

Las neuronas espejo. Un caso de estudio de la relación neurociencia-filosofía

Mirror Neurons. A Case Study of the Neuroscience-Philosophy Relationship

Diana I. Pérez

Instituto de Investigaciones Filosóficas- SADAF-CONICET,

Universidad de Buenos Aires, Argentina

dperez@filo.uba.ar

Resumen

El descubrimiento del sistema de neuronas espejo, que acaeció hace 25 años, fue considerado por algunos autores como prueba definitiva de la superioridad de una teoría filosófica (la Teoría de la Simulación) por sobre otra (la Teoría de la Teoría). Sin embargo, esta pretensión de haber encontrado una respuesta definitiva al problema filosófico de la comprensión de las otras mentes a partir de datos neurocientíficos dista mucho de ser aceptable. En este trabajo mostraré que existe una multiplicidad de interpretaciones posibles relativas al rol de las neuronas espejo, y que una de las más plausibles parece congruente con la perspectiva de segunda persona (Pérez y Gomila 2021). La conclusión que busco extraer de este caso de estudio es que no hay caminos directos desde la neurociencia a la solución de (al menos algunos) problemas filosóficos.

Palabras clave: segunda persona, otras mentes, cognición social, acción.

Abstract

The discovery of the mirror neuron system, which occurred 25 years ago, was considered by some authors as a definitive proof of the superiority of one philosophical theory (the Simulation Theory) over another (the Theory of Theory). However, the claim to have found a definitive answer to the philosophical problem of understanding other minds from neuroscientific data is far from acceptable. In this work I will show that there is a multiplicity of possible interpretations regarding the role of mirror neurons, and that one of the most plausible seems consistent with the second-person perspective



Received: 05/05/2022. Final version: 07/10/2022

eISSN 0719-4242 – © 2022 Instituto de Filosofía, Universidad de Valparaíso

This article is distributed under the terms of the

Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 Internacional License



CC BY-NC-ND

(Pérez and Gomila 2021). The conclusion I seek to draw from this case study is that there are no direct paths from neuroscience to the solution of (at least some) philosophical problems.

Keywords: second person, other minds, social cognition, action.

1. Introducción

El área de la cognición social ha tenido un gran desarrollo en los últimos 50 años. Desde que Premack y Woodruff (1978) plantearan la pregunta acerca de si los chimpancés tenían una teoría de la mente, es decir una capacidad para comprender las mentes de los demás, las investigaciones interdisciplinarias que involucra áreas como la etología, psicología del desarrollo, psicología cognitiva, neurociencia cognitiva y filosofía de la mente se multiplicaron. A lo largo de los primeros 20 años, la polémica se centró en la oposición entre dos aproximaciones teóricas: la teoría de la teoría (TT) y la teoría de la simulación (TS). Con el inicio de un nuevo siglo, una multiplicidad de propuestas, muchas de ellas enmarcadas en una perspectiva postcognitivista de la mente, sumaron alternativas teóricas a este panorama: el interaccionismo (Gallagher 2001), el enactivismo (De Jaegher y Di Paolo 2007), el narrativismo Hutto (2008) y la perspectiva de segunda persona (Gomila 2001, Scotto 2002, Pérez y Gomila 2021), entre otras. Un hito central de la historia reciente del desarrollo de este tema ha sido el descubrimiento del sistema de neuronas espejo, que acaeció hace 25 años. Inicialmente fue considerado por algunos autores como prueba definitiva de la superioridad de una teoría filosófica (la TS) por sobre otra (la TT). Sin embargo, esta pretensión de haber encontrado una respuesta definitiva al problema filosófico de la comprensión de las otras mentes a partir de datos neurocientíficos no fue aceptada por muchos de los participantes de este debate. En este trabajo me propongo evaluar esta cuestión. El plan del artículo es el siguiente: en primer lugar, presentaré brevemente el marco filosófico de discusión vigente al momento del descubrimiento del sistema de neuronas espejo: el debate TT-TS (sección 1). A continuación, reseñaré brevemente en qué consiste el descubrimiento neurocientífico del sistema de neuronas espejo, así como los argumentos que se esgrimieron inicialmente para favorecer la TS (sección 2). Seguidamente repasaré las críticas que esta propuesta ha tenido focalizando en las dificultades y confusiones conceptuales generadas por la búsqueda de explicaciones de fenómenos relativos a las mentes humanas a partir de datos neurocientíficos (sección 3). Finalmente, mostraré que a la multiplicidad de interpretaciones posibles del rol de las neuronas espejo se le puede agregar una lectura alternativa que se encuentra en consonancia con mi propia propuesta de respuesta al problema de la comprensión de las mentes humanas, la perspectiva de segunda persona (Pérez y Gomila 2021). La conclusión que busco extraer de este caso de estudio es que no hay caminos directos desde la neurociencia a la solución de (al menos algunos) problemas filosóficos.

2. La disputa entre teóricos de la teoría y teóricos de la simulación en el último cuarto del siglo XX

La cuestión acerca de la comprensión de la mente propia y ajena tiene una larga historia filosófica. Sin duda, aun en nuestros días seguimos hechizados por la visión moderna de la mente expuesta de forma paradigmática en las *Meditaciones Metafísicas* de Descartes. De acuerdo con esta visión, hay dos caminos de acceso a las mentes. Por un lado, la introspección como vía directa e indubitable de acceso a la propia mente, a la subjetividad, a los pensamientos en primera persona. Por otro lado, el conocimiento de las *otras* mentes, que es indirecto porque está mediado por inferencias producidas a partir de la observación de la conducta pública, y constituye un conocimiento objetivo, de tercera persona. En los últimos veinte años del siglo pasado, las discusiones acerca de los mecanismos inferenciales de acceso a las mentes -o, en términos de esa época, acerca de la naturaleza de la *psicología folk* - reflejó esta dualidad entre primera y tercera persona propia del pensamiento cartesiano, al incluir dos propuestas en pugna: la de los teóricos de la teoría y la de los teóricos de la simulación (Davies y Stone 1995a, 1995b, Carruthers y Smith 1996).

La idea central de la teoría de la teoría es que los seres humanos poseemos teorías (explícitas o implícitas) acerca de diversos dominios de la realidad (psicología, física, biología, aritmética) que guían nuestra acción y nuestra comprensión del mundo que nos rodea. Así, la psicología folk es una teoría más como otras (Lewis 1972, Churchland 1981) porque satisfacen las siguientes dos características. Primero, la psicología folk es una *teoría* de la mente en el sentido de que los conceptos psicológicos en términos de los cuales se formulan sus “leyes” son considerados los términos inobservables/ teóricos de la teoría, en tanto que las conductas son la base empírica correspondiente al ámbito de lo observable. Así, en tanto teoría, la psicología folk postula una serie de entidades inobservables que sirven primariamente para explicar y predecir la conducta de los demás. Los conceptos psicológicos propios de la teoría son conceptos de tercera persona que se aplican en base a la conducta públicamente observable, y sólo derivativamente a la primera persona; en este marco teórico el caso de la autoatribución psicológica es cualitativamente similar (tal vez con alguna diferencia de grado) al de la tercera persona (Gopnik 1993). En segundo lugar, el significado de los conceptos psicológicos se agota en el conjunto de conexiones teóricas en las que están inmersos. Así como los términos teóricos de una disciplina científica adquieren su significado por el rol que juegan en la teoría, lo mismo ocurre con los conceptos psicológicos.¹

La segunda propuesta de acceso a las otras mentes es la teoría de la simulación. Esta teoría está basada en la idea de que la psicología folk no es un conjunto de generalizaciones que aplicamos al caso particular cuya mente queremos entender, sino que el mecanismo básico supone “ponerse en el lugar del otro”, una inferencia de un caso singular -yo-, a otro caso

¹ Esta concepción de los conceptos psicológicos está en línea con el funcionalismo dominante de la época, y puede rastrearse claramente a Lewis (1972), entre otros.

particular -el otro. Hay versiones muy diferentes de la teoría de la simulación,² me centraré en la versión de A. Goldman (1989, 1993, 2006, 2009) por razones que quedarán claras en breve. La teoría de la simulación sostiene que el proceso de atribución mental depende de una inferencia que hacemos a partir del caso de primera persona al de tercera, es decir que para comprender a los demás ponemos en funcionamiento nuestros propios mecanismos deliberativos (nuestro propio mecanismo de razonamiento práctico) y lo hacemos correr offline para obtener como output la acción que yo realizaría en el lugar del otro, prediciendo así la conducta del otro agente. En este caso los conceptos usados en la descripción del caso de primera persona son básicos, y se aplican por proyección a la tercera persona. Goldman (1993) es muy claro acerca del proceso de formación de conceptos psicológicos: el punto de partida es la introspección de un estado cualitativo que es etiquetado en este acto introspectivo, etiqueta que posteriormente se extiende a los otros (las terceras personas) bajo el supuesto de que tienen una vida cualitativa y psicológica similar a la nuestra. Así las adscripciones de estados mentales recurrentes de primera persona son previas a cualquier otra atribución psicológica.

La teoría de la teoría y la teoría de la simulación han sido los dos contendientes en disputa durante más de veinte años, por lo que los argumentos -tanto empíricos como filosóficos- a favor y en contra de cada una de estas propuestas se popularizaron en la literatura de finales del siglo pasado. Es importante destacar que desde sus inicios con el texto de los etólogos Premack y Woodruff (1978) esta polémica fue claramente interdisciplinaria: la etología, la psicología del desarrollo, la psicología cognitiva, y la filosofía estuvieron involucradas en argumentos y estudios experimentales varios para dar apoyo a uno o el otro de los contendientes. No es extraño entonces, que en 1998 Goldman proclamara que había sido hallado el hecho empírico que permitiría inclinar definitivamente la balanza a favor de la teoría de la simulación: las neuronas espejo.

3. El descubrimiento del sistema de neuronas espejo.

En 1996 se publicaron dos artículos que marcarían una bisagra en las discusiones precedentes acerca de los mecanismos de acceso a las mentes. En efecto en Rizollatti et. al. (1996) y Gallese et. al (1996) se da a conocer el hallazgo de las “neuronas espejo” (en adelante NE). El grupo de investigación del que ambos formaban parte estudiaba el funcionamiento del sistema motor en monos, en particular las neuronas involucradas en la coordinación visuo-motora para la realización de acciones básicas como agarrar, sostener, manipular y colocar un objeto con la mano. La novedad que introducen las NE es que la activación de estas neuronas no se produce solamente cuando el mono realiza alguna de estas acciones motoras, sino que exactamente las mismas neuronas se activan cuando observa a otros actuar. Es decir, no sólo se coordinan sistema motor y visual para realizar efectivamente una acción en el mundo, sino que también se activan estas neuronas en la observación/compresión de las acciones ajenas.

² Véase Pérez (2013, Capítulo 4) para un análisis detallado de las diversas versiones de la teoría de la simulación.

Estas neuronas se activan en configuraciones muy específicas, pudiéndose distinguir neuronas-espejo-agarrar de neuronas-espejo-sostener, neuronas-espejo-manipular etc. (Rizzolatti y Sinigaglia 2006, p. 86). Y se activan las mismas neuronas en las mismas configuraciones específicas: por ejemplo, la NE que se activa cuando un mono agarra una banana, es la misma que se activa cuando ese mono ve a un ser humano (o a otro mono) agarrar una banana, pero no se activa cuando el humano mueve la mano exactamente en el mismo ángulo, pero sin la presencia de la banana. Tampoco se activa con la sola presencia de la banana. Se trata entonces de neuronas que se activan exactamente de la misma manera cuando el agente realiza una acción determinada con sus manos sobre un objeto, y cuando el agente ve a otro agente realizar la misma acción con un objeto similar.

Algunas de las investigaciones subsiguientes se volcaron a la búsqueda de NE en humanos. Dejando de lado los detalles de las dificultades y limitaciones éticas que plantearon estos estudios, se ha podido constatar la existencia de un “sistema de neuronas espejo” (SNE) en humanos, con características similares a las halladas en monos. En humanos, este sistema parece ser más sofisticado: permite no sólo codificar el objetivo del acto motor sino además aspectos temporales de los movimientos que componen ese acto (Rizzolatti y Sinigaglia 2006, p. 119), codifica tanto actos motores transitivos como intransitivos, tanto acciones como movimientos básicos que se combinan en acciones más complejas, y se activa con el solo movimiento corporal, en ausencia del objeto, es decir, cuando la acción es imitada (Rizzolatti y Sinigaglia 2006, p. 124). También se activa este sistema con la percepción de expresiones emocionales, permitiendo el reconocimiento de emociones básicas, no sólo de acciones intencionales (Rizzolatti y Sinigaglia 2006, p. 179).

Mucho se ha discutido acerca de la significación de este hallazgo. Jeannerod (véase Jacob y Jeannerod 2003 cap. 7) ha desatacado su rol en la preparación de acciones y en la imitación, siendo a su juicio la base del aprendizaje de acciones habilidosas con herramientas e instrumentos. Sin embargo, para Rizzolatti y Sinigaglia es posible asignarles un rol mucho más básico: “el reconocimiento y la comprensión del significado de los “eventos motores”, es decir de los actos de los demás. ... en cuanto [el mono] ve la mano del experimentador ... dirigirse hacia la comida, *percibe* inmediatamente su significado, *comprendiendo* dichos “acontecimientos motores” en términos de *un determinado tipo de acto.*” (Rizzolatti y Sinigaglia 2006, p. 101, sus itálicas). Sin duda es esta interpretación de la significación de este hallazgo empírico el que le permitió a Goldman sostener que la teoría de la simulación había triunfado sobre la teoría de la teoría.

En efecto, Goldman y Gallese (1998) sostienen que el hallazgo de las neuronas espejo está en concordancia con la teoría de la simulación pero que su hallazgo no podría haber sido previsto desde la teoría de la teoría. Argumentan que las rutinas de simulación que nos permiten comprender a los demás habrían evolucionado a partir de este sistema que correlaciona la ejecución de una acción con la observación de ese mismo tipo de acción, el SNE. El descubrimiento de las NE en monos, por otra parte, permite entender por qué ciertas especies exhiben una comprensión de la acción intencional de sus congéneres como los humanos, y no

son simplemente lectores “conductistas” de los movimientos corporales ajenos. Finalmente, sostienen, la TT no podría explicar por qué en el caso humano no solo el mismo conjunto de neuronas se activa, sino los mismos grupos musculares resultan estimulados al realizar la acción que al percibirla (aunque, obviamente, el output motor está inhibido en el caso de la mera percepción) (Goldman y Gallese 1998, p. 498).

La adopción de esta evidencia empírica a favor de la teoría de la simulación llevó al desarrollo de la idea de simulación corporizada (Gallese 2005), y finalmente a la idea de que en realidad hay dos tipos de simulación involucrada en la comprensión de las mentes, una simulación de bajo nivel (la simulación corporizada) y una simulación de alto nivel (Goldman 2006). Goldman distingue en los trabajos más recientes dos formas de simulación mental. Dice:

hay dos formas de simulación mental. La primera es el espejamiento (*mirroring*) que es automática, casi enteramente inconciente, involucra típicamente estados mentales comparativamente “primitivos” y no descansa en conocimiento o información específica para la tarea (*task-specific knowledge or information*). Este tipo de simulación, denominada “simulación de bajo nivel” debe distinguirse de la “simulación de alto nivel” que corresponde a la idea original de una simulación guiada por la imaginación, o guiada por la simulación (*pretense-driven*). La simulación de alto nivel requiere más esfuerzo, es a veces conciente, involucra estados mentales más complejos y es guiada por conocimiento o información específica para la tarea. (Goldman 2009, p. 246)

Así, y a pesar de que inicialmente se sostuvo que el SNE sería una evidencia a favor de la TS, parece que este descubrimiento neurocientífico sólo es asimilable a uno de los dos tipos de simulación que involucra esta nueva versión de la TS. Pero no resulta evidente por qué ambos tipos de procesos deben ser incluidos bajo el rótulo común de “simulación” ni cómo se conectan estos dos tipos de simulación.

Veremos a continuación que hay quienes han sospechado que esta asociación entre el SNE y la teoría de la simulación descansa, en realidad, en una confusión conceptual, y que es posible dar un lugar al SNE en el marco de otras visiones alternativas acerca de nuestra capacidad para comprender mentes.

4. Problemas para la adopción del SNE como evidencia en favor de la TS

En los últimos 15 años ha habido varios trabajos en los cuales se discute esta asociación entre la teoría de la simulación y el SNE. Desde una visión enactivista de lo mental, Gallagher (2007) distingue tres tipos de teorías de la simulación y propone argumentos específicos en contra de cada una de ellas. Allí se distingue: (1.) la simulación como algo explícito, un ejercicio conciente de imaginación e inferencia deliberativa (la propuesta inicial de la TS en Goldman (1989), la simulación de “alto nivel” en los trabajos más recientes) de (2.) la simulación como un mecanismo subpersonal, un mecanismo funcional explicitado en términos de

procesos neuronales (Hurley) o los procesos neuronales mismos (Gallese).³ Como vimos en el apartado anterior, la propuesta simulacionista de Goldman busca incorporar ambos bajo el rotulo común de teoría de la simulación (distinguiendo la de bajo nivel de la de alto nivel), aunque no explicita cuál es exactamente esta relación.

Quisiera detenerme en los argumentos que propone Gallagher para cuestionar la identificación del SNE con algo que pudiera llamarse simulación corporizada o simulación de bajo nivel. En primer lugar, Gallagher sostiene que no hay razones para considerar que haya un proceso de simulación cuando se disparan las NE al ver a un agente que realiza una acción en frente nuestro. En su lugar, sostiene Gallagher, podemos interpretar que estamos ante un proceso de percepción-acción, que sería la lectura más apropiada del fenómeno neurofisiológico si adoptamos una teoría enactiva de la percepción (véase Gallagher 2007, p. 71). Es por ello que podríamos pensar que la activación del SNE es el proceso neurofisiológico que se correlaciona con nuestra experiencia de nivel personal de *percepción directa* de las acciones intencionales y emociones ajenas, no con nuestros juicios acerca de las mentes ajenas. Es decir, la activación del SNE es lo que está detrás de nuestro acto psicológico de ver la acción realizada o la emoción padecida por el otro, ya que cuando observamos a nuestros congéneres somos capaces de percibir directamente (es decir sin la mediación de ningún proceso inferencial) sus acciones y emociones. No percibimos su movimiento corporal y simulamos el estado mental del otro en nuestro sistema visuomotor, simplemente vemos su estado mental.

Pero Gallagher avanza un paso más allá y no sólo considera que los fenómenos de nivel personal y subpersonal no parecen ser procesos que merezcan un único y mismo nombre -simulación- sino que además argumenta en contra de usar esta expresión para hacer referencia a los mecanismos subpersonales. El argumento de Gallagher, dicho brevemente, es que toda simulación, en algún sentido, supone irrealidad, pero en este caso no la hay (Gallagher 2007, p. 72), y tampoco realiza ninguna acción específica el agente cuyo SNE se activa: la activación de las neuronas espejo en ningún sentido es algo que nosotros hacemos, en tanto que una simulación presupone un agente que simula, es un tipo de acción psicológica (Gallagher 2007, p. 73).

Una línea alternativa de crítica a la idea de que la evidencia proporcionada por el estudio del SNE resulta decisiva a favor de la TS la presentan Spaulding (2012) y Borg (2007, 2012), quienes muestran cómo esta evidencia empírica no sólo es consistente con, sino que además sirve de apoyo para, la teoría de la teoría, es decir la idea de que nuestra comprensión de las mentes ajenas depende de una teoría de la mente, y que involucra una actividad conceptual. Siguiendo a Borg, podemos llamar a esta línea de reflexión la “objeción intencional”. La idea, dicho brevemente, es que las NE nos dan información sobre los movimientos corporales, pero no nos dan información sobre las intenciones que llevaron a esos movimientos.

³ Gallagher considera una tercera opción: la simulación como un proceso explícito pero que no involucra inferencia deliberativa (la simulación radical de R. Gordon). No tomaré en cuenta esta opción dado que no se relaciona con el tema de este trabajo.

El supuesto detrás de esta línea argumental es que puede haber diferentes interpretaciones intencionales de los mismos movimientos corporales, por lo tanto, un movimiento corporal dado sólo contará como encarnando un cierto tipo de acción intencional si es interpretado como tal. Y esta interpretación supone la posesión de una teoría de la mente, es decir de un conjunto de conceptos que se ponen en juego al percibir las acciones ajenas. Como sostiene Spaulding, las neuronas espejo pueden formar parte de los muchos mecanismos de bajo nivel que son causas contribuyentes de la cognición social (Spaulding 2012, p. 532) pero no pueden ser identificadas con la comprensión de la acción, que involucra otros elementos de alto nivel (es decir, una teoría de la mente).

Si atendemos ahora al caso de las NE involucradas en el reconocimiento emocional, podemos encontrar argumentos similares en Goldie 2000. Si bien no menciona explícitamente las NE, y a pesar de haber declarado su simpatía por la versión de J. Heal de la TS, en el capítulo 7 busca responder a la pregunta “¿Cómo es que pensamos acerca de las emociones ajenas?”. Y para responder a esta pregunta distingue la comprensión y explicación de las emociones ajenas, que a su juicio es un fenómeno “personal y normativo” (Goldie 2000, p. 177), de otros fenómenos con los que erróneamente -a su juicio- se ha identificado, tales como el contagio emocional, la empatía, la imaginación-en-sus-zapatos, la simpatía. Goldie sostiene que la explicación del fenómeno personal de comprensión de las emociones ajenas no puede explicarse en términos de mecanismos subpersonales. De acuerdo con Goldie, dado que la comprensión de las emociones supone la ubicación del fenómeno emocional en una narrativa apropiada, la mera observación /reconocimiento de la emoción en el otro (que, agregó yo, sería lo único que debemos a las NE), no puede ser suficiente para que comprendamos cabalmente los procesos emocionales por los que la otra persona está pasando. Una vez más, parece que el SNE es el mecanismo subpersonal que permite ver las emociones en el rostro ajeno, pero de ninguna manera por sí solo nos provee de una comprensión de las emociones ajenas.

En síntesis, vemos que – a pesar de la pretensión de Goldman y Gallese- no ha quedado resuelto el problema filosófico que consiste en explicitar cuál es el rol que un mecanismo subpersonal, neuronal, juega en la comprensión de las mentes. La lectura inicial que ellos propusieron fue que la existencia de este mecanismo era la evidencia necesaria para apoyar la TS, pero como vimos no queda claro ni cómo se relaciona este fenómeno subpersonal con el proceso psicológico de nivel personal de simulación que según ellos subyace a la comprensión de las mentes, ni por qué denominar “simulación” al fenómeno subpersonal. Tampoco se ha establecido aun que tal mecanismo neurofisiológico logre dar cuenta por sí solo del proceso psicológico de comprensión de las mentes en ausencia de otros procesos psicológicos ajenos al SNE, tales como la teoría de la mente, la capacidad narrativa, o la simulación de alto nivel. En última instancia, parece que la activación del SNE puede “leerse” como evidencia de diversos fenómenos psicológicos de nivel personal: en principio hasta aquí puede leerse -como hacen Goldman, Gallese y Rizzolatti y Sinigaglia- como lo que habilita a la *comprensión* de las acciones y emociones ajenas o, de una manera más deflacionaria, como lo que permite la *percepción* de intenciones y acciones ajenas, como propone Gallagher. También puede leerse

como un mecanismo de bajo nivel que, conjuntamente con otros mecanismos psicológicos tanto subpersonales como conceptuales o de alto nivel, habilitan la comprensión de las mentes humanas

5. La perspectiva de segunda persona y las neuronas espejo

En este apartado intentaré esbozar una manera de incorporar el SNE a la cognición social, que toma como punto de partida la propuesta teórica que defendimos con A. Gomila sobre la cuestión de la comprensión de las mentes: la perspectiva de segunda persona, en adelante P2P (Pérez y Gomila 2021). Paso entonces, a presentar brevemente esta propuesta.

La P2P adopta explícitamente los postulados del postcognitivism.⁴ La P2P propone un marco para la comprensión de las mentes en el que ocupan un lugar central las *atribuciones* de segunda persona. Podríamos resumir esta propuesta en las siguientes dos tesis: (1) las interacciones de segunda persona están mediadas por atribuciones de segunda persona, i.e. atribuciones automáticas, prácticas, implícitas, transparentes, recíprocamente contingentes y dinámicas y (2) estas atribuciones son las más básicas filogenética, ontogenética y conceptualmente. Por interacciones de segunda persona entendemos a aquellas situaciones en las que dos (o más) seres humanos⁵ se encuentran en una interacción directa, cara a cara (o tal vez deberíamos decir cuerpo a cuerpo). En estas ocasiones percibimos directamente algunos de los estados mentales del otro individuo (por ejemplo, algunos de sus estados emocionales expresados en su rostro y tono de voz, las intenciones en el movimiento que se despliega, etc.).

⁴ El postcognitivism busca liberarse del hechizo cartesiano, y suele resumirse en 5 E, es decir la idea de que la mente humana es Extendida, Encarnada, Enactiva, situada (Embedded) y Emocional. Se opone a la visión cognitivista clásica heredada de la modernidad que entendía a la mente como aquello que está adentro del cráneo, que podría existir sin cuerpo o en cualquier otro cuerpo – natural o artificial-, que funciona como lo hace con independencia del entorno y de su historia, y que es racional (tanto para generar nuevas creencias a partir de las anteriores, como para tomar decisiones acerca de qué acción realizar a la luz de sus deseos y creencias).

⁵ Describo aquí el caso paradigmático. Podría extenderse la idea de interacción de segunda persona a formas de interacción con otros animales, o con seres artificiales (robots), o con seres de ficción; no voy a entrar en estos ejemplos en esta ocasión.

Estas percepciones directas del otro como triste o con miedo o como agarrando una banana involucran una atribución implícita del estado psicológico en cuestión.⁶ Para la P2P las emociones son la puerta de entrada a las otras mentes.⁷

Es en el contexto de interacciones de segunda persona que se dan las condiciones para la aparición de las primeras atribuciones psicológicas, así como para la adquisición de los primeros conceptos psicológicos. En estas interacciones está presente una enorme cantidad de información multimodal, tanto acerca del otro como de uno mismo. Pensemos en una típica interacción entre un bebé y su figura de crianza, en las sonrisas -tanto del bebé con su carga propioceptiva e interoceptiva, como la sonrisa percibida en el rostro ajeno, el sonido emitido en ocasión de la sonrisa-, los sonidos emitidos por ambos individuos, la mirada dirigida a los ojos del otro individuo, y -hacia el final del primer año de vida- a objetos del entorno compartido, alternando la mirada entre el otro individuo y los objetos del entorno. Notemos que estas interacciones son asimétricas, en el sentido de que la figura de crianza pertenece a una cultura determinada, ha adquirido ciertas específicas pautas de comportamiento, habla una lengua determinada, etc. A lo largo del primer año se va forjando una historia de interacciones que incrementa paulatinamente la comprensión mutua, en un principio entre el bebé y sus figuras de crianza. Con el paso de los meses, y la pluralidad de interacciones intersubjetivas de las que participa, el infante logra interactuar haciéndose comprender y comprendiendo a individuos con los que no había interactuado previamente. Esto es así porque va adquiriendo una comprensión cada vez más profunda de las pautas de interacción propias de la cultura en la que crece, y al mismo tiempo, va comprendiendo progresivamente los estados mentales de los otros, es decir va adquiriendo una competencia conceptual para realizar atribuciones psicológicas al individuo concreto con el que está interactuando en las situaciones concretas en las que se encuentra primero de manera implícita y más tarde explícita. Como vemos, la información disponible al sujeto en ocasión de estas interacciones es tanto información acerca de sí mismo, como acerca del otro; también hay información acerca de las modificaciones que las acciones del otro producen en uno mismo y viceversa. Es decir, es información tanto de primera, como de segunda, como de tercera persona, información acerca de la propia subjetividad, de la reactividad ante el otro, y de la conducta pública del otro.

⁶ Hablamos en general de estados mentales implícitos en aquellos casos en los que el sujeto revela en su conducta tal estado psicológico, aun cuando no articule verbalmente ni para sí mismo ni para los demás la posesión de dicho estado. Así, reaccionar consolando a alguien que está llorando revela (muestra) que atribuimos un estado de tristeza a esa persona, dirigir la mirada al objeto que el otro mira revela que entendemos que ese es el objeto hacia el cual el otro está dirigiendo la atención, etc. La idea central de la P2P es que todas las interacciones de segunda persona suponen este tipo de atribución implícita (que, por supuesto, en el caso de los seres humanos adultos hablantes competentes de una lengua pueden explicitarse en ocasiones).

⁷ La centralidad de la afectividad en la comprensión de las mentes es uno de los rasgos distintivos de la P2P frente a otras teorías postcognitivist. En estas interacciones hay involucramiento afectivo, y una actitud participativa (no distanciada). (véanse capítulos 5 y 6 de Pérez y Gomila 2021).

Estas ricas situaciones de interacción permiten la adquisición de los conceptos psicológicos, por ejemplo “estar alegre” que involucran una pluralidad de habilidades cognitivas, centralmente: (1) ser capaz de aplicar ese concepto a sí mismo y a los demás⁸; (2) ser capaz de entender las expresiones de alegría -las sonrisas, tonos de voz, etc.- como tales; (3) poseer ciertos guiones (*scripts*) básicos relacionados con la alegría, típicamente responder con una sonrisa ante la sonrisa del otro, acercarse, tocarse, emitir/escuchar cierto tipo de sonidos articulados, con cierta entonación, etc.⁹ Este tercer aspecto recoge, entre otras cosas, lo que Goldie concibe como el carácter normativo y narrativo del concepto en cuestión, que en el caso del bebé es muy modesto, y se va complejizando a lo largo de la vida.

Esta visión de los conceptos mentales permite comprender cómo es que los conceptos de emociones básicas, como por ejemplo la alegría, y de intenciones en la acción, como por ejemplo extender un brazo para alcanzar un objeto en el entorno compartido, son los primeros conceptos que se adquieren y por lo tanto las primeras atribuciones psicológicas implícitas que se pueden realizar¹⁰, en primer lugar en los contextos de interacción y luego se pueden realizar más allá de la interacción efectiva. Esta propuesta supone que algunos estados mentales (los recién mencionados) son directamente observables.

Teniendo en mente el marco que nos ofrece la P2P, creo que es posible dar cuenta más cabalmente del rol del SNE en la cognición social humana. Varios autores han propuesto una función conceptual para las NE, dado que satisfacen varias de las características que los conceptos poseen: integran información multimodal, identifican patrones relativamente abstractos y tienen un rol inferencial (Meini y Paternoster 2012). Para algunos autores las activaciones de esta parte del sistema sensorio-motor se consideran conceptuales porque se asume que todo el sistema conceptual humano está fundado en representaciones sensoriomotoras (Gallese, V. & Lakoff, G. 2005). En otros casos, se asume que dados su nivel de abstracción y capacidad de integración de información multimodal estas representaciones son en sentido estricto, amodales (Jacob 2008). Meini y Paternoster (2012) por su parte, proponen algo intermedio: que los conceptos motores son parcialmente constituidos por las activaciones de

⁸ De acuerdo con la P2P las atribuciones de primera y de tercera persona son igualmente primitivas, no hay un concepto psicológico de primera persona que se proyecta a la tercera, ni viceversa.

⁹ Esta forma de entender los conceptos como conjuntos de habilidades que se van adquiriendo a lo largo del desarrollo, permite tener una visión gradualista de la posesión de los conceptos psicológicos. Así, la pregunta acerca de cuáles y cuántos *scripts* debe poseer un individuo, cuáles y cuantas conductas debe ver como expresivas, etc. resulta difícil de responder. No hay una línea clara de corte entre una respuesta conductual que está guiada por la posesión de un concepto y la que no. Para más detalles sobre esta concepción de los conceptos psicológicos y su impacto en el debate TT/TSP/P2P, véase Pérez 2013

¹⁰ Para la P2P no es la creencia el estado psicológico más básico, en términos de los cuales deben comprenderse los demás. Por el contrario, la adquisición del concepto de creencia, que supone la conceptualización de los objetos del entorno y la posibilidad de captar contenidos proposicionales se dará posteriormente en el desarrollo y dependerán de la adquisición de competencias lingüísticas por parte del infante.

las áreas pre-motoras. Esta propuesta es más modesta porque no supone que todo nuestro esquema conceptual esté fundado en representaciones motoras, ni que estas representaciones motoras por sí solas constituyan nuestras habilidades conceptuales.

En concordancia con esta última propuesta, la P2P otorga un rol especial a un tipo muy específico de conceptos psicológicos, los conceptos de emociones básicas, con expresiones públicas manifiestas, y las intenciones en acción, que son justamente el tipo de situación en las que se activa el SNE. Las características que he descrito de comportamiento del SNE nos permite entender además por qué estos conceptos involucran habilidades de auto y de heteroatribución, tal como postula la P2P. En efecto, como hemos visto, el mismo circuito neuronal se activa cuando realizamos una acción determinada y cuando percibimos a otros realizando exactamente esa misma acción (como en el ejemplo de agarrar la banana, o de expresar alegría en nuestro rostro). También se activa cuando escuchamos los sonidos típicos que emitimos al realizar dicha acción (por ejemplo, si masticamos, si hacemos un esfuerzo y exhalamos de una manera característica, si adoptamos un tono alegre en nuestra voz). Así, parece que se trata más bien de un mecanismo que subyace a la identificación de acciones tipo, no de una acción específica de un agente determinado. El mismo Gallese sugiere que una lectura posible del rol de la activación del sistema de neuronas espejo en este sentido. Ya en Gallese (2005) sostenía que, dado que las NE se activan de la misma manera en contextos diversos, y en el caso de la primera y de la tercera persona -es decir no distinguen a yo de otro, sino que simplemente tienen un “parámetro” de agente-, habría que pensarlas como “predicados” (Gallese 2005 p. 33) es decir como codificando acciones tipo, en lugar de ser los correlatos neurales de los pensamientos particulares acerca de la propia acción o de la acción ajena en una situación dada.¹¹

Por otra parte, la activación del SNE se verifica en dos casos: en el caso de las acciones transitivas es decir las acciones intencionales dirigidas a objetos del entorno (por ejemplo, agarrar una banana), y en las acciones intransitivas, expresivas, especialmente en las expresiones emocionales (un caso bien estudiado es el del asco Rizzolatti y Sinigaglia 2006, cap. 7; Gallese 2005, p. 38). El movimiento de las manos, de la boca y las expresiones faciales parecen ser los estímulos que activan el SNE tanto ante la percepción de este tipo de movimientos como en la ejecución de los mismos, es decir tanto en el caso de la tercera persona como de la primera. En Pérez y Gomila (2021) argumentamos en extenso por qué son justamente los conceptos de acciones y emociones básicas las que primero se adquieren en las interacciones de segunda persona, en las que, como dije más arriba, la experiencia de primera persona y de tercera están simultáneamente presentes. Que sean este tipo de estados psicológicos aquellos en los que se verifica la simultánea activación de la experiencia propio e interoceptiva con la experiencia perceptual de la conducta asociada, resulta coherente con nuestra propuesta de que este tipo de conceptos psicológicos constituyen el primer tipo de atribución psicológica (que denominamos atribuciones de segunda persona) y que tienen un carácter similar a lo

¹¹ Jeannerod 2006, pp. 111-112 sostiene algo similar.

que Gallese denomina “simulación corporizada”¹² son implícitas, automáticas y transparentes (“no conceptuales”, dice Gallese, pero prefiero otra terminología¹³). Pero en Pérez y Gomila (2021) agregamos que se trata de atribuciones que son prácticas, dinámicas y recíprocamente contingentes¹⁴: de su existencia depende la posibilidad misma de dar continuidad y fluidez a la interacción de segunda persona. Nuestro foco no está puesto en la resonancia entre las activaciones neuronales de los dos sujetos, ni en la imitación ni en la empatía, ni en el conocimiento compartido. El paradigma de situación humana que adopta la P2P es, por el contrario, el de una suerte de “dialogo” (no lingüístico) que se entabla entre los agentes en este tipo de interacciones. Y para que este “dialogo” tenga continuidad lo que se necesitan son mecanismos que permitan predecir la acción (o reacción) del otro.¹⁵

Así, resulta acertado sostener que el SNE habilita la comprensión de las mentes ajenas y de la propia, como señalaron Goldman y Gallese, porque esta comprensión se asienta en la posesión de los conceptos psicológicos relevantes, que involucran el caso de la primera y de la tercera persona (de yo y del otro) tal como lo hacen las neuronas espejo. Esta comprensión es implícita en el caso de la percepción de las acciones ajenas, por lo que mi propuesta resulta también compatible con la idea de Gallagher. Dicho en otros términos: la percepción del movimiento corporal ajeno es percibido directamente *como* exhibiendo ciertos estados psicológicos (intención de agarrar algo, asco, alegría etc.) en la medida en que se percibe el movimiento en cuestión *como* una acción o emoción, y este *percibir como* supone la posesión del concepto psicológico relevante. Sin embargo, estos conceptos psicológicos no se agotan en la mera categorización del movimiento corporal percibido como un cierto tipo de acción, sino que además involucran habilidades que permiten incorporar dicho estado mental en un en-

¹² Prefiero no usar la idea de simulación tanto por las razones que esgrime Gallagher como por el hecho de que simulación suele confundirse con “imaginación” (el mismo Gallese confunde estos términos, véase por ejemplo Gallese y Lakoff 2005) y no quiero comprometerme con la idea de que la imaginación esté en juego en el tipo de atribución de segunda persona al que me estoy refiriendo.

¹³ Gallese (2005 p. 31) dice que esta “simulación corporizada” es un mecanismo no-conceptual, aunque resulta inconsistente dado que en Gallese y Lakoff 2005 parece decir que se trata de una “teoría corporizada de los conceptos” al describir el concepto de la acción de agarrar (pp. 3-4). Creo que es más adecuado decir que se trata de una atribución psicológica transparente en el sentido de que la posesión del concepto “agarrar” no supone conceptualizar el objeto al que la acción está dirigido de ninguna manera particular. El concepto “agarrar” recoge así abstractamente la acción que un agente (yo, tú, ella, etc.) realiza sobre un objeto (no importa como se lo describa).

¹⁴ Cook et. al 2014 argumentan que el SNE se origina en procesos de aprendizaje asociativo sensoriomotor y sostienen que esta concepción permite dar cuenta del carácter sensitivo al contexto y a las contingencias que exhiben las NE, en consonancia con la propuesta de la P2P. El rol de las interacciones en este aprendizaje es evidente a lo largo de este texto, aunque no esté tematizado con nuestra terminología.

¹⁵ Gallese 2005 usa reiteradamente la idea de que estos mecanismos neurales tienen una función predictiva, sin embargo, nuevamente identifica la predicción con la simulación, es decir con la posibilidad de “imaginar” lo que ocurrirá. Creo que se puede seguir asumiendo la función predictiva de estos mecanismos sin comprometerse con la idea de que se trata de un proceso de simulación/ imaginación.

tramado más amplio de acciones, emociones y pensamientos, que se encuentran explicitados en los *scripts* que vamos adquiriendo y que dan sustancia nuestra comprensión de la mente humana. Más aun, esta comprensión que habilitan las NE no es sólo una comprensión de las mentes ajenas, sino también de la propia acción/emoción, es la comprensión de la realización de un tipo de acción por parte del agente que la realiza, incorporando su movimiento en este entramado significativo.

6. Coda: ¿se puede resolver un problema filosófico con evidencia neurocientífica?

Si mi lectura de la historia reciente acerca del descubrimiento del sistema de neuronas espejo y su impacto en las discusiones filosóficas y psicológicas en torno a la cognición social es correcta, parece claro que el descubrimiento neurocientífico por sí mismo no alcanzó para resolver el problema filosófico planteado. No deberíamos extrañarnos en este caso: como mencioné oportunamente el área misma de la cognición social que tiene al problema de la comprensión de las (otras) mentes en el centro, ha sido desde sus orígenes un emprendimiento multidisciplinario. En este sentido sería extraño que una sola de las disciplinas que forman este conjunto multidisciplinario pudiera dar con la respuesta buscada por si sola. Así, no es la falta de mayor evidencia neurocientífica lo que se necesita en este caso para dirimir la cuestión, sino que la naturaleza misma del emprendimiento intelectual supone una multiplicidad de miradas en donde cada mirada aporta distinto tipo de información valiosa e ineludible para la comprensión del fenómeno. Si queremos entender cómo es que nos comprendemos los unos a los otros como seres que tienen estados mentales y actuamos en consecuencia, tenemos que entender cómo es que llegamos a tener maestría en el uso de los conceptos psicológicos y, como dije, ellos involucran una serie de habilidades normativas, narrativas, reconocitivas, lingüísticas, expresivas, interactivas, que no son objeto de estudio de las neurociencias. La razón es bastante simple: nuestra concepción de nosotros mismos involucra un dominio conceptual irreducible a conceptos no-psicológicos. No estoy siendo muy original en este punto: todas las posiciones materialista no-reduccionistas desde los años '70s y '80s han tratado de fundamentar este punto, incluyendo a Davidson 1970, Fodor 1987, Dennett 1987, Kim 1993 entre muchos otros.¹⁶ En diversos puntos del espectro desde el más cognitivista hasta el más postcognitivista, desde el más normativista hasta el más descriptivista, parece haber un amplio acuerdo respecto de que el lenguaje mentalista no puede ponerse en correspondencia uno a uno con ningún estado o proceso cerebral, ni -en rigor- con ningún estado o proceso de nivel subpersonal. Ni la acción intencional, ni las emociones, ni las creencias, ni las sensaciones, ni la voluntad tienen un correlato subpersonal estable, identificable por métodos que no supongan algún grado de interpretación psicológica del fenómeno. Es por eso que, al menos en este terreno, los hallazgos de la neurociencia siempre serán interpretables de diver-

¹⁶ No quiero incluir al segundo Wittgenstein entre los materialistas no reduccionistas porque seguramente esta -como cualquier otra- etiqueta metafísica le hubiera molestado, pero sin duda sus reflexiones detrás de la dicotomía razón-*causa* está detrás del talante no reduccionista de lo mental.

sas maneras, aunque, por supuesto, estas interpretaciones no están ajenas a su confrontación con ulteriores hallazgos empíricos. Sin duda, podremos seguir conociéndonos cada vez más a nosotros mismos en esta espiral virtuosa de interacciones entre hallazgos neurocientíficos e hipótesis interpretativas que apelan a nuestros conceptos psicológicos.

Queda pendiente la cuestión acerca de la plausible generalización de la conclusión que estoy extrayendo en este caso, a otros problemas filosóficos. ¿Habrá problemas filosóficos de otro tipo que puedan quedar resueltos por la acumulación de datos acerca de cómo funciona nuestro cerebro? Mi sospecha es que, en la medida en que nuestra imagen antropológica esté en juego, estaremos en la misma situación que estamos cuando queremos resolver el problema de cómo nos comprendemos los unos a los otros. Y parece que todos los problemas filosóficos que involucran a los seres humanos suponen una imagen acerca de nuestra forma cotidiana de conocernos, y de interactuar. Así, estudiar una parte de lo que somos (nuestro cerebro) sin combinar este conocimiento con otros provenientes de otros aspectos de nuestra humanidad (nuestra psicología, nuestro desarrollo ontogenético y filogenético, nuestras variadas formas culturales de estar con otros, nuestras variadas formas de interacción con los nichos materiales en los que nos encontramos y con los que nosotros mismos diseñamos, nuestra historia) sospecho que nunca será suficiente para responder las preguntas que nos planteamos acerca de nosotros mismos, nuestra existencia y el significado de nuestras vidas.¹⁷

Referencias bibliográficas

- Borg, Emma (2007). If mirror neurons are the answer, what was the question?. *Journal of Consciousness Studies*, 14(8), 5-19.
- Borg, E. (2012). More questions for mirror neurons. *Consciousness and Cognition*, 22 (3), 1122-1131. <http://doi.org/10.1016/j.concog.2012.11.013>
- Carruthers, P., Smith, P. (1996). *Theories of theories of Mind*. Cambridge: Cambridge University Press.
- Churchland, P. (1981). Eliminative Materialism and the Propositional Attitudes. *The Journal of Philosophy*, 78(2), 67-90.
- Cook, R., Bird, G., Catmur, C., Press, C., Heyes, C. (2014). Mirror neurons: from origin to function. *Behavioral and Brain Sciences*, 37(2), 177-241.
- Davidson, D. (1970). Mental Events. En Davidson, D. (1980). *Essays on Actions and Events*, pp. 207-228. Oxford: Oxford University Press.
- Davies, M., Stone, T. (Eds.) (1995a). *Folk Psychology*. Oxford: Blackwell.
- Davies, M., Stone, T. (Eds.) (1995b). *Mental Simulation*. Oxford: Blackwell.
- Dennett, D. (1987). *The Intentional Stance*. Cambridge: MIT Press.

¹⁷ Agradezco los comentarios de Federico Burdman, Diego Lawler, Marina Trakas y Agostina Vorano a una versión previa de este trabajo que me permitieron mejorarlo.



- Fodor, J. (1987). *Psychosemantics*. Cambridge: MIT Press.
- Gallagher, S. (2007). Logical and phenomenological arguments against simulation theory. In Hutto, D., Ratcliffe, M. (Eds.), *Folk Psychology Re-assessed*, pp. 63-78. Dordrecht: Springer. <https://doi.org/10.1007/978-1-4020-5558-4>
- Gallese, V. (2005). Embodied simulation: From neurons to phenomenal experience. *Phenomenology and the Cognitive Sciences*, 4, 23-48. <https://doi.org/10.1007/s11097-005-4737-z>
- Gallese, V. & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, 22(3), 455-479. <https://doi.org/10.1080/02643290442000310>
- Gallese, V., & Goldman A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 12(2), 493-501. [https://doi.org/10.1016/s1364-6613\(98\)01262-5](https://doi.org/10.1016/s1364-6613(98)01262-5)
- Gallese, V., Fadiga, L., Fogassi, L., Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119(2), 593-609. <https://doi.org/10.1093/brain/119.2.593>
- Goldie, P. (2000). *The Emotions. A philosophical Exploration*. Oxford: Clarendon Press. <https://doi.org/10.2307/3182582>
- Goldman, A. (1993). The Psychology of Folk Psychology. *Behavioral and Brain Sciences*, 16(1), 15-28. <https://doi.org/10.1017/S0140525X00028648>
- Goldman, A. (2006) *Simulating Minds. The philosophy, psychology and neuroscience of mindreading*. Oxford: Oxford University Press.
- Goldman, A. (1989). Interpretation psychologized. *Mind and Language*, 4(3), 161-185.
- Goldman, A. (2009). Mirroring, simulating and mindreading. *Mind & Language*, 24(2), 235-252.
- Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences*, 16, 1-15.
- Jacob, P., Jeannerod, M. (2003) *Ways of Seeing: The Scope and Limits of Visual Cognition*. Oxford: Oxford University Press.
- Jacob, Pierre (2008). 'Mirror neurons' or 'concept neurons'?. *Mind and Language*, 23(2), 190-223.
- Jeannerod, M. (2006). *Motor Cognition. What actions tell to the Self*. Oxford: Oxford University Press.
- Kim, J. (1993). *Supervenience and Mind*. Cambridge: Cambridge University Press.
- Lewis, D. (1972). Psychophysical and Theoretical Identifications. *Australasian Journal of Philosophy*, 50, 249-58. reimpresso en Block 1980.
- Meini, C., Paternoster, A. (2012). Mirror neurons as a conceptual mechanism?. *Mind and Society*, 11(2), 183-201.
- Pérez, D. (2013). *Sentir, desear, creer. Una aproximación filosófica a los conceptos psicológicos*. Buenos Aires: Prometeo.



- Pérez, D., Gomila, A. (2021). *The second person: Social cognition in human interaction*. London: Routledge.
- Premack, D. G., Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and Brain Sciences*, 1(4), 515-526.
- Rizzolatti, G., Sinigaglia, C. (2006). *Las neuronas espejo. Los mecanismos de la empatía emocional*. Barcelona: Paidós.
- Rizzolatti, G., Fadiga, L., Gallese, V., Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3(2), 131-141.
- Spaulding, S. (2012). Mirror Neurons are not Evidence for the Simulation Theory. *Synthese*, 189(3), 515-534.

