólo indique la acción de mover algo de un punto a otro, las especificaciones sobre los brazos del robot para que puedan tomar la caja, etc., sino también que las propiedades de la caja y del ambiente, como el color, el tamaño, la forma, la temperatura de la habitación, etc., se mantienen idénticas cuando el robot ejecuta la tarea. Si esto no sucede, el robot entraría en *loop* “pensando” qué información debe tener en cuenta para llevar a cabo la acción. El problema del marco también es un problema representacional ya que se relaciona con cómo se representa el hecho de que el mundo es estable y no con cómo se especifican los procesos responsables de determinar que no hay cambios. Muchos autores (Pearl, 1988; Thielscher, 2001; Shanahan, 1997, 2009, entre otros) sostienen que han resuelto este problema mediante el uso de diferentes cálculos estadísticos y que, por lo tanto, ya no es un obstáculo dentro de la inteligencia artificial.

Sin embargo, al considerar las implicaciones que tiene el problema del marco dentro de las ciencias cognitivas, distintos autores como Fodor (1983, 1987, 2000), Dennett (1987), Samuels (2010), Chow (2013), etc., lo presentan de manera diferente. Llamo a estas interpretaciones “el problema de la relevancia” para distinguirlo de su formulación original ya que estos autores apelan a la noción de relevancia en sus formulaciones del problema. Siguiendo a Chow (2013), se puede definir el problema de la relevancia de la siguiente manera: dada toda la información disponible en un sistema, ¿cómo éste determina qué información es relevante para realizar una tarea particular? Es posible distinguir dos aspectos del problema de la relevancia. Por un lado, el problema computacional: en sistemas cognitivos humanos es implausible que toda la información sea considerada para una determinada tarea. El problema es sobre cómo un sistema delimita tratablemente la información relevante que se requiere para realizar una tarea e ignora aquello que no lo es. Esto se soluciona sólo si un sistema cognitivo emplea un método para diferenciar entre representaciones relevantes e irrelevantes y luego ignora estas últimas. Pero todo sistema que implementa esta estrategia, al mismo tiempo que ejecuta demasiadas computaciones, infiere conclusiones innecesarias y, además, debe determinar cuándo las representaciones consideradas son suficientes. En otras palabras, el sistema es computacionalmente intratable. Según Fodor (1983, 2000), los sistemas centrales, los encargados de la cognición superior, sufren este problema porque deben revisar toda la información de *background*, i.e., no están encapsulados informacionalmente.

Por otro lado, puede distinguirse un aspecto epistemológico: “¿cómo sabemos sólo lo que es relevante?” (Chow, 2013, p. 314). Es decir, ¿por qué tenemos un éxito “razonable” en determinar lo que es relevante? Los sistemas centrales son isotrópicos y no es posible dividir *a priori* en categorías qué información es relevante: algo puede ser relevante para una situación mientras que puede que no lo sea para otra. En consecuencia,

(i) Cualquier información, en principio, puede ser relevante para cualquier tarea (por el principio de isotropía).

(ii) No es posible revisar toda la información disponible en un sistema para saber qué es lo relevante por constricciones temporales, i.e., no se puede saber qué cosas son sólo las relevantes.

(iii) Por lo tanto, el problema de la relevancia no puede resolverse.

Este problema epistemológico “no tiene tanto que ver con el costo computacional de delimitar lo que es considerado en una tarea sino con considerar las cosas que son adecuadas” (Chow, 2013, p. 314). Sin embargo, en la neurociencia computacional, una de las teorías con mayor aceptación en la actualidad, la teoría del espacio de trabajo global (Baars, 1988, 1997; Dehaene, 2015; Dehaene y Naccache, 2001) intenta dar una respuesta más acabada al problema de la relevancia, aunque sus resultados sólo se implementan dentro de la inteligencia artificial. En la sección 4, se verá cómo esta teoría también puede utilizarse para resolver el aspecto computacional y el epistemológico del problema de la relevancia.

### **3. La teoría del espacio de trabajo global**

La teoría del espacio de trabajo global le atribuye a la conciencia la función principal de ser un fenómeno integrador que distribuye información a lo largo del cerebro y que recluta recursos neuronales para la resolución de problemas. Por conciencia entiendo un fenómeno compuesto de tres estados diferentes: un estado de vigilancia, i.e., un estado de alerta; la atención selectiva, i.e., la capacidad de la mente de focalizar sobre información específica, y esencialmente el acceso consciente, i.e., este estado ocurre cuando la información a la que se le presta atención accede a una posición prominente y se vuelve comunicable a otros. A continuación, delinearé los principales elementos de la teoría del espacio de trabajo global de acuerdo a cómo la presenta Baars, uno de sus principales exponentes.

Para explicar el rol de la conciencia, Baars (1988, 1997) utiliza la metáfora del teatro: el “escenario” del “teatro” es una memoria de trabajo (MT). Esta memoria opera serialmente y posee una capacidad limitada debido a que sólo se puede ser consciente de una experiencia a la vez. Esta limitación no constituye un obstáculo porque se ve compensada por el gran acceso que crea la MT a una enorme cantidad de información. Neurobiológicamente, se encuentra realizada en los lóbulos frontales y temporo-parietales. Los “actores” son procesadores inconscientes (poblaciones neuronales especializadas) que se encuentran en la corteza cerebral o en las áreas de proyección motora y sensorial y compiten por el acceso a la conciencia o al espacio de trabajo global porque aquel que gana el acceso al espacio de trabajo puede transmitir el mensaje a otros procesadores inconscientes que reclutan nuevos componentes y que interpretan una situación o resuelven un problema. Baars entiende a la competencia como

La disminución de los niveles de activación de los mensajes globales y a la cooperación, como aquella que aumenta los niveles de activación. La disminución y el aumento de los niveles de activación son análogos a la excitación y a la inhibición de las neuronas (Baars, 1988, p. 360).<sup>1</sup>

---

<sup>1</sup> Las redes neuronales con el mayor nivel de activación inhiben a las redes cercanas y se convierten en el foco dominante. Este proceso dura milisegundos, debido a que otra red puede convertirse en la más activa y por lo tanto, volverse el foco dominante. La corriente unificada de conciencia está compuesta de muchas regiones de activación diferentes que son independientes y que están en constante cambio. Estas activaciones pueden entrar en conflicto o apoyarse entre ellas y por lo tanto, cambiar el foco dominante, el cual incorpora los contenidos de la atención (Kinsbourne, 1978, 1988, 2006).

Cuando el “foco de atención” o “reflector” que se correlaciona con la atención selectiva, cae sobre un actor, el contenido consciente emerge. Otro elemento del teatro son los “operadores contextuales” que se ubican detrás de escena, también son inconscientes y moldean los eventos en el escenario, esto es, tienen un poderoso efecto sobre la conciencia (el mapa parietal moldea las células visuales que subyacen a la percepción del color). Incluido en estos operadores está el “director” que realiza las funciones ejecutivas (control sobre la MT). Por último, se encuentra la “audiencia” que corresponde a los sistemas de memoria (léxico, creencias, conocimiento del mundo y de uno mismo), a la interpretación del contenido consciente (reconocimiento de objetos, rostros, habla), automatismos (detalles del lenguaje, lectura, pensamiento) y sistemas motivacionales (expresiones faciales, manejo de los conflictos entre objetivos). La “audiencia” son coaliciones de procesadores que *ya* están comprometidos con una manera de procesar información y que *ya* tuvieron acceso al espacio de trabajo global por lo que son reclutados para realizar tareas rutinarias. Mediante este modelo, se especifica el modo en que funciona la conciencia para la teoría del espacio de trabajo global. En la próxima sección se verá cómo esto permite resolver el problema de la relevancia.

#### 4. El problema con el problema de la relevancia

En esta sección analizo si la teoría del espacio de trabajo global permite resolver el problema de la relevancia en relación con los dos aspectos mencionados en la sección 2. En primer lugar, considero que antes de determinar si es posible o no dar por acabado el problema de la relevancia, es necesario notar que no se sabe lo que es la relevancia. Ningún autor hasta el momento elucidó este concepto de manera adecuada.<sup>2</sup> Dentro de la neurociencia, los distintos autores presuponen que se sabe de qué están hablando pero esto solo genera más confusión a la hora de proponer una solución. No obstante, Baars realiza un intento de aclarar la noción de relevancia al afirmar que las coaliciones más relevante son las más “importantes, urgentes, novedosas, inesperadas, ruidosas, radiantes, etc.” (Baars y Franklin, 2007). Sin embargo, no se logra entender de qué está hablando pues ¿cómo una población de neuronas puede ser la más ruidosa, o la más radiante? O incluso ¿cómo una población neuronal relevante es la más inesperada?

Cabe destacar que hay en juego dos sentidos de lo que se entiende por cognición: un sentido normativo, i.e., cómo debería ser la cognición y un sentido descriptivo, i.e., cómo de hecho es la cognición. El problema de la relevancia apela a un sentido normativo, se debería revisar toda la información disponible para que nuestras inferencias sean confiables mientras que *de hecho* no revisamos toda la información del sistema para llevar a cabo una acción sino sólo la información que se requiere para resolver de manera eficaz un problema, i.e., el sentido descriptivo de la cognición (las propuestas que intentan resolver este problema de la cognición humana recurren a las heurísticas, por ejemplo, “use la tres primeras reglas”, etc.). Creo que el sentido descriptivo de la cognición es el que se debe tener en cuenta para resolver el problema de la relevancia.

De acuerdo a la teoría del espacio de trabajo global, la competencia de procesadores o de las poblaciones neuronales, es lo que permite determinar qué es lo relevante. Ahora bien, Baars afirma que la teoría del espacio de trabajo global le permite resolver el problema de la relevancia en inteligencia artificial. Pero ¿qué sucede con el problema computacional y epistemológico de la relevancia? Con respecto al problema computacional, Fodor considera que lo que evita la intratabilidad computacional es el encapsulamiento informacional por lo que los sistemas centrales que son isotrópicos y no encapsulados sufren del problema de la relevancia. Pero esto es un error (Carruthers, 2006a, 2006b; Samuels, 2006, 2010; Chow, 2013) porque lo que

---

<sup>2</sup> Si bien Sperber y Wilson (1986) plantean una teoría de la relevancia, esta se aplica a la comunicación y no a cómo la mente determina qué información debe utilizar para ejecutar una acción o resolver un problema.

se requiere es que los procesos sean frugales, i.e., rápidos y económicos, lo cual se logra recurriendo a poca cantidad de información y a algoritmos poco complejos que se implementan al procesar la información. Una respuesta adecuada a este problema de la relevancia debe especificar los procesos computacionales involucrados en la tarea cognitiva ¿cómo se logra esto en la teoría del espacio de trabajo global?

El sistema determina lo que es relevante de manera inconsciente, esto es, determinar lo relevante no es un proceso que se pueda reportar. De esta manera, la solución al problema computacional se encuentra en el procesamiento distribuido, inconsciente y en paralelo de las poblaciones neuronales. Frente a distintos estímulos, el hecho de que la atención se enfoque solo en uno de ellos, causa una ganancia que amplifica las ondas cerebrales. Debido a que el filtro que impone la atención sucede en paralelo, los recursos y la cantidad de información son mínimos (los esperables en la cognición humana). La competencia de los distintos *inputs* que también ocurre en paralelo y de modo inconsciente, ayuda a evitar el problema de la relevancia computacional. El sistema no debe buscar entre muchas representaciones sino sólo entre aquellas que responden a determinados *inputs*. Las distintas partes del sistema nervioso operan al mismo tiempo y con un grado de independencia unas de otras. Bajo estas circunstancias es fácil pensar que el cerebro es un sistema distribuido que funciona en paralelo. Esta combinación de procesos paralelos e inconscientes es lo que le permite que el proceso sea tratable, i.e., sea frugal y por lo tanto, asegura que el problema computacional de la relevancia no surja en los sistemas centrales.

Considero que este problema computacional se relaciona con el problema del marco original. Si bien este último hace referencia a la dificultad de representar la información relevante mediante fórmulas lógicas, para que el problema computacional se solucione, la especificación del mecanismo computacional implica que haya una representación de fórmulas lógicas de las funciones de activación. Los algoritmos que se encargan de la relevancia están escritos en una lógica no monótona y utilizan un cálculo de eventos (Shanahan, 1997). En otras palabras puesto que la solución en inteligencia artificial es de grano más fino que la solución computacional, si el aspecto de la inteligencia artificial soluciona el problema también lo hace el aspecto computacional.

En lo que respecta al problema epistemológico, según Baars

Existe una competencia entre los sistemas perceptuales por el acceso al espacio de trabajo global. Sólo un *input* puede ganar, y es el que logra más apoyo de los potenciales sistemas receptivos, especialmente aquellos que están “interesados” en el sistema ganador. “Ganador” quiere decir que un sistema obtiene acceso al tálamo y quizás a la formación reticular, permitiendo una transmisión general de algún aspecto central del sistema ganador – sus propiedades espaciotemporales, su significancia, su relevancia a las metas actuales, etc. (Baars, 1988, p. 126-127).

En otras palabras, las redes neuronales inconscientemente evalúan constantemente el entorno y le asignan valores a muchos factores que guían la acción. Para sobrevivir se le debe atribuir un valor positivo o negativo a un estímulo y esto debe ocurrir de modo rápido.<sup>3</sup> Los rasgos más valorados inhiben a los rasgos menos influyentes, i.e., los procesadores inconscientes se asocian y votan quien será el ganador teniendo en cuenta diversos elementos, como el valor que se le asigna a distintos factores, la jerarquía de objetivos, las emociones, el hecho de que el problema a resolver o la acción a ejecutar sea nueva o no (cuanto más se repita una acción, menos competencia hay entre los procesos). Por lo tanto, los factores que ayudan a considerar al ganador de una competencia que se produce paralela e inconscientemente son múltiples. La atención también tiene un rol central ya que se tiene acceso consciente rápido a

---

<sup>3</sup>Se cree que ocurre dentro de los ganglios basales (Dehaene, 2015).

todas las alternativas a las que se les puede prestar atención. La memoria a corto plazo tiene un directorio o menú de opciones disponibles ya preparadas. Una vez que las opciones dominan el espacio de trabajo global se presenta este directorio de contenido consciente y se elige el más relevante al objetivo actual mediante un proceso de decisión que puede incluir estos “votos” de los especialistas. La competencia se da en varios momentos: cuando dos *inputs* compiten para acceder al espacio de trabajo global y una vez que sólo uno ingresó, los procesadores compiten para ver cuál de ellos es reclutado. Un *input* puede ganar la competencia porque el procesador especializado para él ya está listo. Los procesadores que reciben el mensaje retroalimentan sus intereses al espacio de trabajo global lo que fortalece o debilita el mensaje. Pero nuevamente, no parece que esto aclare mucho el panorama. Es, entonces, la neurobiología la que saca de este apuro a la teoría de Baars. Como se mencionó, la noción de competencia es esencial. Una de las ideas centrales en este sentido es la de inhibición. El modelo del espacio de trabajo global plantea que un solo objeto del pensamiento se selecciona y se lo convierte en el foco de nuestra conciencia. Y como el espacio del trabajo global depende de una densa red de regiones cerebrales interconectadas cuando algo accede a este espacio de trabajo coincide con la activación repentina de la conciencia. Las neuronas activas comunican su mensaje por toda la corteza al enviar picos de activación a lo largo de sus axones extensos. Pero en la mayoría de los casos estas señales aterrizan en neuronas inhibitorias que actúan como un silenciador que acalla a un grupo de neuronas. Un *input* es codificado por una pequeña área de células activadas y sincronizadas, junto con una enorme cantidad de neuronas inhibidas. Si no hubiera inhibición las células que se excitan cuando forman las coaliciones necesarias para que surja la conciencia seguirían excitándose, lo cual crearía inestabilidad dentro de estas coaliciones. La mezcla de activación e inhibición es lo que crea oscilaciones que permiten coordinar el mensaje transmitido. De esta manera, y mediante ciertas leyes, las neuronas compiten por la activación y las ganadoras son las que se llevan todo, son las que permiten que actuemos en el mundo con un nivel de éxito razonable.

## 5. Conclusiones

La activación en paralelo y de manera distribuida permite que la información fluya rápidamente por el cerebro. Como los procesadores especializados trabajan mejor juntos que separados (generando mayor estabilidad en la conexión), para cumplir con su función deben ser capaces de dominar el espacio de trabajo el tiempo suficiente como para lograr transmitir el mensaje de reclutamiento a los otros procesadores. Cuanto más poderosa sea esta activación, mayor será la probabilidad de que las redes neuronales ganen el acceso al espacio de trabajo y permanezcan el tiempo necesario para seguir reclutando recursos. Mediante el procesamiento en paralelo y la competencia neuronal, la teoría del espacio de trabajo global parece dar una solución al problema de la relevancia tanto es su aspecto computacional como en su aspecto epistemológico.

## Referencias

- BAARS, B. *A cognitive theory of consciousness*. Cambridge: Cambridge University Press, 1988.
- BAARS, B. *In the theater of consciousness: the workspace of the mind*. New York: Oxford University Press, 1997.
- BAARS, B.; FRANKLIN, S. An architectural model of conscious and unconscious brain functions: global workspace theory and IDA. *Neural Networks*, v. 20, n. 9, p. 955-961, 2007.
- CARRUTHERS, P. The case for massively modular models of Mind. In: STANTON, R. (Ed.) *Contemporary debates in cognitive science*. Massachusetts: Blackwell, 2006. p. 3-21. (2006a)

- CARRUTHERS, P. Simple heuristics meet massive modularity. In: CARRUTHERS, P.; LAURENCE, S.; STICH, S. (Eds.) *The innate mind: culture and cognition*. Oxford: Oxford University Press, 2006. p. 181-198. (2006b)
- CHOW, S. What's the problem with the frame problem? *Review of Philosophy and Psychology*, v. 4, n. 2, p. 309-331, 2013.
- DEHAENE, S. *La conciencia en el cerebro*. Buenos Aires: Siglo Veintiuno, 2015.
- DEHAENE, S.; NACCACHE, L. Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition*, v. 79, n. 1, p. 01-37, 2001.
- DENNETT, D. Cognitive wheels: the frame problem of AI. In: PYLYSHYN, Z. (Ed.) *The robot's dilemma*. Boston: Ablex, 1987. p. 41-74.
- FODOR, J. A. *The modularity of mind*. Massachusetts: MIT Press, 1983.
- FODOR, J. A. Modules, frames, fridgeons, sleeping dogs and the music of the spheres. In: PYLYSHYN, Z. (Ed.) *The robot's dilemma*. Boston: Ablex, 1987. p. 139-149.
- FODOR, J. A. *The mind doesn't work that way*. Massachusetts: MIT Press, 2000.
- KINSBOURNE, M. Integrated cortical field model of consciousness. In: MARCEL, A.; BISIACH, E. (Eds.) *The concept of consciousness in contemporary science*. London: Oxford University Press, 1988. p. 239-256.
- KINSBOURNE, M. From unilateral neglect to the brain basis of consciousness. *Cortex*, v. 42, n. 6, p. 869-874, 2006.
- KINSBOURNE, M.; HICKS, R. Mapping cerebral functional space: competition and collaboration in human performance. In: KINSBOURNE, M. (Ed.) *The asymmetrical function of the brain*. New York: Cambridge University Press, 1978. p. 267-273.
- MCCARTHY, J.; HAYES, P. Some philosophical problems from the standpoint of artificial intelligence. In MELTZER, B.; MICHIE, D. (Eds.) *Machine intelligence*. Edinburgh: Edinburgh University Press, 1969. p. 463-502.
- PEARL, J. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. 2.ed. California: Morgan Kaufmann, 1988.
- SAMUELS, R. Is the human mind massively modular? In: STAINTON, R. (Ed.) *Contemporary debates in cognitive science*. Massachusetts: Blackwell, 2006. p. 37-56.
- SAMUELS, R. Classical computationalism and the many problems of cognitive relevance. *Studies in History and Philosophy of Science*, v. 41, n. 3, p. 280-293, 2010.
- SHANAHAN, M. *Solving the frame problem: a mathematical investigation of the common sense law of inertia*. Massachusetts: MIT Press, 1997.
- SHANAHAN, M. The frame problem. In: ZALTA, E. (Ed.) *The Stanford Encyclopedia of Philosophy*, 2009. Disponible en: <<http://plato.stanford.edu/archives/win2009/entries/frame-problem>>.
- SPERBER, D.; WILSON, D. *Relevance: communication and cognition*. 2.ed. Cambridge: Harvard University Press; Blackwell, 1986.
- THIELSCHER, M. *The fluent calculus*. Dresden: Dresden University of Technology, 2001.