

## Cognición moral y cognición psicológica: las intuiciones vienen primero

*Moral Cognition and Psychological Cognition: Intuitions Come First*

Carolina Scotto

Universidad Nacional de Córdoba, Argentina  
carolinascotto@gmail.com

### Resumen

La comprensión psicológica es una capacidad requerida para la competencia moral en el sentido en que entender las intenciones, creencias e intereses de otros es un *input* crítico para evaluar las responsabilidades involucradas en sus comportamientos y comprender, a su vez, cómo interactuar con ellos para alcanzar nuestros propósitos. Por su parte, la interacción con otros está en el corazón de ambas capacidades, dado que una y otra son componentes esenciales y estrechamente relacionados de la vida social humana. Mi objetivo en este trabajo, en relación con ambos supuestos, será poner de relieve una semejanza estructural entre la cognición psicológica y la cognición moral ordinarias, mostrando cómo ciertos tipos diferenciados de explicaciones en uno y otro dominio, permiten identificar un marco teórico consistente que, a su turno, puede dar cuenta de las relaciones entre ambas. Para ello, me referiré en primer lugar, sintéticamente, al *Modelo Intuicionista Social* (MIS) de J. Haidt (2001), en tanto ejemplifica un enfoque no clásico y dual de la cognición moral en el cual las intuiciones morales desempeñan un papel más básico que las razones. En el ámbito de la cognición psicológica, me referiré con más detenimiento al *enfoque interactivo* o de *segunda persona*, enriquecido por el enfoque dual de la cognición, en particular, por las propuestas sobre variedades implícitas o “mínimas” de atribución mental. En ese contexto, identificaré los llamados *comportamientos expresivos*: su naturaleza y su papel en la conducta y las interacciones humanas y su importancia como base evidencial de la comprensión mental. Ellos constituyen los objetos primarios de las *intuiciones psicológicas*. Espero mostrar que las señales expresivas son “legibles” por medio de las capacidades intuitivas que las detectan y comprenden su significación psicológica de maneras eficientes. Una vez expuestas las



Received: 05/11/2021. Final version: 25/03/2022

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

afinidades teóricas entre el MIS y el enfoque interactivo-dual, pondré de relieve que, así como para el primero, las intuiciones morales vienen primero y las razones después, también para el enfoque interactivo-dual de la comprensión psicológica, las intuiciones psicológicas vienen primero y las razones después.

**Palabras clave:** intuicionismo moral, enfoque dual de la cognición, enfoque interactivo-dual, cognición implícita, intuiciones psicológicas.

### Abstract

Psychological understanding is a required capacity for moral competence in the sense that understanding the intentions, beliefs, and interests of others is a critical input for evaluating the responsibilities involved in their behaviors and understanding, in turn, how to interact with them to achieve our purposes. For its part, interaction with others is at the heart of both capacities, since both are essential and closely related components of human social life. My aim in this paper, in relation to both assumptions, will be to highlight a structural similarity between ordinary psychological cognition and moral cognition, showing as certain differentiated types of explanations on one domain and the other allows to identify a theoretical framework consistent that, in turn, it can account for the relationships between the two. For this, I will refer first, synthetically, to the *Social Intuitionist Model* (MIS) of J. Haidt (2001), as it exemplifies a non-classical and dual approach to moral cognition in which moral intuitions play a more basic role than reasons. In the field of psychological cognition, I will refer with more detail to *the interactive or second-person approach*, enriched by the dual approach to cognition, in particular by the proposals on implicit or “minimal” varieties of mental attribution. In this context, I will identify the so-called *expressive behaviors*: their nature and role in human behavior and interactions, and their significance as the evidential basis for psychological understanding. They constitute the primary objects of psychological intuitions. I hope to show that expressive signals are “readable” through the intuitive abilities that detect them and understand their psychological significance in efficient ways. Having exposed the theoretical affinities between the MIS and the interactive-dual approach, I will emphasize that, just as for the first, moral intuitions come first, and reasons later, also for the interactive-dual approach of psychological understanding, psychological intuitions come first, and reasons later.

**Keywords:** moral intuitionism, dual-approach of cognition, interactive-dual approach, implicit cognition, psychological intuitions.

## 1. Cognición psicológica y cognición moral

[...] highly specialized subcultures such as that of philosophy,  
[...] provides years of training in unnatural modes of human thought.

Haidt y Bjorklund (2008, 193)

La literatura filosófica reciente sobre diversos tópicos concernientes a los fenómenos mentales y a los fenómenos morales abunda en visiones de conjunto, muchas veces críticas, sobre la *imagen clásica* de la mente y la moralidad humanas. Según aquella imagen, aún a riesgo de simplificar el cuadro, la racionalidad y la autonomía confieren su marca distintiva a las criaturas humanas por sobre otras criaturas. Así, las “facultades” y las operaciones cognitivamente superiores, bajo el control consciente del sujeto, tienen un papel preeminente en la explicación de la vida mental y moral de las personas. Entre ellas, la capacidad para poseer y combinar conceptos en estructuras articuladas más complejas, realizar inferencias, tener meta-representaciones y, en suma, teorizar sobre la base de evidencias fragmentarias y presupuestos racionales, tanto acerca de los estados mentales ajenos como propios (y sobre otras materias también). Según aquella misma imagen, los humanos se distinguen de otras criaturas por su capacidad para deliberar de modo racional con el objeto de realizar evaluaciones normativas y orientar responsablemente sus decisiones bajo la guía de principios morales, en el marco de la autonomía de la voluntad.

Las limitaciones y distorsiones inherentes a esa imagen altamente idealizada<sup>1</sup> y los significativos hallazgos proporcionados por las investigaciones recientes en ciencias cognitivas impulsaron sucesivas e incluso drásticas revisiones de ella<sup>2</sup>. Estas abrieron las compuertas al reconocimiento del papel de fenómenos mentales y morales con características muy dispares y estimularon la elaboración de nuevas maneras de comprender nuestras capacidades para “mentalizar” y orientar normativamente la conducta. Así, para muchos psicólogos y filósofos, el intento de encontrar los fundamentos de la moralidad en la razón o en principios que la razón pueda capturar, ligado al intento de “crear un realismo moral defendible”, finalmente

---

<sup>1</sup> En muchos aspectos se trata de una imagen idealizada: el carácter “desencarnado” de las capacidades críticas involucradas; el papel nulo o secundario de las experiencias sensoriales, las sensaciones, los afectos y las emociones; el carácter “puro” de las distintas capacidades, sean las superiores o las otras: ninguna de ellas está contaminada por las otras o, al menos, no debería estarlo; la visión jerárquica o de control que las capacidades racionales ejercen sobre las otras, son, entre otros, rasgos característicos de los modelos mentales y morales clásicos. Aproximadamente estas mismas notas salientes son atribuidas por diversos intérpretes al “cartesianismo”, al racionalismo y al intelectualismo, tanto en la filosofía de la mente como en la filosofía moral.

<sup>2</sup> Como es usual, la imagen opuesta tiene también ilustres antecedentes en la historia de la filosofía. En el caso de la moral, el “combate de fondo” se libró entre las ideas racionalistas de Kant (1785) y las concepciones sentimentalistas de Hume (1739). Por su parte, en ciertos aspectos, la psicología de Hume incluía algunos de los componentes críticos que permitirían socavar el modelo racionalista de la mente, como veremos en seguida.

fracasó (Haidt y Kesebir 2009, 216). Se impuso, entonces, la necesidad de dar una respuesta al papel de la diversidad de las experiencias psicológicas y morales, incluidas aquellas debidas a las pautas y las características culturales variables de las comunidades humanas. Nuestras preocupaciones en este trabajo guardan alguna relación con estas cuestiones filosóficas mayores, pero como se verá enseguida, se enfocan hacia un objetivo mucho más acotado y preliminar.

Es bastante evidente que comprender los fenómenos morales requiere o presupone comprender, en algún grado, la naturaleza y el funcionamiento de la mente. Pareciera también evidente que un sistema moral psicológicamente realista debería descansar en fundamentos cognitivos informados y razonables que nos permitan comprender cómo pensamos, evaluamos y decidimos qué hacer. No obstante, hasta hace relativamente poco tiempo, la psicología y la filosofía se ocupaban de manera independiente de los fenómenos mentales y morales: una de los aspectos empíricos, la otra de los aspectos normativos. Como un eco de la clásica dicotomía entre *hechos* y *valores* se consolidó en la práctica la idea de que los conocimientos y afirmaciones acerca de los hechos (naturales o culturales) son irrelevantes para comprender la naturaleza de nuestra relación con las normas y los valores. El supuesto era que esta división del trabajo permitía preservar un nivel de explicación racional de los fenómenos psicológicos y morales, de otro nivel en el que las razones y las normas no desempeñaban papel alguno. Así, al igual que la psicología empírica y la psicología filosófica (i.e., las teorías racionales de la interpretación), la psicología empírica y la psicología moral discurrieron por carriles independientes hasta hace relativamente poco tiempo atrás.

Diversos intentos por superar las restricciones que impusieron las perspectivas no naturalistas acerca de la mente y la moral emanadas de aquella visión clásica, revitalizaron el papel que la tradición humeana había concedido a los sentimientos y las emociones - “a la fábrica y la constitución particular de la especie humana”- para dar cuenta del funcionamiento de la mente, el razonamiento práctico y la evaluación moral. En esa perspectiva, también las intuiciones y los factores evolutivo-naturales encontraron un papel explicativo importante. En lo que refiere a la cognición moral, según algunos especialistas es actualmente posible avanzar hacia un consenso que reconcilie ambas dimensiones, la intuitivo-emocional y la racional, sin suprimir el papel de ninguna de ellas, es decir, reconociendo que “el juicio moral es el producto de la interacción y la competencia entre distintos sistemas psicológicos” (Cushman et al. 2010, 47). Esfuerzos teóricos en esa misma dirección abundan en la filosofía post-cognitivista de la mente. Finalmente, un rasgo común a los enfoques no clásicos ha sido el papel creciente concedido a los factores interactivos y sociales en la explicación de la génesis y el contenido de las atribuciones psicológicas y de los juicios morales, poniendo en cuestión el marco individualista bajo el cual había sido concebida la comprensión de las “otras mentes” y la evaluación moral. En la feliz expresión de Haidt y Kesebir, “los hechos morales son hechos sociales” (2009, 217). Lo propio cabe afirmar sobre los “hechos mentales”, especialmente

sobre aquellos que son objeto de las capacidades más básicas involucradas en las interacciones humanas ordinarias. En este amplio contexto, tanto la cognición psicológica como la cognición moral encuentran un campo más fértil para explorar su interrelación.

Explicitaré primero las asunciones en las que se apoya este trabajo y presentaré luego los objetivos que se propone. En primer lugar, asumiré que la comprensión psicológica es una capacidad requerida para la competencia moral en el sentido que entender las intenciones, creencias e intereses de otros es un *input* crítico para evaluar las responsabilidades involucradas en sus comportamientos y comprender, a su vez, cómo interactuar con ellos para alcanzar nuestros propósitos. Sea que se defienda la tesis que la atribución mental está al *servicio primario* de la cognición moral (Young y Waytz 2013<sup>3</sup>), o que “está en el corazón del juicio moral” (Guglielmo, Monroe y Malle 2009), o que sólo puede influir sobre ella de diversas maneras, o que, por el contrario, la cognición moral impregna la interpretación psicológica (Knobe 2004, 2006), lo cierto es que una y otra son componentes esenciales y estrechamente relacionados de la vida social humana<sup>4</sup>. Por lo tanto, asumiré también que *la interacción con otros* está en el corazón de ambas capacidades: tanto de la atribución psicológica como de la cognición moral<sup>5</sup>. Es verdad que no muchos especialistas han elaborado modelos que den cuenta de las diversas vías a través de las cuales cada una de estas capacidades depende de, o influye sobre la otra, quizás debido a que algunos de esos vínculos han parecido más o menos obvios.<sup>6</sup> Mi objetivo en este trabajo, en relación con ambos supuestos, será poner de relieve una semejanza estructural de índole más general entre la cognición psicológica<sup>7</sup> y la cognición moral: cómo ciertos tipos diferenciados de explicaciones, en uno y otro dominio, permiten identificar un *marco teórico consistente* que puede, a su vez, dar cuenta de las relaciones entre ambas.

<sup>3</sup> Según esta tesis, las capacidades para comprender estados mentales se tornan menos relevantes en los contextos donde no hay en juego ninguna cuestión moral, tanto respecto de terceros actuando como jueces como de uno mismo en tanto actor moral o en interacción con otros. En suma, el carácter *moralmente saliente* de un contexto hace más relevante la atribución mental.

<sup>4</sup> Mi propuesta acerca de la interrelación y las influencias recíprocas entre ambas, apuntada por una manera similar de entender dichas capacidades, no depende de aceptar una u otra tesis acerca de cuál es el tipo de relación predominante entre ellas o cuál es la capacidad más básica. En mi opinión, la investigación empírica contribuirá decisivamente a esclarecer la naturaleza de esos vínculos.

<sup>5</sup> Como se verá, los aspectos sociales son constitutivos de la cognición psicológica en un sentido diferente a como lo son en el caso de la cognición moral. Sin embargo, no me detendré en esta cuestión.

<sup>6</sup> No obstante, la literatura empírica sobre aspectos específicos de ese vínculo es abundante. En cuanto a la literatura de carácter teórico, basada en evidencias empíricas, además del mencionado capítulo de Young y Waytz (2013), distintas propuestas pueden encontrarse en Knobe (2004, 2006); Malle (2006); Pettit y Knobe (2009); Guglielmo, Monroe y Malle (2009); Waytz et al. (2010); Gray, Young y Waytz (2012).

<sup>7</sup> Las expresiones *comprensión* o *cognición psicológica*, así como la más usual *cognición social*, se emplean en este trabajo de manera intercambiable a *psicología popular*, *atribución psicológica* o *mental*, *teoría de la mente* y *lectura de mentes*. En un sentido estricto, sin embargo, la expresión *cognición psicológica* o *social* refiere a los procesos cognitivos involucrados en la interacción entre personas, mientras las demás refieren más bien al modo ordinario, accesible a las personas, de comprender la vida mental de otros. Más adelante justificaré el empleo preferencial

En relación con la cognición moral, me referiré muy sumariamente a los rasgos centrales del *Modelo Intuicionista Social* (en adelante: MIS) propuesto por Jonathan Haidt (2001), reelaborado y enriquecido en su obra posterior con abundante evidencia empírica en distintos ámbitos de aplicación (política, educación, religión, entre otros), en tanto ejemplifica un modelo no clásico de la cognición moral, e incluso, como se ha afirmado, es “el más importante desafío empírico al racionalismo acerca de la psicología del juicio moral” (Sauer 2011, 708). En particular, me interesa señalar cómo el MIS entiende la naturaleza y el papel de las *intuiciones morales* en el marco de la adopción de un enfoque dual de la cognición humana, claramente contrario a la imagen clásica de la mente y la moral antes descripta (Sección 1.) Respecto de la cognición psicológica, a la que dedicaré la mayor parte de este trabajo, me referiré, en primer lugar, a los rasgos distintivos del *enfoque interactivo* o de *segunda persona*, y señalaré, luego, de qué manera se ve enriquecido por el enfoque dual de la cognición, en particular, por las propuestas sobre variedades implícitas o “mínimas”, que a la vez fundamentan y complementan las capacidades más sofisticadas de atribución mental. En ese contexto, identificaré a los llamados *comportamientos expresivos*: su naturaleza y su papel en la conducta y en las interacciones humanas y, en consecuencia, su importancia crítica como base evidencial de la comprensión mental. Ellos constituyen los objetos primarios de las *intuiciones psicológicas*.<sup>8</sup> Espero mostrar que, en el marco interactivo en el que primariamente emergen y ocurren las atribuciones psicológicas, los comportamientos expresivos son señales “legibles” por medio de capacidades intuitivas que permiten comprender su significación psicológica de maneras eficientes (Sección 2.). A continuación, dedicaré una sección especial a repasar las principales caracterizaciones de las intuiciones que se han ofrecido en la literatura psicológica y filosófica, con el objeto de hacer lugar a las variedades que son relevantes para este trabajo (Sección 3.). En la sección siguiente, después de sintetizar las afinidades teóricas entre el MIS y el enfoque interactivo de la comprensión psicológica, pondré de relieve que, así como para el MIS, las intuiciones morales vienen primero y las razones después, también para el enfoque interactivo de la comprensión psicológica, las intuiciones psicológicas vienen primero, y las razones después (Sección 4.). En la literatura filosófica y psicológica sobre la atribución mental ordinaria no es frecuente la referencia a las intuiciones, y la tesis que *las intuiciones psicológicas vienen primero* no ha sido formulada explícitamente por nadie. Esta es la razón por la que dedico más espacio a bosquejar su formulación y a motivar su interés, y a integrar dicha tesis al enfoque interactivo-dual sobre la comprensión psicológica. En esta ocasión, me desentenderé de los múltiples flancos para la crítica que pueden presentar cada uno de estos enfoques y de los argumentos que podrían esgrimirse para enfrentarlos, por lo mismo que, dados mis propósitos, no se requiere un compromiso estrecho con cada una de

---

de *lectura de mentes* para referir a esta capacidad específica.

<sup>8</sup> Entenderé por *intuiciones psicológicas* las intuiciones referidas a las características psicológicas manifiestas en los comportamientos o señales expresivas, en el mismo sentido que las *intuiciones morales* refieren a las características morales de ciertos comportamientos o acciones, y por extensión, en ambos casos, a los procesos, estados o actitudes psicológicas o morales que los explican.

sus tesis particulares. Mi objetivo, además de general, es más bien positivo: busca iluminar de qué modo la cognición psicológica y la cognición moral ordinarias están signadas por el mismo tipo dual de procesos y estados cognitivos, y más específicamente aún, que en ambos casos las intuiciones, psicológicas y morales, preceden a otras formas más racionales de cognición.

## 2. El modelo intuicionista social (MIS)

Según el *Modelo Intuicionista Social* (MIS), el juicio moral sigue un proceso de dos vías (*two-track*), uno intuitivo y el otro racional:

La intuición moral se refiere a procesos rápidos, automáticos y (usualmente) con carga afectiva, en los cuales un sentimiento evaluativo de bueno-malo o de gusto-disgusto (acerca de las acciones o el carácter de una persona) aparece en la conciencia sin ningún acceso consciente de haber atravesado pasos de buscar, sopesar evidencia, o inferir una conclusión. El razonamiento moral, en contraste, es un proceso controlado y “más frío” (menos afectivo); es una actividad mental consciente que consiste en transformar información acerca de las personas y sus acciones para alcanzar un juicio o una decisión moral” (Haidt 2007, 998; cfr. también 2001, 818). (Yo subrayo)

La idea es simple: las intuiciones, acompañadas usualmente de una valencia afectiva o “sentimiento evaluativo”, se manifiestan regularmente de manera directa y sin esfuerzo en la conciencia y producen un juicio moral intuitivo<sup>9</sup>. Luego, aunque no siempre, las decisiones que emanan de esos juicios morales intuitivos son acompañadas por esfuerzos reflexivos (intencionales, esforzados y controlables) de justificación racional, una forma de razonamiento *post-hoc* ante los otros. En este último sentido, el razonamiento moral es sobre todo un proceso social<sup>10</sup>. Ahora bien, sólo en ocasiones excepcionales (donde las intuiciones son débiles y la capacidad de incorporar información nueva es alta) el razonamiento moral se vuelve sobre los propios los juicios intuitivos y sobre las intuiciones que los originan. El diagnóstico de Haidt (2001) es que buena parte de la filosofía moral occidental, al adoptar un modelo deductivo de razonamiento moral para derivar los juicios morales, cometió el error de confundir los procesos de justificación *post factum* sobre los juicios intuitivos (un mero “*afterthought*”), con la base misma de la cognición moral. Así, el MIS se apoya en dos tesis descriptivas: (1) los juicios morales están influenciados por procesos intuitivos, y (2) el razonamiento moral es un

<sup>9</sup> El MIS planteó dos desafíos al modelo clásico: las intuiciones preceden a los procesos deliberativos o controlados, y los mecanismos afectivos a los cognitivos. En este trabajo voy a enfocarme en la primera dualidad, aunque algunos autores sostienen que las respuestas intuitivo-afectivas serían el fruto de los mismos procesos psicológicos (Cushman et al. 2010).

<sup>10</sup> “Las personas son muy malas en cuestionar su propios supuestos y juicios iniciales, pero en el discurso moral otras personas hacen esto por nosotros... El juicio moral tendría que ser estudiado como un proceso social, y en un contexto social el razonamiento moral importa” (Haidt y Bjorklund 2008, 193).

fenómeno social de justificación (o racionalización) de las intuiciones y juicios morales, por lo cual no tendría que ser estudiado como el acto privado de un razonador solitario (Haidt 2001; Haidt y Joseph 2004; Haidt y Bjorklund 2008). Estos dos aspectos, además de otros más específicos, convierten al MIS en una aproximación convergente con el modelo de la cognición psicológica al que me referiré luego.

Muchos debates en torno a la propuesta intuicionista han puesto el foco no tanto en el carácter dual de la cognición moral como en el peso conferido en él a las intuiciones. En cuanto al *enfoque dual de la cognición*, ofreció al MIS un marco teórico acerca de la cognición en general, así como evidencias que lo apuntalan, distinguiendo a los procesos de tipo reflexivo o racional (el “sistema 2”) de las intuiciones (del “sistema 1”). Según Haidt, los primeros tienen un papel menos relevante en el dominio de la cognición moral que en otros dominios. Así, los procesos racionales son muchas veces causalmente inertes respecto a los juicios morales intuitivos y la acción moral. Haidt (2012) ilustra esa división con la metáfora del “jinete” racional que está mayormente al servicio del “elefante” intuitivo. Así, el primer principio del MIS reza: “las intuiciones vienen primero, las razones después” (Haidt 2012; también Haidt 2013).

Respecto del fundamento intuitivo de los juicios morales, Haidt avanzó, incluso, en el intento de identificar cuáles son las intuiciones morales fundamentales, más allá de las diferentes versiones con las que estas se manifiestan en diferentes culturas. Basadas en sistemas perceptuales y afectivos innatos (incluso quizás modulares), las intuiciones morales “están innatamente preparadas” (Haidt y Joseph 2004), y son relativas a distintas “propiedades emergentes” o fundamentos (daño/cuidado, justicia/reciprocidad, autoridad/respeto, *ingroup*/lealtad y pureza/santidad<sup>11</sup>). Así: “Los cinco fundamentos intuitivos dejan un gran espacio para que las culturas desarrollen diferentes conjuntos de virtudes, pero constriñen el universo de moralidades posibles.” (Haidt y Kesebir 2009, 220). En síntesis, los fundamentos de la moralidad en una “ética intuitiva” resultan ser tanto innatos como aprendidos culturalmente, y más variados que los que se habían identificado en los sistemas morales clásicos. En la Sección 3. me referiré con más detenimiento a la naturaleza de las intuiciones en general y, por lo tanto, también a las intuiciones morales.

En estos 20 años, el MIS ha mostrado ser muy fructífero en recoger evidencias en su favor, y ha logrado agitar la discusión en la psicología y la filosofía moral de un modo inusitadamente intenso. Mi conjetura es que este impacto se explica, al menos en parte, por la importancia que el modelo confiere a las intuiciones sobre el razonamiento moral en la formación del juicio moral, que es el punto en el que me interesa establecer un puente con la cognición psicológica, de la que voy a ocuparme en lo que sigue.

---

<sup>11</sup> En una versión previa, los patrones de eventos relacionados con el comportamiento humano que generan aprobación o desaprobación intuitiva están relacionados con los siguientes dominios nucleares de la moralidad: el sufrimiento, la autoridad, la reciprocidad y la pureza (Haidt y Joseph 2004). Otras modificaciones menores a esta clasificación aparecen en escritos posteriores.

## 2.1. De la “teoría de la mente” a los enfoques interactivos de la “lectura de mentes”

Se define a la “Teoría de la Mente” (ToM) como la capacidad para atribuir estados mentales (en adelante: EMs) a otros. En una de sus versiones clásicas, se trata de la capacidad para hacer inferencias acerca de los EMs de otras personas, a partir de evidencia comportamental, con el objeto de explicar, predecir y justificar su conducta. Desde sus inicios se identificó dicha capacidad con la satisfacción del “test de falsa creencia”<sup>12</sup>, cuyo diseño original presuponía competencia lingüística y sólo era sorteado por niños entre 3 años y medio y 4 años de edad. El nombre que adquirió esta capacidad crítica estuvo fuertemente asociado a un *modelo teoricista y epistémico*: conocer otras mentes era identificar evidencias relevantes, postular entidades inobservables, realizar o emplear generalizaciones, producir inferencias, predecir y explicar comportamientos. El modelo clásico impuso la *perspectiva del observador*: la mente de otros es enfocada como una entidad separada e inaccesible de manera directa para el intérprete. En cuanto al objeto de la atribución psicológica, en los tempranos modelos de la ToM, quedó restringido a las así llamadas “actitudes proposicionales”: estados mentales como las creencias y los deseos, con contenidos estructurados proposicionalmente. Las actitudes proposicionales, a su vez, impusieron sobre la atribución psicológica un modelo lingüístico y, con ello, todos los requisitos asociados a él, sobre todo un cierto enfoque semántico para dar cuenta de las atribuciones psicológicas: un enfoque descriptivista acerca de una actitud que relaciona un agente con un contenido (cf. Castro 2017). El modelo canónico conformado por estos elementos se conoce como *full-blown propositional attitude psychology* (Apperly y Butterfill 2009) y puede resumirse, aproximadamente, así: dados ciertos presupuestos normativos en el marco de otras asignaciones de actitudes proposicionales, la teoría debe identificar actitudes psicológicas y contenidos proposicionales (vgr. creencias y deseos), en el marco de una estructura causal y normativa, por medio de inferencias hacia la mejor explicación entre EMs y comportamientos.

Como señala acertadamente Apperly (2011), ToM era una denominación “tendenciosa” porque dejaba afuera otros tipos de mecanismos de reconocimiento y atribución y otros tipos de EMs<sup>13</sup>. “Lectura de mentes” (en adelante: LM) (*mindreading*) (Nichols y Stich 2003), de uso algo menos frecuente, pareciera una denominación más apropiada para referir a una capacidad cognitiva antes que al formato (cuestionable) del producto de dicha capacidad: una “teoría”. Además, la analogía con la *lectura* sugiere varios rasgos comunes a ambas competencias: *leer un texto* supone un abanico de capacidades para procesar un *input* complejo:

<sup>12</sup> Se trata de una prueba en la que los niños tienen que predecir el comportamiento de una persona sobre la base de la creencia falsa de esa persona, es decir, no tomando en consideración su propia creencia verdadera.

<sup>13</sup> Apperly (2011) considera que la expresión “lectura de mentes” es preferible a ToM en virtud de una acertada analogía con la lectura: el proceso de identificar los pensamientos y las acciones de las personas sobre la base de los movimientos corporales es análogo al proceso de identificar proposiciones sobre la base de inscripciones. Zawidzki (2013), en cambio, objeta la analogía por el peso que confiere a las motivaciones epistémicas y a los logros individuales. En mi opinión, la analogía es más rica que en una u otra interpretación, tal como lo sugiere en el texto.

procesos perceptuales, procesamiento semántico, memoria, inferencias, etc. Por su parte, la competencia lectora admite gradaciones, desde el aprendiz al lector diestro. El “lector de mentes” enfrenta su tarea con similares y variadas capacidades. En estos sentidos, la LM es también una competencia compleja en la que caben gradaciones y distinciones.

La investigación experimental y los intensos debates entre los “teóricos de la teoría” y los defensores de la teoría de la “simulación”, y entre las perspectivas de tercera y primera persona, respectivamente, fueron minando la visión homogénea inicial. Sin embargo, la disparidad de contextos y los tipos de criaturas estudiadas (niños, animales no humanos y en menor medida, adultos), así como de las áreas disciplinares y los estilos investigativos involucrados, obstaculizó la elaboración de lineamientos satisfactorios que fueran adecuados para los distintos tipos de agentes y capacidades examinados. No es difícil advertir que el modelo clásico no estaba preparado para tomar en cuenta el papel de otro tipo de EMs y otro tipo de conductas o rasgos de ellas (otro *explanandum*), ni tampoco podía dar cuenta de modos eficientes, rápidos y confiables de rastrear dichos EMs. No obstante, las constantes revisiones de los tempranos modelos teoristas deflacionaron gradualmente los requisitos cognitivos, p.e., evaluándolos con versiones no lingüísticas del test de falsa creencia. En el mismo sentido, se identificaron una variedad de procesos y capacidades precursoras de la ToM. En suma, la reformulación de las distintas hipótesis y modelos comenzaron a mostrar signos de convergencia hacia una concepción menos restrictiva o estrecha de los fenómenos abarcados por la LM (Spaulding 2019) y hacia variedades cada vez más híbridas (Spaulding 2018).

Actualmente resulta curioso que hayan pasado inadvertidos, dentro del alcance de la LM, diversos rasgos que poseen significados psicológico-sociales y cuya detección es no sólo inevitable sino casi automática, y que influyen de maneras críticas sobre la identificación de EMs. Me refiero a la identidad personal, los rasgos del carácter, las categorías y los roles sociales (vgr. el género, la edad, la raza, la clase, la nacionalidad, etc.) y los estereotipos asociados a ellos (Spaulding 2019). Se trata de rasgos individuales y sociales de las personas que poseen una fuerte carga normativa antes que un contenido puramente descriptivo, y que se integran a un sistema social de expectativas y regulaciones del que el intérprete usualmente forma parte, antes que a una teoría predictiva de la conducta de los otros. Estos rasgos, por otra parte, sugieren una reconfiguración de las funciones de la LM, desplazando a la interpretación por otras tales como influir, manipular, regular o “dar forma” a la conducta o a las disposiciones a la conducta de otros, en virtud de algún interés propio o común. Como señala Zawidzki (2013), “dar forma” (*mindshaping*)<sup>14</sup> a las mentes de otros es más bien un logro *social*, asociado a fines prácticos, antes que un logro *epistémico*. Es incluso razonable pensar que las funciones y capacidades involucradas en *mindshaping* y *mindreading*

---

<sup>14</sup> Zawidzki (2013) emplea el concepto de *mindshaping* para referir a una variedad de capacidades y prácticas (la imitación, el aprendizaje, el conformismo social, entre otras) orientadas a “dar forma” a las mentes de los demás, más que a representarla fidedignamente, para hacerlas interpretables. Estas capacidades de LM también harían posible el desarrollo de otras más sofisticadas.

se complementen: las primeras permiten que las segundas desempeñen un rol específico como componentes de una práctica justificatoria que depende del previo ajuste de las expectativas y los comportamientos en la interacción. Los motivos sociales también permiten explicar la influencia de los juicios morales intuitivos y los objetivos prácticos en activar, modular e incluso dar forma a la atribución psicológica. Así, además de la influencia más obvia que la atribución mental ejerce sobre la formación de juicios morales, se ha comprobado que estos también influyen sobre la identificación de los EMs de los demás (e incluso sobre la comprensión de otro tipo de información no psicológica) (cfr. Knobe 2004, 2005)<sup>15</sup>.

Ahora bien, en mi opinión, el cambio más significativo en la LM fue introducido por la propuesta de distintos *enfoques interactivos* (Gallagher y Hutto 2008; Di Jaegher, Di Paolo y Gallagher 2010), también llamados *enfoques de segunda persona* (Gomila 2001; Scotto 2002), que propusieron poner el foco en la dinámica de las *interacciones humanas*, más específicamente, en los *encuentros cara-a-cara*, en los que el centro de la atención teórica se desplaza hacia una variedad diferente de EMs: *emociones y estados afectivos, intenciones y fines* y el amplio repertorio de *comportamientos expresivos* que los vehiculizan, es decir, estados que no son actitudes proposicionales y evidencias que no son primariamente verbales<sup>16</sup>. Butterfill (2013) destaca la detección de *fines* para la acción conjunta, mediante evidencias como las que consideraremos más abajo: gestos y expresiones faciales, es decir, indicadores comunicativos no verbales que no requieren atribución de creencias u otras actitudes proposicionales. Así pues, los aspectos más participativos de la comprensión psicológica desplazaron de su papel excluyente en la LM a los mecanismos cognitivos individuales, diseñados bajo el modelo del “teórico”, incorporando otros de rastreo, detección y reconocimiento de señales expresivas, que son más rápidos y menos demandantes cognitivamente, y que son críticos para dar cuenta de una variedad de interacciones en los contextos prácticos en los que se desenvuelven.

El contraste es claro: los modelos clásicos no asignaban ningún papel a los mecanismos y las bases evidenciales que podrían operar *sólo* en contextos interactivos, aquellos en los que la *reciprocidad* no sólo estrecha el rango de las interpretaciones, sino que hace posible las primeras manifestaciones de LM: más que generar una “escalada” de adscripciones de nivel superior y creciente complejidad, en la interacción ambas “mentes”, en primer lugar, simplemente se encuentran. En ese encuentro interactivo, evalúan de maneras recíprocamente contingentes los comportamientos expresivos del otro, lo que les permite reconocerse, reaccionar, comprometerse, coordinar la acción, etc. El “lector de mentes” no es ya un observador pasivo ni desinteresado con finalidades sólo o primariamente epistémicas. Dicho de otra forma,

<sup>15</sup> Según Knobe (2006), algunos conceptos empleados en la *folk psychology* podrían ser, en realidad, herramientas especializadas para la cognición moral, vg., el concepto de *acción intencional*. Knobe da por sentada, no obstante, la visión clásica acerca de la función primaria de la atribución psicológica: la explicación y la predicción de la acción.

<sup>16</sup> Por razones de espacio, daré por supuesta la adecuación de los enfoques interactivos para todo el planteo que sigue. Referencias a distintas variedades de enfoques interactivos y un examen de sus virtudes explicativas en distintos ámbitos, pueden encontrarse en el volumen de Pérez y Gomila (2022).

cuando la LM es recíproca, cada uno de los miembros de esa interacción es recíproco en convertir al otro en el destinatario de la LM. Bajo este marco, *los mismos rasgos que caen bajo el alcance de estas competencias sirven también como herramientas para la comprensión recíproca*<sup>17</sup>.

Por otra parte, en estos enfoques, la emergencia de la LM se sitúa en un nivel ontogenéticamente más básico, es decir, en etapas del desarrollo previas a la adquisición de la competencia lingüística y las capacidades meta-representacionales. Ahora bien, cuando las personas interactúan mediadas por este tipo de comprensión recíproca, se generan las condiciones que posibilitan el desarrollo de las capacidades requeridas para variedades más complejas de LM, aplicables también a fenómenos que no ocurren en contextos interactivos y que no dependen (o no lo hacen tanto) de los comportamientos expresivos que la suscitan<sup>18</sup>. Otro aspecto de enorme importancia, es que se trata de capacidades que persisten y complementan a las que son más sofisticadas, en la cognición madura. En consecuencia, los enfoques interactivos no reemplazan a los enfoques teoristas e individualistas, sino que unos y otros dan cuenta del empleo de capacidades cognitivas diferentes pero complementarias en la cognición social adulta. Enseguida veremos cómo.

Como afirman Apperly y Butterfill (2009), para dar cuenta de todo tipo de casos de comprensión psicológica se requiere satisfacer dos demandas de procesamiento cognitivo que son contradictorias: *eficiencia* y *flexibilidad*. Sin ellas, la LM no podría facilitar la resolución de los diversos problemas sociales de “cooperación” o “competencia” para las cuales es una herramienta cognitiva necesaria. Así, los modelos más sobre-intelectualizados están mejor adaptados, en principio, para dar cuenta de la flexibilidad de las atribuciones, pero no para explicar cómo podrían ser también eficientes. Por su parte, las atribuciones eficientes parecen imprescindibles en el caso de las interacciones cooperativas, por ejemplo, en las específicamente comunicativas en las que están involucrados EM, porque requieren interpretación *online*<sup>19</sup>. Volveremos en seguida sobre este punto.

---

<sup>17</sup> Una manifestación ubicua del mismo tipo de fenómenos en contextos interactivos, tanto en los intérpretes como en los interpretados, es el “doble rol” de las emociones morales y las evaluaciones morales implícitas. Ambas proporcionan el fundamento intersubjetivo de moralidad (cfr. Isern-Mas y Gomila 2020, 795 y ss.). El MIS, por el contrario, sitúa la interacción al nivel de la racionalización de las intuiciones.

<sup>18</sup> Una variedad de hipótesis precisas y resultados experimentales dan cuenta de las interacciones que permiten a los niños adquirir gradualmente capacidades más refinadas para la comprensión psicológica, pero no tengo espacio para referirme a ellas.

<sup>19</sup> Como señalan estos autores, las tempranas propuestas innatistas y modularistas de Leslie (1994), entre otras, pretendían dar cuenta de ese tipo de procesos *eficientes* para la atribución desde la temprana infancia. Por su parte, quienes proponían la intervención de mecanismos inferenciales y conocimiento acumulado hacían lugar a la *flexibilidad* de la ToM (Gopnik y Meltzoff 1997).

## 2.2. Enfoques duales de la cognición

El giro interactivo mencionado recibió un impulso indirecto, aunque vigoroso, de las *teorías de los procesos o sistemas duales* de la cognición humana, porque estas teorías ofrecieron un modelo alternativo al de la arquitectura cognitiva unificada que había sido presupuesto en las explicaciones clásicas de LM. De acuerdo a este nuevo modelo, a cada capacidad cognitiva pueden subyacer distintos procesos cognitivos y distinto tipo de representaciones. De acuerdo a una versión de estas teorías, existirían *dos sistemas* de “razonamiento”, cada uno con estructuras, funciones e historias evolutivas diferentes (Frankish 2010): uno, así llamado sistema Tipo 1 o *intuitivo*, sería de procesamiento holístico, no conceptual, irracional e inconsciente, evolutivamente más antiguo y compartido con otras especies. El otro sistema, llamado sistema de Tipo 2, sería de procesamiento *analítico*, conceptual, racional y consciente, evolutivamente más reciente y únicamente humano. Mientras el primero involucraría procesos automáticos, el otro, en cambio, recurre a procesos controlados. Así, el sistema 1 sería más rápido, se iniciaría de modo espontáneo, funcionaría de modo paralelo, tendría carácter asociativo y consistiría en conocimiento inconsciente, contextualizado y de carácter intuitivo; el segundo, por el contrario, sería más lento y abstracto o descontextualizado, secuencial y basado en reglas, controlado y reflexivo (Glöckner y Witteman 2010; Frankish y Evans 2009). Mientras el sistema 2 sería distintivamente humano, el Sistema 1 sería evolutivamente antiguo y compartido con los animales no humanos.

Ahora bien, ambos sistemas operarían en paralelo: a veces colaborando y a veces compitiendo, según los constreñimientos que estén operando y los requerimientos de las conductas a interpretar. Incluso hay versiones de los enfoques duales más matizadas que no distinguen entre dos *sistemas* sino entre dos tipos de *procesamiento cognitivo*, que pueden llevar tanto a resultados conflictivos como también a operar de maneras complementarias (cfr. Frankish 2010). Por ejemplo, un tipo de *procesamiento* sería *asociativo, heurístico o intuitivo*, es decir, guiado por los principios de similaridad y contigüidad temporal, mientras que el otro estaría *basado en reglas*, sería *proposicional, analítico o reflexivo*, y elaboraría representaciones con formato conceptual y valores de verdad. Como señala Frankish (2010), el núcleo de estas teorías se asienta en la distinción pre-teórica entre *intuición* –una capacidad cuasi-perceptual, inmediata, sensible a claves inconscientes y frecuentemente sesgada– y *razón* –una capacidad más lenta, explícita, que requiere esfuerzo y es más prudente. Como veremos en la Sección 3., casi todos los rasgos de las intuiciones se atribuyen a los procesos y los productos del sistema 1 y al tipo de procesamiento cognitivo que lo caracteriza.

Caben aquí algunas precisiones. Según ciertas propuestas recientes, mientras algunos de estos rasgos serían “nucleares” para cada tipo de procesamiento o sistema, otros deberían verse sólo como típicos, pero no esenciales. Así, en el caso de las intuiciones morales, por ejemplo, aun cuando involucren procesos de tipo 1 como los ya señalados, estarían, sin embargo, basados en reglas. Las características culturales específicas de las normas morales, además, requieren de procesos de aprendizaje cultural que sólo con el tiempo se convierten en juicios morales automáticos (cfr. Saunders 2009, 341 y ss.). Este punto se relaciona con el

siguiente: mientras algunos estados del sistema 1 son relativos a los procesos de maduración cognitiva normales (como detectar emociones o estados de ánimo en las expresiones faciales), otros dependen de la acumulación de experiencia y aprendizaje (vgr. las evaluaciones morales aprendidas culturalmente en la infancia). Así, una intuición puede tener en su origen procesos cognitivos diferentes, aun cuando los estados resultantes pertenezcan al sistema de tipo 1: una sería la “cognición madurativamente natural” y la otra, la “cognición natural experta” (De Cruz 2015).

La *teoría dual de la cognición* fue propuesta para estudiar los diversos procesos cognitivos automáticos que influyen en la conducta social, tales como los estereotipos, los sesgos y las actitudes implícitas, entre otros, que son especialmente relevantes para explicar tanto la cognición social como la cognición moral. Saunders (2009), por ejemplo, sostiene que las intuiciones morales y la reflexión moral deberían ubicarse cada una en un sistema distinto, una idea adoptada por el MIS. En ese marco, argumenta que los enfoques duales podrían explicar mejor que una visión unificada (y racionalista) de la mente los frecuentes hiatos (*gaps*) que se producen entre las intuiciones y los juicios morales intuitivos y las convicciones morales razonadas o las teorías conscientemente adoptadas, vgr. el caso de la estupefacción moral (*moral dumbfounding*) (cfr. Haidt 2001). En filosofía de la mente, los enfoques duales se emplearon para explicar los casos difíciles para la atribución intencional: la disociación, la akrasia y el auto-engaño, y, como veremos en seguida, aunque en menor medida, para comprender la psicología *folk* (Frankish 2010). Veamos entonces versiones de este tipo de enfoques aplicados a la LM.

### 2.3. “La lectura mínima de la mente” en un enfoque dual

En relación con la LM, se ha propuesto la hipótesis que existirían dos tipos de capacidades cognitivas (Apperly y Butterfill 2009; Butterfill y Apperly 2013; Apperly 2013): una más eficiente pero inflexible, es decir, que involucra procesos de “más bajo nivel”, y otra más flexible pero cognitivamente más demandante que requiere la intervención de procesos de “más alto nivel”. La primera se ha identificado con las siguientes denominaciones: *teoría de la mente implícita, automática o mínima* (Butterfill y Apperly 2013). La teoría “mínima” sería una cierta capacidad (*theory of mind ability*) para “rastrear” las percepciones y los registros de otros, para “explotar o influenciar hechos acerca de esos estados” (2013, 607), “ninguno de los cuales son representaciones” (2013, 621). Es decir, no requiere “representar representaciones” (2013, 614). Existe evidencia experimental acerca de la capacidad para rastrear estados perceptuales y creencias sin representarlos, es decir, de manera más o menos automática, en adultos, niños y animales no humanos con tareas implícitas: “...la mera presencia de agentes sociales es suficiente para activar *automáticamente* “*online belief computations*” (o sólo intencionalidad y fines), tanto en adultos como en niños pequeños desde los 7 meses, las que podrían mantenerse como “representaciones alternativas del entorno”, incluso en ausencia del

agente (Kovács, Téglás, y Endress 2010, 1834)<sup>20</sup>. La teoría “plena” (*theory of mind cognition*), en cambio, supone la capacidad para elaborar representaciones de actitudes proposicionales con el objeto de construir explicaciones causales de la acción. Como se verá luego, esta versión de enfoque dual es todavía demasiado restrictiva: presta atención predominante a EMs con significación epistémica y parece suscribir la tesis que la primera carece de las herramientas que la segunda finalmente alcanza, sugiriendo, en consecuencia, que esta podría reemplazarla.

Apperly y Butterfill (2009) examinan diferentes propuestas que también buscan explicar el *funcionamiento eficiente* de las atribuciones psicológicas. Una de ellas recurre a los procesos de *automatización* adquiridos con la práctica, a los que nos hemos referido en el apartado anterior. Sin embargo, dado que suponen tanto la intervención previa de procesos inferenciales conscientes como suficiente práctica acumulada, no permitirían explicar la atribución eficiente en casos nuevos. En tal sentido, los procesos de *automatización* no pueden explicar *todos* los casos de *automaticidad* (Butterfill y Apperly 2013). Sin embargo, son interesantes para nuestros fines, porque algunas intuiciones psicológicas y morales serían efectos de la experiencia acumulada, el aprendizaje e incluso la intervención de procesos inferenciales. Otro mecanismo propuesto podría ser aún más simple: la combinación de *asociaciones aprendidas*, es decir, de regularidades estadísticas entre comportamientos con el empleo de reglas, que permiten formar expectativas acerca del comportamiento futuro. Ahora bien, este tipo de mecanismos tampoco permitiría lidiar con todo tipo de casos, al depender de regularidades comportamentales previamente observadas. Finalmente, una variante de las teorías duales propone diferenciar las capacidades para la “percepción social”, que detectan, mediante “juicios inmediatos *online*”, las características expresivas y emocionales de una persona a través de diversas señales -los llamados “componentes socio-perceptuales” de la LM-, de las capacidades inferenciales sobre sus estados representacionales más complejos, como las creencias y similares -los “componentes socio-cognitivos”- (Tager-Flusberg y Sullivan 2000). El primer componente incluye capacidades muy básicas pero críticas, tales como distinguir personas de no personas, a una de otra persona, y una variedad de otras características personales (volveremos sobre esto). Según Apperly y Butterfill (2009), esta versión no es incompatible con la suya, pero no sería suficiente puesto que está limitada a rasgos expresivos perceptibles. La alternativa que proponen bajo la denominación de *teoría mínima de la mente*, como lo anticipamos, sería la postulación de mecanismos y estados no tan complejos: la capacidad para detectar comportamientos como resultados del ejercicio de una función: detectar las percepciones de otros, en particular, de aquello con lo que estos se “encuentran” (*encounterings*), *rastreando*

---

<sup>20</sup> No todos acuerdan que las primeras sean capacidades para mentalizar: al involucrar procesos de nivel más bajo, meras asociaciones entre estímulos, sostienen que se trataría, más bien, de capacidades para “submentalizar” (Heyes 2014). Sin embargo, como veremos en lo que sigue, los procesos cognitivos abarcados por la “mente implícita” tienen un papel sustantivo en la comprensión psicológica que no puede quedar fuera del alcance de la LM.

relaciones intencionales primitivas a través de hechos acerca de esos EMs. Ello le permitiría al “intérprete” influir o explotar esos hechos. Estas capacidades no requieren la formación de representaciones.

En mi opinión, la “teoría mínima” debería incluir todos los componentes mencionados hasta aquí: las asociaciones perceptuales, la percepción de rasgos expresivos, la capacidad para rastrear estados de registro, las habilidades automatizadas: todos ellos permiten detectar, de maneras eficientes, automáticas e implícitas, rasgos personales que son psicológica y socialmente relevantes y EMs como las emociones, las intenciones y los fines. Veamos ahora, brevemente, cómo caracterizar la base evidencial requerida para una LM mínima.

#### 2.4 Rasgos y comportamientos expresivos

Las personas tenemos múltiples maneras de expresarnos o dar expresión a nuestros EMs y a nuestras maneras de “ver” el mundo a nuestro alrededor. Lo hacemos, además, todo el tiempo y con una variedad de vehículos, es decir, no sólo ni primariamente con recursos lingüísticos. Dicho de otra forma, disponemos de un vasto repertorio de formas de comunicación corporizada. A pesar de su ubicuidad y significación para las interacciones humanas, los *rasgos y comportamientos expresivos* no han merecido suficiente atención teórica ni en la filosofía de la mente ni en la filosofía del lenguaje. Con escasas excepciones, los comportamientos expresivos<sup>21</sup> (en adelante: CEs) fueron confinados al rango de otros fenómenos biológicos (como los instintos, los síntomas o las claves) o al de meras conductas en respuesta a estímulos (como los reflejos o los hábitos).

El filósofo Mitchell Green (2007) elaboró una teoría detallada de las *expresiones*, muchos de cuyos rasgos nos servirán para identificar la base evidencial de la LM mínima. Según Green, los CEs deben satisfacer ciertos criterios. No puedo examinar aquí su propuesta integral, por lo que mencionaré sólo aquellos que, en mi opinión, son *constitutivos* de todos los CEs que nos interesan para la LM. Según Green, los CEs *muestran* los estados de las personas (en vez de ocultarlos, de actuar sobre ellos o de describirlos); son *auto-expresiones*, es decir, expresan los propios estados (en vez de dar expresión a estados que no son propios); *no son criaturas lingüísticas* (aunque algunas expresiones lingüísticas puedan ser o contener vehículos expresivos<sup>22</sup>); son *señales*, es decir, rasgos diseñados, natural o culturalmente, por su capacidad para transmitir información (y no meramente claves); pueden ser *voluntarias o involuntarias*;

<sup>21</sup> Me referiré en adelante a los *comportamientos expresivos*, pero con la intención de referir también a los *rasgos expresivos*, los que, además, pueden ser empleados para realizar comportamientos expresivos.

<sup>22</sup> El repertorio expresivo de la comunicación vocal humana incluye vocalizaciones convencionalizadas (como las interjecciones), otras escasamente convencionalizadas y las que son fonológicamente anómalas (como los suspiros, los gemidos, los quejidos, etc.), así como los rasgos prosódicos (cambios en la entonación, el volumen y la duración) que acompañan regularmente el uso del lenguaje hablado. Por cierto, la comunicación expresiva es multimodal, no sólo vocal.

son acerca de estados internos (aunque sólo de los propios y no de estados fingidos)<sup>23</sup>; característicamente *se dirigen a una audiencia*; tienen una dimensión “adverbial”, es decir, no sólo comunican mediante un comportamiento sino también mediante un *modo de realización*; pueden o no ser exitosos; pueden ser naturales, es decir, no necesitan estar convencionalizados. El repertorio de señales expresivas humanas es variado: las expresiones faciales, los gestos, las actitudes corporales, las vocalizaciones no verbales y los rasgos prosódicos de la comunicación verbal.

Así, donde A es un agente y B un estado experiencial, afectivo o cognitivo, *A expresa B* si y solo si A está en el estado B, y alguna acción o conducta de A *tanto muestra como señala* su estado B. Green distingue, además, distintos tipos de mostrar: *mostrar hechos* (o *mostrar-que*: lo mostrado posee contenido proposicional: “se muestra que algo es así”); *mostrar cosas* (o *mostrar-x*: lo mostrado es un objeto singular y mostrar es *hacer perceptible algo*), y *mostrar cómo algo se experimenta* (o *mostrar-cómo*: un conocimiento experiencial). Cada uno de los tres tipos de “mostrar” reflejan distintas formas de “conocimiento” de quien muestra y emplea “señales” de tipos distintivos. Los tres tipos *expresiones* intervienen en la activación de las intuiciones psicológicas y las morales: las que *hacen perceptible algo*, por ejemplo, un estado emocional con un contenido intencional específico; las que muestran algo, por ejemplo, una intención agresiva, y las del tercer tipo: aquellas que *muestran cómo se experimenta algo*, por ejemplo, con qué intensidad se experimenta un estado de repugnancia moral.

En este contexto, el rostro humano merece una atención especial. Los rostros son *funcionalmente especiales* para muchas criaturas sociales en tanto vehiculizan información valiosa para quienes los miran. Así, son un componente fundamental de la comunicación no-verbal, incluso cuando la comunicación verbal está disponible. Por esta razón no es posible caracterizar a los rostros sólo en base a las funciones biológicas de sus rasgos, sino que se requiere comprenderlos también como objetos visuales de carácter social, más específicamente, como sofisticados “órganos de comunicación social” (Bruce y Young 2012). En tal sentido, comunican señales que sirven a los distintos propósitos vitales de quienes los observan. Muchas de ellas son señales naturales que no codifican estructuras conceptuales y poseen un “significado” altamente dependiente del contexto. En este nivel funcional, se hace evidente la riqueza informacional de los contenidos de la percepción facial: la identidad individual; el foco atencional, a través de la detección de la dirección de la mirada, y también los movimientos de la boca que ayudan a la percepción del habla u otros que complementan la información verbal comunicada; distintos estados mentales, sobre todo estados emocionales y estados de ánimo; así como información más estable, tales como rasgos de personalidad, el sexo, la edad y el grupo racial. Los rasgos distintivos de cada rostro individual, generan sutiles diferencias con otros rostros, además de una amplia dinámica expresiva, para cuya identificación se requieren

<sup>23</sup> Contra Green, los rasgos expresivos no son *sólo* manifestaciones de estados internos, sino también de rasgos más integrales de las personas, vgr. actitudes cordiales u hostiles, modos pausados y serenos de hablar, maneras penetrantes o inquisidoras de dirigir la mirada, etc.

capacidades perceptuales altamente especializadas, “innatamente preparadas”. Ahora bien, el reconocimiento de las expresiones faciales es también susceptible de refinamiento: los adultos pueden discriminar expresiones faciales más finamente que los niños y detectar una mayor variedad de información psicológica y moralmente relevante.

Una propuesta que pone de relieve el carácter interactivo y la intervención de capacidades implícitas para la percepción de rostros y el reconocimiento de las expresiones faciales es el *enfoque ecológico-social de la percepción facial* (Zebrowitz 2011). Este enfoque abarca, además, no sólo las expresiones emocionales básicas, sino también de intenciones para la acción y la *percepción social de rostros* a la que ya nos hemos referido: las cualidades faciales más estables y socialmente significativas. La idea es que “los rostros comunican *affordances* conductuales” muy diversas. No se trata, pues, de propiedades de un objeto visual externo para un “observador”, sino de rasgos del rostro en tanto *affordance* social. Así pues, por ejemplo, observamos rostros humanos como siendo amigables, hostiles, colaboradores, etc. Este enfoque acentúa el carácter directo de la percepción de la información facial para los percipientes y, con ello, agrega fundamentos a los enfoques interactivos, al subrayar la importancia de los contenidos que emergen en los encuentros de “segunda persona”.

### 3. ¿Qué son las intuiciones?

Aunque según la opinión más aceptada entre los filósofos, las intuiciones son un sub-tipo específico de estados mentales (Pust 2017), hay quienes afirman, por el contrario, que no constituyen un constructo homogéneo (Glöckner y Witteman 2010; Nado 2015) y que sería recomendable “una disección de grano más fino de la taxonomía intuitiva” (Schukraft 2016, 334). Complica las cosas el hecho que suele abordarse el análisis de las intuiciones, al mismo tiempo, aunque en ocasiones de manera inadvertida, a la vez como *proceso* y como *producto*, es decir, sin distinguir entre *el intuir* y *lo intuido*. Las funciones, los dominios (moral, psicológico, filosófico<sup>24</sup>) y los efectos cognitivos y prácticos de las intuiciones, finalmente, también han servido para intentar comprender sus rasgos peculiares.

También según la opinión mayoritaria de los filósofos, las intuiciones sólo son un tipo de juicios o creencias (o disposiciones a la creencia), cuyos contenidos son proposicionales, aunque con algunos rasgos distintivos: un elemento de espontaneidad y una característica fuerza asertiva. Para otros, se trataría, en cambio, de un tipo muy peculiar de estados mentales: “pareceres” de diversos tipos (perceptuales, afectivos o intelectuales), con diferentes formatos (no conceptual, conceptual o con una estructura proposicional). En el caso de las intuiciones basadas en la percepción o en fenómenos afectivos o emocionales, su contenido

<sup>24</sup> Haidt y Kesebir (2009) proponen identificar diferentes subtipos de intuiciones según estas se relacionen con estudios acerca de la personalidad, acerca del comportamiento o acerca de los procesos mentales, y en este último caso, según se apliquen a fenómenos no humanos o a fenómenos humanos, sociales, morales u otros. En ese contexto, afirman, las intuiciones morales constituyen un subtipo diferenciado y más complejo que las demás.

podría ser objetual y no proposicional y su componente evaluativo podría tener la forma de una sensación de aprobación o de desaprobación. Otros filósofos, finalmente, caracterizan a las intuiciones como “sensaciones no teóricas”, “corazonadas”, “alarmas internas” o “vagas sospechas” (MacGahhey y van Leeuwen 2018), es decir, las conciben como estados híbridos, a la vez, sub-doxásticos, afectivos y experienciales. Finalmente, es propio de la fenomenología de las intuiciones la sensación de “saber lo que uno tiene que hacer” o de confianza subjetiva (Thompson et al. 2011).

En cuanto a su etiología cognitiva, se acepta que se trata de estados irreflexivos, no originados por procesos mentales conscientes y que, por ello mismo, son más resistentes a las evidencias o las creencias en sentido contrario. Así, logran tener credibilidad *prima facie* para quienes las poseen (De Cruz 2015). Hay coincidencia, pues, en contrastar el proceso de formación de las intuiciones, que sería sub-personal, con el modo de recepción, que es consciente y genera respuestas espontáneas (la definición que propone Haidt destaca, justamente, ese contraste). Por otra parte, aunque las intuiciones diferirían de las creencias al menos en este aspecto, podrían, sin embargo, modificarlas: las experiencias mencionadas podrían fortalecer o debilitar, según el caso, los contenidos creídos (Schukraft 2016). La influencia en la dirección contraria no parece tan fácilmente admisible. Al menos en algunos casos, la combinación de componentes afectivos y experienciales, sensoriales y doxásticos, evaluativos y motivacionales con las que se presenta el contenido de muchas intuiciones justificaría, según algunos autores, caracterizarla como un “constructo multidimensional” (Sinclair 2010).

En distintos contextos teóricos, se ha señalado la utilidad de distinguir entre *intuiciones de primer nivel e intuiciones de segundo nivel* (Cohnitz y Haukioja 2015). Esta distinción puede ser útil, a su vez, para diferenciar entre dos funciones diferentes de las intuiciones: su *rol constitutivo* en el caso de las primeras, y su *rol evidencial* en el caso de las segundas. Una instanciación de ellas sería la distinción entre intuiciones psicológicas e intuiciones meta-psicológicas. Las primeras serían las manifestaciones (*outputs*) de una capacidad, que se consideran constitutivas de dicha capacidad, por ejemplo, ciertas reacciones espontáneas ante la percepción de un estímulo con un significado psicológico. Por cierto, la conexión de las intuiciones psicológicas con los rasgos y comportamientos expresivos depende de tomar a estos, a su vez, como *señales constitutivas* de los EM. Ahora bien, dado el predominio de las funciones sociales y prácticas de la LM, el papel evidencial de las intuiciones psicológicas sólo se manifiesta claramente en los casos problemáticos de interpretación, aquellos originados por la activación de sesgos, prejuicios y estereotipos que operan implícitamente. En este caso, las intuiciones de segundo nivel, por ejemplo, la creencia explícita y consciente que alguien está en un cierto estado psicológico (vgr. está feliz), podría colisionar con las intuiciones de primer nivel activadas de manera implícita (vgr. está riendo nerviosamente). En el caso de las intuiciones morales, serían de primer nivel las respuestas espontáneas ante una emoción moral como el desprecio (de uno hacia otros) o la vergüenza (hacia uno mismo); las intuiciones de segundo nivel, en cambio, serían esas mismas capacidades en tanto conducen a la formación de creencias sobre las primeras, por ejemplo, la creencia o el juicio moral que una

acción, ajena o propia, es moralmente despreciable. Es posible que, según el MIS, las intuiciones morales sean intuiciones de primer nivel, en tanto tienen una función constitutiva para la cognición y el comportamiento moral, mientras que los juicios morales basados en ellas serían las intuiciones de segundo nivel, aquellas que pueden ser referidas para dar cuenta del comportamiento de otro o para someter a reflexión moral el propio.

En la *literatura psicológica*, la visión predominante acerca del papel epistémico de las intuiciones fue primero más bien pesimista, al atribuírseles un mero papel heurístico, en el que resultarían guías o atajos poco confiables para la toma de decisiones o la resolución de problemas, en virtud de los sesgos irracionales que las afectan. Esta visión ha cambiado en los últimos años, asignándose a las intuiciones también una función positiva en la economía cognitiva (Pretz 2008). Ello se vio claramente reflejado en los enfoques duales la cognición humana, a los que me referí antes. La distinción entre pensamiento intuitivo y pensamiento analítico (o procesos intuitivos y procesos racionales) se ha hecho frecuente en la literatura en ciencias cognitivas (Pretz 2008). Varios autores han reconocido distintas formas de procesamiento cognitivo involucradas en la etiología de las intuiciones (Sinclair 2011). Teniendo en cuenta ese criterio, se ha propuesto diferenciar entre dos tipos diferentes: las *intuiciones holísticas* y las *intuiciones inferenciales* (Pretz 2008, 2011). De acuerdo a dicha clasificación, todas ellas serían “juicios” rápidos y automáticos, realizados sin esfuerzo, esto es, que no requieren atención ni memoria de trabajo y están fuera del control consciente de quien los produce. Es decir, los dos tipos de intuiciones ejemplificarían procesos de Tipo 1. Ahora bien, las *intuiciones holísticas* son aquellas que integran o sintetizan información perceptual compleja, mediante procesos inconscientes y más primarios, es decir, no basados en procesos analíticos. Estas intuiciones resultan de la “integración holística de claves”<sup>25</sup> (Sinclair 2011, 18), y no dependen de la experiencia previa. Las *intuiciones inferenciales*, en cambio, serían juicios basados en inferencias automáticas, altamente confiables, que derivan tanto de la experiencia perceptual acumulada como de los procesos analíticos de su adquisición. Abarcan, por ejemplo, a los así llamados *juicios* o *creencias expertas* (Pretz 2008), que se tornan automáticos con la práctica, y son más claramente adquiridos que el tipo anterior de intuiciones<sup>26</sup>. Mientras algunas intuiciones psicológicas serían del tipo holístico, al integrar diferentes señales en una configuración expresiva, vgr. una actitud postural, una expresión facial y cierto tono de voz nos permiten detectar, por ejemplo, un estado de tristeza, estupor o miedo; otras, en cambio, podrían ser inferenciales, combinando información personal acumulada con señales expresivas percibidas, por ejemplo, atribuir rasgos de personalidad agresivos o poco confiables a alguien por su pertenencia a un cierto grupo social, asociándolos a cierta manera de caminar, de mirar y de dirigirse a otros.

<sup>25</sup> Dada nuestra interpretación de los rostros y de cada uno de sus rasgos como *señales expresivas*, deberíamos decir que las intuiciones psicológicas holísticas integran señales, no claves.

<sup>26</sup> La mayor parte de los juicios intuitivos expertos caen fuera de nuestro interés porque es al menos discutible que las intuiciones psicológicas y morales ordinarias puedan tornarse expertas *en el mismo sentido* en que lo son las intuiciones expertas de un músico, un gran cocinero o un atleta.

Por otra parte, distintos estilos de procesamiento producen intuiciones: procesos de aprendizaje meramente asociativos; de aprendizaje a partir de casos, en los que se compara la situación actual con esquemas mentales, prototipos o imágenes almacenados; la integración de ambos y, finalmente, procesos *top-down* más constructivos, incluyendo el uso de representaciones mentales. Mientras las intuiciones que tienen un fuerte componente afectivo estarían sólo basadas en procesos asociativos y las inferenciales lo estarían en procesos constructivos, las intuiciones holísticas serían el fruto de la integración de procesos *bottom-up* (Glöckner y Witteman 2010). Este tipo de procesos combinan capacidades cognitivas de ambos tipos, Tipo 1 y Tipo 2, pero que, con la práctica acumulada y la automatización, adquieren las características de los primeros. En síntesis, la distinción entre intuiciones holísticas e intuiciones inferenciales abarca tanto distintos tipos de contenidos como distintos estilos de procesamiento. Ambos dan cuenta de diferentes tipos de intuiciones psicológicas: las que detectan señales expresivas y las que activan sesgos, prejuicios y estereotipos, respectivamente; y también morales: las que guían la propia conducta o las que permiten detectar y evaluar las propiedades morales en las conductas de otras personas, y las que originan juicios morales basados en ellas, respectivamente.

En suma, las intuiciones psicológicas resultan de procesar, almacenar y/o refinar la información psicológica manifiesta en las señales ofrecidas por los rasgos y los comportamientos expresivos de las personas, en los contextos interactivos o cara-a-cara. Las intuiciones morales incorporan contenidos más complejos: emociones morales y patrones normativo-culturales en diferentes dominios, relativos a sí mismo y a otros. Aunque el origen de ambas puede deberse a rasgos adaptativos innatos para los cuales constituyen una heurística apropiada, otras necesitan ser aprendidas con la experiencia. Las intuiciones psicológicas tienen un importante papel en la LM porque permiten identificar a las personas por sus rasgos expresivos y por sus comportamientos distintivos y/o estereotípicos, así como categorizar y evaluar los EMs y detectar la valencia de las conductas expresivas. Así, el comportamiento expresivo y la comprensión psicológica intuitiva son las dos caras de un mismo constructo interactivo, mayormente implícito, que está en la base de la LM entre las personas, incluidas las personas adultas. Así, las intuiciones psicológicas facilitan y modulan las interacciones personales de maneras automáticas, espontáneas y sin el control consciente de los sujetos. De ese modo, condicionan o facilitan la LM respecto de otros estados mentales que no tienen una relación directa y más o menos transparente con los comportamientos expresivos, para los cuales se requieren otras evidencias proporcionadas por los agentes y otros mecanismos utilizados por el intérprete. Su eficiencia para evaluar el tipo y carácter de las señales expresivas, no impide, como veremos en seguida, que también puedan conducir a interpretaciones distorsivas o sesgadas.

De una manera tentativa quiero a referirme ahora a otros fenómenos cognitivos, relevantes para la comprensión psicológica y moral y que, según el caso, comparten rasgos similares a los descritos aquí acerca de las intuiciones. Haidt (2013) sugiere la pertinencia de esa relación al referirse positivamente a la “revolución de la automaticidad” como una de las diferentes fuentes teóricas que confluyeron en provocar un giro intuicionista en la psicología moral

hacia fines de los años 90. Se trata de los fenómenos mentales automáticos o implícitos que recibieron una atención preferencial en la investigación experimental en los últimos años. Estos fenómenos se conocen bajo el concepto general de *cognición social implícita* (Payne y Cameron 2010<sup>27</sup>; Greenwald y Lai 2020). Esta forma de cognición abarca una familia de procesos y estados que intervienen de modos característicos tanto en el comportamiento de las personas como en su interpretación. Se entiende que un proceso cognitivo o un EM es *implícito* cuando es *automático* (no está bajo el control de quien lo tiene), es típicamente *inconsciente*<sup>28</sup> (inconsciente puede ser el estado, el proceso que lo ha originado y/o su influencia sobre el comportamiento) (Holroyd, Scaife y Stafford 2017)<sup>29</sup> e *inarticulado*, es decir, posee rasgos introspectivamente inaccesibles: la persona “sabe” más que lo que puede reportar. Por esta razón, los EMs implícitos se estudian por medio de técnicas de medición indirecta, que no requieren reportes verbales de los sujetos estudiados, especialmente el muy conocido Test de Asociación Implícita (IAT) (Greenwald 1998). Se consideran también implícitos los conocimientos que consisten en habilidades o destrezas, y todos los aprendizajes que generan comportamientos habituales automáticos. Estos rasgos, que suelen presentarse relacionados entre sí, aunque no siempre todos a la vez, hace pensar que ninguno es esencial para caracterizar a un proceso o un estado como implícito. Ello sugiere que los criterios adecuados para su identificación no deberían ser los típicos de la vieja psicología de las actitudes proposicionales, y que, incluso podrían tener un perfil disposicional, en el sentido que podrían ser más bien “actitudes” antes que estados, vgr. preferencias para evaluar o sentir de una cierta manera acerca de ciertas entidades en ciertas situaciones (cf. Machery 2016).

Lo cierto es que algunos de los fenómenos que integran la cognición social implícita son relevantes tanto para la LM como para el juicio moral. Me refiero a las *impresiones implícitas* y a los *sesgos implícitos*, denominaciones que en ocasiones se usan de manera intercambiable y que refieren fenómenos que, en parte, se solapan. Las *impresiones implícitas* se definen como “los residuos no episódicos en la memoria de nuestras observaciones de, interacciones con, e inferencias acerca de otros” y que no dependen de la memoria explícita de encuentros pasados. Constituyen las bases intuitivas de las descripciones que haríamos de otras personas y que “guían nuestros pensamientos explícitos, nuestras emociones, y nuestras conductas” hacia ellos, sin hacerse explícitos (Uleman et al. 2008). Las impresiones se forman mediante “inferencias espontáneas” a partir de la percepción de una variedad de características de las personas: los rasgos estáticos y dinámicos, que provienen sobre todo de la percepción visual de las

<sup>27</sup> Estos autores atribuyen justamente al MIS el haber sido “el primero en incorporar seriamente las lecciones de la cognición social implícita” (p. 454).

<sup>28</sup> Es usual la distinción, en el caso de las intuiciones, entre un proceso inconsciente y un estado consciente. Dependiendo de los contextos, sin embargo, su influencia podría ejercerse de ambas maneras.

<sup>29</sup> Algunos resultados experimentales evidencian que las personas atribuyen mayor responsabilidad a un agente cuando su conducta está causada por un proceso inconsciente que si lo está por uno automático (cf. Cameron, Payne y Knobe (2010). Así, los procesos cognitivos automáticos no necesariamente son también inconscientes, aunque sí lo sean usualmente.

personas: facial, gestual y postural, a los que ya hemos hecho referencia. También incluyen la detección de fines, creencias y valores morales asociados al comportamiento observado. En cuanto a los *sesgos implícitos* son asociaciones automáticas, no conscientes y difíciles de controlar, que pueden entrar en conflicto con las creencias y normas que una persona defiende de manera consciente. Incluyen estereotipos, prejuicios y sesgos sobre ciertos grupos sociales, que pueden dar lugar a juicios morales discriminatorios y atribuciones psicológicas influidas por ellos. Estas formas de cognición implícita podrían estar basadas en procesos asociativos, pero también podrían emplear reglas combinando estructuras de procesamiento diferentes (Holroyd, Scaife y Stafford 2017).

Más allá de sus diferentes denominaciones y caracterizaciones, los procesos, EMs o actitudes implícitas abarcan una variedad de fenómenos: las asociaciones perceptuales y las inferencias espontáneas sobre rasgos de las personas, los estereotipos, prejuicios y sesgos evaluativos, las preferencias, con valencias positivas o negativas, acerca de esos rasgos y comportamientos. Aunque en la literatura sobre cognición implícita estos fenómenos no son asimilados a las intuiciones, comparten con ellas varias de las notas que señalamos sobre los procesos que las originan y los contenidos a los que refieren. Así, la “mente implícita” ocupa la misma región que la “mente intuitiva”: es el asiento de los procesos que originan las intuiciones psicológicas y las intuiciones morales.

#### 4. Las intuiciones vienen primero, también en la cognición psicológica

Según se desprende de lo expuesto en las secciones 1. y 2., hay una sustantiva convergencia entre la explicación de la cognición moral ofrecida por el MIS y la explicación de la cognición psicológica o de la LM expuesta en este trabajo. En ambos casos, se trata de una caracterización “dual”, es decir, una caracterización que reconoce un papel diferenciado a cada tipo de capacidades, vgr. intuiciones y emociones frente a creencias y razones basadas en principios morales, en el primer caso; y modos de atribución psicológica mínimos e implícitos, también intuitivos (entre otros) vs. atribuciones explícitas, inferencias teóricas y razones que las justifican. Se advierte, también, un claro contraste entre ambos enfoques y los enfoques clásicos: por una parte, el racionalismo y el individualismo moral y, por la otra, los enfoques teoricitas, epistémicos e individualistas de la atribución mental. Como lo expusimos en la sección 3., las características específicas de las intuiciones en los ámbitos de la cognición psicológica y la cognición moral, y de otros fenómenos cognitivos implícitos que operan de manera similar, así como su papel en orientar las interacciones y en guiar las evaluaciones morales, respectivamente, son indicativas de semejanzas estructurales entre ambas capacidades. Esta manera de entender sus semejanzas, sin embargo, no equivale a afirmar que uno y otro dominio no posean, además, otros rasgos específicos y modos de operar distintivos.

Ahora bien, según el MIS, “las intuiciones morales vienen primero”, es decir, guían la conducta en el caso propio y evalúan moralmente la conducta de otros, mientras que las razones justifican, persuaden o revisan dichas intuiciones y juicios morales, pero llegan después

(Haidt 2012, 2013). De la exposición realizada sobre las variedades “mínimas” o implícitas de comprensión psicológica, que preceden, fundamentan y luego acompañan otras variedades más sofisticadas de LM, se sigue que también las intuiciones de naturaleza psicológica vienen primero. En tal sentido, las razones que podemos ofrecer acerca del comportamiento de otros, basadas en el papel que atribuimos a los EMs postulados, o bien no logran dar cuenta de la comprensión intuitiva que las precede, o bien en ocasiones la contradicen, aunque difícilmente la pueden corregir, o bien, de manera similar al caso moral, sólo logran racionalizarla. Es decir, también en el dominio de la cognición psicológica, las razones llegan después. Si es correcto afirmar que existe una semejanza estructural entre la cognición psicológica y la cognición moral, es decir, que ciertos tipos diferenciados de explicaciones, en uno y otro dominio, se apoyan en un marco teórico ampliamente consistente, es posible conjeturar que este también permitirá dar cuenta de las relaciones estructurales entre ambas.

### Referencias bibliográficas

- Apperly, I. A., Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, 116, 953-970.
- Apperly, I. A. (2011). *Mindreaders. The cognitive basis of “theory of mind”*. Hove: Psychology Press.
- Apperly, I.A. (2013). Can theory of mind grow up? Mindreading in adults, its implications for the development and neuroscience of mindreading. In S. Baron Cohen, H. Tager-Flusberg, M.V. Lombardo (Eds.), *Understanding Other Minds. Perspectives from the Developmental Social Neuroscience*, pp. 72-92. Oxford: Oxford University Press.
- Butterfill, S. A., Apperly, I. A. (2013). How to construct a minimal theory of mind, *Mind & Language*, 28, 606-637.
- Butterfill, S. A. (2013). Interacting mindreaders, *Philosophical Studies*, 165(3), 841-863.
- Brownstein, M. (2018). *The Implicit Mind. Cognitive Architecture, The Self, and Ethics*. Oxford: OUP.
- Bruce, V., Young, A. (2012). *Face perception*. New York: Psychology Press.
- Cameron, C. D., Payne, B. K., Knobe, J. (2010). Do theories of implicit race bias change moral judgments? *Social Justice Research*, 23, 272-289.
- Castro, V. F. (2017). The expressive function of folk psychology. *Unisinos Journal of Philosophy*, 18(1), 36-46.
- Cohnitz, D., Haukioja, J. (2015). Intuitions in Philosophical Semantics. *Erkenntnis*, 80(3), 617-641.

- Cushman, F., Young, L., Greene, J. D. (2010). Multi-system Moral Psychology, In Doris, J. M. and the Moral Psychology Research Group (Eds.), *The Moral Psychology Handbook*, pp. 47-71. Oxford: OUP.
- De Cruz, H. (2015). Where philosophical intuitions come from? *Australasian Journal of Philosophy*, 93(2), 233-249.
- Di Jaegher, H., Di Paolo, E., Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Science*, 14(10), 441-447.
- Frankish, K., Evans, J. (2009). The duality of mind: An historical perspective. In J. Evans, & K. Frankish (Eds.), *In Two minds: Dual processes and beyond*, pp. 1-29. Oxford: OUP.
- Frankish, K. (2010). Dual-process and dual-system theories of reasoning. *Philosophy Compass*, 5/10, 914-926.
- Gallagher, S., Hutto, D. (2008). Understanding Others through Primary Interaction. In J. Zlatev et al. (Eds.), *The Shared Mind: Perspectives on Intersubjectivity*, pp. 17-38. Amsterdam: John Benjamin Pub.
- Glöckner, A., Witteman, C. (2010). Beyond dual-process models: a categorization of processes underlying intuitive judgements and decision making. *Thinking and Reasoning*, vol. 16(1), 1-25.
- Gray, K., Young, L., Waytz, A. (2012). Mind Perception Is the Essence of Morality. *Psychological Inquiry*, 23(2), 101-124. <https://doi.org/10.1080/1047840X.2012.651387>
- Green, M. (2007). *Self-Expression*. Oxford: OUP.
- Greenwald, A. G., Lai, C. K. (2020). Implicit Social Cognition. *Annual Review of Psychology*, 71, 419-445.
- Gomila, A. (2001). La perspectiva de segunda persona: mecanismos mentales de la intersubjetividad. *Contrastes*, 6, 65-86.
- Gopnik, A., Meltzoff, A. (1997). *Words, Thoughts, and Theories*. Cambridge: MIT Press.
- Guglielmo, S., Monroe, A. E., Malle, B. F. (2009). At the Heart of Morality Lies Folk Psychology, *Inquiry*, 52(5), 449-466, <https://doi.org/10.1080/00201740903302600>
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108, 814-34.
- Haidt, J., Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, 133(4), 55-66.
- Haidt, J. (2007). The New Synthesis in Moral Psychology. *Science*, 316(5827), 998-1002.

- Haidt, J., Bjorklund, F. (2008). Social intuitionists answer six questions about moral psychology. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Volume 2: The Cognitive Science of Morality: Intuition and Diversity*, pp. 181-217. Cambridge, MA: MIT Press.
- Haidt, J., Kesebir, S. (2009). In the forest of value: Why moral intuitions are different from other kinds. In H. Plessner, C. Betsch, T. Betsch (Eds.) *A new look on intuition in judgment and decision making*, pp. 209-229. New York: Taylor & Francis Group.
- Haidt, J. (2012). *The righteous mind. Why good people are divided by politics and religion*. New York: First Vintage Book Ed.
- Haidt, J. (2013). Moral psychology for the twenty-first century. *Journal of Moral Education*, 42(3), 281-297. <https://doi.org/10.1080/03057240.2013.817327>
- Heyes, C. (2014). Submentalizing: I am not really reading your mind. *Perspectives on Psychological Science*, 9, 131-143.
- Holroyd J., Scaife, R., Stafford, T. (2017). What is implicit bias? *Philosophy Compass*, 12(10), e12437.
- Hume, D. (1739/1978). *A Treatise of Human Nature*. New York: Oxford University Press.
- Isern-Mas, C., Gomila, A. (2020). Naturalizing Darwall's Second Person Standpoint, *Integrative Psychological and Behavioral Science*, 54: 785-804.
- Kant, I. (1785/1959). *Foundations of the Metaphysics of Morals*. New York: Macmillan.
- Knobe, J. (2004). Intention, Intentional Action and Moral Considerations. *Analysis*, 64, 181-187.
- Knobe, J. (2005). Theory of Mind and Moral Cognition: Exploring the Connections. *Trends in Cognitive Science*, 9(8), 357-359.
- Knobe, J. (2006). The concept of intentional action: a case study in the uses of folk psychology. *Philosophical Studies*, 130, 203-231.
- Kovács, Á. M., Téglás, E., Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults, *Science*, 330, 1830-1834.
- Leslie, A. M. (1994). Pretending and believing: Issues in the theory of ToMM. *Cognition*, 50, 211-238.
- McGahhey, M., Van Leeuwen, N. (2018). Interpreting Intuitions. In J. Kirsch and P. Pedrini (Eds.), *Third Person, Self-Knowledge and Self-Interpretation, and Narrative*, pp. 73-98. Switzerland: Springer.
- Machery, E. (2016). De-Freuding implicit attitudes. In M. Brownstein, J. Saul (Eds.), *Implicit bias and philosophy, Vol. 1: Metaphysics and epistemology*, pp.104-129. Oxford: Oxford University Press.

- Malle, B. (2006). Intentionality, morality, and their relationship in human judgment. *Journal of Cognition and Culture*, 6, 87-113.
- Nichols, S., Stich, S. P. (2003). *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. Oxford: Oxford University Press.
- Payne, K. B., Cameron, C. D. (2010). Divided Minds, Divided Morals. How Implicit Social Cognition Underpins and Undermines our Sense of Social Justice. In B. Gawronski & B. Keith Payne (Eds.), *Handbook of Implicit Social Cognition: Measurement, Theory, and Applications*, pp. 445-462. New York: Guilford Press.
- Pérez, D. I., Gomila, A. (2022). *Social Cognition and the Second Person in Human Interaction*. New York: Routledge.
- Pettit, D., Knobe, J. (2009). The pervasive impact of moral judgment. *Mind & Language*, 24(5), 586-604.
- Pretz, J. E. (2008). Intuition vs Analysis: Strategy and experiences in complex everyday solving problems. *Memory & Cognition*, 36(3), 554-566.
- Pretz, J. E. (2011). Types of intuition: Inferential and holistic. In M. Sinclair (Ed.), *Handbook of intuition research*, pp. 17-27. Northampton, MA: Edward Elgar Publishers.
- Pust, J. (2017). Intuition, *Stanford Encyclopedia of philosophy*.
- Sauer, H. (2011). Social Intuitionism and the Psychology of Moral Reasoning. *Philosophy Compass*. 6/10, 708-72.
- Saunders, L. F. (2009). Reason and Intuition in the Moral Life: A Dual-Process Account of Moral Justification. In J. St. B. T. Evans, K. Frankish (Eds.), *Two Minds: Dual Processes and Beyond*, pp. 335-354. Oxford: Oxford University Press.
- Scotto, C. (2002). Interacción y atribución mental: la perspectiva de la segunda persona. *Análisis Filosófico*, 22(2), 135-151.
- Sinclair, M. (2010). Misconceptions about intuition. *Psychological Inquiry: An International Journal for the Advancement of the Psychological Theory*, 21(4), 378-386.
- Sinclair, M. (2011). An Integrated Framework of Intuition. In M. Sinclair (Ed.), *Handbook of Intuition Research*, pp. 3-16. Northampton, MA.: Edward Elgar Publ. Co.
- Sinnott-Armstrong, W., Young, L., Cushman, F. (2010). Moral Intuitions. In Doris, J.M. and the Moral Psychology Research Group, pp. 246-271. *The Moral Psychology Handbook*. Oxford: OUP.
- Schukraft, J. (2016). Carving Intuitions at its Joints, *Metaphilosophy*, 47(3), 1026-1068.
- Spaulding, S. (2018). *How We Understand Others: Philosophy and Social Cognition*. New York: Routledge.

- Spaulding, S. (2019). What is mindreading? *WIREs Cognitive Science*, e1593.
- Strack, F., Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, 8, 220-247.
- Tager-Flusberg, H., Sullivan, K. (2000). A componential view of theory of mind: Evidence from Williams syndrome. *Cognition*, 76, 59-89.
- Thompson, V.A., Prowser Turner, J. A., Pennycook, G. (2011). Intuition, Reason, and Metacognition. *Cognitive Psychology*, 63(3), 107-40.
- Uleman, J. S., Adil Saribay, S., Gonzalez, C. M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology*, 59, 329-360.
- Waytz, A., Gray, K., Epley, N., Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in cognitive sciences*, 14(8), 383-388.
- Young, L., Waytz, A. (2013). Mind attribution is for morality. In Baron-Cohen, S., Tager-Flusberg, H. y Lombardo, M. V., *Understanding Other Minds. Perspectives of Developmental Social Neuroscience*, pp. 93-103. Oxford, Oxford University Press.
- Zawidzki, T. (2013). *Mindshaping. A New Framework for Understanding Human Social Cognition*. Cambridge Mass: The MIT Press.
- Zebrowitz, L.A. (2011). Ecological and Social Approaches to Face Perception. In: A. Calder, G. Rhodes, M. Johnson, J. Haxby (Eds.), *Oxford Handbook of Face Perception*, pp. 31-50. Oxford: Oxford University Press.