

Temas de filosofía de la mente

Atribución psicológica

TOMÁS BALMACEDA y KARINA PEDACE **compiladores**

colaboradores

TOMÁS BALMACEDA, LUCAS BUCCI,
FEDERICO BURDMAN, ANTONI GOMILA,
DIEGO LAWLER, ANDREA F. MELAMED,
KARINA PEDACE, DIANA PÉREZ,
ALEJANDRO ZÁRATE.

Temas de filosofía de la mente

Temas de filosofía de la mente

Atribución psicológica

Tomás Balmaceda y Karina Pedace (compiladores)

SADAF

Temas de filosofía de la mente : atribución psicológica / Tomás Balmaceda ... [et al.] ;
compilado por Tomás Balmaceda ; Karina Pedace. - 1a ed. - Ciudad Autónoma de Buenos
Aires : SADAF, 2018.

236 p. ; 20 x 14 cm.

ISBN 978-987-46708-4-7

1. Análisis Filosófico. I. Balmaceda, Tomás II. Balmaceda, Tomás, comp. III. Pedace,
Karina, comp.
CDD 190

© 2018, por los textos: Tomás Balmaceda, Lucas Bucci, Federico Burdman,
Antoni Gomila, Diego Lawler, Andrea F. Melamed, Karina Pedace, Diana
Pérez, Alejandro Zárate

© 2018, por esta edición: SADAF

SADAF

www.sadaf.org.ar

Diseño de tapa: Iñaki Jankowski

Desarrollo y producción editorial: Recursos Editoriales

isbn: 978-987-46708-4-7

Queda prohibida la reproducción total o parcial de esta obra, por cualquier
medio o procedimiento, sin permiso previo del editor y/o autor.

Índice

Agradecimientos	8
Introducción	9
<i>Tomás Balmaceda y Karina Pedace</i>	
Corporalidad y atribución psicológica	15
<i>Tomás Balmaceda</i>	
¿Cognición social sin atribución psicológica?	41
Una propuesta enactiva	
<i>Federico Burdman</i>	
La atribución mental y la segunda persona	69
<i>Diana Pérez y Antoni Gomila</i>	
Atribución y teorías de emociones:	99
escenas de la vida cotidiana	
<i>Andrea F. Melamed</i>	
La atribución psicológica a personajes de ficción	127
<i>Lucas Bucci</i>	
Atribución de estados mentales conscientes:	151
cartesianismo vs. expresivismo	
<i>Alejandro Zárate</i>	
La atribución psicológica:	181
entre la primera y la tercera persona	
<i>Diego Lawler</i>	
La atribución psicológica	205
y la normatividad de lo mental	
<i>Karina Pedace</i>	
Sobre los autores	231

Agradecimientos

Este libro surge de las reuniones organizadas por el PICT 2013-1419, que tuvieron lugar en el Instituto de Investigaciones Filosóficas de SADAFA. Agradecemos a todos sus participantes y, muy especialmente, a aquellos que generosamente nos enviaron sus trabajos para que sean incluidos en este volumen.

El financiamiento de este libro fue posible gracias al PICT 2013-1419: “La atribución psicológica: perspectivas y problemas”.

Introducción

Tomás Balmaceda y Karina Pedace

La concepción de que los seres humanos somos “animales que piensan” recorre la filosofía desde sus mismos orígenes. Encontramos esta idea pensada y modelada de diferentes maneras, desde el *zoon politikón* aristotélico hasta la comunión cartesiana entre *res extensa* y *res cogitans*; pero no se detiene ahí, ya que es una concepción que aún sigue presente en varios ámbitos. Sin embargo, es interesante comprobar que esta definición fue recibida por varias tradiciones filosóficas con el acento puesto en los pensamientos, como si en lo mental se escondiese aquello que es relevante de analizar. Pero somos animales: somos esta carne y estos huesos. Somos animales con un cuerpo, con instintos, con necesidades y con pasiones. Y no sólo eso: experimentamos sensaciones y emociones que no siempre van de la mano con un juicio tranquilo y razonado; estamos todo el tiempo en relación con los demás, en interacciones cara a cara y dándole sentido entre todos a lo que vivimos; distinguimos entre nuestros propios estados y los de los demás pero, a pesar de nuestros esfuerzos, en ocasiones no podemos evitar que lo que sentimos aparezca en nuestras acciones.

Los seres humanos somos animales que bailamos, que sentimos la sangre hervir cuando presenciamos una injusticia y que damos pequeños saltos de alegría cuando nos dan una buena noticia. Somos animales que comemos aunque no sintamos hambre si alguien que nos quiere nos cocinó, que brindamos pensando en proyectos futuros que quizá nunca se concreten y que si hace falta podemos intentar fingir una sonrisa aunque tengamos el corazón roto. Lloramos en la soledad de nuestra casa leyendo una novela y compartimos lágrimas con desconocidos mientras pasan los créditos de una película en el cine. Nos quedamos despiertos hasta entrada la madrugada si un problema

nos quita el sueño o fantaseando con una vida diferente a la que tenemos. Cruzamos miradas con alguien en un bar y creemos encontrar allí una chispa compartida para empezar una aventura de a dos, estamos muy seguros de aquello en lo que creemos pero a veces un pequeño gesto hace trizas todas nuestras certezas... ¿cómo dar cuenta de este campo increíblemente vasto de las atribuciones psicológicas?

En el seno de la reflexión filosófica acerca de las atribuciones psicológicas que nos hacemos cotidianamente, el cognitivismo clásico u ortodoxo ha detentado una posición hegemónica. En la medida en que se lo presentó con frecuencia como atravesado por dos enfoques fundamentalmente diferentes (la “teoría-teoría” y la “teoría de la simulación”), se tornó un tanto difuso hasta qué punto ambos grupos de teorías abrazaban una serie de supuestos y definiciones filosóficamente fundamentales en su concepción misma del fenómeno a explicar. En particular, el debate clásico dentro del campo cognitivista se encaramó sobre alguna versión del supuesto de que lo que hacemos en nuestra práctica cotidiana es atribuir estados mentales a los demás para explicar y predecir su conducta, que es conceptualizada como opaca en un sentido fundamental. De manera concomitante, lo mental fue visto como algo a ser hipotetizado o desentrañado a partir de los signos observables disponibles.

Frente a esta posición del cognitivismo clásico u ortodoxo, en los últimos años se multiplicaron las voces críticas, que denunciaron las omisiones de las posturas tradicionales y sus defectos a la hora de concebir la atribución psicológica. Uno de estos defectos es el “olvido del cuerpo”, tal como señala **Tomás Balmaceda** en su capítulo, frente al interés casi exclusivo puesto en el ámbito de lo mental en las teorías clásicas sobre la atribución psicológica. Este desdén por lo corporal queda expuesto en la teoría de la teoría y la teoría de la simulación, dos posiciones que entraron en pugna a finales de la década del XX, y cuya disputa llega a un callejón sin salida en el arranque del nuevo siglo. Es allí cuando aparecen autores que se interesaron por la dimensión corporizada de la cognición. En particular, Balmaceda analiza la teoría de la interacción de Shaun Gallagher y la perspectiva de se-

gunda persona y traza líneas de vínculos y de tensiones con otras dimensiones involucradas en el rechazo al cognitivismo clásico, como el enactivismo, la tesis de la mente situada y la propuesta de la mente extendida.

Justamente una de estas alternativas a las posturas ortodoxas es el enactivismo, que es elucidado en esta compilación por **Federico Burdman**. En su trabajo, se consigna el profundo cambio filosófico que representa este nuevo tipo de abordaje, a partir de la presentación de la teoría sobre la construcción participativa de sentido, propuesta por Hanne De Jaegher y el argentino Ezequiel Di Paolo. Se trata de un estudio de la cognición a partir de un replanteo de nuestro modo de relacionarnos con los demás desde una perspectiva corporizada pero, a diferencia de lo expuesto por Balmaceda, con el foco puesto en la dinámica de la coordinación y la interacción social antes que en las estrategias de postulación y atribución de estados mentales más usuales en las opciones cognitivistas tradicionales. La presentación del capítulo no sólo muestra una útil génesis de las ideas de De Jaegher y Di Paolo, sino que también ofrece una visión crítica de las limitaciones y alcances del proyecto.

Quizá la propuesta más reciente y original para enfrentar las nociones del cognitivismo clásico en la atribución mental sea la perspectiva de segunda persona, que ha sido defendida por autores como Antoni Gomila, Carolina Scotto, Evan Thompson y Diana Pérez, entre otros. De hecho, **Pérez y Gomila** escriben en conjunto en este libro sobre estas ideas, aclarando las bases conceptuales sobre las que se fundan y situando su origen en el rechazo tanto a la teoría de tercera persona, la teoría de la teoría, y la de primera, la teoría de la simulación, que fueron abordadas críticamente por Balmaceda y Burdman en sus respectivos trabajos. En este caso, Gomila y Pérez pretenden en su capítulo ir más allá de esta disputa de perspectivas para sostener que las atribuciones mentales de segunda persona tienen una primacía ontogenética, filogenética y conceptual frente a las de primera y tercera persona; además de defender que la maestría de los conceptos psicológicos se adquiere en las interacciones de segunda persona.

Uno de los fundamentos sobre los que se basa el cognitivismo clásico u ortodoxo, en estrecha vinculación con el mencionado

desdén por el cuerpo, es el foco puesto en estados mentales como los deseos y las creencias, dejando de lado otras variedades mentales como las emociones. En las últimas tres décadas, sin embargo, varios autores se interesaron por su naturaleza, su historia filogenética y su vinculación con la cognición. Ese es el punto que trata **Andrea Melamed** en su contribución a este volumen, analizando cómo se da cuenta de la atribución emocional en las teorías cognitivas y no cognitivas de las emociones. Se trata de un debate de estricta actualidad de dos posiciones que suelen verse como incompatibles, pero que la autora sostiene que mantienen un espíritu compartido que podría servir para un abordaje más fructífero del tema.

Precisamente, parece que la adscripción de estados mentales tales como las emociones –junto con creencias y deseos– a personajes de ficción es un fenómeno crucial en nuestro consumo de películas, piezas teatrales, obras de ficción y todo tipo de obras de arte que conlleven una narrativa. Tal es el tópico que aborda **Lucas Bucci** en su contribución. A su juicio, sin tales atribuciones psicológicas se tornaría imposible entender de manera cabal una historia protagonizada por personajes. En particular, aborda el problema relativo a la recepción o contemplación de arte narrativo a partir de la llamada “paradoja de la ficción”, señalada por Radford, i.e. el fenómeno aparentemente paradójico de que los espectadores se emocionan con eventos que saben que son puramente ficticios. Este punto ha moldeado gran parte de la discusión sobre la recepción de obras de arte narrativo en los últimos cuarenta años, vertebrada por el interrogante acerca de si es plausible que los espectadores experimenten emociones reales por fenómenos que saben que son ficcionales. En el marco de este debate, Bucci revisa las posturas contrapuestas de Gregory Currie y Noel Carroll para esclarecer si en nuestras atribuciones psicológicas a personajes ficticios utilizamos o no los mismos mecanismos psicológicos que empleamos para tales adscripciones a personas en nuestro entorno cotidiano.

Por su parte, en nuestras interacciones diarias, se suele aceptar que al menos ciertas afirmaciones y pensamientos acerca de nuestros propios estados mentales son mucho más seguros y

confiables que las afirmaciones que podemos hacer acerca de los estados mentales de los demás. En su contribución, **Alejandro Zárate** indaga la tarea a la que se han dado muchos filósofos a la hora de intentar dar cuenta de este punto y de determinar si nuestra confianza en dichas afirmaciones está verdaderamente justificada y, en caso de ser así, cuál es su alcance y origen. Específicamente, confronta y pondera críticamente dos maneras distintas de dar cuenta de la presunta asimetría entre las atribuciones de estados conscientes en primera persona del singular (como “Tengo un dolor”) y en tercera persona (como “Él siente dolor”). Para ello, por un lado, presenta una perspectiva que denomina “cartesiana”, según la cual, la asimetría en cuestión se funda en una asimetría epistémica y explicita las dificultades a las que, en su opinión, se enfrenta un cartesiano contemporáneo como el filósofo David Chalmers. Por otro lado, ofrece una explicación diferente de esta asimetría desde una perspectiva “expresivista”, tal como la que presenta la filósofa Dorit Bar-On, opción que promete poder dar cuenta de las asimetrías mencionadas sin que se susciten los problemas que enfrentan los partidarios de la mente cartesiana.

En sintonía con el interés que suscita la atribución psicológica cifrada entre las perspectivas de la primera y la tercera persona, la contribución de **Diego Lawler** articula dos historias. Una de ellas abreva en una tradición cuyas raíces se remontan a Kant. El punto de partida de esta narración consiste en que los seres humanos nos autodeterminamos cognitiva y prácticamente como criaturas racionales. Esta historia enlaza, pues, el autoconocimiento con la agencia racional a partir de los trabajos de Shoemaker, Moran y Boyle. Desde esta perspectiva, que Lawler denomina de “primera persona”, se le confiere una importancia filosófica especial a la presencia de la asimetría entre la primera persona y la tercera persona. No sólo reflejaría un modo de conocimiento fundamentalmente diferente (una asimetría epistémica), sino que, además, exhibiría que hay una manera distintiva en que los estados mentales se presentan a los sujetos que los tienen y que su descripción es necesaria para entender la mente. Frente a ella, opone la que denomina como “perspectiva reductiva o de la tercera persona” y que

articula desde los trabajos de Ryle y Dennett. Se trata de una historia que no le otorga ninguna relevancia filosófica a la asimetría presente en la vida ordinaria cuando informamos sobre nuestra mente y la de los otros. A la vera de esta perspectiva, nuestros modos para saber cosas sobre nuestros deseos, creencias, intenciones, etc., son los mismos que empleamos para saber estas mismas cosas de otras personas. No habría, pues, una diferencia de clase, sino meramente de grado entre el autoconocimiento y el conocimiento de los otros. Finalmente, de lo que se trata es de contrastar las dos historias en cuestión con el propósito de mostrar la especificidad de la perspectiva de la primera persona.

Frente a la tradicional preeminencia que detentó la perspectiva de primera persona, a mediados de siglo XX tuvo lugar una inflexión con la irrupción en escena de la perspectiva del intérprete a la hora de elucidar la naturaleza de la atribución psicológica. En efecto, en las huellas de la querella que suscitó la aparición decimonónica de las “Ciencias del Espíritu” y a través de su impacto en los trabajos de Wittgenstein, Quine y Davidson –entre otros– tuvo lugar un giro en la concepción de lo mental. Precisamente, **Karina Pedace** aborda en su contribución el que denomina como “giro interpretativo” en la filosofía de la mente, como un paso conceptualmente indispensable para comprender de modo cabal el derrotero que lleva al encuentro de lo mental con la normatividad propia de la atribución psicológica. Argumenta que esta suerte de “giro copernicano” deja de lado la clásica concepción cartesiana de la mente como un espectáculo privado y accesible sólo al “ojo interior” de cada sujeto, i.e. aquella con la que se comprometió el cognitivismo clásico u ortodoxo. Tal concepción de lo mental resulta, en su opinión, desacreditada por el análisis de los supuestos necesarios de la interpretación y la comunicación intersubjetiva, entre los cuales la normatividad juega un papel central. En tal sentido, su trabajo trata de precisar de qué hablamos cuando hablamos de “normas de lo mental”, sus variedades y alcance, así como de mostrar algunos puentes entre la filosofía de la mente y la filosofía práctica, que serían posibilitados por la normatividad de la atribución psicológica.

Corporalidad y atribución psicológica

Tomás Balmaceda

*Nací con vos, te acompaño a todos lados,
yo te guío y te aliento y, aunque a veces te atormento,
si te faltase, amigo, no contarías más el cuento.
¿Quién soy?*

No resulta difícil descubrir el resultado de esta adivinanza infantil. Nuestro cuerpo nació con nosotros, nos acompaña a todos lados y es la fuente de guía, aliento y, en ocasiones, tormentos. Además, sin él no sería posible estar vivos. Sin embargo, sorprende que no siempre en las reflexiones filosóficas estos hechos que nos parecen evidentes hayan estado presentes. De hecho, la corporalidad es un dominio espinoso en la historia de la filosofía, que ha pasado por distintas fases a lo largo del tiempo. En este artículo quisiera realizar un breve recorrido por algunas de las ideas que considero relevantes y que serán útiles para mi objetivo principal, que es mostrar cómo cambió el rol del cuerpo en las teorías de atribución psicológica al pasar del marco del cognitivismo clásico u ortodoxo a ideas más recientes, en donde se lo rescató como un elemento central para entender nuestra cognición.

Para eso, en la sección 1 analizaré un punto de partida ineludible para cualquier reflexión acerca del rol del cuerpo en el ámbito filosófico: las ideas de René Descartes. Si bien, como señalaré, él no fue el primero en distinguir entre la mente y el cuerpo, su obra quedó ligada a una visión dualista de la realidad, un mote que en los últimos años fue puesto en duda, ya que su planteo es más complejo que el simple postulado de un fantasma en la máquina. Me parece relevante regresar sobre sus ideas y dejar planteados al-

gunos interrogantes sobre su posición con respecto al problema, ya que son temáticas que aparecerán en lo que sigue.

Luego hablaré en 2 sobre el “olvido del cuerpo en las teorías clásicas sobre la atribución de lo mental”. Es menester realizar varias aclaraciones. Prefiero hablar de “olvido” y no de “ocultamiento” porque no estoy del todo seguro de que el movimiento que realizó la corriente que denominaré el cognitivismo clásico haya sido deliberado, sino que puede ser adjudicado simplemente a un desdén no intencional. Por otro lado, al hablar de “teorías clásicas de la atribución de lo mental” intento dejar en claro que lo que realizo es un recorte muy preciso dentro del amplio campo de reflexión sobre el vínculo entre lo corporal y lo mental. Para quien esté familiarizado con una tradición continental contemporánea le sorprenderá descubrir que mientras que la obra de autores como Maurice Merleau-Ponty ya eran leídas y discutidas en muchos círculos, en el área analítica el interés seguía puesto en una razón incorpórea. En primer término en 2.1 explicaré a qué me refiero por cognitivismo clásico, un marco difuso de reflexión filosófica sin límites precisos pero que ofrece una forma de abordaje de la realidad, y en 2.2 me referiré específicamente a dos estrategias de atribución mental concebidas bajo su sombra, la teoría de la teoría y la teoría de la simulación.

En 3 me referiré al “regreso del cuerpo en las teorías sobre la atribución mental”, explicando lo que se suele llamar la dimensión corporizada de la cognición, un posición que aparece con autores que abandonan el cognitivismo clásico y se liberan de sus constricciones. El término, sin embargo, a veces se utiliza con liviandad y explicaré dos posibles sentidos, muy diferentes, de “embodiment”. Luego, en 3.1 explicaré dos maneras de entender la atribución psicológica bajo la tesis de la corporización, la teoría de la interacción y la perspectiva de segunda persona.

Finalmente, en 4, cerraré este capítulo esbozando algunas consecuencias de poner al cuerpo en el centro de la escena filosófica a la hora de hablar de la atribución mental y cuáles son sus posibles vínculos con otras dimensiones involucradas, como el enactivismo, la tesis de la mente situada y la propuesta de la mente extendida.

1. El legado cartesiano

Desde los inicios de la reflexión filosófica, los hombres y mujeres se han interesado por pensarse a sí mismos. El vínculo entre nuestros pensamientos y nuestro cuerpo fue de inmediato un motivo de debate. Uno de los ejemplos más claro son las ideas de Platón, en las que el cuerpo es un obstáculo para el acceso al verdadero conocimiento, al que sólo alcanza el alma intangible. El célebre juego de palabras de Platón entre *sôma* (“cuerpo”) y *sêma* (“tumba” y “signo”) (cfr. Fedro 65a, 66a-d; 79c, 82e-83c), ya muestra que para él el cuerpo es “un estorbo con el cual el alma batalla” (Fedro 67e), aunque ésta desea vivir de forma corporizada porque lo considera mejor (Broadie, 2001: 304). Así, pese a ser un obstáculo para el conocimiento, el cuerpo también es un vehículo necesario para satisfacer ciertos deseos, aunque a veces la unión entre alma y cuerpo genera aspectos irracionales en la primera (Fedro 248a).

A pesar de la larga tradición de abordajes y perspectivas en este tema, la tradición filosófica ubica a René Descartes como el representante icónico de la postulación dicotómica excluyente entre el cuerpo y el alma o espíritu, en donde esta última se encuentra por encima del primero en muchos aspectos. Como señalaré en esta sección, esta caracterización es simplista y, en ocasiones, parece reflejar más la recepción de las ideas cartesianas que lo que efectivamente el filósofo francés afirmó, ya que sus ideas han sido revisitadas con el tiempo y hoy existen numerosas voces que cuestionan una interpretación simple del asunto (cfr. Clarke, 2003). Pero repensar esta doctrina tiene un fuerte desafío por delante, ya que es la visión dominante, y que ha pregnado con fuerza en varios niveles, dentro y fuera de la filosofía.

Incluso con estos reparos, podemos asegurar que Descartes plantea un dualismo de sustancias, es decir, postula la existencia de dos dominios ontológicamente separados, uno que es mental y otro que es físico. El primero es el ámbito de aquello cuya propiedad es el pensar, la *res cogitans*; y el segundo tiene como propiedad la extensión, la *res extensa*. Aunque planteado así podría parecer sencillo, la relación entre estas dos sustancias es central

en toda la filosofía cartesiana y aparece tematizada en varias de sus obras.

En la segunda de sus *Meditaciones metafísicas* comprueba que él tiene un cuerpo pero rápidamente aclara que su identidad no se encuentra atada a él: “No soy esta reunión de miembros llamada cuerpo humano; no soy un aire sutil y penetrante, difundido por todos esos miembros” (*Meditaciones metafísicas*: II, 38). El cuerpo, como *res extensa*, es una suerte de máquina de huesos y carne, plausible de ser divisible. El espíritu, en cambio, es *res cogitans*, una sustancia indivisible y única: “hay gran diferencia entre el espíritu y el cuerpo; pues el cuerpo es siempre divisible por naturaleza y el espíritu es enteramente indivisible” (*Meditaciones metafísicas*: II, 38). El espíritu, entonces, es fundamentalmente distinta del cuerpo pero está unida a él. Es en el espíritu en donde reside mi esencia, no en mi cuerpo:

Como sé de cierto que existo, y, sin embargo, no advierto que convenga necesariamente a mi naturaleza o esencia otra cosa que ser cosa pensante, concluyo rectamente que mi esencia consiste sólo en ser una cosa que piensa, o una substancia cuya esencia o naturaleza toda consiste sólo en pensar. Y aunque acaso (o mejor, con toda seguridad, como diré en seguida) tengo un cuerpo al que estoy estrechamente unido, con todo, puesto que, por una parte, tengo una idea clara y distinta de mí mismo, en cuanto que yo soy sólo una cosa que piensa –y no extensa–, y, por otra parte, tengo una idea distinta del cuerpo, en cuanto que él es sólo una cosa extensa –y no pensante–, es cierto entonces que ese yo (es decir, mi alma, por la cual soy lo que soy), es enteramente distinto de mi cuerpo, y que puede existir sin él (*Meditaciones metafísicas*: II, 42).

Al igual que Platón, Descartes toma como un *factum* que espíritu y cuerpo se hallan unidos. Pero, a diferencia de lo que propone el autor de *República*, aquí no hay un entramado íntimo de dos sustancias que se entrecruzan, sino que para el francés están “unidos”, con una acción conjunta paralela en la que no hay fusión pero sí íntima conexión. La fusión platónica, que resolvería varios inconvenientes, no podría darse porque se trata aquí de dos sustancias radicalmente heterogéneas. Si

el espíritu no es extenso y la causalidad mecánica es función de la extensión, este no podría entrar en relaciones de causalidad mecánica. Este es considerado el origen –no cronológico porque, como ya mencioné, Descartes se inscribe en una larga tradición dualista que existe mucho antes que él y que aún continúa aunque no en ámbitos académicos– de lo que se conoce como “el problema mente cuerpo”, una temática fundacional de la Filosofía de la Mente y sobre la cual se ha escrito mucho: la polémica acerca de cómo se relacionan las propiedades mentales y las propiedades físicas.

El complejo vínculo entre las dos sustancias es problematizado en la célebre comparación con el piloto y su barco, que está tanto en la sexta de las *Meditaciones metafísicas*, en donde señala que:

La naturaleza me enseña también por medio de estos sentimientos de dolor, de hambre, de sed, etc., que yo no estoy alojado en mi cuerpo como un piloto en su navío, sino que estoy muy estrechamente unido y confundido y mezclado de tal modo que formo como un único todo con él (*Meditaciones metafísicas*: VI, 81).

Así como también en el *Discurso del método*, en donde aclara que:

No basta que [el alma] esté alojada en el cuerpo humano, como un piloto en su navío, a no ser acaso para mover sus miembros, sino que es necesario que esté junta y unida al cuerpo más estrechamente, para tener sentimientos y apetitos semejantes a los nuestros y componer así un hombre verdadero (34).

La unión espíritu-cuerpo no es una unión que se pueda explicar desde un punto de vista ontológico porque no puede constituir una tercera sustancia. Esto es así porque según su visión lo que nos permite distinguir algo como una sustancia es el hecho de que se le presenta a la mente como algo realmente distinto. Si la definición de sustancia es su concepción con claridad y distinción, el vínculo entre el espíritu y el cuerpo tiene que ser explicado en clave epistemológica. La cuestión ya fue objetada por los primeros lectores de estas obras (como los intercambios epistolares con Elizabeth, Arnould y Mesland) y

las dudas recorrieron los siglos y se consolidaron como una de las críticas más usuales a este modelo. Una de las respuestas que dio Descartes fue advertir que muchas de estas acusaciones cometían el error de comprender la interacción psico-física como si fuese similar a las relaciones físico-físico, ya que la causalidad mecánica no es la única variedad de causa posible (cfr. Richardson, 1982). Sin embargo, sus respuestas nunca fueron del todo claras o, al menos, no fueron tan contundentes como para dejar satisfechos a sus críticos. En esa línea, algunos intérpretes contemporáneos de Descartes han descartado este tipo de objeciones al considerar que bajo la óptica del francés su concepción de la causalidad no involucraría una afectación real entre entidades, sino que sería una suerte de concepción humeana *avant la lettre*, basada en la mera regularidad entre eventos. De este modo, la heterogeneidad deja de ser problemática (cfr. Loeb, 1981; O'Neill, 1987; Della Rocca, 2008). Pero esta salida no convenció a todos. Jaegwon Kim aseguró que ni siquiera una visión humeana podría salvar a Descartes, ya que el problema está radicado en la inespacialidad de la mente, lo que le impide cualquier tipo de vínculo con algo corporal (cfr. Kim, 2005). Pero estos debates aún no han logrado esmerilar la imagen con la que ha quedado asociado el dualismo de Descartes, que fue resumida por Ryle con la metáfora de un fantasma que habitaba y controlaba una máquina (cfr. Ryle, 1967), que borra la complejidad y sofisticación inherente en sus ideas.

Más allá del problema mente-cuerpo –que, como señalé, es una disputa muy transitada y con múltiples abordajes y soluciones en las últimas décadas– el dualismo cartesiano deja al ámbito de lo mental en un lugar privilegiado frente al cuerpo, una impronta que, mostraré en la próxima sección, seguirá vigente incluso cuando ya no haya corrientes que adhieran a un dualismo de sustancias como el cartesiano. Aunque sin el desprecio presente en el discurso platónico, aquí también el espíritu es el origen fiable del conocimiento, ya que el cuerpo puede mostrar cosas de una forma equivocada, como en el experimento mental del trozo de cera, y sólo es con el entendimiento como llego a conocer sus verdaderas cualidades. Esta supremacía queda también

patente en la regla XII de sus *Reglas para la dirección del espíritu*, cuando asegura que:

Es preciso servirse de todos los recursos del entendimiento, de la imaginación, de los sentidos y de la memoria: ya para intuir distintamente las proposiciones simples; ya para comparar debidamente lo que se busca con lo que se conoce, a fin de reconocerlo; ya para descubrir aquellas cosas que deben ser comparadas entre sí de modo que no se omita ningún elemento de la habilidad humana.

Se trata de la culminación del tratado, en donde aprovecha para volver a remarcar la distinción ontológica entre ambas sustancias y plantear la supremacía de la mente sobre el cuerpo. Un cuerpo del que, como explica en la tercera parte de *Discurso del método*, puede prescindir:

Examiné después atentamente lo que yo era, y viendo que podía fingir que no tenía cuerpo alguno y que no había mundo ni lugar alguno en el que yo me encontrase, pero que no podía fingir por ello que yo no fuese, sino al contrario, por lo mismo que pensaba en dudar de la verdad de las otras cosas, se seguía muy cierta y evidentemente que yo era, mientras que, con sólo dejar de pensar, aunque todo lo demás que había imaginado fuese verdad, no tenía ya razón alguna para creer que yo era, conocí por ello que yo era una sustancia cuya esencia y naturaleza toda es pensar, y que no necesita, para ser, de lugar alguno, ni depende de cosa alguna material; de suerte que este yo, es decir, el alma, por la cual yo soy lo que soy, es enteramente distinta del cuerpo y hasta más fácil de conocer que éste y, aunque el cuerpo no fuese, el alma no dejaría de ser cuanto es (24).

Queda entonces claro que aquí hay una primacía absoluta de mi espíritu, que bien utilizado me permite acceder a un conocimiento libre de errores. Y si bien Descartes no tematiza explícitamente en su obra la capacidad de atribución mental, creo que a partir de este análisis se pueden desprender dos cuestiones relacionadas. Por un lado, la actividad de entender e interpretar la conducta de los demás es, para quien la realiza, una facultad absolutamente mental que no necesita de manera esencial al cuer-

po para ser realizada. Por otro lado, lo que sucede en la mente del otro también es, en principio, separable de lo que ocurre en su cuerpo, ya que no hay fusión de ambas sustancias, sino en todo caso acción coordinada. Puedo utilizar los movimientos de esta suerte de máquina para inferir qué es lo que sucede en el espíritu. Es justamente esta intuición la que recorrerá muchas de las teorías sobre atribución psicológica a partir del siglo XXI.

2. El olvido del cuerpo en las teorías clásicas sobre la atribución mental

Aunque es un término que terminó abarcando propuestas que, vistas en detalle son muy diversas y en sentidos profundos incluso llegan en ocasiones a ser incompatibles, solemos referirnos al cognitivismo como la corriente de pensamiento que entre la tercera y la cuarta década del siglo XX desplazó al conductismo y ocupó el centro de la escena de la reflexión filosófica y psicológica. Luego de algunos años de desdén del ámbito de lo mental, que había considerado ajeno al alcance de lo científico, con el cognitivismo se consolidó la idea de que el pensamiento inteligente debía ser abordado como un proceso de cálculo formal, descriptible en términos de un conjunto finito de operaciones mecánicas comparables a las que puede instanciar una computadora (cfr. Burdman, 2016: cap. 1). Esta intuición original dio lugar tanto al estudio de la mente bajo el ramillete de disciplinas que suele ser denominada “ciencias cognitivas” como las bases de las tecnologías de la computación actuales.

Existen muchas maneras de caracterizar esta corriente. A los fines de este capítulo, consideraré al “cognitivismo clásico” como la postura que reúne a autores que adopten la metáfora computacional para dar cuenta del funcionamiento de la mente y de su vínculo con el cuerpo. Su formulación de esta metáfora es compleja y se gestó a lo largo de la década del 40 y el 50 a partir de tres tradicionales: la cibernética, la inteligencia artificial simbólica (IA) y la investigación con redes (cfr. Bechtel, Abrahamsen y Graham, 1999), pero podemos acordar que lo central de

la misma es que habilitó pensar a la cognición como procesos computacionales que manipulan representaciones internas de acuerdo a reglas. De este modo, el ámbito de lo psicológico deja de ser un reino inmaterial, intangible y absolutamente heterogéneo, como planteaba Descartes, y tampoco era más el misterio del que el científico debía abstenerse de pensar como creían los conductistas.

Gracias a la metáfora computacional, los sistemas que permiten los procesos psicológicos como sistemas implementados en última instancia en el cerebro, los vuelve susceptibles de estudio científico. El objeto de las ciencias cognitivas, entonces, es el plano computacional, que es diferente pero está relacionado con el resto del proceso. Un buen ejemplo de esto es el modelo que propuso para la visión el neurocientífico David Marr en la década del 80. En su modelo hay un nivel funcional, en donde se da el análisis funcional de la conducta del sistema; el algorítmico, correspondiente al diseño de los procesos computacionales propiamente dichos y el nivel de la implementación física de esos procesos computacionales. Cada nivel es “bastante independiente” del otro a la hora de ofrecer la explicación de un fenómeno:

Estos tres niveles se encuentran acoplados, pero no estrechamente. La elección de un algoritmo está influenciada, por ejemplo, por lo que tiene que hacer y por el hardware en el que debe correr. Pero hay un amplio rango de opciones disponibles en cada nivel, y la explicación en cada nivel presenta asuntos bastante independientes de los demás. Cada nivel de descripción tendrá su lugar en una eventual comprensión completa del procesamiento de la información perceptual, y por supuesto que están lógica y causalmente relacionados. Sin embargo, es importante notar que dado que los tres niveles no están estrechamente acoplados, algunos fenómenos pueden ser explicados haciendo referencia sólo a uno o dos de ellos (Marr, 1982: 25).

Al igual que lo que sucede con las computadoras, los procesos computacionales son formales, es decir que podrían instanciarse en principio por diferentes sistemas físicos. Como bien señala Burdman, esto determinó que para el cognitivismo ortodoxo no hayan constituido un tema de interés “los detalles del modo en

que se hallaban implementados los sistemas computacionales” (2016: 19-20). Así, el cuerpo queda relegado al plano de la implementación física de Marr, pudiendo ser suplementado por cualquier otra entidad física que cumpla los requisitos necesarios para su instanciación.

Otra manera de entender cómo funciona el cuerpo en el esquema del cognitivismo clásico es con la imagen de la cognición como sándwich, que propone Susan Hurley:

Una visión de la mente sostenida con mucha frecuencia tiene dos componentes principales. El primero es una visión de la percepción y la acción como separadas la una de la otra y como periféricas. El segundo es una visión del pensamiento o la cognición como el núcleo central de la mente, al menos para las criaturas con habilidades cognitivas. La cognición es virtualmente central, aun si la mera implementación de los procesos cognitivos fuese distribuida. La mente se descompone verticalmente en módulos: la cognición hace de interfaz entre la percepción y la acción. Percepción y acción no sólo están separadas la una de la otra, sino también de los procesos superiores de la cognición. La mente es un tipo de sándwich, y la cognición es el relleno (Hurley, 1998: 401).

En esta descripción comprobamos que los procesos cognitivos son independientes de la percepción y la acción y que su función es mediar entre ellos. En esos extremos está el cuerpo, que se encarga de brindar información sobre el mundo exterior y de actuar de acuerdo a lo que dispone la mente. La mente, por su parte, está lejos del mundo, en un espacio interior en el que procesa los datos de la percepción como un *input*, a partir del cual obtiene la acción como *output*. Los procesos cognitivos son centrales y se los entiende como crucialmente distintos de los procesamiento sensoriomotores, ya que su función es la manipulación y transformación de estructuras con contenido representacional.

2.1. Teoría de la teoría y teoría de la simulación

En el marco del cognitivismo clásico aparecieron y crecieron muchas áreas de estudio y de interés. Una de ellas, vinculada con la Filosofía de la Mente, es la que se ocupa de las habilidades

involucradas en nuestros vínculos cotidianos con los demás, en donde asumimos que el resto de las personas cuentan con pensamientos parecidos a los nuestros y que sus conductas están motivadas por estos pensamientos. La manera en que se producía ese proceso de atribución psicológica generó diferentes posturas. Existen dos que revisten especial importancia y que fueron, durante el último cuarto del siglo XX, las únicas opciones en juego. En esta sección, intentaré exponerlas y señalar cuál es el rol otorgado al cuerpo.

Al área de este debate se la suele conocer como Psicología de Sentido Común y su consolidación se dio a comienzos de la década del 80, poco tiempo después de que se publicara el trabajo considerado seminal, un escrito de los primatólogos Premack y Woodruff en que abordan el fenómeno y acuñan el término “teoría de la mente” (Premack y Woodruff, 1978). Esa investigación generó un inmediato y fértil debate (Dennett, 1978a; Bennett, 1978; Harman, 1978), del que surgió un paradigma experimental, lo que se conoce como “el test de la falsa creencia”, que ya es utilizado pocos años después (Wimmer y Perner, 1983). En los años siguientes aparecen una serie de libros y artículos en los que podemos encontrar esbozados los lineamientos básicos de las opciones teóricas que dominarían los debates en el área, la llamada teoría de la teoría (Gopnik y Astington, 1987; Wellman y Bartsch, 1988) y la teoría de la simulación (Gordon, 1986; Heal, 1986; Goldman, 1989). Puede considerarse que el marco clásico del debate en el área queda ya definido en sus puntos esenciales con la publicación, en 1992, de un número doble de la revista *Mind and Language* en el que los principales filósofos y científicos de cada corriente expusieron sus ideas y establecieron los lineamientos principales sobre los que se desarrollarían las discusiones posteriores.

El núcleo compartido por aquellos autores que se identifican con la teoría de la teoría es que las prácticas de Psicología de Sentido Común están basadas en nuestro manejo de una teoría acerca de la estructura y funcionamiento de la mente. Existen diferentes opiniones acerca de cómo es esta teoría: si innata, modularizada, aprendida individualmente o adquirida

a través de un proceso de inculturación, sólo por nombrar algunos puntos en discusión. Los teóricos de la Simulación, por su parte, creen que en estas prácticas la simulación, entendida como un proceso de ponerse en el lugar del otro, ocupa un rol central. Como señalé más arriba, se trata de posiciones antagónicas y excluyentes, que entraron en debate por varios años pero que, eventualmente, también dieron luz a “modelos híbridos”, movidos por la premisa de conjugar lo mejor de ambos grupos teóricos (ver, por ejemplo, Carruthers, 1996 y Heal, 1996). Sin embargo, y sin desconocer estos desarrollos, puede decirse sin dudas que fue la oposición dicotómica entre teoría de la teoría y teoría de la simulación la que definió el espacio del debate teórico entre 1978 y 1995.

Pero por encima de este enfrentamiento, existen supuesto comunes que comparten ambas estrategias de atribución. Argumenté a favor de estas ideas en otros artículos (Balmaceda, 2014; Balmaceda y Burdman, 2015), pero a los fines de este artículo, en el que recorro el rol del cuerpo a lo largo de algunas posturas filosóficas, será suficiente con que señale tres. El primero es que ambas posiciones coinciden en que la conducta puede explicarse y predecirse mediante la atribución de estados mentales. Según ellos, se trata de “un campo específico de investigación cuyo objetivo es proveer una explicación a la habilidad –que podría no ser exclusiva de los seres humanos– de explicar y predecir acciones, tanto para uno mismo como para otros agentes inteligentes” (Carruthers y Smith, 1996: 1) y que “las nociones mentales son inferencias mediatas que permiten entrelazar y relacionar conductas y situaciones del mundo, dar cuenta de las relaciones de aquéllas y éstas y predecir las conductas” (Rivière, 2000: 276). En la introducción a una compilación de trabajos sobre el tema, los editores Martin Davis y Tony Stone aseguraron:

Un hecho impactante sobre los seres humanos es que, con poco o ningún entrenamiento formal, desarrollan la capacidad de desplegar conceptos psicológicos tales como creencia y deseo en predicciones y explicaciones de las acciones y los estados mentales de otros miembros de la especie. Estas predicciones y explica-

ciones se supone que “racionalizan” las acciones de los sujetos o estados mentales (Davies y Stone, 1995: p. 2).

De este modo, explicación y predicción de la conducta se vuelven los objetivos de la atribución psicológica, que siempre manipula estados mentales, entendidos como entidades inobservables. Esto constituye el segundo supuesto que comparten y es una consecuencia de haber adoptado el cognitivismo clásico como marco. Bajo esta concepción, los estados mentales intencionales no están disponibles para la percepción, sino que son entidades que se pueden inferir a partir de un estado de cosas en el mundo. Para los teóricos de la teoría, la inaccesibilidad de estos estados vuelve necesario postular una manera de poder extraer esos estados de la conducta observable:

Al igual que la teoría genética opera a un nivel más profundo que las generalizaciones fenotípicas de las generalizaciones sobre la herencia, la Psicología de Sentido Común introduce el pensar acerca de estados representacionales causalmente activos e internos (Botterill, 1996: 108).

Pero este supuesto es compartido por todos los modelos vistos. Baron-Cohen comienza su libro *Mindblindness* dando por sentado que los estados mentales son entes inobservables que están a la base de cualquier conducta:

Imagina lo que sería tu mundo si pudieses estar consciente de las cosas físicas pero ciego a la existencia de cosas mentales. Me refiero, por supuesto, a estar ciego a cosas como pensamientos, creencias, conocimiento, deseos e intenciones, que para la mayoría de nosotros subyacen a la conducta (Baron-Cohen, 1995: 1).

No existe, sin embargo, acceso directo a los estados mentales que expliquen nuestras acciones y las demás y que nos permitirían también predecirlas. En el artículo con el que introduce –de manera contemporánea con Gordon– la simulación como mecanismo central en Psicología de sentido Común, Heal parte de este supuesto:

Solemos decir que vemos a las personas como vemos estrellas, nubes o formaciones geológicas. Las personas son sólo objetos complejos en nuestros medios cuyo comportamiento deseamos anticipar pero cuyas causas internas no podemos percibir. Procedemos, entonces, a observar las complejidades de su conducta externa y formular algunas hipótesis acerca de cómo su interior está estructurado (Heal, 1986: 135).

Esto será lo que lleve a los autores a construir las distintas variantes sobre cómo es la adscripción de estos estados mentales. Por el lado de teoría de la teoría, la atribución de estados mentales se explica con el despliegue del conocimiento teórico con el que cuentan los sujetos normales, mientras que los teóricos de la simulación postulan que este proceso se lleva adelante poniéndonos en el lugar de quien queremos explicar o predecir. Aunque el proceso de este “ponerse en el lugar del otro” varía de acuerdo a cada autor, una manera de pensarlo es que la adscripción de estados mentales es alimentar con estados mentales fingidos en nuestro sistema de toma de decisiones, que es puesto a trabajar fuera de línea. Este resultado no llega a efectivizarse, sino que es utilizado para entender o predecir. En este sentido es una “simulación”.

Tanto en el planteo de teoría de la teoría como en el de la teoría de la simulación, una vez más comprobamos que el ámbito de lo corporal queda relegado a un papel marginal, sirviendo sólo de dominio en el cual encontramos “pistas” de lo que le está sucediendo al otro, pero que deben aún ser captadas, procesadas e interpretadas antes de poder entender qué es lo que le sucede. Hay una mediación a la hora de la adjudicación de los estados mentales que es una consecuencia esperable de estar en el marco del cognitivismo clásico: todo el trabajo lo realiza una mente que no se encuentra en el mundo exterior, sino que toma de él la información y la convierte en estructuras no observacionales que puede manipular y transformar. Sin esta clase de atribución mentalista, cuya habilidad se alcanza entre los tres años y los cinco, todo episodio conductual que percibimos resultaría perfectamente opaco.

3. El regreso del cuerpo en las teorías sobre la atribución mental

Luego de años de dominio del cognitivismo clásico en los trabajos vinculados a la Filosofía Analítica, desde la década del 90 comenzó a gestarse un movimiento de signo contrario, que buscó poner en tela de juicio muchos de los presupuestos de ese marco. Una de sus principales banderas era volver a poner en el centro de escena al ámbito de lo corporal. No se trató de un movimiento ordenado y sistemático, pero no hay dudas de que respondió a cuestionamientos que incomodaban a varios autores. Como sucede en estos casos, es difícil hablar de fechas precisas y cualquier cronología cometería injusticias porque estos “cambios de clima” se van gestando lentamente y siempre existen pioneros que adelantan estas movidas y autores más rezagados que insisten con viejas fórmulas incluso cuando sus colegas ya han aceptado el cambio.

Es por eso que en esta sección analizaré las posturas dedicadas a la atribución mental que nacieron bajo el paraguas del post-Cognitivism. Presentaré la teoría de la interacción de Shaun Gallagher y la perspectiva de segunda persona, defendida por Toni Gomila, Carolina Scotto, Evan Thompson y Diana Pérez entre otros. Como Gomila y Pérez escriben en este mismo libro sobre estas ideas, me dedicaré a relevar sus lineamientos principales haciendo hincapié en un trabajo de Scotto que adelantó algunas cuestiones que hoy se discuten.

Pero antes me gustaría hacer una salvedad acerca del alcance de este “regreso del cuerpo”. En la bibliografía tradicional, este interés por incluir a la corporalidad en la reflexión sobre la cognición es conocida como “la tesis de la corporalidad” o la “dimensión corporizada”. En ocasiones, es vista como una faceta de un abordaje mucho más general que incluye las ideas sobre mente situada, enactiva y extendida. En la sección 4 me encargaré de señalar algunos posibles vínculos en esta línea pero quisiera mencionar cierta complejidad que esconde esta tesis, ya que suele englobar posiciones muy diferentes que pueden confundirse entre sí.

Un autor que tematizó bien las diferencias entre las posturas rotuladas como de “cognición corporizada” fue Andy Clark, quien postula tensiones difíciles de resolver entre dos concepciones predominantes acerca del rol del cuerpo:

Propongo subrayar lo que a mí me parecen dos tesis centrales diferentes que organizan y dan forma al cúmulo de la bibliografía reciente.

La primera tesis es que aspectos del cuerpo y del mundo pueden, en ocasiones, ser partes de mecanismos más extensos cuyos estados y perfil operativo determinan (o, puesto en términos mínimos, ayudan a determinar) nuestros estados mentales y propiedades. Llamaré a esto una “historia de mecanismos amplios”. La segunda tesis es que detalles específicos de la corporización humana hacen una contribución especial e [...] ineludible a nuestros estados mentales y propiedades. Llamaré a esto una “historia de contribución especial” (Clark, 2008: 39).

Así, se postula un rol causal del cuerpo en la cognición en la primera postura y un rol constitutivo del mismo en la segunda. Según la visión del autor, esto genera, respectivamente, una reflexión que es una mera extensión de la perspectiva funcionalista acerca de lo mental o un “cuerpo-centrismo de acuerdo al cual la presencia de una mente como la humana depende de modo bastante directo de la posesión de un cuerpo como el humano” (Clark, 2008: 40). Para el primero, se presenta al cuerpo como intrínsecamente especial. El cuerpo con un rol único, no trivial y no computable. En esta contraposición, escondido bajo el uso de una terminología común y confusa existe

una tensión entre el viejo y querido funcionalismo, en versión extendida, amigable con la razón situada, y fundamentalmente más carnal: la noción que características del cuerpo realizan una contribución especial y en algún sentido no-negociable a la mente y la mentalidad (2008: p. 51).

Un funcionalismo que coloca al cuerpo en pie de igualdad con el cerebro y el mundo. Lo que consideramos “mente” es el resultado del equilibrio de estos tres factores. Un todo funcional

en el que el cuerpo juega un rol importante pero compartido y computable.

Si bien Clark boga por esta clase de funcionalismo, que se puede encontrar ejemplificado en su propuesta de una dimensión extendida de lo cognitivo, el análisis que sugiere es útil para ilustrar las diferencias internas de la corriente corporizada, que dista de ser homogénea.

3.1. Teoría de la interacción y Perspectiva de la Segunda Persona

Luego de casi tres décadas de la disputa entre la posición de la teoría de la teoría y la teoría de la simulación que mencioné en 2.2, en el inicio del nuevo siglo el campo de la discusión por las estrategias de atribución mental comenzó a ver nuevas contribuciones que dejaban de lado muchos de los supuestos que imponía el cognitivismo clásico. Dentro de estos modelos, me gustaría destacar aquí dos, en los que queda claro el papel central que desempeña el cuerpo.

Uno de los autores que más tiempo y esfuerzo le ha dedicado a erosionar y socavar las bases cognitivistas clásicas es el filósofo Shaun Gallagher, quien busca refundar los estudios sobre Psicología de Sentido Común y atribución mental destacando los errores de la teoría de la teoría y la teoría de la simulación (cfr. Gallagher, 2002, 2006, 2007 y 2010) y poniendo el foco tanto en el rol del cuerpo como en los aportes que la fenomenología continental puede hacer al modo en que los seres humanos nos reconocemos y entendemos cotidianamente como sujetos con mentes. La tesis principal de Gallagher es que, en oposición con lo que sostuvo durante años en el área, es el cuerpo el que moldea a la mente, ya que las capacidades de percepción, cognición y la conducta son moldeadas por los movimientos corporales. Se trata de una tesis fuerte que defiende que desde el momento mismo del nacimiento hay una influencia del cuerpo sobre la manera en que captamos el mundo y lo entendemos (incluso los movimientos prenatales están organizados de acuerdo al cuerpo humano, con registros propioceptivos y multimodales por parte del sujeto) y se extiende a lo largo de toda la vida.

En ese sentido, cuando abrimos por primera vez nuestros ojos no sólo podemos ver, sino que nuestra visión –aún siendo imperfecta– ya se encuentra en sintonía con aquellas formas que se parecen a nuestras formas: vemos nuestras posibilidades en la cara de los demás (Gallagher, 2005: 1).

Uno de los puntos destacados de las ideas de Gallagher es borrar la separación entre cuerpo y mente, sugiriendo que “el movimiento corporal prefigura las líneas de la intencionalidad y los gestos, mientras que el contorno de la cognición social y el cuerpo moldean a la mente” (2005: 3). Esto lo lleva a postular que a la hora de las atribuciones mentales se prescinda de cualquier elemento teórico o simulación, sino que nuestro vínculo con los demás debe entenderse en términos de una “práctica corporizada”. Esto no significa que las personas no desarrollen eventualmente habilidades teóricas o de simulación ni que jamás echen mano a estos recursos, sino que simplemente no constituyen la manera habitual en la que nos relacionamos. Las inferencias a partir de generalizaciones teóricas y las simulaciones se utilizan de manera ocasional y sólo pueden aplicarse a un conjunto muy acotado de experiencias.

Gallagher bautizó a su modelo teoría de la interacción y se fundamenta en el supuesto de que a la interacción comunicativa se la comprende como alcanzable en la misma acción de la comunicación. Existe un movimiento expresivo en el habla y gesto que determina en sí mismo la interacción y por lo tanto ya no es necesario para comprender teorizar sobre una creencia no visible o realizar una lectura de mente en base a recreaciones imaginativas. En la mayoría de las situaciones intersubjetivas tenemos un conocimiento directo de las intenciones de la otra persona porque sus intenciones están explícitamente plasmadas en sus acciones corporizadas y se espejan en nuestras propias capacidades para la acción. Esta comprensión no requiere la postulación de ninguna creencia o deseo que estén escondidos de la mente del otro, ya que todo se refleja directamente en la manera de actuar. Su modelo, al menos en la formulación más completa que tenemos, se compone de tres componentes centrales, la intersubjetividad primaria, la intersubjetividad secundaria y la compe-

tencia narrativa. Se trata de etapas que van apareciendo desde el momento mismo del nacimiento y hasta la adultez pero que no reemplazan unas a otras, sino que una vez que se hacen presentes en el desarrollo de un sujeto los acompañan el resto de sus vidas.

La segunda de las posturas sobre atribución psicológica surge del post-cognitivismo clásico y que pone el acento en la corporalidad es la perspectiva de segunda persona. Es un enfoque que toma de la gramática de los pronombres personales un elemento que hasta ahora no había sido tenido en cuenta en las posturas del cognitivismo clásico. Mientras que la teoría de la teoría se sitúa desde una perspectiva de tercera persona, una posición meramente de observación y sin participación, la teoría de la simulación adopta el punto de vista de la primera persona, desde el yo que realiza las simulaciones mentales. La segunda persona, en cambio, es la caracterización gramatical con la que se hace referencia a un otro que está interviniendo en una acción conmigo. Este libro incluye un completo artículo de dos de los autores que están trabajando en el área, Antoni Gomila y Diana Pérez. Sin embargo, no quiero dejar de delinear sus fundamentos básicos aquí, enfocándome en uno de los primeros trabajos que tematizó esta incipiente corriente, “Interacción y atribución mental: La perspectiva de la segunda persona”, publicado en 2002. En la visión de la filósofa argentina, el mecanismo de adscripción de estados mentales debe ser entendida como “un conjunto de habilidades o una competencia compleja para la comprensión recíproca, cuyo desarrollo y expresión se da en contextos interactivos, es decir, a la vez públicos, sociales y prácticos y cuyos propósitos, dependiendo de esos contextos, son evaluativos” (cfr. Scotto, 2002: 140).

Se desprenden de esto tres características de la atribución psicológica: su manifestación en conductas públicas, su desarrollo como habilidad y su esencia indubitavelmente social. En cuanto al primer rasgo, la idea es que en esta interacción aparecen componentes públicos o componentes mentales con expresiones públicas, como los gestos del rostro y las posturas corporales. Esto no sucede con la atribución tal es caracterizada por teoría de la teoría y teoría de la simulación, en donde tanto la inferencia a

partir de generalizaciones teóricas como simulaciones mentales se realiza mediante mecanismos internos que necesitan de entidades mentales no observables. La segunda nota relevante es que la estrategia de atribución desde una perspectiva de segunda persona es entendida como una práctica y no como una teorización o la aplicación de una analogía inferencial a partir del autoconocimiento. Finalmente, la Perspectiva de la Segunda Persona tiene una impronta social, porque se despliega en la interacción que ya está presente en las relaciones más comunes. La atribución intencional bajo esta óptica es recíproca, inmediata, dinámica y situada. Estos rasgos precisan algunas breves aclaraciones. La reciprocidad postulada no es normativa ni tiene un sesgo moral, sino que es la primitiva sensibilidad al estímulo, una forma de “impacto causal intersubjetivo” en palabras de Radu Bogdan (Scotto, 2002: 143), que permite dar respuestas adecuadas de acuerdo a las acciones que las ocasionan, posibilitando sincronizar los intercambios sociales (cfr. Hatfield, Cacioppo y Rapson, 1992). Son vínculos que mantienen los sujetos entre sí, en sus relaciones específicas. Y por esto mismo los intercambios son inmediatos, en tanto no reflexivos y en la mayoría de los casos no controlables. Los sujetos perciben e identifican lo que le está sucediendo al otro sin necesidad de una mediación con hipótesis, mecanismos de simulación o generalizaciones teóricas. Como la atribución desde la segunda persona se da en estas interacciones que son tan variadas, su marca es que son variables y siempre con una comprensión situada, es decir, necesita ser flexible a las variaciones contextuales, a la información disponible y a la conducta de los demás.

Así, existen situaciones muy precisas que son iluminadas y explicadas claramente por este enfoque. La perspectiva de segunda persona ofrece la capacidad de involucrarse con otros agentes en patrones de interacción intersubjetivos, con un conocimiento implícito de carácter práctico de las configuraciones expresivas de aquellos con quienes interactuamos y viceversa. Es una estrategia atributiva que se pone en juego en las situaciones de interacción cara a cara, aunque puede activarse por condiciones artificiales como las del cine y el teatro. Se basa en aspectos ex-

presivos como la posición corporal, la orientación, el tono de voz, la configuración facial, etc. Estos aspectos son percibidos directamente como significativas, como parte constitutiva de la emoción que se adscribe y no como síntomas que hay que interpretar y cuyo resultado no se formula conscientemente ni requiere una metarrepresentación conceptual, sino que constituye la base de una reacción correspondiente. Pero sin dudas los casos específicos de perspectiva de segunda persona son los casos de intersubjetividad recíproca, que están presente desde muy temprano en la vida de los sujetos y que se mantienen a lo largo de su adultez, con mayor sofisticación. Se trata de atribuciones con un fuerte carácter práctico, porque orientan la acción, del mismo modo involuntario en que se producen las expresiones emocionales. Esto explica la dimensión comunicativa de la expresividad y le da un sentido evolutivo a la reacción emocional.

Esta perspectiva también cuestiona y rechaza el modo en que se caracteriza a los estados mentales en el cognitivismo clásico, que ya no son privados, internos e inobservables, sino que ahora se los ve con una fuerte carga expresiva, a través de configuraciones y relaciones. Los estados mentales que captamos a través de la segunda persona son expresivos, con configuraciones corporales específicas, con un papel protagonista del rostro: “vemos” que alguien está contento, molesto, triste o eufórico. Así, las situaciones en el ámbito emocional son clave pero esta postura va más allá de este marco e incluye en las formas de interacción comunicativa pre intencional, en la que se detecta una atribución implícita de estados mentales, concebidos en términos expresivos, públicos.

4. La tesis de la mente corporizada y las otras dimensiones de la cognición

Como una suerte de breve epílogo, quisiera consignar en dónde pueden situarse estas ideas sobre corporización y atribución psicológica. Existe una bibliografía creciente que suele referirse a las nuevas dimensiones de la cognición como “las cuatro E”, en

referencia a los términos con los que son designadas en inglés, *embodied*, *embedded*, *enactive* y *extended*. Tomadas en tándem, estas reflexiones buscan ofrecer alternativas al cognitivismo clásico. Pero existen muchas confusiones e imprecisiones en la definición y alcance de estas dimensiones, una deuda que motivó algunos de los artículos de este libro. Se trata de tesis independientes entre sí y que deben ser tratadas como tales, pero se pueden trazar algunas relaciones. Me gustaría concluir este capítulo esbozando estas ideas.

Por un lado, una cognición corporizada está vinculada con la dimensión situada porque el medio “entra en juego” gracias a las capacidades sensoriomotoras, como la visión, el reconocimiento de patrones y la manipulación de objetos (cfr. Robbins y Aydede, 2009). En cuanto a las ideas enactivistas, es claro que la percepción está esencialmente corporeizada, en el sentido de que las categorías y la estructura de nuestra percepción y cognición están constreñidas y moldeadas por el tipo de agentes que somos. Podemos percibir y entender el mundo porque este está dispuesto de tal manera que es percible por nosotros, lo que a la vez señala que para el enactivista, la percepción y la cognición están situadas al menos en el sentido de que la actividad que es constitutiva de la percepción y la cognición depende de la existencia de un ambiente en el cual estas se den. Finalmente, plantear la posibilidad de una cognición extendida es hacer partícipe de manera esencial de nuestro cuerpo en los procesos que están involucrados en el pensamiento.

Bibliografía

- Balmaceda, T. (2014), “Mente, cuerpo y sociedad. Una mirada desde la filosofía analítica”, en: *Revista de Psicoanálisis*, Tomo LXXI, N° 4.
- Balmaceda, T. y F. Burdman (2015), “Precisiones en torno al debate entre las posiciones tradicionales y los nuevos enfoques por la Psicología de Sentido Común”, en: Skidelsky, L., D. Pérez y D. Lawler (comps.), *Temas de filosofía de la psicología*, Buenos Aires: Eudeba.
- Baron-Cohen, S. (1995), *Mindblindness: An Essay on Autism and Theory of Mind*, Cambridge: MIT Press.
- Bechtel, W., A. Abrahamsen y G. Graham (1999), “The life of cognitive science”, en: Bechtel, W. y G. Graham (eds), *A companion to cognitive science*, Oxford: Blackwell, pp. 1-104.
- Bennett, J. (1978), “Commentary on three papers about animal cognition”, *Behavioral and Brain Sciences*, 4, pp. 556-560.
- Botterill, G. (1996), “Folk Psychology and Theoretical Status”, en: Carruthers, P. y P. Smith, *Theory of theories of mind*, Cambridge University Press.
- Broadie, S. (2001), “Soul and body in Plato and Descartes”, *Proceedings for the Aristotelian Society*, N° 101, pp. 295-308
- Burdman, F. (2016), “El post-cognitivismo en cuestión: extensión, corporización y enactivismo”, *Principia*, Vol. 19, N° 3, pp. 475-495.
- Carruthers, P. (1996), *Language, Thought and Consciousness: an essay in philosophical psychology*, Cambridge: Cambridge University Press.
- Carruthers, P. y P. Smith (eds.) (1996), *Theories of Theories of Mind*, Cambridge: Cambridge University Press.
- Clark, A. (2008), *Supersizing the Mind. Embodiment, Action, and Cognitive Extension*, Oxford University Press.
- Clarke, D. M. (2003), *Descartes's Theory of Mind*, Oxford: Clarendon Press.
- Davies, M. y T. Stone (eds.) (1995), *Folk psychology: the theory of mind debate*, Oxford: Blackwell.
- Della Rocca, M. (2008), “Causation without Intelligibility and Causation without God in Descartes”, en: Broughton, J. y J. Carrierio (eds.), *A Companion to Descartes*, Oxford: Blackwell.

- Dennett, D. (1978), "Beliefs about beliefs", *Behavioral and Brain Sciences*, 4, pp. 568-570.
- . (1978a) *Brainstorms: philosophical essays on mind and psychology*, Brighton, Hamester Press.
- Descartes, R., (1973). *Meditaciones metafísicas*, México: Aguilar.
- . (1974), *Reglas para la dirección de la mente*, Madrid: Tecnos.
- . (2003), *Discurso del método*, Madrid: Tecnos.
- Gallagher, S. (2002), "The self: Philosophical problems", en: *Encyclopedia of Cognitive Science*, London: Macmillan.
- . (2005). *How the Body Shapes the Mind*, Oxford: Oxford University Press.
- . (2006), "The Narrative Alternative to Theory of Mind", en: Menary; R. (ed.), *Radical Enactivism: Intentionality, Phenomenology, and Narrative*, Amsterdam: John Benjamins, pp. 223-229.
- . (2007), "Simulation trouble", *Social Neuroscience*, 2 (3-4), pp. 353-365.
- . (2010), "Joint attention, joint action, and participatory sense making", en: *Alter: Revue de Phénoménologie*, 18, pp. 111-124.
- Goldman, A. I. (1989), "In defense of the simulation theory", *Mind & Language*, 7(1-2), pp. 104-11.
- Gopnik, A. y J. W. Astington (1988), "Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction", *Child development*, pp. 26-37.
- Gordon, R. M. (1986), "Folk psychology as simulation", *Mind & Language*, 1(2), pp. 158-171.
- Harman, G. (1978), "Studying the Chimpanzee's Theory of Mind", *Behavioral and Brain Sciences*, 1(4), pp. 560-576.
- Hatfield, E., J. T. Cacioppo y R. L. Rapson (1992), "Primitive emotional contagion", en: Clarck, M. S. (ed.), *Emotion and Social Behavior. Review of Personality and Social Psychology*, 14, pp. 151-177.
- Heal, J. (1986), "Replication and Functionalism", en: Davies, M. y T. Stone (eds.), *Folk Psychology*, pp. 45-59.
- . (1996), "Simulation, Theory, and Content", en: Carruthers, P. y P. Smith (eds.), *Theories of Theories of Mind*, Cambridge: Cambridge University Press.
- Hurley, S. L. (1998), *Consciousness in Action*, London: Harvard University Press.

- Kim, J. (2005), *Physicalism or Something Near Enough*, Princeton: Princeton University Press.
- Loeb, L. (1981), *From Descartes to Hume: Continental Metaphysics and the Development of Modern Philosophy*, Monograph Collection.
- Marr, D. (1982), *Visión: una investigación basada en el cálculo acerca de la representación y el procesamiento humano de la información visual*, Madrid: Alianza.
- O'Neill, E. (1987), "Mind-Body Interaction and Metaphysical Consistency: A Defense of Descartes", *Journal of the History of Philosophy*, 25, 2, pp. 227-245
- Platon (1957), *Fedro*.
- Premack, D. y G. Woodruff (1978), "Does the chimpanzee have a theory of mind?", *Behavioral and Brain Sciences*, 4 (4), pp. 515-629.
- Richardson, R. C. (1982), "The 'Scandal' of Cartesian Interactionism", *Mind*, 91 (January), pp. 20-37.
- Rivière, Á. (2000), "¿Cómo aparece el autismo? Diagnóstico temprano e indicadores precoces del trastorno autista", *El niño pequeño con autismo*, Madrid: APNA-IMSERSO, pp. 13-32.
- Robbins, P. y M. Aydede (eds.) (2009), *The Cambridge Handbook of Situated Cognition*, Cambridge, UK: Cambridge University Press.
- Ryle, G. (1967), *El concepto de lo mental*, Buenos Aires: Paidós.
- Scotto, C. (2002), "Interacción y Atribución Mental: la perspectiva de la Segunda Persona", *Análisis Filosófico XXII*, (2), pp. 135-155.
- Wellman, H. M. y K. Bartsch (1988), "Young children's reasoning about beliefs", *Cognition*, 30(3), pp. 239-277.
- Wimmer, H. y J. Perner (1983), "Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception", *Cognition*, 13(1), pp. 103-128.

¿Cognición social sin atribución psicológica? Una propuesta enactiva

Federico Burdman

Dentro del cognitivismo ortodoxo, el intento por comprender nuestras capacidades de cognición social estuvo centrado durante años en el debate entre dos grupos principales de teorías. Por un lado, la llamada *teoría de la teoría* planteó que nuestra capacidad para atribuir estados mentales a los demás es el reflejo de nuestro manejo de una teoría general acerca de la conducta de las personas, teoría cuyas leyes aplicamos a situaciones particulares para explicar y predecir la conducta de los individuos en contextos determinados. En contraposición, los defensores de la *teoría de la simulación* propusieron la hipótesis de que nuestra capacidad para atribuir estados mentales a los demás es el reflejo de una habilidad de ‘simulación’ de los estados de los demás, proceso mediante el cual evaluamos qué sucedería con nosotros mismos si estuviésemos en tales condiciones, utilizando luego esta simulación para explicar y predecir la conducta de los demás. Sin embargo, a pesar de que tal discusión se presentó con frecuencia como un debate entre dos enfoques fundamentalmente diferentes, eso sirvió al mismo tiempo para oscurecer hasta qué punto ambos grupos de teorías compartían una serie de supuestos y definiciones filosóficamente fundamentales en su concepción misma del fenómeno a explicar. En particular, todo el debate clásico dentro del campo cognitivista fue construido sobre alguna versión del supuesto de que lo que hacemos en nuestra práctica cotidiana es atribuir estados mentales a los demás para explicar y predecir su conducta, que es conceptualizada como *opaca* en un sentido fundamental; correlativamente, lo mental es visto como algo que debe ser hipotetizado o desentrañado a partir de los signos observables disponibles.

En contraste, en años recientes han proliferado intentos por construir una alternativa que difiera de esas posiciones ortodoxas no sólo en sus principios explicativos sino en su orientación filosófica fundamental. El presente capítulo se centrará en uno de ellos: la teoría sobre la *construcción participativa de sentido* propuesta por Hanne De Jaegher y Ezequiel Di Paolo (2007, entre otros). Montada sobre las ideas básicas del enactivismo para el estudio de la cognición, la propuesta de De Jaegher y Di Paolo pretende ofrecer un modo corporizado de pensar nuestra manera de relacionarnos con los demás, poniendo en primer plano la dinámica de la coordinación y la interacción social antes que las estrategias de postulación y atribución de estados mentales más usuales en las opciones cognitivistas tradicionales. Las características de este planteo harán necesario que en primer lugar nos detengamos brevemente en el análisis de los rasgos centrales del enactivismo y su hipótesis de la continuidad mente-vida. En la última sección, a modo de conclusión, presentaremos algunos puntos de discusión en torno a la propuesta enactivista.

1. El enactivismo como marco teórico

Si bien muchas de sus líneas centrales se encuentran todavía en desarrollo, el enactivismo es uno de los proyectos de investigación que suponen una ruptura más radical con la visión ortodoxa de lo mental. Se trata de un enfoque que podemos pensar como *naturalista*, en cuanto busca anclar la intencionalidad, el sentido y el valor en procesos biológicamente básicos. Al mismo tiempo, su concepción de los procesos biológicos es algo heterodoxa, de modo que el resultado es una propuesta distante de las usuales dentro del naturalismo. Más allá de eso, el enactivismo conecta con otras líneas críticas de la concepción cognitivista ortodoxa de lo mental, en cuanto supone ver a la cognición en términos de capacidades de acción del agente en acoplamiento estructural con su entorno, asigna un papel central a los bucles de percepción-acción y piensa a los procesos cognitivos como esencialmente corporizados y valorativos. Se trata, además, de

una de las opciones más decididamente anti-computacionalistas ya que lo usual es considerar que una perspectiva enactiva involucra abandonar la apelación a representaciones mentales. Adicionalmente, su ontología es usualmente discutida en términos de sistemas y procesos que parecen difícilmente reconciliables con el análisis estándar en términos de objetos y propiedades, especialmente cuando la visión procesual es aplicada a la individuación de los propios agentes cognitivos.

En la literatura hay al menos dos sentidos relacionados en que el término ‘enactivo’ y sus derivados son usados con frecuencia (Torrance, 2006; la misma distinción hace Noë, 2001: nota 9). Por un lado, se llama de ese modo a un conjunto de enfoques o teorías en el estudio de la percepción, que parten de una concepción de lo perceptivo como esencialmente imbricado con la acción corporizada. Como ejemplo de una teoría enactiva en este primer sentido podemos citar la teoría de las contingencias sensoriomotoras de Alva Noë (2004, 2009). En un segundo sentido, más amplio que el anterior, el enactivismo es una teoría o un proyecto de investigación acerca de lo mental que parte de la visión de los agentes cognitivos desde el punto de vista de su *autonomía* y asigna un papel central a la hipótesis de la *continuidad mente-vida* (Thompson, 2004, 2007; Varela, 1997). Naturalmente, los dos sentidos de ‘enactivo’ no se hallan desconectados, aunque no parecen equivalentes. En ambos sentidos del término, el trabajo fundacional para este enfoque se halla en *The Embodied Mind*, de Varela, Thompson y Rosch (1991).

El corazón de la visión enactivista está dado por un intento de pensar a los procesos cognitivos como emergentes de la actividad de los organismos biológicos en sus intercambios con el entorno. Y en la caracterización de los organismos biológicos que el enfoque supone, el rasgo central que caracteriza a estos es su *autonomía*. Esto implica en primer lugar –en consonancia con un sentido estándar del término– que el organismo ‘se da su propia ley’, en cuanto la actividad propia de un organismo vivo depende en un sentido fundamental de la dinámica de los propios procesos que lo constituyen (Varela, 1991; Weber y Varela, 2002). Desde ya, en otro sentido, todo organismo/sistema vivo depende

de modo no menos crucial de su relación con el entorno y de sus intercambios con él (en particular, de sus intercambios materiales y energéticos), pero su autonomía radica en que el impacto que tienen sobre el sistema tales intercambios no está determinado directamente por los estímulos externos sino que depende de sus propios procesos. Según la formulación de Maturana y Varela, para un ser vivo las interacciones con el medio son fuente de “perturbaciones” pero no de “instrucciones” (1987: 81 y ss.).

Luego, en un sentido todavía más básico que subyace a la noción anterior, la autonomía de los seres vivos se traduce en el modo en que estos constituyen su *identidad*. Y aquí un aspecto central del enfoque enactivo es que la visión de la identidad que adopta es esencialmente *dinámica y temporal*. Como resultado, la identidad de un organismo/sistema no es vista como una propiedad meramente dada sino como el resultado emergente de un proceso extendido en el tiempo. En particular, la identidad es pensada como constituida por una red dinámica de procesos, lo que hace de ella no sólo algo *logrado* activamente por el organismo sino que implica verla como un logro siempre precario, continuamente amenazado por la falla o la disrupción de los procesos que la mantienen. Esta noción implica entonces un segundo aspecto de la noción de autonomía, en cuanto la identidad es dada por la propia actividad del organismo que se auto-constituye dinámicamente a sí mismo (Di Paolo, Rohde y De Jaegher, 2010). Luego, como antes señalamos, la autonomía de los seres vivos no debe entenderse como una carencia de restricciones sino como subrayando la capacidad del organismo para regular las condiciones de su acoplamiento necesario con el entorno. De este modo, la relación de un ser vivo con su entorno resulta ser una de “libertad con necesidad” (Jonas, 1966).

A partir de esta concepción de la autonomía y la identidad, el enactivismo presenta una concepción de la actividad cognitiva como actividad de *construcción de sentido* (*sense-making*; Varela, 1997). En el sentido básico en que este concepto es desarrollado por el enactivismo, las actividades de construcción de sentido pueden ser pensadas como una característica básica de todos los seres vivos, y de hecho como un proceso inherente a la vida mis-

ma, de donde resulta la idea de una continuidad entre las categorías de ‘ser vivo’ y ‘ser cognitivo’ o ‘ser con mente’ (Thompson, 2004, 2007; Varela, 1997). El proceso mismo del sostenimiento de la vida es pensado, en estos términos, como ‘trayendo con su acción (*enact*) un mundo de sentido’. La actividad cognitiva consiste entonces primariamente en un modo de actuar sobre el entorno o acoplarse a él, y por tanto es vista desde el principio como una actividad necesariamente corporizada. De hecho, desde esta perspectiva, esto puede identificarse ya en el nivel más básico de la organización celular, donde estar vivo equivale a verse involucrado en interacciones discriminativas con el entorno, mediadas por las formas de acoplamiento sensoriomotor propias de tal nivel de organización. Esto además supone que esas interacciones discriminativas están siempre atravesadas por una perspectiva valorativa, orientada teleológicamente (Colombetti, 2009; Di Paolo, 2005).

2. Un enfoque enactivo para la cognición social

Volviendo al plano de la cognición social, el punto de partida para De Jaegher y Di Paolo (2007) es una crítica a los supuestos propios de las opciones cognitivistas tradicionales, de modo consonante con otras críticas similares que encontramos en otros trabajos en tiempos recientes. Luego, en particular, su propuesta gira en torno a dos movidas centrales: en primer lugar, consideran a la interacción entre dos agentes como un nivel emergente de fenómenos y, en particular, como la formación transitoria de un *sistema autónomo*, al que se trasladarán muchos de los rasgos que poseen tales sistemas desde una perspectiva enactivista, junto a conceptos derivados de la teoría de sistemas dinámicos (ver Chemero, 2009). En segundo lugar, proponen una extensión del concepto básico de ‘construcción de sentido’ hacia el nivel de la interacción, incluyendo tanto los modos en que la interacción afecta los procesos de construcción de sentido de los agentes individuales como los modos en que los agentes en interacción deben acoplar y coordinar sus actividades de construcción de

sentido para sostener la interacción misma, de modo conjunto. Ese proceso es lo que denominan la ‘construcción participativa de sentido’.

2.1. La crítica a los enfoques cognitivistas y el foco en la dinámica de la interacción

Las propuestas teóricas basadas en diferentes versiones de la *teoría de la teoría* y la *teoría de la simulación* han recibido una serie de críticas a lo largo de los años que se han vuelto relativamente estandarizadas dentro de los enfoques que avanzan propuestas alternativas al cognitivismo clásico u ortodoxo. Algunas de esas críticas, además, se centran en aspectos de tales propuestas que no son más que trasposiciones al ámbito específico de la cognición social de principios teóricos y filosóficos más generales que subyacen al marco cognitivista como un todo. Así, por ejemplo, De Jaegher y Di Paolo (2007) critican a los enfoques tradicionales por pensar a los procesos que subyacen a nuestras habilidades de cognición social como procesos de resolución de problemas que no involucran de modo central o constitutivo a la dimensión corporal de nuestra relación con los demás. En ese mismo primer nivel general de problemas podemos ubicar igualmente a los problemas de ‘homuncularidad’ que resultan de explicar fenómenos de nivel personal, descriptos en términos intencionales, por medio de la apelación a mecanismos subpersonales de atribución que operarían, como muchas veces se supone, con términos y conceptos igualmente intencionales. Lo mismo sucede con los diferentes problemas sobre cómo entender los constreñimientos de relevancia pragmática para la aplicación computacional de reglas generales a contextos particulares (una variante de lo que los computacionalistas clásicos denominaron el ‘problema del marco’), que parece resurgir como un elemento especialmente incómodo para una propuesta basada en reglas como la *teoría de la teoría*.

Al margen de estos puntos generales, hay toda una serie de críticas que apunta a un conjunto de supuestos o tesis básicas interrelacionadas en torno a la forma en que los enfoques ortodoxos definen el propio objeto de estudio de la cognición social.

Muchos de esos puntos en discusión quedan abarcados bajo la etiqueta de “enfoques ‘Ventana indiscreta’ para pensar lo social” (De Jaegher y Di Paolo, 2007: 489), en alusión a la película clásica de Hitchcock en que un fotógrafo, confinado al aburrimiento y la soledad de su propio departamento a causa de una lesión que le impide caminar, se dedica a observar conductas sospechosas de los habitantes de un departamento vecino y a elaborar sofisticadas conjeturas sobre la naturaleza de sus actos e intenciones.

Bajo un modelo de tal tipo, en primer lugar, la cognición social es presentada como un ámbito más de resolución de problemas, que sólo contaría como ‘social’ en la medida en que el tópico sobre el cual se aplican las habilidades interpretativas y de resolución de problemas es definido como el desciframiento de las intenciones de otras personas. Más allá de eso, sin embargo, éste no es pensado como un ámbito inherentemente social, sino básicamente como una tarea más de extracción de información e interpretación de sucesos del entorno. Luego, de modo más problemático, los enfoques ortodoxos parten del supuesto crucial de que entre los agentes hay una “desconexión total” (Di Paolo, Rohde y De Jaegher, 2010), en cuanto se piensa a las mentes de los demás como fundamentalmente *opacas*. Para De Jaegher y Di Paolo la apelación a las habilidades de ‘lectura de mente’ puede ser necesaria para dar cuenta de algunos fenómenos, pero no puede ser considerada como el foco central del fenómeno a explicar, que debe ser entendido en cambio en términos fundamentalmente interactivos. Este punto se complementa luego con la hipótesis adicional de que las propias habilidades de ‘lectura de mente’ que ponemos en práctica en los aspectos no-interactivos de la cognición social pueden depender ellas mismas de los factores interactivos como un antecedente necesario, al menos en un sentido ontogenético (De Jaegher y Di Paolo, 2013; De Jaegher, Di Paolo y Gallagher, 2010).

Como señalamos, toda esta línea de críticas a los enfoques ortodoxos implica una consecuencia epistemológicamente crucial, en cuanto supone un recorte diferente del objeto de estudio mismo de la cognición social. De este modo, el alcance de la cognición social es definido como incluyendo la habilidad para inten-

tar adivinar o entender lo que hace el otro, pero también para el fenómeno de entender *junto a* un otro, involucrando así también a las habilidades que nos permiten sostener las interacciones, formar relaciones, entendernos mutuamente y actuar en conjunto (De Jaegher, Di Paolo y Gallagher, 2010: 442). En un sentido más amplio aún, De Jaegher y Di Paolo proponen la etiqueta de “intersubjetividad corporizada” para aludir al campo de estudios interdisciplinarios que aborda “el actuar y percibir juntos, el involucramiento verbal y no-verbal, las experiencias de (des-) conexión, las relaciones dentro de un grupo, la construcción de significados conjunta, la intimidad, la confianza, el secreto, el conflicto, la negociación, las relaciones asimétricas, la mediación material de la interacción social, la acción colectiva, el involucramiento contextual con normas socio-culturales, etc.” (Di Paolo y De Jaegher, 2015: 1). De este modo, la adopción de esta perspectiva teórica supone ampliar considerablemente el horizonte de lo que se entiende normalmente por cognición social.

La referencia anterior a la prioridad de la interacción en el enfoque de Jaegher y Di Paolo merece ser subrayada con cuidado ya que es entendida en términos de una tesis particular, que tendrá consecuencias centrales para el marco enactivo. En efecto, muchos de los trabajos que parten de críticas similares a los enfoques ortodoxos adoptan luego una visión de la cognición social centrada en la dinámica propia de las interacciones de ‘segunda persona’, en contraposición con la tendencia de los enfoques ortodoxos a partir de una desconexión entre los agentes y una perspectiva donde la conducta del otro es objeto de interpretación desde un marco distanciado, desapegado. En este primer sentido el enfoque enactivo de Jaegher y Di Paolo parece seguir la misma línea que ya encontramos en otras propuestas de abordaje de la cognición social desde el punto de vista de una práctica interactiva, corporizada, donde el fenómeno central a considerar se da en el encuentro con el otro en una interacción cara a cara y cuerpo a cuerpo que debe entenderse de modo dinámico como una serie de acciones y reacciones bidireccionales (Gallagher 2001, 2004; Hutto, 2004; Ratcliffe, 2007; Thompson, 2001; Gomila, 2002; Scotto, 2002; Pérez, 2013; y los capítulos de Balmaceda y Pérez y

Gomila de esta compilación). Dentro de ese contexto, sin embargo, la propuesta de De Jaegher y Di Paolo supone un paso adicional y postula que el encuentro interactivo entre dos agentes puede dar lugar a un dominio de fenómenos nuevo, emergente, constituido por la dinámica de la interacción misma. Desde esta perspectiva, el reconocimiento frecuente dentro de la literatura del carácter crucial de la interacción para la cognición social resultaría sin embargo insuficiente si no va acompañado por un énfasis en el estudio de la *dinámica* misma de la interacción, como un dominio propio respecto de las acciones e intenciones que pueden atribuirse a los agentes participantes considerados individualmente. En otros términos, aun cuando muchos han planteado la necesidad de pensar al nivel interactivo como central, esto no ha ido de la mano con la adopción de una perspectiva acerca de la interacción como un *sistema autónomo*.

Omitir este punto tendría como consecuencia el confinamiento de las explicaciones del fenómeno interactivo al ámbito de los mecanismos psicológicos individuales, propios de cada agente que interviene en el encuentro. Por el contrario, la consideración de la interacción como un dominio autónomo no sólo volvería disponible un nivel nuevo para la descripción del fenómeno a explicar sino que también habilitaría la apelación a elementos interactivos, trans-individuales, como factores genuinamente explicativos y no-reducibles a las acciones o intenciones atribuibles a los individuos participantes.

2.2. Coordinación, interacción e interacción social

La idea de poner un nuevo énfasis en la dinámica de la interacción entre los sujetos supone ya cambiar significativamente las coordenadas conceptuales con que se movieron los enfoques ortodoxos. Ahora bien, De Jaegher y Di Paolo proponen entender la interacción a través de categorías de la teoría de sistemas dinámicos, lo que da como resultado un punto de vista aún más alejado del tradicional (Di Paolo, Rohde y De Jaegher, 2010). Según este enfoque, para pensar la cognición social debemos en primer lugar visualizar que el fenómeno básico que estamos abordando involucra a dos sistemas –en el sentido más general

de la palabra– que establecen una interacción –en el sentido más general de la palabra. En ese sentido conceptualmente mínimo, no se trata de un asunto que involucre necesariamente a *personas* sino que puede ser descripto de modo abstracto como un patrón de interacción entre dos sistemas dados. El resultado, como dijimos, es un acercamiento a los problemas de la cognición social con un aspecto notablemente diferente del que ofrecen otros enfoques, al punto que a veces puede resultar difícil reconocer un terreno común.

Según De Jaegher y Di Paolo (2007), el nivel más básico en que puede relacionarse el comportamiento de dos sistemas dados es descripto como una *correlación*, que se da cuando se establece una similitud o coherencia entre el comportamiento de dos sistemas dados, más allá de lo que podría esperarse simplemente como resultado de la dinámica interna a cada sistema. Tal correlación puede darse de modo accidental o no-accidental. En este último caso, tendremos lo que se denomina una *coordinación* entre ambos sistemas, como resultado de uno o más factores conectores o comunes en el comportamiento de ambos. Tales factores pueden ser de diversa naturaleza. Si se trata, por ejemplo, de un evento externo a los dos sistemas que provoca en ambos una misma respuesta, se dirá que los sistemas en cuestión se hallan en ese caso externamente coordinados. De otro modo, si el comportamiento similar o coherente de ambos sistemas se da como resultado de una predisposición a comportarse de tal manera que ambos sistemas comparten –sea por compartir mecanismos internos similares o como resultado de una historia común–, se tratará de lo que se denomina una *pre-coordinación*.

Sin embargo, los casos más interesantes para pensar la cognición social parecen exceder lo que puede alcanzarse mediante alguna forma de pre-coordinación, aun si ésta también estuviera siempre presente, como sería esperable al aplicar estos conceptos de sistemas dinámicos a la coordinación entre individuos de la misma especie y enculturizados bajo las mismas o similares normas acerca de cómo responder durante encuentros con otro agente. Bajo este análisis, los agentes en interacción deben hallar formas de coordinar su comportamiento en vivo, durante

el encuentro mismo. Ese tipo de coordinación supone abrir el nivel emergente de la *interacción*. La característica distintiva de las coordinaciones interaccionales es que el proceso mismo de la interacción juega un papel en la generación, facilitación o sostenimiento de la dinámica de la coordinación; al mismo tiempo, en la otra dirección, la propia coordinación puede tener una función interaccional, aumentando las probabilidades de que la interacción tenga lugar o se sostenga en el tiempo.

Este tipo de retroalimentación entre coordinación e interacción sería, de acuerdo a este enfoque, el rasgo distintivo de las *interacciones sociales*. La coordinación misma, tal como se la definió antes, es un fenómeno ubicuo en todo tipo de sistemas físicos y biológicos. Y de modo acorde con la perspectiva de sistemas dinámicos, De Jaegher y Di Paolo proponen adoptar la hipótesis de que típicamente las coordinaciones entre dos sistemas se alcanzan por medio de mecanismos sencillos, aun cuando estamos hablando de la coordinación entre agentes cognitivamente sofisticados.

De ahí resulta una directriz metodológica crucial para los defensores de esta propuesta: explicar la coordinación observada entre dos sistemas no vuelve necesario postular un mecanismo (interno, individual) encargado de *mantener* la coordinación; por el contrario, es esperable que ésta emerja bajo ciertas condiciones si los sistemas poseen propiedades similares (De Jaegher y Di Paolo, 2007). Más allá de ese nivel básico, la interacción ‘social’ se presenta como un dominio diferente debido a que exhibe una característica no tan común en el dominio físico, dada por el hecho de que los patrones de coordinación pueden influir directamente sobre la disposición de los agentes involucrados a sostener o modificar su acoplamiento. Y la dinámica que se desarrolla en tal interacción entre dos (o más) agentes adquiere así, como antes señalamos, las notas de un sistema autónomo (aunque transitorio) y muestra una auto-organización que De Jaegher y Di Paolo proponen pensar, desde el marco enactivo, como un nuevo nivel de fenómenos en donde se exhibe la clausura operacional, definida en este caso por el modo en que se alimentan mutuamente coordinación e interacción: la coordina-

ción hace posible el desarrollo del acoplamiento y la dinámica de la interacción hace posible el sostenimiento de la coordinación, durante el tiempo que la interacción dure. Cuando se da esa dinámica, los agentes involucrados co-emergen como ‘interactores’ en la interacción misma. Adicionalmente, el que los ‘interactores’ mantengan su autonomía fundamental como agentes dentro del contexto de la interacción es presentado por De Jaegher y Di Paolo como una condición para que la interacción sea de tipo *social*. En el caso contrario, si la autonomía de uno de los individuos se viese eliminada o recortada (por ejemplo, a través de una forma de coerción física), el proceso se reduciría al involucramiento cognitivo de un agente con un dominio no-social, donde el otro individuo sólo figuraría como una herramienta o un elemento más del entorno.

De este modo, tenemos ya todos los elementos involucrados en la definición propuesta para el concepto central de *interacción social*:

La interacción social es el acoplamiento regulado entre al menos dos agentes autónomos, donde la regulación alcanza a aspectos del acoplamiento mismo, de modo que éste constituye una organización autónoma emergente en el dominio de la dinámica relacional, sin destruir con ello la autonomía de los agentes involucrados (aunque ésta pueda ver aumentado o disminuido su alcance) (De Jaegher y Di Paolo, 2007: 493).

De este modo, para que tengamos un caso de interacción social debemos contar con dos agentes autónomos (en el sentido enactivo) que puedan co-regular su acoplamiento y las condiciones de la coordinación de su conducta durante el transcurso de la interacción misma. El requisito de la autonomía implica en ese contexto que la regulación del acoplamiento esté en manos de ambos agentes, y excluye por tanto escenarios donde un sólo agente impone sus condiciones. En general, excluye de ese modo a todos los casos donde no hay mutualidad, como la mera observación de una escena social desde la distancia, o la mera presencia del otro (o la creencia en que el otro está presente) (De Jaegher, Di Paolo y Gallagher: 2010). Tales situaciones podrían

razonablemente definirse como “sociales” pero no caerían dentro del ámbito de la interacción.

Debe notarse que, en principio al menos, la definición tal como recién la citamos apunta a capturar la más amplia gama posible de casos de interacción y no está centrada en actuaciones cognitivamente sofisticadas. Así, uno de los ejemplos discutidos por De Jaegher y Di Paolo es el de dos personas que deben coordinar o co-regular su tono de voz o la distancia a la que están hablando para poder escucharse mejor en una conversación en un ambiente con mucho ruido (De Jaegher y Di Paolo, 2013). Otro de los ejemplos recurrentes con que se ilustra esta noción es el de un encuentro entre dos personas que transitan en direcciones opuestas por un pasillo estrecho. En tales circunstancias puede darse con facilidad una situación en la que ambas personas intenten apartarse de la trayectoria de la otra mediante movimientos en espejo en la misma dirección. Ese caso, de acuerdo al análisis propuesto, serviría como una ilustración interesante de la autonomía de la interacción porque mostraría cómo la dinámica de la coordinación puede resultar en el sostenimiento de la interacción precisamente a pesar de los esfuerzos de los individuos por romperla (a diferencia de lo que sucede, por ejemplo, durante una pieza de baile compartida o una conversación en la que ambos individuos tienen, por el contrario, un interés en sostener la dinámica interaccional y alimentarla). En el encuentro trabado en el pasillo estrecho, en cambio, la coordinación no-intencional entre los movimientos en espejo de ambos agentes puede fácilmente resultar en un aumento de la probabilidad de que el movimiento siguiente de ambos sea igualmente en espejo, forzando a los individuos a mantener (transitoriamente) la dinámica de la interacción, a pesar de sus esfuerzos por salirse de ella. A su vez, en ese tipo de casos, la interacción misma promueve acciones individuales que tienden a mantener la coordinación simétrica.

Ejemplos como estos ilustran también una inclinación en el enfoque de De Jaegher y Di Paolo que los lleva a plantear modelos que partan de considerar la mínima expresión posible del fenómeno a analizar, para extraer de tales casos los principios teóricos fundamentales con los que se proponen luego abordar

los casos más complejos. Como ellos mismos lo formulan, su propuesta se basa en “enfocar los patrones comunes presentes en el más amplio rango posible de interacciones sociales” (2007: 486). De este modo, si bien la dinámica de la interacción ‘social’ supone un nivel diferente del de otros tipos de coordinación o interacción, el fenómeno de la interacción social en términos de la definición anterior no pone como condición que la interacción esté mediada por una comprensión de las acciones del otro, ni requiere que los agentes puedan dominar conceptos intencionales ni recursos lingüísticos, ni siquiera que se trate de individuos humanos ni de otros tipos de animales cognitivamente sofisticados. Desde ya, una conversación entre dos personas sobre algún evento significativo contará como una ‘interacción social’ bajo la definición propuesta, pero también lo será, por ejemplo, el intercambio de ladridos entre perros vecinos por la noche.

En consonancia con las ideas enactivas acerca de la continuidad entre los fenómenos básicos de la vida y los fenómenos más complejos del orden cognitivo o existencial, también en el abordaje de la cognición social encontramos aquí una apuesta fundamental para construir un modelo que comience desde abajo hacia arriba, identificando en el nivel más básico posible el origen de los rasgos centrales de los encuentros típicamente humanos, cargados de significados más profundos y atravesados por las contingencias de la normatividad lingüística y las idiosincrasias culturales. El rasgo clave que se hallaría ya en los casos más sencillos de coordinación interaccional como el de la co-regulación del tono de voz durante una conversación o el encuentro trabado en el pasillo estrecho es el de la autonomía, presente en el hecho de que tales interacciones dependen de las contribuciones individuales pero no están completamente determinadas por ellas, ya que dependen igualmente de la dinámica relacional entre los sujetos y de otros factores. Al mismo tiempo, y también en consonancia con las ideas enactivistas, el nivel interactivo no sólo emerge como un nivel autónomo a partir de las contribuciones individuales, sino que puede a su vez alterar y tener un efecto en la conducta de los individuos involucrados en la interacción. De ahí resulta también la importancia que tiene para la propues-

ta enactiva la discusión contra la suposición tácita, presente en otros enfoques, según la cual todas las explicaciones aceptables han de ser en términos de mecanismos individuales.

En ese contexto debemos entender también que los ejemplos de los casos más sencillos de coordinación interaccional sean de especial interés para el enfoque enactivo y que muchos de los estudios discutidos por De Jaegher y Di Paolo se basen en simulaciones computacionales que pretenden modelar los rasgos mínimos indispensables para la emergencia de la interacción social. Así, por ejemplo, Di Paolo (2000) propone estudiar la dinámica de la interacción entre dos células robóticas mínimas que se buscan una a otra mediante señales en un medio acústico. Otros estudios empíricos muy discutidos en la literatura, como el de Auvray, Lenay y Stewart (2009), trabajan con sujetos humanos pero con tareas de interacción que buscan imponer demandas cognitivas mínimas. Esos casos de interacciones mínimas serían relevantes como terreno fértil para explorar las complejas relaciones de causación en ambas direcciones que se dan, incluso en tales modelos mínimos, entre el nivel individual y el social.

La sugerencia es que muchas de las propiedades del nivel interaccional y, en particular, muchos de los modos en que puede alcanzarse la coordinación, podrían ser simplemente el resultado esperable de la interacción entre dos sistemas dinámicos, en lugar de ser el producto de algún mecanismo individual especializado en alcanzar la coordinación (Di Paolo, Rohde y De Jaegher, 2010). Esto supone no sólo la eficacia causal de variables transindividuales sino también que muchos de los factores individuales que contribuyen al patrón resultante de la interacción podrían ser explicados por “mecanismos subpersonales” sencillos, muy alejados de la “terminología de nivel intencional” que supusieron típicamente las explicaciones en psicología del sentido común (De Jaegher y Di Paolo, 2013).

Los estudios a través de simulaciones computacionales con agentes virtuales mínimos tienen, por otro lado, la limitación de excluir todo aspecto del encuentro entre los agentes que no sea capturable en el código en términos objetivos. Se reconoce que fenomenológicamente un rasgo de las interacciones en el

mundo humano está dado por un sentido de involucramiento que experimentan los agentes, y también por el hecho de que, en tales contextos, la dinámica de la interacción puede ‘tomar vida propia’ y puede que ninguno de los agentes participantes tenga la sensación de estar controlando el curso de los acontecimientos. Sin embargo, a pesar de la importancia que se atribuye en las presentaciones estándar del marco enactivo al análisis fenomenológico de la experiencia como un compromiso metodológico central y distintivo de tales enfoques, el concepto de interacción social de De Jaegher y Di Paolo está planteado a partir de una preferencia metodológica por “centrarse en los aspectos objetivos, ya que sólo entonces podemos explicar el vínculo entre los patrones interactivos y la experiencia de la interacción, en lugar de simplemente darlo por supuesto” (De Jaegher, Di Paolo y Gallagher, 2010: 442). La idea detrás de ello parece ser que, a partir del estudio de la interacción en términos puramente objetivos podrían reconstruirse, desde una perspectiva de ‘tercera persona’, los mismos elementos centrales que revelaría un análisis fenomenológico de la experiencia de la interacción (Froese y Di Paolo, 2009). Al margen de la suerte que tenga esa tesis adicional, el resultado del enfoque propuesto por De Jaegher y Di Paolo es presentar un análisis des-subjetivizado de la interacción social que no incluye siquiera el que los agentes en interacción se reconozcan unos a otros como agentes ni sean conscientes de la naturaleza social de su acoplamiento.

2.3. La construcción participativa de sentido

Al comienzo de nuestra presentación destacamos que la propuesta enactiva para la cognición social involucraba dos movidas conceptuales básicas: el reconocimiento de la centralidad de la interacción desde el punto de vista de los sistemas autónomos, y la extensión del concepto enactivista de ‘construcción de sentido’ (*sense-making*) al entorno de la interacción social. Ese segundo movimiento es el que De Jaegher y Di Paolo presentan bajo la etiqueta de ‘construcción participativa de sentido’ (*participatory sense-making*).

Según vimos en secciones anteriores, el marco enactivista supone ver siempre a los agentes vivos como involucrados en la

construcción de sentido, regulando su acoplamiento al entorno a través del ejercicio de sus capacidades sensoriomotoras. Eso, por lo demás, es algo que sucede siempre en un plano corporizado. Por lo tanto, en este sentido básico del término, toda coordinación y toda interacción ‘sociales’ involucran necesariamente a agentes que ya se hallan inmersos cada uno en sus propias actividades de ‘construcción de sentido’. Y como tales actividades son pensadas en este marco básicamente como procesos que involucran al cuerpo, todo proceso de coordinación de la conducta con otro agente está condenado, por definición, a tener algún grado de efecto sobre las actividades cognitivas (de ‘construcción de sentido’) que cada agente desarrolla. Con esto tenemos un primer sentido en que la coordinación interactiva puede tener un impacto cognitivo.

Sin embargo, lo que verdaderamente podría abrir otro nivel de fenómenos está dado por la posibilidad de que dos (o más) agentes coordinen no sólo su conducta descriptible en términos físicos sino también sus propias actividades cognitivas (de ‘construcción de sentido’). Esta posibilidad parece seguirse de modo más o menos directo a partir del marco de ideas que venimos considerando, en la medida en que las actividades de ‘construcción de sentido’ no se distinguen de forma tan clara de otros tipos de actividades sensoriomotoras que puedan atribuirse a cada agente. En la medida entonces en que los movimientos de dos agentes puedan ser coordinados entre ambos durante un encuentro, y en la medida en que sus actividades de construcción de sentido son pensadas como procesos corporizados, la propuesta resultante es que las actividades cognitivas de los agentes pueden ser ellas mismas objeto de coordinación y co-regulación durante la interacción. De allí resulta la definición de *‘construcción participativa de sentido’* que proponen De Jaegher y Di Paolo como “la coordinación de la actividad intencional en la interacción, a través de la cual los procesos individuales de ‘construcción de sentido’ (*sense-making*) se ven afectados y pueden generarse nuevos dominios de ‘construcción de sentido’ social que no están disponibles a cada individuo por sí mismo” (2007: 497).

Como sugiere la cita, la ‘construcción participativa de sentido’ viene en diferentes formatos, que De Jaegher y Di Paolo

pretenden ubicar como diferencias de grado en un eje continuo, que varía según el grado de participación involucrado. En el polo menos participativo, tendríamos los casos donde el ritmo y la dinámica de la coordinación afectan los procesos de construcción de sentido de los participantes del encuentro, aunque estos podrían sin embargo ser entendidos básicamente como procesos individuales. En el polo opuesto, de mayor participación, tendríamos a las actividades de construcción de sentido esencialmente interactivas y co-reguladas entre dos (o más) agentes. Como ilustración del primer tipo de casos, De Jaegher y Di Paolo discuten el modo en que la coordinación de los turnos de habla y las pausas en una conversación entre dos personas pueden modular de diferentes maneras el modo en que cada uno entiende lo que el otro pretende decir (2007: 498). Cerca de ese polo se ubicarían también los casos que discuten bajo el nombre de *‘orientación’*, donde la coordinación es puesta al servicio de la guía de uno de los agentes en un dominio cognitivo nuevo que ya era parte de las actividades de construcción de sentido del *‘orientador’*. En esos términos proponen analizar, por ejemplo, muchas interacciones tempranas entre niños y madres/cuidadores, donde el adulto propone como juego diferentes formas de coordinación regidas por algún patrón particular, orientando de ese modo al bebé al reconocimiento del patrón y a la respuesta ante él. En un sentido más amplio, se discuten bajo el modo de la *‘orientación’* a todas las interacciones en que una persona puede dirigir a otra a notar algún aspecto de una escena común, afectando de ese modo su actividad cognitiva a través de la interacción coordinada. Así podríamos pensar, por ejemplo, el señalar a otro un aspecto determinado de un cuadro que ambos estamos viendo o de un ambiente que compartimos. El señalamiento que da lugar a este tipo de *‘orientación’* puede tomar diferentes formas, desde las más explícitas en su objeto y alcance hasta formas más sutiles, como el caso de quien se para delante de una ventana a contemplar un paisaje y suspira sonoramente, sabiendo que la actitud adoptada y el suspiro emitido serán captados por la persona con la que se está interactuando (De Jaegher y Di Paolo, 2007: 499).

Avanzando hacia grados mayores de participación, los agentes pueden no sólo coordinar de diferentes modos sus actividades cognitivas individuales sino verse involucrados en una verdadera actividad conjunta de construcción de sentido. En los casos de mayor participación y mutualidad, pueden abrirse incluso posibilidades de construcción de sentido que no estén disponibles para los sujetos individualmente sino que sólo sucedan en el nivel de la dinámica colectiva. Podemos entender en estos términos un proceso de negociación entre varias partes, u otro de los ejemplos favoritos de De Jaegher y Di Paolo: una colaboración académica o científica entre pares. Allí tenemos una interacción que puede ir más allá de la orientación hacia una *co-regulación* de las actividades cognitivas de los agentes. Si pensamos, en particular, en colaboraciones interdisciplinarias –un caso especialmente significativo para las ciencias cognitivas– es fácil advertir cómo el resultado de la colaboración puede ser en parte regulado por los diferentes participantes sin que ninguno tenga realmente la potestad de controlar totalmente el resultado final, que sólo puede entenderse como producto de la dinámica colaborativa.

Hay algunos rasgos del modo enactivo de pensar la construcción de sentido que se trasladan a este análisis de las interacciones sociales que debemos destacar. Por un lado, De Jaegher y Di Paolo insisten en que, a pesar de la centralidad que tiene el concepto de ‘coordinación’ para su enfoque, eso no implica desde un punto de vista enactivo dejar de lado a las fallas en la coordinación (o a la dimensión del conflicto, en otros términos). Por el contrario, como antes vimos, el marco enactivo supone que la actividad destinada a regular el acoplamiento y mejorar sus condiciones es en un sentido *alimentada* precisamente por la fragilidad del proceso (de coordinación, en este caso). Eso sucedería en las formas más elementales de construcción de sentido a nivel celular, tanto como en las formas complejas de interacción entre agentes humanos. Los problemas que surjan en la dinámica de la coordinación son precisamente los que deben ser abordados por los participantes del encuentro ajustando sus procesos y respuestas a los del otro y, en ese sentido, co-regulando sus propios procesos de construcción de sentido con los del otro. Como re-

sultado, puede sostenerse que la comprensión social del otro no sería posible sin al menos un elemento mínimo de conflicto (De Jaegher y Di Paolo, 2013).

Otro aspecto importante de la propuesta de De Jaegher y Di Paolo es que deja espacio para la posibilidad de diferentes tipos de tensiones entre el nivel individual y el interactivo. Como vimos antes, la interacción es pensada en este enfoque desde la perspectiva de la autonomía, y en particular como un proceso que tiene lugar entre dos agentes que son, ellos mismos, autónomos. Esto no supone por sí mismo un problema para el marco emergentista del enactivismo, ya que los agentes autónomos mismos pueden ser analizados como colecciones de sistemas de diferentes tipos, de acuerdo a diferentes niveles de descripción orientados por diferentes objetivos pragmáticos. Así, por ejemplo, podemos pensar al sistema nervioso como un sistema autónomo e incluso, en el nivel más básico, a cada célula del cuerpo como un sistema autónomo si optamos por ese nivel de descripción. En la medida en que la tesis de la continuidad entre los procesos propios de la vida misma y el ámbito de lo mental es formulada en términos del concepto de ‘construcción de sentido’, puede hablarse de ese tipo de procesos en todos estos diversos niveles de análisis. Del mismo modo, la perspectiva de la autonomía de cada nivel implica una normatividad propia. La posibilidad abstracta de que las normas relevantes para la construcción de sentido de un nivel dado puedan entrar en conflicto con las de otro cobra un sentido más claro cuando consideramos la interacción entre el nivel individual y el nivel interactivo en un encuentro social. Esa tensión latente entre la lógica de la ‘construcción de sentido’ individual y la dinámica autónoma de la ‘construcción de sentido’ en conjunto surge así, de acuerdo a este análisis, como un rasgo ineliminable de los encuentros sociales (Cuffari, Di Paolo y De Jaegher, 2014).

3. Desafíos para la propuesta enactivista

Como ya señalamos, el enfoque propuesto por De Jaegher y Di Paolo para la cognición social resulta en un cuadro de situación

notablemente diferente del que encontramos en otras alternativas en la literatura. Una vez que la discusión ha llegado a este punto, quizás una reacción natural podría ser plantear que De Jaegher y Di Paolo han corrido poco a poco el asunto de discusión hasta el punto en que el tema que supuestamente debíamos discutir –la cognición social– ha desaparecido del horizonte. Se nos propone conceptualizar a las actividades que normalmente llamaríamos ‘cognitivas’ como procesos de ‘construcción de sentido’, y muchos de los casos de ‘construcción participativa de sentido’ discutidos parecen concernir a rasgos de las interacciones entre individuos que quizás podríamos capturar en términos de una variante de la idea de que estamos ante un proceso *extendido*, en el sentido en el cual se habla normalmente de la ‘extensión’ de la cognición. De Jaegher y Di Paolo por momentos parecen centrarse en casos en que la colaboración cognitiva entre dos sujetos da lugar a un proceso cuya dinámica no puede reducirse a las actividades de cada uno por separado. Ante ese análisis, una reacción posible podría ser pensar que se trata ante todo de una tesis acerca de procesos cognitivos que involucran la participación de más de un agente (de ahí la referencia anterior a la tesis de la extensión) pero que en el camino hemos olvidado que es difícil llegar a ver la conexión de *esa* discusión con la cognición social.

Esa podría ser una reacción enteramente natural para quien esté habituado a abordar estas discusiones en términos de ‘psicología del sentido común’, dada la transformación radical de los términos del debate que supone adoptar el punto de vista que proponen De Jaegher y Di Paolo. Uno puede criticar por más de un motivo al punto de vista que ellos proponen; sin embargo, también puede señalarse que si uno adopta su punto de vista para abordar la cognición social, entonces ese cambio resulta motivado. Desde ya que en un sentido De Jaegher y Di Paolo *están* hablando de otra cosa, pero poner el foco donde ellos lo ponen parece lógico si uno acepta la perspectiva desde la cual están formulando su propuesta. Incluso la acusación de haber desplazado el debate hacia un análisis de la coordinación de actividades cognitivas no-sociales resulta infundada si analizamos el asunto desde el punto de vista de De Jaegher y Di Paolo, ya

que –en términos de su planteo– no habría realmente una distinción tajante entre comprender al otro y comprender un asunto junto a otro. Estaríamos, en ambos casos, ante dimensiones de la ‘construcción participativa de sentido’, entendida como la coordinación interaccional de las actividades de ‘construcción de sentido’ de cada agente.

Para el enactivismo, en las diversas formas que puede adoptar la ‘construcción de sentido’ es siempre el acoplamiento activo con el entorno el que da lugar a la generación de formas de valoración, sentido o significado como un rasgo interno al proceso mismo de regulación del acoplamiento. En la propuesta de De Jaegher y Di Paolo, es ese mismo rasgo el que se extiende al ámbito de la comprensión del otro social. La comprensión del otro como agente no es pensada como un problema cognitivo más a solucionar ni, en particular, como algo independizable del proceso mismo de coordinación e interacción, sino que es en la regulación o co-regulación a lo largo de la interacción misma, en la búsqueda de mayor coordinación con el otro de los procesos de cada agente (búsqueda que es motorizada, como vimos, por las fallas o los conflictos que pueden surgir en el acoplamiento entre ambos), que el otro aparece para mis actividades de ‘construcción de sentido’ como un otro social. De este modo, el contraste parece máximo con los modelos cognitivistas centrados en la idea de atribución de estados intencionales a otros desde un punto de vista desapegado, con el fin de explicar o predecir su conducta. Desde este modelo, en cambio, el otro aparece para mí como un otro en la medida en que nos involucramos en un proceso de coordinación interaccional de nuestras actividades, y en particular cuando influimos mutuamente o co-regulamos nuestras actividades de ‘construcción de sentido’.

Un elemento potencialmente importante para el análisis de la interacción al que la propuesta enactiva presta escasa atención es el involucramiento emocional entre los participantes y el papel que éste tiene tanto en el sostenimiento de la interacción misma como en la generación de sentido o la actividad cognitiva misma. El resultado paradójico de esa directiva no es otro que el

dejar de lado el hecho mismo de que nuestras habilidades son las de animales corporizados. Puede señalarse a ese resultado como ‘paradójico’ porque la adopción de una perspectiva corporizada es precisamente una de las premisas fundamentales del enactivismo. Y otro tanto podría decirse acerca del reconocimiento teórico del papel de las emociones en la cognición. De hecho, según la posición oficialmente defendida por De Jaegher y Di Paolo, todo el proceso de la construcción participativa de sentido debería ser visto como emocionalmente cargado. Esto resultaría ya del concepto mismo de ‘construcción de sentido’ que se halla a la base de la tesis de la continuidad, ya que nos presentaría un acercamiento a la cognición que “no promueve una fisura entre lo cognitivo y lo afectivo” (De Jaegher y Di Paolo, 2007: 488). Tanto como eso es correcto, e importante, pero no parece ser suficiente. Es cierto que pensar a la cognición desde el concepto enactivista de ‘construcción de sentido’ implica pensarla en términos de habilidades constitutivamente corporizadas, que involucran siempre un acercamiento valorativo hacia el entorno; y podemos conceder en esa medida que no se trata de un enfoque que cree un *hiato* conceptual insalvable entre cognición y emoción. Sin embargo, ese punto conceptual general acerca de la idea de ‘construcción de sentido’ no alcanza por sí sólo para promover una visión del proceso de interacción social que incorpore en toda su dimensión al involucramiento emocional de los agentes. En otros términos, el modo en que De Jaegher y Di Paolo caracterizan el concepto de ‘construcción participativa de sentido’ alcanza a asegurar que no haya un hiato conceptual por principio entre ese tipo de actividad cognitiva y la dimensión emocional, pero es claro al mismo tiempo que mucho queda por decir acerca del papel del involucramiento emocional en la interacción social más allá de ese punto conceptual general.

Finalmente, otro conjunto de preocupaciones podría plantearse en relación con el compromiso de la propuesta enactiva con la tesis de la continuidad fundamental entre vida y cognición, dado que junto con ese compromiso se heredan también algunos de los problemas de esa posición. En particular, el enactivismo debe defenderse ante la acusación de vacuidad al extrapolar el vocabulario

intencional/cognitivo desde su entorno ordinario a los eslabones más básicos de los fenómenos biológicos o de los fenómenos interactivos, que en principio no parecen exigir un análisis en tales términos. La cuestión que se plantea entonces es dar cuenta de la pregunta: ¿cómo llegamos desde *aquí* hasta *allí*?

Parte al menos de lo que queremos entender se ubica específicamente en el nivel de las interacciones humanas, aunque probablemente incluya también más que ello. En este punto entra en juego una de las apuestas teóricas características de De Jaegher y Di Paolo, inspirada por las teorías de sistemas dinámicos: no toda forma de coordinación requiere ser explicada mediante mecanismos que supongan una demanda cognitiva fuerte por parte de los agentes. Como principio general, puede ser útil en muchos contextos apelar a mecanismos de coordinación más sencillos que se expresen en la acción de las personas y puedan resultar explicables en términos deflacionarios. Parece probable en tal sentido que muchas formas de coordinación del movimiento o del ritmo entre dos personas a lo largo de una interacción no requieran explicaciones en términos de mecanismos cognitivamente sofisticados. Es posible, por ejemplo, que la coordinación interaccional del movimiento en un encuentro entre dos personas en un pasillo estrecho no requiera de formas recargadas de ‘lectura de mente’ ni requieran amplios recursos conceptuales y la conciencia del otro como un agente cognitivo. Sin embargo, el fenómeno que queremos entender ha de poder incluir también, en otros niveles, aspectos de la interacción entre personas que sí requieren tales recursos cognitivamente más sofisticados y que sí implican, al menos en algunos casos, una conciencia del otro como un agente cognitivo o como una persona. Y el enfoque enactivo de De Jaegher y Di Paolo se halla bajo presión entonces para mostrar que es capaz de dar lugar a lo que podríamos llamar también, forzando apenas los términos, *‘historias de continuidad’* que muestren el camino común que une las manifestaciones mínimas del fenómeno interaccional en dos sistemas cualesquiera que se influyen mutuamente y las formas más sofisticadas cognitivamente (incluso, con frecuencia, conceptualmente) que pueden adoptar las interacciones humanas. Es clara la apuesta de De Jaegher y Di Paolo por identificar

los rasgos mínimos de la construcción participativa de sentido y la apuesta por estudiar sus propiedades dinámicas a través de sus expresiones más sencillas. Esto nos deja, sin embargo, ante un panorama incierto respecto de cómo podría extenderse el modelo aplicado primariamente a los casos ‘sencillos’ hasta poder dar cuenta de la gama completa de fenómenos que esperamos que la teoría pueda abarcar.

Es importante destacar, en este sentido, que otras propuestas críticas de los modelos cognitivistas ortodoxos proponen finalmente incorporar algunos elementos de TT y TS. Por ejemplo, la llamada ‘perspectiva de segunda persona’ (Gomila, 2002; Pérez, 2013; y el capítulo escrito por ambos en este mismo libro) afirma la necesidad de complementar las estrategias de atribución en primera y tercera persona con otros tipos de mecanismos que entrarían en juego en la interacción cuerpo a cuerpo en tiempo presente, pero no pretende con ello reducir todo el campo de la cognición social a lo que pueda explicarse en términos de ese tipo de mecanismos. Se plantea, en ocasiones, que los mecanismos de segunda persona pueden ser considerados primarios en el plano ontogenético (o incluso filogenético) pero se asume al mismo tiempo la necesidad de contar una historia respecto de cómo las perspectivas de primera y de tercera podrían ser derivadas del terreno inicialmente marcado por las interacciones de segunda persona. Ese mismo tipo de compromiso es el que puede serle exigido al enfoque enactivo de De Jaegher y Di Paolo. El problema que podemos anticipar, en el caso de este último, es que las ventajas –que antes señalamos– que derivan de tomar un punto de partida tan ‘atrasado’ en el orden de la complejidad, parecen volverse ahora inconvenientes para visualizar un camino que nos permita abarcar también a los casos cognitivamente más complejos.

Finalmente, si pensamos que no es posible ignorar el papel que desempeñan en estos asuntos los conceptos psicológicos que usamos, la estrategia explicativa adoptada por De Jaegher y Di Paolo también nos deja ante un camino difícil, en cuanto no contaríamos con muchas pistas acerca de cómo podría incluirse a las habilidades conceptuales dentro del cuadro propuesto los autores.

Bibliografía

- Auvrey, M., C. Lenay y J. Stewart (2009), "Perceptual interactions in a minimalist virtual environment", *New Ideas in Psychology*, 27(1), pp. 32-47.
- Balmaceda, T. (2014), *Psicología de sentido común. Pasado, presente y futuros*, Buenos Aires: Título.
- Balmaceda, T. y F. Burdman (en prensa), "Precisiones en torno al debate entre las posiciones tradicionales y los nuevos enfoques por la Psicología de Sentido Común", en prensa en: Skidelsky, L., D. Lawler y D. Pérez (eds.), *Temas de Filosofía de la Psicología*.
- Boden, M. (2000), "Autopoiesis and Life", *Cognitive Science Quarterly*, 1, pp. 117-145.
- Brunsteins, P. (2010), *La Psicología Folk: teorías, prácticas y perspectivas*, Ediciones del Signo.
- Burdman, F. (2015), "El postcognitivismo en cuestión: extensión, corporización y enactivismo", *Principia*, Vol. 19, Nº 3, pp. 475-495.
- Chemero, A. (2009), *Radical Embodied Cognitive Science*, Cambridge: MIT Press.
- Clark, A. y D. Chalmers (1998), "The Extended Mind", *Analysis*, 58, pp. 10-23.
- Colombetti, G. (2009), "Enaction, sense-making and emotion", en: Stewart, J., O. Gapenne y E. Di Paolo (eds.), *Enaction: towards a new paradigm for cognitive science*, Cambridge: The MIT Press.
- Cuffari, E., E. Di Paolo y J. De Jaegher (2014), "From participatory sense-making to language: there and back again", *Phenomenology and the Cognitive Sciences*, 14 (4), pp. 1089-1125.
- De Jaegher, H. y E. Di Paolo (2007), "Participatory Sense-Making: An Enactive Approach to Social Cognition", *Phenomenology and the Cognitive Sciences*, 6, pp. 485-507.
- . (2013), "Enactivism is not Interactionism", *Frontiers in Human Neuroscience*, 6, 345:1-2.
- De Jaegher, H., E. Di Paolo y S. Gallagher (2010), "Can social interaction constitute social cognition?", *Trends in Cognitive Sciences*, 14 (10), pp. 441-447.

- De Jaegher, H. y T. Froese (2009), "On the role of social interaction in individual agency", *Adaptive Behavior*, 17(5), pp. 444-460.
- Di Paolo, E. (2000), "Behavioral coordination, structural congruence and entrainment in acoustically coupled agents", *Adaptive Behavior*, 8, pp. 27 -47.
 —. (2005), "Autopoiesis, adaptivity, teleology, agency", *Phenomenology and the Cognitive Sciences*, 4, pp. 429-432.
- Di Paolo, E. y H. De Jaegher (2015), "Toward an embodied science of intersubjectivity: widening the scope of social understanding research", *Frontiers in Psychology*, 6: 234.
- Di Paolo, E., M. Rohde y H. De Jaegher (2010), "Horizons for the enactive mind: Values, social interaction, and play", en: Stewart, J., O. Gapenne y E. Di Paolo (eds.), *Enaction: Towards a New Paradigm for Cognitive Science*, Cambridge: MIT Press.
- Froese, T. y E. Di Paolo (2009), "Sociality and the life-mind continuity thesis", *Phenomenology and the Cognitive Sciences*, 8, pp. 439-463.
- Gallagher, S. (2001), "The practice of mind. Theory, simulation or primary interaction?", *Journal of Consciousness Studies*, 8 (5-7), pp. 83-108.
 —. (2004), "Understanding interpersonal problems in autism: Interaction theory as an alternative to theory of mind", *Philosophy, Psychiatry and Psychology*, 11 (3), pp. 199-217.
- Gomila, A. (2002), "La perspectiva de segunda persona de la atribución mental", *Azafea*, 4, pp. 123-138.
- Hutto, D. (2004), "The limits of spectatorial folk psychology", *Mind and Language*, 19, pp. 548-573.
- Jonas, H. (1966), *The phenomenon of life: Towards a philosophical biology*, Evanston: Northwestern University Press.
- Maturana, H. y F. Varela (1980), *Autopoiesis and Cognition: The realization of the living*, Dordrecht: Reidel Publishing.
 —. (1987), *El árbol del conocimiento*, trad. esp., Buenos Aires: Lumen, 1996.
- Noë, A. (2001), "Experience and the Active Mind", *Synthese*, 129, pp. 41-60.
 —. (2004), *Action in Perception*, Cambridge: MIT Press.
 —. (2009), *Out of Our Heads. Why You Are Not Your Brain and Other Lessons from the Biology of Consciousness*, New York: Hill and Wang.
- Pérez, D. (2013), *Sentir, desear, creer. Una aproximación filosófica a los conceptos psicológicos*, Buenos Aires: Prometeo Libros.

- Ratcliffe, M. (2007), "From Folk Psychology to Commonsense", en: Hutto, D. y M. Ratcliffe (eds.), *Folk Psychology Re-Assessed*, pp. 223-243.
- Scotto, C. (2002), "Interacción y atribución mental: la perspectiva de segunda persona", *Análisis Filosófico*, XXII, pp. 135-151.
- Thompson, E. (2001), *Between ourselves: Second-person issues in the study of consciousness*, Exeter: Imprint Academic.
- (2004), "Life and Mind: From Autopoiesis to Neurophenomenology", *Phenomenology and the Cognitive Sciences*, 3, pp. 381-398.
- (2007) *Mind in Life. Biology, Phenomenology and the Sciences of Mind*, Cambridge: Harvard University Press.
- Thompson, E. y M. Stapleton (2009), "Making Sense of Sense-Making: Reflections on Enactive and Extended Mind Theories", *Topoi*, 28, pp. 23-30.
- Torrance, S. (2006), "In search of the enactive: Introduction to the special issue on enactive experience", *Phenomenology and Cognitive Sciences*, 4, pp. 357-368.
- Varela, F. (1991), "Organism: A meshwork of selfless selves", en: Tauber, A. (ed.), *Organism and the origins of self*, Netherlands: Kluwer Academic Publishers.
- (1997), "Patterns of life: Intertwining identity and cognition", *Brain and Cognition*, 34, pp. 72-87.
- Varela, F., E. Thompson y E. Rosch (1991), *The Embodied Mind. Cognitive Science and Human Experience*, Cambridge: MIT Press.
- Weber, A. y F. Varela (2002), "Life after Kant: Natural purposes and the autopoietic foundations of biological individuality", *Phenomenology and the Cognitive Sciences*, 1, pp. 97-125.

La atribución mental y la segunda persona

Diana Pérez y Antoni Gomila

*“Strangers in the night exchanging glances
Wond’ring in the night what were the chances
We’d be sharing love before the night was through
Something in your eyes was so inviting
Something in your smile was so exciting
Something in my heart told me I must have you...”*

La perspectiva de segunda persona apareció como alternativa a las clásicas teoría de la teoría (de tercera persona) y teoría de la simulación (de primera persona) que se disputaban la hegemonía de los mecanismos cognitivos en juego en los procesos de atribución de estados mentales (Davies y Stone, 1995a, 1995b). En efecto, la disputa a fines de siglo pasado estaba planteada en términos de cuáles son los mecanismos por *default* presentes cuando explicamos y predecimos la conducta de nuestros congéneres apelando para ello (como de hecho lo hacemos) a la atribución de estados mentales internos del individuo que actúa. Si bien la perspectiva de segunda persona surgió en este ámbito de discusiones, y aparentemente se presentaba como alternativa explicativa de los mismos fenómenos, resulta difícil identificar atribuciones psicológicas en muchas interacciones de segunda persona y no está claro en qué medida se trata de una perspectiva desde la cual sea posible la atribución de estados mentales de una manera genuina y no reducible a las formas usuales que adoptan estas atribuciones en las teorías de tercera y primera persona. Esta es la pregunta que trataremos de responder en este trabajo.

Para introducir la idea básica de la segunda persona, podemos considerar un ejemplo: la letra de la canción “Strangers in the Night”, que describe un encuentro entre dos personas, un encuentro que se da a través de las miradas, las sonrisas... en suma, un encuentro de segunda persona. Estas interacciones de

segunda persona tienen las siguientes características, descriptas originalmente en Gomila, 2002:

1. Hay una interacción dinámica, directa, cara a cara (o tal vez deberíamos decir cuerpo a cuerpo).
2. Los aspectos expresivos del cuerpo son *vistos como significativos* (no interpretados), por ello no pueden darse si no es de cuerpo presente. Son estados constitutivamente corporales. En otras palabras, los estados mentales del individuo con el cual estamos en interacción no son inferidos a partir de conductas, sino que sus acciones son vistas (directamente) como significativas. Tal como sostuvo Wittgenstein, vemos la emoción en el rostro del otro (1967: 220, 225).
3. Hay reciprocidad: cada uno atribuye estados psicológicos al otro y en el mismo acto hay modificación de los propios estados mentales.
4. El ejemplo paradigmático donde se pone en juego esta perspectiva no son las actitudes proposicionales, como deseos y creencias, sino las emociones (aunque no es el único caso).
5. Las interacciones de segunda persona no suponen que haya una actividad “meta”, no hay un estado mental *acerca del* estado mental del otro, sino un estado mental causado por el estado mental-expresado-corporalmente del otro y así sucesivamente.
6. Este proceso está ligado a la acción en el sentido de que involucra acción corporal pública, abierta, conducta emocional, o sea alteraciones del rostro, y otras conductas como huida, consuelo, etc. (aunque no necesariamente acción propositiva/intencional).
7. No es indispensable la existencia de un mundo compartido.
8. No requiere del lenguaje.

En este trabajo nos proponemos explicitar de qué forma se da la atribución psicológica en las interacciones de segunda persona caracterizadas de la manera precedente. Para ello, en primer lugar, consideraremos cuáles son los requisitos conceptuales de

la atribución psicológica en general; en segundo lugar, rechazaremos la idea de que “creencia” es el paradigma de todo estado mental y por lo tanto la idea de que el caso paradigmático a explicar en una teoría de la atribución psicológica sea el de la atribución de creencias, mostrando la heterogeneidad de los conceptos mentales y consecuentemente las variedades de atribución psicológica; en tercer lugar, explicaremos el rol del lenguaje en las atribuciones mentales; en cuarto lugar, explicitaremos los diversos niveles de complejidad de las atribuciones psicológicas y finalmente explicitaremos la relación entre las atribuciones mentales de segunda, primera y tercera persona.

1. Los requisitos conceptuales de la atribución psicológica

La atribución psicológica es un fenómeno que se produce cuando hay dos individuos en juego, uno de los cuales es el sujeto de la atribución (llamémoslo S), aquel a quien se le atribuye un estado mental M, y el otro es quien hace la atribución, el atributor (llamémoslo A). A atribuye un estado psicológico M a S cuando A juzga/crea/afirma que S está en M. Es importante notar que el sujeto de atribución psicológica puede ser un objeto inanimado (por ejemplo, una muñeca), o animado y en este caso puede no ser humano (por ejemplo, una mascota). Atribuir un estado psicológico no nos compromete con la verdad de lo atribuido, así como cualquier afirmación o juicio que se realiza acerca de cualquier objeto del mundo puede ser verdadera o falsa. Vamos a dejar de lado entonces, al menos por el momento, la cuestión de la existencia efectiva del estado psicológico atribuido a S, y vamos a concentrarnos en lo que hace A, cuando realiza una atribución psicológica. Si bien en la segunda persona los roles de S y A se van intercambiando continuamente, desde el punto de vista conceptual en cada específica atribución uno de los dos individuos en interacción es quien atribuye (A) y el otro es quien resulta el sujeto de la atribución (S).

Es un lugar común en la filosofía de la mente contemporánea distinguir en toda atribución de un estado psicológico entre el

estado mental o actitud, y el contenido intencional del estado atribuido. También es un lugar común reconocer que el contenido puede ser objetual o proposicional. La tesis que buscamos desarrollar en este capítulo es que la atribución de segunda persona queda restringida a estados mentales sin contenido o con contenido objetual, y por tanto, no es posible –si nos restringimos a las herramientas de la segunda persona– la atribución de estados de actitud proposicional como la creencia, dado que la atribución de creencias implica necesariamente un contenido proposicional. Esto es así porque para que A pueda atribuir una creencia debe poseer el concepto de creencia, y A sólo puede tener el concepto de creencia –para poder atribuir las– cuando es capaz de articular proposicionalmente el contenido creído que atribuye a S, y esto es sólo posible recurriendo a habilidades cognitivas que van más allá de las que la segunda persona provee.

Es una condición necesaria para que A atribuya un estado mental M a S, que A posea el concepto M. Por ejemplo, si A le atribuye un dolor a S, A tiene que poseer el concepto “dolor”; si le atribuye una intención a S, A tiene que poseer el concepto de intención; y así sucesivamente. La maestría de un concepto es la maestría de las varias habilidades involucradas en sus condiciones de posesión. Así, A posee el concepto “silla” cuando sabe que la función de las sillas es sentarse en ellas, y puede discriminar a las sillas de las cosas que no son sillas cuando quiere sentarse. Estas habilidades no se adquieren todas juntas, por lo que la maestría de un concepto no es cosa de todo-o-nada, por el contrario, es posible tener una comprensión parcial o implícita de un concepto. Lo mismo ocurre con los conceptos mentales: para hacer una atribución psicológica a S, A tiene que poseer los recursos conceptuales necesarios para articular el pensamiento “S está en M”. Para que A le atribuya implícitamente tristeza a S cuando percibe directamente que S está triste, A tiene que tener alguna comprensión del concepto tristeza, aunque sea una comprensión parcial o implícita.

Pongamos un ejemplo. A un niño de 18 meses le pedimos que nos traiga la pelota roja, y la trae (hay, supongamos, varias pelotas de diversos colores, y otros objetos rojos que no son pelotas

en el entorno inmediato del niño). Si el niño realiza esta tarea exitosamente en reiteradas ocasiones, aun cuando en algunos casos, por ejemplo con una pelota de rugby no tenga éxito, estaríamos dispuestos a aceptar que maneja los conceptos “pelota” y “rojo” y que cree que lo que tiene en la mano es una pelota roja. De la misma manera, si a ese mismo niño le decimos que su hermanita está triste, y se acerca y la abraza, o la acaricia en actitud consolatoria (de nuevo, en reiteradas ocasiones, en diversos contextos), diríamos que sabe que su hermanita está triste, y que posee el concepto de “tristeza”. Lo mismo diríamos si espontáneamente se acerca a consolar a alguien que llora, sin que le digamos nada en absoluto. En este caso estamos autorizados a decir que el niño está efectivamente realizando una atribución psicológica, es decir, le está atribuyendo un estado emocional específico (tristeza) a su hermanita.

Este punto de vista acerca de los conceptos tiene dos consecuencias para la atribución psicológica que deben remarcarse. En primer lugar, vuelve la posesión de conceptos independiente de las habilidades de expresión lingüística (aun cuando el desarrollo lingüístico tenga un impacto en ellos volviéndolos más determinados y de grano fino, como veremos en el parágrafo 3.). La conducta lingüística es solo una parte de las múltiples maneras en las que la comprensión conceptual se manifiesta. De esta manera, las criaturas pre-lingüísticas o no-lingüísticas no quedan excluidas del mundo cognitivo (contra Davidson, 1982). En segundo lugar, esta visión de los conceptos mentales permite que sus condiciones de posesión sean diversas y se adquieran de diferentes formas, algo que discutiremos en las próximas secciones.

2. La diversidad de los conceptos psicológicos y las peculiaridades del concepto “creencia” (o por qué no considerar a las creencias como el paradigma de lo mental)

Una cuestión que es importante remarcar es que el rango de estados mentales o psicológicos que los humanos típicamente atribuimos a nuestros congéneres es muy amplio y heterogé-

neo. Hay una gran variedad de estados y procesos psicológicos y una gran variedad de conceptos psicológicos que usamos para hacer referencia a estos estados. Hay una marcada heterogeneidad entre estos conceptos, tal como sostiene Wittgenstein en su plan para el tratamiento de estos conceptos (Wittgenstein, 1967: §472); y no son todos reducibles a un único tipo; es decir no todos los conceptos psicológicos tienen el mismo comportamiento gramatical. Como se puede ver a través de una serie de ejemplos, hay diferencias muy notables entre estos diferentes conceptos. Veamos algunos casos de atribuciones de estados que podemos realizar involucrando distintos tipos de conceptos psicológicos:

- (a) Juan *cree* que las clases empiezan el 1 de marzo.
- (b) María *desea* que termine la ola de calor.
- (c) Sandra *quiere* agua.
- (d) Pepe *sabe* que el 9 de julio es el día de la independencia de Argentina.
- (e) Astérix *teme* que el cielo se caiga sobre su cabeza.
- (f) José *tiene miedo* a las arañas.
- (g) Otelo *está celoso* de Desdémona.
- (h) Antonio *está resentido* con Paula.
- (i) Ana *vio* que el auto atropelló a la mujer.
- (j) Sofía *vio* la película *Django desencadenado*.
- (k) Celia *vio* una sombra al costado del camino.
- (l) Juana *siente* un intenso *dolor* de muelas.
- (m) Pedro *tiene hambre*.
- (n) Alan *está angustiado*.

Como se puede observar, los ejemplos son muy diversos: algunos verbos psicológicos requieren de complementos directos, otros no; algunos requieren complementos con oraciones subordinadas (“cláusulas-que”), otros un verbo en infinitivo, otros un sustantivo o frase sustantiva (objeto directo) y algunos nada. Algunos verbos admiten las dos formas: (b)-(c), (e)-(f) e (i)-(j)-(k); otros sólo admiten complementos proposicionales (a), (d); otros no admiten ningún complemento, como (l), (m) y (n). En algunos casos, los complementos no-oracionales pueden transfor-

marse en oracionales conservando el significado, por ejemplo en (c) la idea implícita sería que Sandra quiere que Sandra (ella misma) tome agua, o (f) donde podría decirse que lo que Juan teme es que una araña lo pique. Pero en otros casos esto es imposible porque no hay objeto directo alguno (en los últimos tres casos), o porque como en el caso (j) es prácticamente imposible resumir la totalidad de las cláusulas que- relevantes correspondientes a ver una película (incluso tal vez sea imposible, si se agregan apreciaciones personales e interpretaciones de cada espectador en el conjunto, además de las cláusulas que simplemente resumen la trama de la película).

Si resulta confiable basarse en esta guía gramatical, es plausible distinguir al menos tres tipos diferentes de estados mentales y correspondientemente tres tipos diferentes de conceptos psicológicos: los estados y procesos cognitivos (usualmente entendidos en términos de actitudes proposicionales), las sensaciones y las emociones.

La tradición filosófica considera a los conceptos de actitud proposicional como paradigma de lo psicológico. Las actitudes proposicionales son estados mentales con contenido proposicional, esto es con condiciones de verdad, con propiedades semánticas. El ejemplo paradigmático de actitud proposicional es la creencia, y la ortodoxia filosófica suele extender sin más las consideraciones hechas sobre este tipo de estado psicológico al resto de los conceptos mentales. Los casos (a) a (k) presentados arriba son casos de actitudes proposicionales o casos que con más o menos esfuerzo parece que podrían entenderse como casos de estados de actitud proposicional. No así los últimos tres casos.

Las actitudes proposicionales, se sostiene, involucran cuatro elementos:

- (1) un sujeto/organismo/individuo/sistema,
- (2) una actitud psicológica (de creencia, deseo, etc.),
- (3) una referencia, o condiciones de verdad (el mundo representado),
- (4) un modo de presentación de la referencia, o contenido.

De (3) se sigue que quien posee el concepto tiene que poder entender que aquel al que se le atribuye puede representar el mundo erróneamente (esta es la razón para considerar al test de la falsa creencia como la marca de la capacidad de atribución psicológica). De (4) se sigue el principio de opacidad referencial, es decir la posibilidad de que alguien puede referirse al mismo estado en el mundo que uno sin saberlo, por desconocer las identidades “intensionales” relevantes.

Pero, además, el concepto de creencia está conceptualmente ligado a conceptos perceptivos, esto es a la idea de que los humanos tenemos las creencias que tenemos en función de las percepciones que tenemos. Y que, si no percibimos algo que otra persona efectivamente percibió, no tenemos por qué adquirir la creencia del otro. (Y esto también es necesario para pasar el test de falsa creencia). También está conceptualmente ligado a los conceptos de acción intencional y de deseo: el esquema del silogismo práctico es una verdad conceptual (una guía/regla para la atribución intencional, tal como defiende Dennett, 1987), no una generalización empírica que hacemos a partir de casos observados. Es decir, es lo que se “compra” al adoptar la “actitud intencional” de Dennett. En efecto, los conceptos de creencia y deseo son los que estructuran las inferencias prácticas, esto es nuestras formas de comprender como racional la conducta de nuestros congéneres, lo que ocurre al considerar la acción ajena como un elemento en el espacio de las razones. Es por ello que la atribución de deseos y creencias (y de todas las actitudes proposicionales en general) está constreñida por una serie de principios normativos específicos como son el principio de caridad (esto es que interpretemos mayoritariamente al otro como creyente de verdades), y el principio de coherencia (esto es que interpretemos al otro como un ser que no se contradice ni teórica ni prácticamente, esto es que no va a asumir como verdaderas proposiciones contradictorias, ni va a actuar en contra de lo que concluye a partir de los procesos de deliberación racional que lleva a cabo). Tanto Davidson como Dennett sostienen que toda la actividad interpretativa que nos permite atribuir estados mentales está regida por este tipo de principios normativos. Es decir,

que nuestras atribuciones de actitudes proposicionales suponen a la vez un conjunto muy grande de otras actitudes proposicionales atribuidas (esto es lo que suele denominarse el “holismo” de lo mental). Nótese que no ocurre lo mismo con las emociones (básicas) ni con las sensaciones, donde los criterios que guían la aplicación de estos conceptos no incluyen los estados cognitivos del sujeto, sino sus expresiones corporales, por lo que la atribución resulta atomista.

Hay una distinción adicional que vale la pena considerar. Algunos conceptos psicológicos son verbos de logro, verbos factivos; mientras que otros no lo son, son verbos no-factivos. En efecto, verbos como saber o ver, suponen que el complemento intencional es el caso o existe. No puede ser verdadero que alguien sepa que llueve si no es el caso que está lloviendo, y nadie puede ver un gato si no hay gatos allí. Estos verbos funcionan de una manera diferente a verbos como creer o alucinar, dado que puedo creer que llueve sin que llueva efectivamente, o puedo alucinar que hay un duende (o un gato) sin que haya nada allí delante mío. Es por esta razón que podría afirmarse que los conceptos como saber y ver son más básicos conceptualmente, y por lo tanto más simples que el concepto de creencia; un indicio de esto es que se adquiere más temprano en el desarrollo (Nagel, 2015). En efecto, para comprender que alguien está viendo algo, por ejemplo que S está viendo una pelota, alcanza que nosotros mismos estemos viendo a S dirigiendo su mirada a la pelota, y entender que S sabe que hay una pelota allí depende simplemente de entender que quien ve algo sabe que lo que vio es el caso. En cambio, dado que creencia no es un verbo de logro, no podemos ver lo que alguien cree. Claro que si la creencia es verdadera podemos atribuirle tan fácilmente como atribuimos conocimiento, pero si la creencia es falsa, la atribución es mucho más compleja. Es este desacoplamiento entre la mente y el mundo uno de los factores que produce dificultades a la hora de atribuir creencias y de pasar el test de falsa creencia.

Veamos ahora qué ocurre con los estados que tienen contenidos objetuales. Dretske estableció una diferencia entre el “ver simple” (*simple seeing*, ejemplo (k)) y el “ver epistémico” (*epistemic*

seeing, ejemplo (i)). El ver simple es caracterizado de la siguiente manera:

- (1) toma complementos no-proposicionales,
 - (2) crea contextos transparentes,
 - (3) es hecha verdadera por relaciones perceptuales directas entre los receptores y los particulares,
- en cambio, el ver epistémico se caracteriza por que:
- (1') toma complementos proposicionales,
 - (2') crea contextos opacos,
 - (3') es hecho verdadero por relaciones epistémicas basadas perceptualmente entre receptores y proposiciones.

Para Dretske, la distinción depende de considerar que los primeros son estados representacionales no-conceptuales en tanto los segundo sí son conceptuales. En efecto, parece bastante razonable pensar que para que un cierto estado tenga un complemento proposicional debe haber alguna manera en la que el propio sujeto sea capaz de trazar distinciones finas, tales como las que permite el lenguaje. Pero si Dretske está en lo cierto, el caso del “ver simple” no puede ser redescrito en términos del “ver epistémico” (ni viceversa), ya que en el primer caso no hay la distinción trazada entre la referencia y el modo de presentación de la referencia que el segundo requiere (o sea entre los puntos (3) y (4) en nuestra caracterización de las actitudes proposicionales). Todo ver simple es exclusivamente referencial.

Así, no es difícil trasladar esta distinción a otros estados intencionales más allá de “ver”, específicamente sostenemos que esta distinción se puede extender a intenciones y a algunos estados emocionales (específicamente, las emociones básicas) y así reconocer estados mentales dirigidos a objetos, transparentes (por ejemplo (c), (f), (g) y (h)). En efecto, puede ocurrir que alguien tenga miedo a las arañas no importa cómo se las describa, mientras que hay otros casos en los que las emociones se dirigen a estados de cosas complejos y resultan opacas (por ejemplo, si tengo miedo de que mi jefe se entere de que he llegado tarde al trabajo). En casos como el primero, el sujeto no tiene por qué es-

tar considerando ninguna proposición relativa a las arañas, puede no tener creencia alguna acerca de ellas, puede que ni siquiera logre distinguir correctamente a las arañas de otros bichos. Por el contrario, en el segundo caso, el estado emocional involucra una serie de creencias que el sujeto posee, como que al jefe no le gusta que sus empleados lleguen tarde, que ya llegó tarde en ocasiones anteriores y el jefe está a punto de despedirlo, etc.

Nótese, finalmente, que en el caso de las creencias no puede trazarse ni la distinción entre estados simples y epistémicos, ni entre estados factivos y no factivos, toda creencia es epistémica y no-factiva. Tener una creencia implica tener la capacidad de hacer distinciones finas como las que se requieren para generar contextos opacos; tener creencias requiere de la capacidad para recombinar elementos que se dan separadamente, dado que al no presuponer la verdad de lo creído, lo creído debe ser meramente un estado de cosas posible, una recombinación posible de elementos más básicos.

Hechas todas las distinciones que hemos recogido hasta aquí, es fácil ver que el “imperialismo de la creencia” (Pérez, 2013), es decir la idea ortodoxa de que la creencia es el paradigma de estado psicológico y que por lo tanto todos los demás estados psicológicos deben entenderse como ella, debe ser descartada. Este supuesto compartido por la mayoría de los filósofos de la mente (y consecuentemente por los psicólogos del desarrollo y cognitivos en general) durante por lo menos los últimos cincuenta años ha llevado a darle un lugar más importante del que merece en estas discusiones al “test de falsa creencia”. En efecto, la creencia es un estado psicológico complejo, de tardía adquisición en la ontogénesis, pero de su posesión no depende la posesión de muchos otros conceptos psicológicos (contra Davidson, 1970). En particular no dependen del concepto de creencia los conceptos de emociones básicas cuando no involucran un contenido proposicional, ni los conceptos de sensaciones corporales. En el apartado cuarto de este capítulo nos ocuparemos con detalle de la cuestión del contenido; nos basta por el momento remarcar el hecho de que es posible tener miedo, y atribuir miedo con un contenido objetual. Como ejemplificamos más arriba, alguien

puede tener miedo a las arañas, o a una araña presente en su entorno inmediato, sin tener creencias acerca del carácter venenoso de las arañas, ni tener la menor idea de qué es un veneno, ni de que hay arañas venenosas y otras que no lo son. Pero claramente esto no ocurre con las creencias, dado que tener una creencia es aceptar como verdadero un cierto contenido proposicional, contenido que por lo tanto debe estar estructurado predicativamente.

Resumiendo lo dicho hasta aquí: para que podamos decir que S cree algo, deben satisfacerse las siguientes condiciones:

- (1) Lo creído es un contenido proposicional, esto es algo que tiene una estructura predicativa (por lo menos). Así, tiene sentido decir que Juan cree que la luna es blanca, pero no tiene sentido decir que cree en la luna.
- (2) Para creer algo, el individuo tiene que estar en posesión de los conceptos que constituyen el contenido creído. Si Juan cree que la luna es blanca, tiene que estar en posesión del concepto “luna” y el concepto “blanco”. Estar en posesión de estos conceptos quiere decir que es capaz de recombinarlos (*à la* Evans, 1982).

En el caso de una atribución psicológica, A cree que S está en un estado M, por lo que en este caso específico A tiene que estar en posesión del concepto M (pero este concepto no necesariamente tiene que estar en posesión de S).

Todavía, sin embargo, no hemos especificado qué habilidades son las que debe tener un individuo para estar en posesión de un concepto psicológico cualquiera. A nuestro juicio, para estar en posesión de un concepto psicológico cualquiera, digamos M, A debe:

- (a) Poder aplicar M a sí mismo y a los demás apropiadamente (es decir, tener maestría en los usos en primera y en tercera persona de dicho concepto). Así, y a diferencia de la teoría Goldman (1993) que considera que la autoatribución psicológica es anterior a la heteroatribución, y a diferencia de Ryle (1949) y la ortodoxa teoría de la teo-

ría (paradigmáticamente Gopnik, 1993) que sostienen que la atribución de tercera persona es previa, nosotros vamos a argumentar que la posibilidad de realizar los dos tipos de atribución es requisito para la posesión de los conceptos psicológicos. Esta idea quedará justificada cuando veamos en el apartado 4. cómo se adquieren estos conceptos.

(b) Entender las conductas que son expresivas de M como criterios/síntomas/señales de M.

(c) Tener ciertos mínimos guiones (*scripts*) relacionados con M (por ejemplo, si M = miedo, comprender que si alguien tiene M se esconde, o huye; que si alguien ve algo entonces adquiere conocimiento acerca de su existencia, etc.).

Como sostuvimos más arriba, la posesión de conceptos no es una cosa de todo o nada, hay distintas tonalidades de gris: se puede tener desde una comprensión mínima de un concepto hasta una comprensión muy profunda, que va aumentando y complejizándose con el tiempo, la experiencia y la información que vamos adquiriendo a lo largo de nuestra vida (por ejemplo, se pueden incorporar nuevos scripts, o se pueden identificar nuevas conductas expresivas como criterios para la aplicación del concepto). Poseer un concepto es una habilidad o conjunto de habilidades, por eso es que tiene grados y no es necesario que quien posee un concepto mental y por tanto es capaz de atribuir estados mentales sea capaz de articular lingüísticamente la presencia de dicho estado mental. Así, un sujeto A será capaz de atribuir miedo a los demás cuando posea el concepto miedo. Tener miedo no es suficiente para tener el concepto miedo, se requiere además tener una idea de lo que ocurre cuando se tiene miedo, y de identificar casos de miedo tanto en uno mismo como en los demás. De la misma manera tener creencias no es suficiente para tener el concepto de creencia o, dicho a la inversa, tener el concepto de creencia no es necesario para tener creencias (contra Davidson, 1982), pero es necesario tener el concepto de creencia para poder atribuir creencias a uno mismo y a los demás.

3. El rol del lenguaje en la atribución mental

Hay dos modos en los que la maestría en el uso del lenguaje público influyen en nuestra habilidad para realizar atribuciones mentales, porque hay dos maneras en las que esta maestría cambia nuestras habilidades conceptuales. Por un lado, la maestría del lenguaje público nos ayuda a adquirir ciertas habilidades sintácticas, por ejemplo la capacidad para entender la predicación y la subordinación (las cláusulas que-). Estas capacidades sintácticas están involucradas en las condiciones de posesión de algunos conceptos psicológicos, específicamente en las condiciones de posesión de las actitudes proposicionales. Por otro lado, la adquisición léxica de un lenguaje natural particular nos induce a entender nuestra experiencia de un cierto modo, y la diferencia léxica entre diferentes lenguajes puede generar diferencias en las maneras en las que categorizamos nuestros estados psicológicos. Consideraremos estas dos cuestiones en este apartado.

En primer lugar, como dijimos arriba, las condiciones de posesión de un concepto dado no son cuestiones de todo-o-nada. Por el contrario, en el curso del desarrollo nos volvemos capaces de dar una articulación más precisa y sutil de los estados psicológicos (propios y ajenos) y de sus contenidos. Así, en primer lugar, está bien establecido que estados cuyos contenidos son objetuales, y por tanto, más simples, pueden darse desde muy temprano en el desarrollo (miedo a la araña, o querer a la mamá). La atribución de estos estados requiere la posesión del concepto del estado en cuestión y sólo una comprensión muy mínima del objeto intencional: no se requiere tampoco de la posesión del concepto correspondiente (en sentido fregeano, como un modo de presentación, o un sentido), sino la capacidad de identificar (¿y reidentificar?) el objeto en cuestión, es decir estar en una relación epistémica básica con ese objeto particular (o clase de cosas, o patrón perceptivo). Por el otro lado, las actitudes proposicionales en general presuponen la habilidad de articular los conceptos en contenidos proposicionales y subordinarlos a los conceptos psicológicos. Los contenidos proposicionales resultan de la combinación de conceptos en una estructura predicativa, que pare-

ce ser un logro cognitivo básico, y entender a estos contenidos como complemento de los verbos mentales es algo que parece depender de la adquisición de estructuras lingüísticas específicas: las cláusulas *que-*, algo que ocurre hacia el tercer año de vida (De Villiers y De Villiers, 2009; Gomila, 2012; Hinzen, 2013). En otros términos, la maestría de los conceptos psicológicos cuya estructura involucre contenidos proposicionales depende de la posibilidad de ser capaz de pensar contenidos proposicionales. También se ha sostenido que el concepto “saber” se adquiere antes que el concepto “creer”, dado que comprender las creencias supone comprender que diferentes individuos pueden tener distintas perspectivas de la misma situación (Gordon, 1987). Y los conceptos psicológicos factivos –los que presuponen que sus contenidos intencionales son el caso o existen, como en el caso de “saber” o “ver”– se desarrollan antes que los no-factivos o epis-témicos, como “creer” (Nagel, 2015).

En segundo lugar, nuestras prácticas lingüísticas estructuran y estabilizan la comprensión de nuestra experiencia. Para desarrollar nuestra comprensión de los conceptos psicológicos trazaremos un paralelo con el caso del color y nuestras capacidades discriminativas del espectro de color. Para ello comencemos por repasar brevemente algunos hechos acerca de la influencia que tiene el lenguaje en la forma de recortar el espectro de color que percibimos gracias a las características biológicas de nuestro aparato perceptivo. Veremos que hay ciertos parecidos estructurales entre el dominio del color y el de nuestra vida mental, por lo que comprender el primer dominio de cuestiones nos ayudará a comprender qué ocurre con el menos explorado empíricamente caso de los conceptos psicológicos.

Los lenguajes difieren en su vocabulario de color (D’Andrade, 1989). El color es una dimensión perceptiva saliente para los humanos, que posee un sistema neurofisiológico dedicado que hace posible la percepción del color. Además, el color puede ser descrito físicamente. Esto hace que se puedan testear más fácilmente las diferencias producidas por la influencia del lenguaje en la percepción del color, de una manera no circular, dada la existencia de una descripción independiente, física del estímulo.

lo. Hay evidencia concluyente que muestra que la codificación lingüística disponible influye en nuestra memoria, esto es la capacidad de almacenar información acerca de los colores (Lucy y Shweder, 1979), así como la posibilidad de realizar juicios perceptivos acerca de la similitud respecto del color (Kay y Kempton, 1984). En general si dos colores están incluidos bajo una misma etiqueta lingüística en un lenguaje (en muchos lenguajes hay una única etiqueta para “verde” y “azul”, por ejemplo), los hablantes de ese lenguaje tenderán a encontrar más similares las muestras de esos colores que aquellos que poseen dos etiquetas diferentes en su lenguaje, y tenderán a confundirlos más en sus recuerdos (Davidoff, Davies y Roberson, 1999; Roberson, Davies y Davidoff, 2000). Estas diferencias aparecen tempranamente en el desarrollo y coinciden con la adquisición de los términos de color (Robertson et al., 2004). Asimismo, se ha establecido que hay interferencia lingüística en el procesamiento del color (Philling et al., 2003; Roberson y Davidoff, 2000; Witthoff et al., 2003). Esta cuestión es diferente de la cuestión de la existencia o no de universales de color en los diferentes lenguajes. Las diferencias lingüísticas podrían tener un rol cognitivo en el procesamiento del color sin que haya por eso una categorización de los colores puramente arbitraria y convencional. De acuerdo con los resultados alcanzados hasta ahora (Kay y Regier, 2003; Kay y Maffi, 1999, 2011), parece haber tendencias universales en la categorización (*naming*) de los colores, pero son menos fuertes que las que se pensó originalmente en los trabajos de los sesenta y setenta (Berlin y Kay, 1969; Rosch, 1975). Sin embargo, qué muestra de color se considera prototípica de cada etiqueta lingüística depende del vocabulario de color, en lugar de a la inversa. Además, ha quedado establecido experimentalmente (Lupyan, 2008a, 2008b; Lupyan, Rakison y McClelland, 2007; Lupyan, Thompson-Schill y Swingle, 2010) que las etiquetas léxicas juegan un rol en el aprendizaje de conceptos en general, no sólo en el caso de los colores, haciéndolo más rápido, produciendo conceptos más coherentes, reforzando los rasgos perceptuales correlativos que comprenden las etiquetas y haciéndolos más discretos y definidos. Asimismo, en la búsqueda visual las etiquetas facilitan el

procesamiento visual de una familia de estímulos homogénea, en lo que ha dado en llamarse el “*grouping effect*”. En general las etiquetas léxicas permiten cambios representacionales, esto es maneras diversas de codificar los estímulos (Gomila, 2012: 112). En otras palabras, uno aprende el léxico de los términos de color de su comunidad y así se modifican las formas básicas de categorización del color en línea con el lenguaje adquirido.

Consideremos ahora el caso del etiquetamiento de nuestros estados mentales. Como en el caso del color hay una diversidad de términos en los diferentes lenguajes para describir la vida mental (Dryer y Haspelmath, 2013). También como en el caso del color, hay una base neurofisiológica de nuestros estados mentales, se trata de mecanismos muy bien estudiados en ciertos casos, como el del miedo (LeDoux, 1996), pero sólo hipotetizados en muchos otros casos. No hay, sin embargo, un estímulo físico independiente en este caso, por lo que el testeo de estas ideas resulta más dificultoso (si no imposible).

La tesis que queremos proponer aquí es la siguiente: los primeros etiquetamientos de los estados mentales se producen en interacciones de segunda persona. En efecto, en estos intercambios intersubjetivos los adultos suelen etiquetar los diversos estados en los que los bebés se encuentran; por ejemplo, cuando un bebé pequeño llora la madre se acerca y le dice “¿Qué te pasa? ¿Tenés hambre?”, y cuando es más grande puede acercarse diciéndole “¿Te lastimaste?” o “Sana, sana, colita de rana...”. Así, los primeros etiquetamientos de los estados mentales se dan en una interacción en la cual están presentes las experiencias/sensaciones subjetivas/corporales por las que el bebé está pasando, las conductas abiertas asociadas, incluyendo la propia expresión del estado interno, y también están presentes las expresiones del adulto, y las conductas reactivas del adulto ante el bebé. De esta manera, el bebé transita una experiencia sumamente compleja, multimodal: visual, auditiva, propioceptiva, etc., y es transitando esta compleja experiencia que el bebé adquiere las etiquetas propias de su comunidad lingüística. Estas etiquetas son las que le permiten recortar esta experiencia multimodal, y atender a determinados aspectos de la misma: por ejemplo a su expe-

riencia propioceptiva, sus sensaciones internas, etiquetándolas con el concepto mental en cuestión. Pero a la vez, atiende a las reacciones del adulto propias de quien interactúa con alguien en tal estado, y así adquiere un concepto que le permite identificar un estado psicológico a través de múltiples habilidades que le permiten al mismo tiempo recortar conductas propias, ajenas y sensaciones corporales. A partir de esta compleja experiencia, es capaz de llegar a comprender los correspondientes estados psicológicos, satisfaciendo las tres condiciones mencionadas arriba (i.e. aplicando el concepto a sí mismo y a los demás a la vez, y comprendiendo qué conductas expresivas y guiones están conectados con ese estado mental).

Así, el estado inicial que posibilita la adquisición léxica supone una suerte de comprensión implícita del estado mental, en la interacción; por lo que se puede decir que, en un sentido, antes de poder recortar su compleja experiencia de la manera en que su comunidad lingüística lo hace, el bebé no siente miedo –en el sentido en que miedo es una categoría socialmente establecida; pero en otro sentido sí lo siente: tiene una experiencia subjetiva que está tratando de categorizar, con ayuda externa, y que es similar a aquella que denominará miedo cuando haya adquirido los conceptos correspondientes. Así, la posesión de un concepto psicológico dependerá de haber adquirido el patrón de clasificación propio de la comunidad lingüística que estructura este tipo de experiencia. Se puede afirmar, entonces, que al adquirir las etiquetas de la comunidad lingüística a la que el individuo pertenece a la vez que se aprenden los conceptos relevantes para pensar acerca de la vida mental propia y ajena y se le da forma a la experiencia, al recortarla de una manera y no de otra. En este sentido los conceptos son constitutivos de la experiencia. Este modelo es compatible con la idea de que nuestros conceptos mentales son conceptos universales (esto es que hay algunos conceptos psicológicos universales, como ocurre con el caso de los colores) y también con la idea de que no todos lo son (lo que también ocurre en el caso de los colores, dado que no en todas las comunidades lingüísticas existen las mismas etiquetas para recortar el espectro de color). Lo importante es remarcar que lo

que hace universal un concepto no es que sea biológico/innato, sino que todas las comunidades lo hayan adquirido (aunque no necesariamente lexicalizado) de manera similar. Y este es un hecho contingente, no necesario.

Para resumir: sostenemos que las interacciones de segunda persona hacen posible el enriquecimiento de la maestría de los conceptos psicológicos que se adquieren en interacciones de segunda persona. Le permiten a los niños desarrollar los guiones correspondientes y reconocer los patrones expresivos relevantes de los conceptos psicológicos, conjuntamente con la dimensión interna, de tal manera que la maestría del concepto se desarrolla a la vez para la aplicación a uno mismo y a los demás.

4. Niveles de complejidad en la atribución mental

Como dijimos en el apartado anterior, hay un sentido en el cual la efectiva posesión de un estado mental M por parte de un sujeto S es independiente de la existencia de un A (sea $S = A$, o no) que realiza la atribución del estado mental M a S . Pero en otro sentido vimos que esta independencia es relativa, dado que al categorizar lo que le está ocurriendo a S como un caso de M se producen varios efectos cognitivos y conductuales diferentes. La pregunta de la que queremos ocuparnos en este apartado es la siguiente: ¿cuáles son los indicios/evidencias/señales/criterios que autorizan la atribución de un estado mental a alguien en la interacción? Nótese que la pregunta no es general, esto es no estamos preguntando cuáles son en general los indicios que nos autorizan a una atribución psicológica. Está claro que, por ejemplo, las emisiones lingüísticas de S son indicios para la atribución psicológica (Davidson, 1975; Kripke, 1979). La pregunta que nos interesa ahora es determinar cuáles son los criterios de la atribución psicológica, incluyendo la atribución implícita del tipo que estuvimos describiendo hasta aquí, entre un A y un S cuando están en una interacción de segunda persona en la que el lenguaje puede no estar presente, pensemos en los “extraños en la noche” (recordemos que los roles de A y S se intercambian

constantemente en la interacción: ambos sujetos son atributores y atribuidos, para que la interacción sea exitosa).

La tesis que proponemos es que la base evidencial de las atribuciones de segunda persona es la expresión de S, su corporalidad en contexto: sus expresiones, sus movimientos, sus miradas. Se trata de información no conceptualizada por los individuos en interacción, pero que resulta indispensable en el proceso de categorización por medio de los conceptos psicológicos adquiridos por A. Algunos de los indicios son las expresiones (y micro expresiones) faciales, la dirección de la mirada, el ritmo y las pausas en la interacción, el timbre, el tono y la altura de la voz, la modalidad del estímulo táctil (cuando lo hay) es decir el modo en el que nos tocan y somos tocados en la interacción, entre otros.

Es importante notar que no se trata de información disponible a cualquier observador externo de la interacción entre los dos individuos. Hay un sentido en el que la información producida por los individuos en sus interacciones es relevante sólo para quien participa de la interacción. Y es información producida *por* la interacción, por cuanto los individuos en interacción provocan en el otro las conductas expresivas que permiten este tipo de atribuciones en la interacción. Pensemos en lo que nos ocurre cuando escuchamos a alguien hablar por teléfono y sólo accedemos a la información que brinda uno de los individuos en interacción: está claro que no comprendemos de la misma manera qué es lo que está ocurriendo que quien participa de la interacción, porque no escuchamos la respuesta del otro somos ajenos a la longitud de las pausas en el intercambio lingüístico, al tono de voz o a la acentuación prosódica del interlocutor, etc. Es que la interacción misma es un modo de obtener información contingente a la propia acción; en este sentido sostenemos que la atribución psicológica de segunda persona es dinámica y recíproca: la mirada del “extraño” desencadena la sonrisa de la “extraña”, sonrisa que es leída como una invitación a acercarse.

Por otra parte, también es importante resaltar que si sólo disponemos de esta información corporal tenemos un acceso parcial a la mente de quien está interactuando con nosotros. En efecto, esta información nos autoriza a la atribución de cierto

tipo particular de estados mentales, aquellos en los que no está involucrada la forma peculiar de presentación del mundo para ese individuo, esto es aquellos estados que no tienen contenido proposicional, ni involucran una conceptualización determinada del objeto intencional del estado mental atribuido. Así, intenciones dirigidas a objetos del entorno (por ejemplo la intención de agarrar algo), emociones básicas (miedo a la araña que está en el entorno compartido), y sensaciones corporales (dolor) son el tipo paradigmático de atribuciones de segunda persona. El “extraño” ve la invitación en la sonrisa de la “extraña” pero no sabe en qué consiste exactamente la invitación: ¿sólo sexo?, ¿sólo conversación?, ¿la posibilidad de un amor eterno?... sólo el diálogo, las interacciones lingüísticas permitirán desentrañar estas dudas.

Como dijimos en el apartado anterior, a medida que se adquieren capacidades lingüísticas se complejizan los contenidos atribuibles. Y, como también dijimos más arriba, una misma actitud psicológica admite grados crecientes de complejidad del contenido atribuido (pensemos en la diferencia entre la atribución a S de la intención de agarrar eso, la intención de agarrar un juguete, la intención de agarrar mi juguete favorito, la intención de agarrar el juguete que ayer me regalaron, etc.). Es importante notar que todos estos son ejemplos de atribuciones hechas en la interacción. Es decir, a medida que las capacidades lingüísticas de A se complejizan, A es capaz de realizar atribuciones cada vez más complejas.

Las diferencias que nos parecen significativas en lo que hace al desarrollo de estas capacidades cada vez más complejas de atribución psicológica podrían resumirse de la siguiente manera. En primer lugar, sin más requerimientos conceptuales que los que se necesitan para categorizar estados mentales, se pueden atribuir estados mentales sin contenido, como por ejemplo estar triste (sé que alguien está triste, aunque no sepa por qué), o sentir dolor. En segundo lugar, dependiendo de la capacidad de triangular, esto es de reparar en el hecho de que hay un mundo compartido entre A y S (lo que los bebés son capaces de hacer hacia los 9 meses de vida, Trevarthen y Hubley, 1978), se vuelve

posible atribuir estados con un contenido objetual (*de re*), por ejemplo que S tiene miedo de *eso*, o que S quiere *eso* (y en este caso el indéxico es la única herramienta disponible para identificar el objeto intencional del estado mental, ya que el estado mental en este caso está dirigido al objeto en el mundo, no importa cómo se lo describa). Más adelante, cuando además de lo anterior A tiene disponibles las herramientas conceptuales mínimas indispensables para categorizar los objetos de este mundo compartido, aparece la posibilidad de atribuir estados mentales con un contenido objetual bajo un determinado modo de presentación; en este caso A es capaz de atribuirle a S miedo *a la araña*; y en este paso es donde aparece recién un primer posible desacoplamiento entre el contenido del estado mental atribuido y el mundo efectivo: lo que está presente en el entorno de A y S en este caso puede no ser una araña, sino una araña de juguete o una sombra. El paso siguiente en la complejización del contenido atribuido aparece cuando aparece el juego funcional, en estos casos hay evidencia de que el niño es capaz de realizar atribuciones psicológicas de estados mentales con contenido proposicional mínimo, es decir, es capaz de captar el contraste entre *hay X / no hay X*. Esto se hace evidente en el juego funcional, cuando el niño por ejemplo juega con una taza vacía a tomar la leche. En este caso sus capacidades perceptivas le indican que no hay nada en la taza, pero él atribuye a quien está jugando con él a tomar la leche, la intención de tomar la leche. Finalmente, con el juego de ficción aparece la capacidad de atribuir estados con contenido proposicional con la estructura *S es P* (o sea con la estructura predicativa que el lenguaje nos brinda). Se trata de casos como el famoso juego de que la banana es un teléfono (Leslie, 1987). Pero como puede verse, hay una historia interesante que contar antes de que este tipo de juego aparezca, hacia los 18-24 meses.

Nótese que en todos estos casos estamos hablando de atribuciones psicológicas realizadas en la interacción, atribuciones que paulatinamente requieren de más recursos cognitivos. Una vez que adquirimos el lenguaje, las atribuciones que realizamos en nuestras interacciones con otros humanos poseen toda la complejidad mencionada, pero no por ello dejan de ser relevan-

tes todos los indicios no conceptuales mencionados al comenzar el párrafo. En efecto, si bien las interacciones entre humanos adultos suelen involucrar intercambios lingüísticos, no son sólo las emisiones lingüísticas las evidencias de las que disponemos para realizar las atribuciones psicológicas que hacemos, sino que además tomamos en cuenta todos los elementos corporales, expresivos, como la actitud corporal de nuestro interlocutor, sus expresiones faciales, la dirección de su mirada, su timbre de voz, etc. Y cuando, ya adultos, interactuamos con bebés prelingüísticos o con mascotas, o cuando los niños en edad escolar realizan juegos protagonizados, también se está atendiendo a todos estos complejos elementos al realizar atribuciones psicológicas.

Por último, también es importante destacar que todos aquellos seres que son capaces de “leer” estas señales son capaces de entrar en interacciones de segunda persona, sean las que fueren las actuales o potenciales capacidades cognitivas del individuo: podemos entrar en este tipo de interacciones con bebés recién nacidos dado que son sensibles a este tipo de señales que los adultos emitimos (nuestro tono de voz, nuestra caricia, nuestra sonrisa dirigida a su mirada), y también con otros mamíferos, por ejemplo mascotas, que claramente reconocen el tono de voz con el que nos dirigimos a ellas, entre otras cosas. En otras palabras, una vez que somos capaces de atribuir estados psicológicos con la máxima complejidad, comenzamos a “sobreatribuir” estados mentales con complejos contenidos proposicionales a individuos que no poseen las mismas capacidades cognitivas y lingüísticas que nosotros. Y esta es la manera en la que se abre la puerta a los bebés a este mundo de complejas atribuciones psicológicas.

5. Los distintos tipos de atribución psicológica: atribuciones de segunda, de tercera y de primera persona

Recapitemos lo que hemos desarrollado hasta aquí. Las interacciones de segunda persona son el modo más básico en el que tenemos acceso a las mentes humanas, y estos mecanismos en jue-

go en las interacciones de segunda persona se mantienen activos en toda interacción humana, y aun fuera de ellas, cuando reflexionamos en soledad, vemos una película, etc. Estas interacciones son, además, la ocasión para la adquisición de los primeros conceptos psicológicos (de estados que no requieren contenido proposicional) y por lo tanto para el desarrollo de las capacidades de atribución psicológica. Estos conceptos psicológicos se adquieren en la interacción en segunda persona, y posteriormente se adquieren otros conceptos psicológicos más complejos, gracias al desarrollo del lenguaje, que permite articular contenidos proposicionales y adscribir contenidos y estados más complejos, a los demás y a uno mismo. Así, adquirimos la capacidad de combinar las evidencias que nos llevan a la atribución de segunda con la evidencia lingüística producida por las emisiones lingüísticas realizadas por nuestro interlocutor, para realizar atribuciones complejas, de estados con contenidos proposicionales que con mucho exceden las evidencia corporal, expresiva en la que se fundaban las atribuciones de segunda persona.

Pero no todas las atribuciones mentales que realizamos se dan en el ámbito de la interacción humana. En efecto, a veces nos encontramos en soledad reflexionando sobre nuestros propios estados mentales, sobre nuestra vida emocional, sobre nuestros deseos y compromisos afectivos, sobre qué decisión debemos tomar, sobre las creencias y actitudes que debemos adoptar acerca de los demás, etc. A veces espontáneamente emitimos oraciones acerca de nuestros propios estados internos, en presencia de otros o en soledad, incluso hablamos con nosotros mismos ante el espejo. También nos ocurre en ocasiones que observamos a dos individuos interactuando, y nosotros, como observadores ajenos a la situación tratamos de comprender lo que está pasando y atribuimos estados mentales diversos a los dos individuos. Es decir, no toda atribución psicológica se da en el contexto de una interacción con un semejante. Queremos sostener la idea de que todas estas formas de atribución son derivadas de la forma de atribuir propia de las interacciones de segunda persona. Y que es gracias a la adquisición de los conceptos psicológicos, de la manera en que hemos descripto, que somos capaces de realizar estas otras formas de atribución.

En efecto, en primer lugar, dado que los conceptos psicológicos son etiquetas con las que discriminamos patrones de interacción en los cuales hay involucrados aspectos internos (las sensaciones propioceptivas, etc.) y externos (la conducta expresiva propia y ajena) los conceptos psicológicos suponen criterios de aplicación tanto a la primera persona, como a la tercera persona. Una vez adquiridos, y sobre la base de los criterios correspondientes a cada caso, podemos hacer autoatribuciones basándonos exclusivamente en dirigir nuestra atención a nuestros estados internos y así pensar acerca de nuestra vida mental, o hablar de ella sin audiencia, y podemos basarnos exclusivamente en los elementos conductuales, expresivos públicos de alguien y realizar atribuciones de tercera persona a alguien con quien no estamos interactuando (alguien que vemos en la vereda de enfrente y que ni nos dirige la mirada, o un personaje de la película que estamos viendo).

Más aun, una vez adquiridos los conceptos psicológicos y con ellos la posibilidad de realizar atribuciones desinvolucradas de los estados mentales ajenos, podemos “jugar” con las dos perspectivas, de segunda y de tercera en una interacción con algún semejante. Un caso paradigmático es lo que hacemos al jugar póker. Por un lado, atendemos a los más mínimos indicios expresivos de nuestro oponente, tratamos de adivinar a través de expresiones faciales, el timbre de voz, etc., qué cartas recibió en la mano. Además, atendemos a sus conductas abiertas, y a sus emisiones lingüísticas, indicios que usamos para interpretar sus pensamientos, sus cálculos, acerca de las condiciones en las cuales estará dispuesto a apostar, y cuánto, qué patrones de riesgo asume, etc. O sea, ponemos en juego nuestros mecanismos de segunda y de tercera persona al mismo tiempo para ser eficaces y ganar la partida. Pero además, con respecto a nosotros mismos, realizamos una tarea consciente de bloqueo de las reacciones expresivas, de segunda persona, que el mundo (las cartas que tenemos en la mano) y los otros nos conducirían naturalmente a tener. Ponemos “cara de póker”. Tratamos de no ofrecer a nuestro oponente los indicios que necesita para atribuirnos los pensamientos que de hecho

tenemos (por ejemplo, que tenemos una buena o mala mano). Sin la posibilidad de volver opaca nuestra mente a nuestro oponente no ganaríamos la partida. Es evidente que la vida en las sociedades humanas está organizada sobre esta fluctuación entre nuestra capacidad de leer automáticamente aspectos de la mente ajena, pero a la vez la posibilidad de esconder nuestros pensamientos más íntimos.

6. Recapitulando: los límites de la perspectiva de segunda persona

Para resumir, las tesis principales defendidas en este trabajo son las siguientes:

- (1) Las atribuciones mentales de segunda persona son más básicas ontogenética, filogenética y conceptualmente que las de primera y tercera persona.
- (2) Para realizar una atribución, el atributor debe tener una posesión mínima, parcial, implícita, de los conceptos mentales relevantes.
- (3) La maestría de los conceptos psicológicos se adquiere en las interacciones de segunda persona. Es este hecho el que da lugar a que la maestría de todo concepto psicológico M involucre: (a) la habilidad de aplicar el concepto apropiadamente a sí mismo y a los demás; (b) la habilidad de reconocer las conductas expresivas de M como criterios/síntomas /signos de M, y (c) poseer scripts relacionados con M, al menos aquellos en juego en las interacciones en las que se produce la adquisición del concepto M (véase arriba en la sección 2).
- (4) Los conceptos mentales adquiridos durante estas interacciones de segunda persona son más básicos (es decir, conceptualmente más simples) que los otros (que requieren habilidades cognitivas independientes de la perspectiva de segunda persona).
- (5) Los conceptos mentales más complejos, como los de

actitudes proposicionales, dependen de la adquisición de específicas habilidades lingüísticas.

(6) Cuando los conceptos mentales más complejos son adquiridos, los diferentes niveles de atribución mental pueden estar simultáneamente en juego en las interacciones humanas, y también se vuelve posible la atribución psicológica por fuera de las interacciones efectivas.

Bibliografía

- Berlin, B. y P. Kay (1969), *Basic Color Terms: Their Universality and Evolution*, Berkeley: University of California Press.
- Bermúdez, J. L. (2003), *Thinking without words*, OUP.
- D'Andrade, R. G. (1989), "Cultural cognition", en: Posner, M. (ed.), *Foundations of Cognitive Science*, Cambridge: MIT Press, pp. 795-829.
- Davidoff, J., I. Davies y D. Roberson (1999), "Color categories of a stone-age tribe", *Nature* 398: 203-204.
- Davidson, D. (1970), "Mental Events", en: Davidson, D. (1980), *Essays on Actions and Events*, OUP.
- . (1975), "Thought and Talk", en: Davidson, D. (1984), *Inquiries into Truth and Interpretation*, OUP.
- . (1982), "Rational Animals", en: Davidson, D. (2001), *Subjective, Intersubjective, Objective*, OUP.
- Davies, M. y T. Stone (eds.) (1995a), *Folk Psychology: The Theory of Mind Debate*, Blackwell.
- . (1995b), *Mental Simulation. Evaluations and Applications*, Blackwell.
- De Villiers, J. y P. de Villiers (2009), "Complements enable representation of the contents of false beliefs: the evolution of a theory of theory of mind", en: Foster-Cohen, S. (ed.), *Language Acquisition*, New York: Palgrave Macmillan, pp. 169-195.
- Dennett, D. (1987), *The Intentional Stance*, Cambridge: MIT Press.
- Dretske, F. (1995), *Naturalizing the mind*, Cambridge: MIT Press.
- Dryer, M. S. y M. Haspelmath (2013), "The world atlas of language structures", online, Leipzig: Max planck institute for evolutionary anthropology, retrieved April, 19, 2015.
- Evans, G. (1982), *Varieties of Reference*, OUP.
- Goldman, A. (1993), "The Psychology of Folk Psychology", *Behavioral and Brain Sciences*, 16.
- Gomila, A. (2002), "La perspectiva de segunda persona de la atribución mental", *Azafea*, 1, pp. 123-138
- . (2012), *Verbal Minds. Language and the Architecture of Cognition*, Elsevier.

- Gopnik, A. (1993), "How we know our minds: The illusion of first-person knowledge of intentionality", *Behavioral and Brain Sciences*, 16, pp. 1-15.
- Gordon, R. (1987), *The Structure of Emotions: Investigations in Cognitive Philosophy*, Cambridge University Press.
- Hinzen, W. (2013), "Narrow syntax and the Language of Thought", *Philosophical Psychology*, 26, 1, pp. 1-23.
- Kay, P. y W. Kempton (1984), "What is the Sapir-Whorf hypothesis?", *American Anthropologist*, 86, pp. 65-79.
- Kay, P. y L. Maffi (1999), "Color Appearance and the Emergence and Evolution of Basic Color Lexicons", *American Anthropologist*, 101, pp. 743-760.
- . (2011) "Color terms", en: Dryer, M. S. y M. Haspelmath (eds.), *The World Atlas of Language Structures Online*, Munich: Max Planck Digital Library, accessed August 17, 2011; chps. 132-135.
- Kay, P. y T. Regier (2003), "Resolving the question of color naming universals", *Proceedings of the National Academy of Sciences*, 100, pp. 9085-9089.
- Kripke, S. (1979), "A puzzle about belief", en: Margalit, A. (ed.), *Meaning and Use*, Reidel, pp. 239-283.
- LeDoux J. E. (1996), *The Emotional Brain*, New York: Simon and Schuster.
- Leslie, A. (1987), "Pretence and representation: the origins of "Theory of mind"", *Psychological Review*, 94, pp. 412-426.
- Lucy, J. y R. Shweder (1979), "Whorf and his critics: linguistic and non-linguistic influences on color memory", *American Anthropologist*, 81, pp. 581-615.
- Lupyan, G. (2008a), "The conceptual grouping effect: categories matter (and named categories matter more)", *Cognition*, 108, pp. 566-577.
- . (2008b), "From chair to "chair": a representational shift account of object labeling effects on memory", *Journal of Experimental Psychology: General*, 137, pp. 348-369.
- Lupyan, G., D. H. Rakison y J. K. McClelland (2007), "Language is not just for talking: redundant labels facilitate learning of novel categories", *Psychological Science*, 18, pp. 1077-1083.
- Lupyan, G., S. L. Thompson-Schill y D. Swingley (2010), "Conceptual penetration of visual processing", *Psychological Science*, 21, pp. 682-691.

- Nagel, J. (2015), "Knowledge as a Mental State", *Oxford Studies in Epistemology*, 4, pp. 273-308.
- Pérez, D. (2013), *Sentir, desear, creer. Una aproximación filosófica a los conceptos psicológicos*, Buenos Aires: Prometeo.
- Pérez, D. y S. Español (2014), "Intersubjetividad y atribución psicológica", en: Cepeda, P., C. Mantilla y P. Quintanilla, *Cognición social y lenguaje: La intersubjetividad en la evolución de la especie y en el desarrollo del niño*, Lima: Fondo Editorial de la Pontificia Universidad Católica del Perú.
- Phillings, M. et al. (2003), "Is color categorical perception really perceptual?", *Memory and Cognition*, 31, pp. 538-551.
- Roberson, D. y J. Davidoff (2000), "The categorical perception of colours and facial expressions: the effect of verbal interference", *Memory and Cognition*, 28, pp. 977-986.
- Roberson, D. et al. (2004), "Colour categories and category acquisition in Himba and English", *Journal of Experimental Psychology: General*, 133, pp. 554-571.
- Roberson, D., I. Davies y J. Davidoff (2000), "Color categories are not universal: replications and new evidence from a stone-age culture", *Journal of Experimental Psychology: General*, 129, pp. 369-398.
- Rosch, E. (1975), "Cognitive representation of semantic categories", *Journal of Experimental Psychology*, 104(3), pp. 192-233.
- Ryle, G. (1949/1967), *El concepto de lo mental*, Buenos Aires: Paidós.
- Trevarthen, C. y P. Hubley (1978), "Secondary intersubjectivity: Confidence, confiding and acts of meaning in the first year", en: Lock, A. (ed.), *Action, gesture and symbol: The emergence of Language*, Londres: Academic Press, pp. 183-229.
- Wittgenstein, L. (1967), *Zettel*, Oxford: Basil Blackwell.
- Witthoff, N. et al. (2003), "Effects of language on color discriminability", en: Alterman, R. y D. Kirsh (eds.), *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*. Hillsdale: Lawrence Erlbaum, pp. 1247-1252.

Atribución y teorías de emociones: escenas de la vida cotidiana

Andrea F. Melamed

Estela y Carlos suelen ir a cenar al mismo restaurante todos los domingos. Esta vez, además, celebran que Carlos ha recibido un premio por su trayectoria profesional. Conversan plácidamente mientras esperan su mesa. Sin embargo, la atención de Carlos se desvía cuando ve que un grupo de personas ingresa al establecimiento. Al ver a Carlos se aproximan a saludarlo, entre ellos, unas muchachas muy simpáticas que muy efusivamente lo congratulan. Estela observa tímidamente la situación y sonríe, esperando la debida presentación, pero luego de un breve intercambio de abrazos y apretones de manos, ellos se alejan. Estela se incomoda. Al cabo de unos minutos, son invitados a sentarse. Se acerca el mozo, ofrece un menú, Carlos lo toma, y el mozo se retira. Carlos quiere conversar con su esposa acerca de qué ordenar, pero se encuentra a Estela con el ceño fruncido, un tono de voz elevado y una actitud general molesta. Al preguntarle qué le sucede, Estela dice que está bien, que nada pasa, mientras empuña el cuchillo, rabiosa. Rápidamente reacciona y nota que en efecto se encuentra ‘alterada’ y agrega:

—Sí, estoy disgustada, el mozo me ofendió, al ignorarme y no entregarme un menú, me faltó el respeto.

1. Introducción

En los últimos treinta años, mucho se ha escrito en torno a las emociones, a su naturaleza, su función, su historia filogenética. Entre ellas, la cuestión de la relación de las emociones con la cognición ha dominado gran parte del debate, dando lugar principalmente a dos versiones del fenómeno. Como resultado, la escena actual exhibe dos grandes enfoques o modos de aproxi-

mación al fenómeno que, presentándose como incompatibles, buscan ambos dar cuenta de qué son y cómo funcionan las emociones. Por ejemplo, Morris (2002) recoge el debate en términos de teorías cognitivas y aproximaciones periféricas, mientras que Charland (1997) las presenta en términos de teorías cognitivas y teorías perceptivas, e Izard (1986) lo hace en términos de teorías cognitivas y teorías dinámicas. Esta disputa incluye tanto una tipificación de modos de abordaje de las preguntas acerca de las emociones, como la defensa explícita de alguno de estos modos por parte de los autores que pretenden brindar caracterizaciones de las emociones. De este modo, en la literatura filosófica y psicológica sobre emociones encontramos presentaciones que toman como un hecho la existencia de estos dos enfoques que, siendo presuntamente irreconciliables, buscan dar cuenta del mismo fenómeno.

Este trabajo apunta a revisar el desacuerdo que mantienen los enfoques somáticos (no cognitivo) y cognitivo, en particular, desde el punto de vista de la atribución de las emociones. Este objetivo específico se inserta en el contexto de un propósito más amplio, vinculado a esclarecer el sentido real de tal división entre dos enfoques. A tal efecto, en la primera parte del trabajo me ocuparé de presentar las propuestas de algunos autores paradigmáticos de cada una de las corrientes, a fin exhibir, por una parte, el espíritu compartido entre los autores que forman parte de un mismo enfoque, así como también, en segundo lugar, los puntos de discordia entre las perspectivas. El examen de las posiciones (tanto somáticas como cognitivas) inevitablemente me obliga a efectuar cierto recorte, pues una revisión completa de todos los autores que han ofrecido aportes resulta imposible, más aun a los efectos de este trabajo. No obstante, puesto que la elección de ciertos autores (de los cuales sí ofreceré un análisis más detallado) en definitiva persigue fines más bien ejemplares, nos será igualmente posible ilustrar los conceptos o tesis fundamentales de cada corriente. En particular, la selección resultante obedeció a la intención de ofrecer una presentación de cada perspectiva que fuera variada, pero a la vez no resultara redundante. De modo que a efectos de caracterizar la 'corriente somática'

presentaré en primer lugar la posición de William James, famosa tanto por el carácter fundacional a la vez que revolucionario de su propuesta como por su audacia teórica. Además, expondré brevemente la posición de Robert Zajonc, que está guiada por descubrimientos provenientes de la experimentación en psicología empírica. Con respecto a la ‘corriente cognitiva’, presentaré tres aproximaciones. En la primera, proveniente del ámbito de la filosofía, Robert C. Solomon reflexiona sobre la tradición que relegó a las emociones al terreno de la pasividad, procurando restituir con su análisis el carácter racional que las emociones nunca debieron perder; las otras, provenientes del campo de la psicología, por un lado, la de Richard Lazarus, y por otro, la versión cognitiva mixta proveniente de los experimentos de Schachter y Singer.

En la segunda parte del trabajo tomaré los enfoques de manera global, contraponiéndolos específicamente por sus capacidades atributivas. Al hacerlo, además, veremos explicitados un conjunto de supuestos hasta el momento ocultos.

2. Perspectivas sobre la atribución emocional

2.1. La corriente somática

2.1.1. La teoría de los sentires¹: la audacia de William James

Como tantos otros casos de codescubrimiento científico, la denominada “teoría de la sensación” fue propuesta por dos pensadores de modo independiente: William James (1884) y Carl G.

1 La traducción del término inglés “*feeling*” es problemática. Algunos han optado por traducirlo como “sentimiento” (por ejemplo, la traducción española de Kandel, Schwartz y Jessell (1997), elección que propicia la confusión con el “*sentiment*”, que claramente es otra cosa. Otra traducción posible es “sensación”, pero esta traducción no es completamente satisfactoria, puesto que como señala Hacker (2009) las sensaciones son sólo un subtipo o especie del género “*feeling*”, que también incluye percepciones, afecciones y apetitos como otras subclases. De modo que hablar de sensaciones podría llevar a desatender a las otras subclases de *feelings*. Es por ello que opté por traducir “*feelings*” como ‘sentires’, buscando subrayar la raíz semántica “*feel*”, ‘sentir’, pero evitando su identificación estrictamente ni con sentimientos ni con las sensaciones (que, entre otras cosas, carecen de intencionalidad).

Lange (1885). La propuesta atrajo especial atención en el ámbito científico y académico, puesto que sugería que era erróneo el modo habitual en que las emociones eran concebidas en ese momento.² Este modo tradicional o estándar concebía a la emoción como el resultado de una percepción, y la causa de la expresión física. Es decir, de acuerdo con esta, la percepción de un objeto o evento despertaba un estado emocional que, a su vez, generaba una manifestación a través de cambios corporales.

James se propone argumentar contra este modo de concebir las emociones, y afirma:

Mi tesis es que los cambios corporales siguen directamente la percepción del hecho, y que nuestro sentir [*feeling*] de esos mismos cambios mientras ocurren es la emoción (James, 1884: 189-190, énfasis del autor).

Es decir, el sentido común dicta que, si enfrentamos un peligro, entonces tenemos miedo y (luego o por ello) huimos. La sugerencia de James apunta a invertir la secuencia anterior, sin dejar fuera a ninguna de las variables que intervenían. Según esta: “un estado mental no es inducido inmediatamente por el otro, las manifestaciones corporales deben interponerse”. De modo que lo correcto sería decir que, ante la amenaza de peligro, los cambios corporales surgen de modo inmediato, y es la percepción de tales cambios fisiológicos y conductuales la razón por la que sentimos miedo: “[...] nos sentimos tristes porque lloramos, furiosos porque golpeamos, o asustados porque temblamos; no es que lloremos, golpeemos o temblemos porque estemos tristes, furiosos o asustados, como cabría esperar” (James, 1884: 190).

De esta concepción de las emociones como *sentires*, merecen ser destacadas dos cuestiones que serán centrales para la discusión que sigue.

i. La centralidad del sentir: El rol que la sensación recibe en este

2 Cabe aclarar que para James el dominio de aplicación de la teoría que presenta es el de las comúnmente denominadas “emociones básicas”. Es sobre este subconjunto de emociones que versa la discusión.

esquema es clave, al extremo de que la emoción es ni más ni menos que esa sensación, causada directamente por la respuesta fisiológica desencadenada luego de la percepción del objeto. Ahora bien, esta es una sensación de algo: de los cambios corporales. De modo que el núcleo de una emoción está compuesto de un conjunto exclusivo (único) de cambios corporales, que son a su vez “sentidos” [*felt*] o de los que tenemos una sensación. Lo que resulta doblemente importante, en primer lugar, porque esa univocidad en la conjunción de cambios corporales será lo que explique la variedad de tipos emocionales, i.e. a cada conjunto de cambios corporales, seguidos de la sensación, le corresponde (produce) un tipo emocional distintivo. En segundo lugar, porque como dijera James “sin los cambios corporales que siguen a la percepción [del objeto], la última sería puramente cognitiva en forma, pálida, sosa, desprovista de la calidez emocional”. En efecto, James mantiene que tales cambios corporales son siempre sentidos [*felt*] en el momento en que tienen lugar: “las emociones disociadas de la sensación corporal son inconcebibles” (James, 1890: 745). Todo lo expuesto queda exquisitamente resumido en el siguiente párrafo:

Si imaginamos una fuerte emoción, y luego tratamos de abstraer de nuestra conciencia de ella todos los sentires [*feelings*] de sus síntomas corporales característicos, hallaremos que no nos quedó nada, ninguna “substancia mental” de la cual pueda ser constituida la emoción; todo lo que nos queda es un estado frío y neutro de percepción intelectual (James, 1884: 193).

Es decir, sin los cambios corporales que siguen la percepción de un evento, si tuviéramos un encuentro peligroso, podríamos juzgar que lo mejor es huir, pero en rigor, no sentiríamos miedo.

ii. La inmediatez de la emoción: Las emociones, definidas como la sensación de los cambios corporales, no están constreñidas por ningún tipo de mediador o disparador de corte cognitivo. Esta equiparación entre emoción y sensación visceral ha motivado la exigencia de un subconjunto de cambios corporales marcadamente distinto para cada emoción distinguible, exigencia que,

como veremos a continuación, se convirtió en una sólida réplica a la teoría.

2.1.2. Las preferencias no necesitan inferencias: R. B. Zajonc

Un siglo más tarde, Zajonc (1980) pone en tela de juicio la aproximación dominante de su época que sostiene que todas las reacciones afectivas son consecuencia de un proceso cognitivo previo.³ Como consecuencia, consigue socavar una implicancia directa de esa concepción: la idea de que conocer (categorizar) un objeto sea condición de posibilidad para poder ser afectado por él.

El recorrido que realiza para arribar a la conclusión de que existen (como mínimo, algunas) emociones independientes de la cognición, surge del examen de un tipo particular de emoción: el implicado en las preferencias. Es decir, es a partir del estudio pormenorizado de las preferencias, que sostiene “que es enteramente posible que el primer estadio de la reacción de un sujeto ante un estímulo sea afectivo” (Zajonc, 1980: 154). Su defensa se articula a partir de la presentación de un conjunto de experimentos destinados a mostrar que las preferencias no se producen como consecuencia de la categorización de los objetos (sino, dirá él, todo lo contrario). En particular, el fenómeno denominado “efecto de mera exposición” muestra cómo los sujetos que han sido expuestos a determinadas figuras –por ejemplo a ideogramas chinos– tienden a preferir los previamente expuestos sobre los nuevos, mostrando que la mera exposición es suficiente para crear preferencias. Estos resultados son reforzados con una segunda serie de experimentos, donde la exposición a los estímulos es subliminal. En la última, a pesar de que los sujetos no eran capaces de reportar cuáles objetos habían sido previamente observados, el efecto de mera exposición se mantuvo.⁴

3 Nótese la dinámica que la historia del debate forjó: la defensa no-cognitiva de Zajonc se alza contra la posición cognitiva predominante en la época; que, a su vez, había sido diseñada en respuesta al descrédito que habían recibido las emociones por parte de otras posiciones no cognitivas que dominaran en su momento (por ejemplo, por parte del conductismo).

4 Véase Zajonc (2001) para una versión más reciente del fenómeno de la mera exposición. Otros resultados experimentales que refuerzan la hipótesis de la

Es a partir de estos resultados que Zajonc considera refutada la tesis general de que debemos conocer un objeto para poder establecer si nos agrada o no: “es posible que algo nos pueda agradar, o que le podamos temer a algo, antes de que conozcamos con precisión qué es y quizás incluso sin conocer qué es” (Zajonc, 1980: 154). En pocas palabras, según la perspectiva afectiva que defiende Zajonc:

Compramos los autos que nos “gustan”, elegimos los empleos y las casas que encontramos “atractivas”, y luego justificamos esas elecciones por varias razones que parezcan convincentes a otros, que nos preguntarán “¿Por qué ese auto?” o “¿Por qué esa casa?”. Nosotros no necesitamos convencernos. Nosotros sabemos qué queremos (Zajonc, 1980: 155, énfasis del autor).

La caracterización de las reacciones afectivas (en oposición a las cogniciones frías) se completa con los siguientes rasgos: son ineludibles (no pueden ser controladas voluntariamente por procesos atencionales), irrevocables (una vez formadas, no cesan), inmediatas, imprecisas, difíciles de verbalizar, no obstante, fáciles de comunicar y comprender.

En síntesis, podemos decir que, al menos respecto a este conjunto de emociones, las reacciones afectivas (específicamente las implicadas en las preferencias) tienen primacía y son independientes de la cognición, es decir, pueden existir antes e inclusive sin cognición mediante. Este punto es fundamental y merece ser destacado: Zajonc no pretende con esto dar por demostrado que todos los tipos de emociones sean de esta clase. Por el con-

primacía de la afición son los de Murphy y Zajonc (1993) y Winkielman, Zajonc y Schwarz (1997). También Adolphs et al. (2005) presentan evidencia en favor de la posibilidad de discriminación (y preferencia) entre estímulos que no son conscientemente percibidos (o recordados), sugiriendo una disociación fuerte entre la preferencia y el reconocimiento. Según ellos mismos, esta disociación sería además compatible con la distinción entre, por una parte, una vía gustativa ventral límbica (que incluye a la amígdala, al hipotálamo y a regiones del ganglio basal) que parece ser suficiente para una discriminación conductual básica del gusto y, por otra parte, de una vía cortical dorsal que sería necesaria para un procesamiento y aprendizaje gustativo complejo (Pfaffmann, Norgren y Grill, 1977).

trario, procura que las emociones (en particular las reacciones afectivas) se vean salvadas del error análogo que cometen los teóricos cognitivos, que del hecho de que la cognición pueda dar lugar por sí misma a estados emocionales, pretenden concluir que el componente cognitivo es un componente necesario para toda clase de emoción. En particular, Zajonc destaca que si por definición se pretende que las afecciones tengan una evaluación cognitiva como precondition necesaria, es menester descubrir de qué modo entraría la cognición para cada uno de estos fenómenos. Si uno se empeña en postular precursores cognitivos para todas las emociones “uno está forzado a admitir la reducción de la cognición a procesos tan mínimos como los disparos de las células retinales. Por tanto, si aceptamos la posición [cognitiva] de Lazarus, todas las distinciones entre percepción, cognición y sensación, desaparecen” (Zajonc, 1984: 121).

El balance de la propuesta de Zajonc es haber encontrado un resquicio donde situar a las emociones que carecen de elementos cognitivos que funcionen como mediadores o les den forma. Al tiempo que reconoce que lo anterior es absolutamente compatible con la tesis de que existen otros tipos de emociones que son alcanzadas sólo a través de juicios evaluativos.

2.2. La corriente cognitiva

La teorización de las emociones ha cambiado en últimos cuarenta años, en gran medida debido al interés por el rol de la cognición en las emociones (Frijda, Manstead y Bem, 2000).⁵ Así es que, buscando corregir en cierto modo la suerte que corrieran las emociones en su tratamiento a lo largo de la historia de la filosofía, esto es, como aquello que se oponía a la razón o la distorsionaba –lo que motivó su desprecio y desatención durante tanto tiempo–, algunos pensadores se vieron tentados a definir a las emociones a partir del elemento racional por excelencia, el pensamiento, la actividad cognitiva. De esta manera, se conse-

5 Resulta notable el vínculo que guarda este modo específico de concebir las emociones con el surgimiento de la tradición cognitiva en psicología, fundamentalmente guiada por la reacción contra la ortodoxia conductista (Bearison & Zimiles, 1986: capítulo 1)

guía revertir una limitación decisiva, aunque de la mano de estas especificaciones nació un conjunto de nuevas dificultades. La principal consecuencia de este movimiento resultó en que se estableciera que estudiar las emociones equivalía a estudiar la cognición que las dispara o constituye. De esta manera, la investigación de las emociones se concibe a partir de la tipificación de los tipos de creencias (u otros tipos de estados cognitivos) que constituyen una emoción.

2.2.1. La racionalidad de las emociones: Robert C. Solomon

Solomon (2003: prefacio) subraya su interés por combatir el ‘antiguo prejuicio’ de que las emociones simplemente nos suceden, y de que está absolutamente fuera de nuestro alcance controlarlas, dando por tierra la tradicional rivalidad entre la racionalidad y las emociones. La tesis más fuerte de Solomon es que las emociones son juicios normativos y frecuentemente morales. Estar enojada con Juan porque tomó mi auto implica que yo creo que de algún modo Juan me ha ofendido, lo cual es independiente del hecho de que Juan efectivamente me haya ofendido, o incluso de que haya tomado mi auto.

El juicio (moral) involucrado por mi cólera no es un juicio acerca de mi cólera (aunque alguien más pueda juzgar si mi enojo es justificado o injustificado, racional, prudente, tonto, indulgente, terapéutico, beneficioso, desafortunado, patológico o gracioso). Mi cólera es ese juicio (Solomon, 1973: 8, énfasis del autor).

En otras palabras, tener una emoción es realizar un juicio normativo acerca de la situación presente, pero el objeto de una emoción no puede ser simplemente un hecho: el objeto emocional únicamente puede ser caracterizado de modo completo como objeto de mi ira. Esto quiere decir que un evento o la mera percepción de un evento no es suficiente para producir una emoción: esta involucra necesariamente una evaluación personal de la significación del incidente (Solomon, 1976: 187). Brevemente, “mi cólera-con-Juan-por-robar-mi-auto es inseparable de mi juicio de que Juan me ofendió, mientras que es claro que el hecho de

que Juan robó mi auto es muy distinto de mi cólera o mi juicio” (Solomon, 1973: 8).⁶

De su posición se sigue que ante un cambio en el juicio, concomitantemente se produzca una variación en la emoción,⁷ de modo que bajo esta descripción “soy responsable de mis emociones como lo soy de los juicios que realizo” (Solomon, 1973: 10). Esta afirmación, aunque polémica, resulta comprensible a la luz del espíritu de su propuesta. Su perspectiva cognitiva representa un desafío a la división tajante entre emociones y racionalidad, rechazando la concepción ampliamente aceptada de que las emociones son involuntarias e irracionales (Solomon, 2001: 178). En particular, afirma que somos responsables de nuestras emociones, impugnando fundamentalmente la alternativa que utiliza a las emociones como excusas, como fenómenos que sufrimos, que simplemente nos suceden, que atravesamos con completa pasividad, haciendo imposible la atribución de responsabilidad. Es en este último sentido que debe entenderse su defensa de la elegibilidad de las emociones: las emociones son juicios, juicios que hacemos. Eso no quiere decir que simplemente podamos optar por juzgar una situación como ventajosa (o peligrosa), es decir, no significa que podamos elegir en sentido fuerte qué emoción tendremos en cada momento. Por el contrario, significa que juzgar es algo que hacemos activamente, no algo que padecemos.

Pero si las emociones son juicios ¿tiene sentido conservar la distinción entre emociones y juicios? ¿Debe abandonarse la distinción intuitiva, proveniente del sentido común, entre juicios y emociones, que nos lleva a pensar que pertenecen a ámbitos distintos, como si nombraran realidades mentales diversas? Anticipando esta posible objeción, Solomon traza una distin-

6 Sin embargo, no todos los juicios evaluativos son emociones. Las emociones recortan, sobre el dominio de los juicios que ofrecen valoraciones de los sucesos del entorno, aquellos que son “acerca de uno mismo y relativamente intensos” (Solomon, 1976: 188).

7 Nótese la diferencia con la perspectiva somática, donde es manifiesta la irrevocabilidad de la emoción, es decir, la inmutabilidad de la emoción ante la modificación en el conjunto de creencias.

ción en el seno de los juicios, mostrando que las emociones conformarían un subtipo dentro de los juicios: las emociones deben ser equiparadas a juicios, pero deben ser distinguidas de los “juicios fríos” [“cool” judgments] y la acción normal, racional y deliberada,⁸ en tanto las respuestas emocionales son producidas con urgencia o rapidez, pero principalmente en tanto carecen de “frialidad”. Según esta idea, las emociones son respuestas apremiantes a situaciones inesperadas, para las cuales uno no se encuentra preparado, es decir, para las cuales uno no dispone de patrones habituales de conducta. Una vez más, aquello a lo que le cabe el mote de irracional es a la situación, no a la emoción.⁹

Por último cabe destacar que su concepción de las emociones como juicios no debe ser confundida con otras propuestas, también de índole cognitivas, que parecen articular la noción de emoción a partir del concepto de creencia. Es decir, aunque su afirmación “las emociones son juicios” pueda parecer meramente un slogan, la elección de sus términos no es casual. Según sus propias palabras ‘las creencias parecen ser demasiado articuladas para la reacción no reflexiva que caracteriza a la mayoría de las emociones’ (Solomon, 1993: 12).¹⁰ Puesto que “el juicio parece tener el alcance y flexibilidad para aplicarse a cualquier cosa desde la emoción animal e infantil hasta las emociones humanas más complejas y so-

8 Como veremos, comparte este punto con Lazarus, quien situándose también en el enfoque cognitivo concibe el carácter cognitivo como algo distinto de la actividad mental consciente y deliberada.

9 En consonancia con James (1884), quien se refería al resultado de la abstracción de todos las sensaciones de los cambios corporales característicos de una emoción como “un estado de percepción intelectual frío y neutral”. También Zajonc (1980) equipara la distinción entre juicio (o cognición) fría/caliente a la de cognitivo/afectivo, respectivamente.

10 También Gordon (1987) toma distancia del enfoque según el cual el miedo siempre involucra la *creencia* de que uno está en una situación de peligro, o la creencia de que algo es peligroso. Gordon rescata de esta perspectiva solamente la idea de que las emociones guardan una similitud funcional con las creencias: aunque uno piense que es altamente improbable que aquello que uno teme que suceda, realmente ocurra, uno actúa *como si* uno creyera o supiera que lo hace (Gordon, 1987: 84).

fisticadas como los celos o la indignación moral” (Solomon, 2001: 187).¹¹

2.2.2. Richard Lazarus: Las relaciones entre las emociones y la cognición

La concepción de Lazarus puede sintetizarse bajo la idea de que cierto tipo de pensamiento o cognición es una precondition necesaria para toda emoción. En particular, Lazarus (1982) postula que son las valoraciones cognitivas las que vinculan y median entre los sujetos y el ambiente, provocando emociones particulares como resultado de las evaluaciones específicas que realiza el sujeto, de su relación con ambiente en virtud de su bienestar.¹² La tesis general que guía su posición es que “cognición y emoción están usualmente fusionadas en la naturaleza” (Lazarus, 1982: 1019).¹³

En su argumentación, se vale del punto de vista de Zajonc para discutir contra la posibilidad de que existan episodios emocionales propiamente dichos que prescindan de todo elemento cognitivo entre sus disparadores. El ataque puntual a la perspectiva afectiva de Zajonc le servirá como dispositivo para argumentar a favor de su propia concepción cognitiva, esto es, la que sostiene que cierta actividad cognitiva es condición necesaria y suficiente para toda emoción (Lazarus, 1982: 1019). Aplicando la máxima de que la mejor estrategia de defensa es un buen ataque,

11 No obstante, y muy a pesar de sus esfuerzos por subrayar la distinción entre lo que él llama “juicios” y las creencias, la distinción no queda trazada de modo categórico. Encontramos entre sus afirmaciones que “nuestras emociones son altamente dependientes de nuestras opiniones y creencias”, que un cambio en mis creencias implica modificaciones en mi emoción y que “no puedo estar enojado si no creo que alguien me ha ofendido” (Solomon, 1976: 187).

12 Sigue explícitamente y extiende la perspectiva de Arnold (1960). En Lazarus, Averill y Oprton (1970) afirma que cada reacción emocional (al margen de su contenido) es una “función de un tipo particular de cognición o *valoración*” (218, cursivas de los autores). De igual modo Lyons (1999) dice: “tu emoción es causada por aquello que conoces acerca de la situación y su relevancia para ti, y no por lo que algún observador objetivo especula o sabe de la situación” (39).

13 Más adelante le añadirá un tercer elemento, la motivación, dando lugar a su famosa tesis de la conjunción e interdependencia de emoción, cognición y motivación (Lazarus, 1999: 13).

busca socavar la perspectiva afectiva atribuyéndole una incorrecta comprensión de la cognición. Según Lazarus, el responsable del desacuerdo es el modo inadecuado en que Zajonc concibe a la cognición, según el cual:

Si uno acepta el principio de que el significado se encuentra al final de un proceso cognitivo seriado, entonces, acomodar el hecho de que podemos reaccionar emotivamente de modo instantáneo nos fuerza a abandonar la idea de que la emoción y la cognición están necesariamente conectadas causalmente (Lazarus, 1982: 1021).

No obstante, arguye Lazarus, la cognición debe separarse conceptualmente de la racionalidad y la deliberación: como resultado de la herencia neural y su experiencia, un sujeto tiene esquemas cognitivos que ante un sonido brusco significan peligro instantáneamente. A pesar de que los esquemas requeridos en asuntos humanos pueden ser más complejos, la valoración del miedo no debe ser necesariamente deliberada (Lazarus, 1982: 1022). Asimismo, las evaluaciones cognitivas no deben implicar necesariamente conciencia de los factores sobre los que descansa. Afirma:

Estaría de acuerdo con que una persona no necesita ser consciente de sus evaluaciones cognitivas y puede utilizar una lógica primitiva, pero argumentaría en contra de la idea de que algunas evaluaciones (Zajonc se refiere a las preferencias) no son cognitivas (Lazarus, 1982: 1022).

La cuestión clave, entonces descansa en cómo caracterizar ese prerequisite cognitivo que, como anteriormente fue señalado, no implicaría percatación ni deliberación. Lazarus se pregunta más puntualmente ¿qué podría transformar estados sensoriales en emociones?¹⁴ (1984: 126). Mostrando cierta coincidencia

14 Nótese que esta formulación incluye una serie de presupuestos, de modo que en la mera presentación del problema se desliza una respuesta preliminar. Por un lado, establece que las emociones son estados distintos de los estados sensoriales, estados que, por otro lado, precisan de una ulterior elaboración

con la fórmula de Zajonc, Lazarus distingue *dos modos de valorar*: uno automático¹⁵, no-reflexivo e inconsciente o preconsciente; el otro, consciente y deliberado. Y destaca que “no es posible decir con seguridad qué proporción de valoraciones y emociones están basadas en cada modo de actividad cognitiva, y quizás la mayoría de las valoraciones de los adultos involucra una mezcla de ambas” (1991: 155). Su respuesta será que aquella transformación necesaria para producir una emoción a partir de estados sensoriales “es una valoración de aquellos estados como favorables o dañinos para nuestro bienestar” (Lazarus, 1984: 126). De modo que, habiendo desechado la posibilidad de que una emoción ocurra en ausencia de algún participante cognitivo, es decir, habiendo respondido negativamente a la pregunta que se hacía Zajonc acerca de la posibilidad de que inputs puramente sensoriales, no transformados, generen directamente emociones (Zajonc, 1984: 122), Lazarus propone un reemplazo del interrogante. La verdadera incógnita para él es cómo la cognición moldea las emociones, es decir, qué tipo/s de cognición es/son capaces de despertar emociones de diferentes intensidades y clases, tal como el miedo, la cólera, la culpa, la felicidad, etc.

Debe advertirse que la disputa que Lazarus mantiene con la perspectiva somática en general, y con Zajonc en particular, comprende también la interpretación de la evidencia experimental disponible. De acuerdo con Lazarus, Zajonc no logra realmente eliminar la posibilidad de que alguna actividad cognitiva se halle involucrada y afirma:

La esencia de mi posición es, de hecho, que en esta etapa de la teoría, el conocimiento y el método, Zajonc no puede probar que una cognición no esté presente en una emoción, menos aún an-

(o transformación) para dar lugar al estado emocional. Además, se postula a la cognición (o componente cognitivo) como el operador responsable de aquella transformación –necesaria por definición– de los inputs sensoriales, que da por resultado a los estados emocionales.

- 15 Advierte Lazarus que tal carácter automático no debe ser confundido con una condición primitiva, puesto que el procesamiento automático admite “significados complejos, abstractos y simbólicos, que a través de la experiencia pueden ser condensados en significado instantáneo” (1991: 155).

tes de que ocurra, tanto como yo no puedo probar que esté presente (Lazarus, 1984: 126).

Sin embargo, agrega:

Si la cuestión central fuera ¿las evaluaciones cognitivas afectan a las emociones?, en lugar de ¿las emociones requieren evaluaciones cognitivas?, habría abundante evidencia a favor (Lazarus, 1984: 127).

2.2.3. Una teoría cognitiva impura: el experimento de Schachter y Singer

La experiencia desarrollada por Schachter y Singer (1962) merece ser destacada tanto por su relevancia histórica como por las múltiples interpretaciones que estos resultados han recibido.¹⁶ Esta experiencia ha sido interpretada por sus responsables y autores como evidencia en favor del carácter cognitivo de las emociones. Prinz (2004) llama a esta postura “teoría etiquetadora cognitiva” [*cognitive labeling theory*] y la sitúa entre las teorías cognitivas impuras dado que para ésta una emoción involucra tanto cambios corporales (estados fisiológicos) como *interpretaciones cognitivas* de esos estados.¹⁷ Muchos otros autores, luego, se han apoyado en estos resultados para retratar posiciones de corte cognitivo, a pesar de los matices (Charland, 1997; Lazarus, 1991; Lyons, 1980; Power y Dalgleish, 2008; Prinz, 2004; Solomon, 1976).

Este experimento busca desafiar la tesis jamesiana de la suficiencia del sentir de cambios corporales para caracterizar a las

16 En rigor, el primero en efectuar una aplicación de este paradigma experimental es Marañón (1924). Una de las diferencias entre las experiencias de Marañón y Schachter-Singer es que los últimos enmascararon el procedimiento que llevaban adelante, haciéndole creer a los participantes que se trataba de una prueba de visión. Marañón (según sospechan Schachter y Singer) habría sido menos cuidadoso en ese respecto, y los sujetos que participaron en su experiencia habrían estado al tanto de los procedimientos involucrados y sus efectos.

17 El determinante crucial para el tipo de emoción experimentada era cómo el individuo *explicaba* la activación [*arousal*].

emociones. La hipótesis que lo guía es que para que estos cambios corporales califiquen como estados emocionales deben estar acompañados de un juicio o categorización que vinculen esos cambios a ciertos objetos emotivamente relevantes. Para ello, los experimentadores se propusieron manipular los determinantes de los estados emocionales, que desde su perspectiva incluyen aspectos tanto fisiológicos como situacionales. Para este experimento, se le indicó al total de voluntarios que el experimento tenía como objetivo probar los efectos de una vitamina que mejoraba la visión (Suproxin). Siendo la premisa del experimento que los cambios corporales inducidos artificialmente no generarían por sí solos el estado emocional, los experimentadores procedieron a la manipulación fisiológica, con la inyección intravenosa de epinefrina (adrenalina) –conocida por provocar excitación en el sistema autonómico– en uno de los grupos (GA) y con la de una solución salina (placebo) en el grupo control (GP). Dentro del grupo que recibió la inyección de adrenalina, un subgrupo fue advertido acerca de los efectos colaterales del Suproxin, tal como la aceleración del ritmo cardíaco (GAI_{nf}), mientras que otro subgrupo recibió información falsa acerca de sus efectos, como por ejemplo que producía dolores de cabeza (GAMa_lin_f) y los restantes no recibieron información alguna, ignorando completamente sus efectos posibles (GAI_{gn}). Con esto, se buscó la manipulación del alcance de las explicaciones que el sujeto da de sus propios estados corporales: los sujetos del GAI_{nf} tendrían una explicación apropiada de sus estados internos, mientras que los del GAMa_lin_f tendrían una explicación inadecuada, y los del GAI_{gn}, ninguna explicación. Mientras se los hacía esperar para efectuar una prueba de visión (que, naturalmente, nunca llegaría), en una sala los acompañaba un experimentador que actuaba eufóricamente, haciendo entre otras cosas aviones de papel (S1), mientras que en otra habitación, otro secuaz fingía enojarse por las preguntas contenidas en un cuestionario (S2), dando lugar así a la creación de dos situaciones diferentes, de las que los sujetos derivarán cogniciones explicativas (Schachter y Singer, 1962: 382). Finalmente, luego de la espera, todos los sujetos debieron completar un cuestionario acerca de sus estados físicos y psico-

lógicos. Schachter y Singer, responsables del experimento, observaron que los sujetos que estaban con el experimentador que actuaba eufóricamente (S1) se comportaban alegremente, mientras que los sujetos que interactuaban con el secuaz iracundo, actuaban como si estuvieran enojados. Desde un punto de vista estrictamente fisiológico, la epinefrina actuó de acuerdo a lo esperado, aumentando la frecuencia cardíaca, en comparación con el placebo. Según los autores, en la situación S1 (euforia) quedaba claro que “los sujetos eran más susceptibles al humor del experimentador infiltrado cuando no tenían explicación de sus propios estados”, en comparación con aquellos que sí disponían de esa información. Mientras que en el análisis de la S2 (enojo) se sospechó del informe personal en tanto suponían que el sujeto no expresaría enojo de participar en la prueba. Sin embargo, la conducta observada exhibió que los GAIgn estaban mucho más enojados que los GAInf y los GP.

Lo curioso, empero, es que, desde mi punto de vista, esta experiencia podría ser presentada como evidencia en favor de la tesis contraria, puesto que la excitación inducida artificialmente consiguió despertar emociones en los individuos, aun cuando posteriormente ellos se percataran de que ese estado no fue originado por verdaderos estímulos emotivos, sino que por el contrario fueron generados en una situación anómala, lo que naturalmente produjo una disminución en su estado emocional. En ese sentido, acuerdo con (Frijda, 1994) en que las variables cognitivas pueden servir para limitar las ocurrencias emocionales, en lugar de funcionar como variable en la generación de emociones. En pocas palabras, la cognición quizás tenga más que ver con el control de las emociones que con su producción.

También Zajonc interpreta los resultados de Schachter y Singer como evidencia de que los “estados afectivos pueden ser inducidos por procedimientos no cognitivos y no perceptuales”. La excitación emocional puede ser inducida por drogas, hormonas o hasta por estimulación eléctrica del cerebro: por ejemplo, si se esconde *valium* en determinado alimento, aquel que lo ingiera cambiará su estado de ánimo, tenga o no conocimiento de la droga que ha ingerido (y sus efectos). Si bien es posible que algu-

nas características del estado inducido por el miorrelajante sea alterado por un *input* cognitivo, como Schachter y Singer han mostrado, en el análisis final, al menos algunos aspectos muy significativos del cambio del estado emocional serán causados directamente por el *valium* sin tener en cuenta qué información recibieron, ni qué justificación ellos mismos ofrecen luego (Zajonc, 1984: 120).¹⁸

En resumidas cuentas, una teoría de doble factor, como la que defienden Schachter y Singer, afirma que una emoción involucra (1) tanto cambios corporales (estados fisiológicos) como (2) interpretaciones cognitivas de esos estados. No es sino la cognición, en la forma de una interpretación, la que determina si un estado fisiológico de excitación será etiquetado como “enojo”, “alegría”, “miedo” u otro.

3. Atribución y teorías de emociones: escenas de la vida cotidiana

Existe una idea ampliamente extendida en la literatura según la cual la creencia es el estado psicológico paradigmático. Este “imperialismo de la creencia” (Pérez, 2013; Pérez y Gomila, este volumen) que impera en la psicología cognitiva clásica, ha tenido consecuencias directas sobre la teorización acerca de las emociones. Esta posición, que sitúa a la creencia como paradigma de todo estado mental, y que asume que es a partir de ellas que se deben explicar todos los demás estados psicológicos, configura un modo específico de concebir la comprensión o el acceso a las emociones: si logro identificar la creencia sobre la cual se erige la emoción, puedo definir y/o entender la emoción.

Las creencias son vistas como uno de los principales deter-

18 También están quienes afirman que el experimento no prueba concluyentemente nada. Es decir, no funciona como evidencia a favor ni en contra de ninguna de las dos perspectivas. En ese sentido, Deigh afirma que “nada excluye la posibilidad de que entre las *cogniciones* que los experimentadores inducen en los sujetos haya percepciones jamesianas de eventos *emocionantes*” (Deigh, 1994: n. 4).

minantes de las emociones, y de este modo una parte importante del estudio de las emociones puede ser adecuadamente visto como cayendo dentro del paraguas de la psicología cognitiva (Frijda, Manstead y Bem, 2000: 1). Es por eso que los filósofos cognitivos tienden a preferir el análisis de los pensamientos que constituyen la emoción, “en su entorno social o incrustados [*embedded*] en una narrativa” (Hutchinson, 2009: 61). Es precisamente en ese sentido que Amélie O. Rorty afirma que “cuando las personas actúan o reaccionan de modos que pueden ser explicados por creencias y deseos razonables, tendemos a suponer que estas creencias y deseos son causas de su comportamiento” (Rorty, 1982: 103).

Cuando consideramos situaciones como la retratada al comienzo de este artículo, se hace patente cómo dependiendo del marco teórico que se adopte, la naturaleza de las emociones varía y con ella, las condiciones de corrección para el uso de términos de emociones, tanto para estados propios como ajenos (i.e. varían las condiciones bajo las cuales puedo (es adecuado) usar las palabras ‘miedo’, ‘enojo’ o ‘celos’ para describir estados de cosas). A continuación apuntaré algunas dimensiones desde las cuales abordar las problemáticas que se suscitan en la atribución de emociones, dificultades que se ven profundizadas por la diversidad (e incompatibilidad) de enfoques disponibles.

3.1. Distinción entre (prioridad de la) primera y tercera persona

Lo primero que hay que notar es que se produce cierto ‘choque’ de intuiciones. Por una parte Carlos (observador externo) percibe ciertos rasgos en el rostro, actitud corporal, tono de voz, etc., de Estela, que le sugieren que ella está enojada. Por otra parte, Estela, protagonista del episodio emocional en primera persona, no tomó debida cuenta del evento, y cuando su marido le pregunta qué le sucede, ella niega estar atravesando una emoción (sin estar mintiendo deliberadamente, al menos en esta instancia inicial). La atribución que realiza Carlos entra en conflicto con la autoatribución que Estela realiza. En otras palabras, Carlos detecta –acertadamente diría yo– que su esposa se encuentra

afectivamente alterada. La ‘excitación’ se manifiesta a través de un conjunto de variables corporales. Carlos ‘lee’ en el rostro y la actitud corporal de su esposa signos de enojo, y consiguientemente le atribuye el estado emocional propio: ‘estás enojada’. Pero Estela no se percató de lo que le estaba sucediendo, al punto que, al ser consultada, su primer respuesta es que nada le sucede.

¿Cómo habría de resolverse este conflicto? Desde una perspectiva somática/fisiologista, cabría decir, junto con Carlos, que Estela está enojada, aun en ausencia de reconocimiento de quien lo padece. Pero hacer esto, además, tendría consecuencias sobre la tesis cartesiana de la asimetría de la primera y tercera persona, donde la primera persona tiene prioridad en cuanto al conocimiento de sus estados psicológicos (véase Lawler, este volumen). Sin embargo, las condiciones de atribución varían si uno adopta una teoría cognitiva, haciendo necesaria la intervención de algún elemento de carácter valorativo, Estela está enojada sólo si juzga o valora que algo o alguien la ha ofendido.

3.2. Distinción causa y objeto de la emoción

Afortunadamente, Estela toma conciencia de que algo le pasa y admite estar molesta. Y es que a pesar de no haber reconocido de modo inmediato su afectación emocional, el proceso emocional se encontraba ya iniciado. Sin embargo, al reconocer que está enojada dirige su atención (¿equivocadamente?) hacia el mozo.

Como decía más arriba, de acuerdo a las teorías cognitivas de las emociones, Estela estaría enojada (en sentido propio) sólo si juzgara que algo la ha ofendido; de lo que se sigue que, si posteriormente alguien le señalara que no hay ofensa alguna en el accionar del mozo y ella abandonara la creencia de que el mozo la ofendió, consecuentemente debería desvanecerse la emoción. Es decir, si uno le señalase a Estela que no hay motivo para que se sienta ofendida, su enojo debería disiparse.¹⁹ Sin embargo, resulta esperable que Estela siga molesta. Para dar cuenta de esta

19 Algunos teóricos describen este fenómeno a partir de la noción de revalorización [*reappraisal*], no es sino a través de un nuevo proceso de valoración, que se suspende la emoción en curso.

situación, debe tenerse en cuenta la distinción que realiza Kenny (1963) entre causa y objeto de una emoción (distinción deudora de Wittgenstein (1988)): no es necesario reconocer algo como la causa de mi emoción para que ésta opere como desencadenante de una emoción. El punto aquí es que si bien en algunos casos la causa y el objeto de una emoción coinciden, esto no es necesario en todos los casos. Para establecer la diferencia, Kenny invita a hacerse la siguiente pregunta: (P) ¿por qué Estela siente enojo?²⁰ Esta pregunta acerca del porqué esconde cierta ambigüedad, precisamente porque en rigor encierra *dos* preguntas. (P) puede responderse señalando alguna causa remota, responsable de la adquisición de una disposición a ser afectado de un modo particular. Por ejemplo, si analizamos el dicho popular “quien se quemó con leche ve una vaca y llora”, y nos preguntamos “¿por qué llora?” podemos responder que llora porque se quemó con leche, donde se señala el evento pasado que condiciona los futuros enfrentamientos con seres u objetos similares al que le produjeran el daño. En segundo lugar, puede contestarse que siente miedo de (porque) tiene una vaca cerca. Es decir, mientras la primera respuesta apunta a la **causa** o historia causal de la disposición, la segunda señala el **objeto** de la emoción. Esta diferencia es fácilmente reconocible a partir del examen de las diversas estructuras lingüísticas de las que disponemos para hablar de las emociones. Por ejemplo:

- (1) “estar enojado con *x*”;
- (2) “estar enojado porque *x*”.²¹

Aquí, (1) y (2) se distinguen en cuanto a que (1) se refiere al objeto de la emoción, mientras que (2) se refiere a su causa. Pero esto no siempre es suficiente, puesto que el lenguaje cotidiano presenta muchas imprecisiones. Considérese la proposición (v):

- (v) “La conducta de Juan me causó vergüenza”.

Por la utilización del término “causar” uno tendería rápidamente a asumir que la conducta de Juan es causa de mi emoción. Pero en rigor, algunas aplicaciones del lenguaje ordinario son

20 O dicho de modo más general ¿por qué *x* tiene-la-emoción-e?

21 En inglés “*to be angry at*” versus “*to be angry because*”.

equívocas: la conducta de Juan no es la causa de mi vergüenza, sino el objeto acerca del cual versa mi emoción (Gosling, 1965). Vislumbrando esta opacidad, típica de numerosas expresiones lingüísticas ordinarias, Kenny sugiere que esta distinción resulta mejor trazada si se toman en consideración el conocimiento o las creencias del sujeto. De este modo, introduce una prueba que permite operar la desambiguación: dada una proposición, de la forma “*A tiene-la-emoción-x porque q*”, debemos preguntar si es condición necesaria de la verdad de esta proposición que A conozca o crea que *q*. Si es así, la oración contiene una alusión al objeto de la emoción; si no lo es, a su causa (Kenny, 1963: 75). La distinción apunta a mostrar que es imposible estar enojada por [*because*] el modo en el que alguien habla, si no me percaté del modo en el que habla (objeto), a la vez que sí es posible sentirme enojada porque [*because*] estoy famélica sin percatarme de que estoy famélica (causa).

De manera que incluso en el caso en que Estela se mantuviera firme en la idea de que nada le sucede y no volviera su atención hacia sus sensaciones y la situación que la rodea, podríamos igualmente reconstruir la escena, señalando que la causa de su estado fue el ‘desprecio’ sufrido (por parte de su marido, al no presentarla y excluirla).

3.3. Distinción entre objeto formal (intencional) y objeto material de la emoción que se desarrolla

Estela finalmente reconoce su enojo y lo explica (lo hace inteligible de algún modo) ubicando al mozo como el responsable de la ofensa. De acuerdo a la tesis cognitivista, para que un estado emocional se constituya en enojo, algo del estilo de una ofensa tiene que haber acontecido. De modo que aun si Estela se equivocara al señalar al mozo como el destinatario de su enojo, es posible que igualmente esté ‘acertando’ en cuanto al tipo de emoción. Cuando Kenny dice que “el objeto del miedo es aquello que es temido” (Kenny, 1963: 188) no está refiriéndose a ninguna instanciación particular de ofensa, sino a aquello que todas tienen en común. Lo que es temido, de modo general, representa el objeto formal de la emoción de miedo, lo que no

implica en principio ningún constreñimiento respecto a lo que de hecho funcione como objeto material de la emoción. Por muy singular que pueda parecer, la desatención de un mozo puede ser objeto (particular) del enojo de Estela, precisamente en el mismo sentido en que en el ejemplo de Solomon, bastaba con creer que Juan me ofendió (cuando tomó mi auto). Es suficiente reconocer apropiadamente que existió una ‘ofensa degradante, contra uno y lo suyo’ (Lazarus, 1991) para estar enojado.²² Sin embargo Kenny advierte también:

Asignarle un objeto formal a una acción es aplicar restricciones sobre lo que puede acontecer como el objeto directo del verbo que describe la acción (Kenny, 1963: 189).

Las restricciones que Kenny tiene en mente son restricciones propias de la lógica o la gramática de los conceptos de emoción, y se vinculan con cuestiones temporales (sólo lo pasado puede ser recordado o vengado; sólo lo que no ha pasado puede temerse o anhelarse), espaciales (sólo lo que está presente puede ser disfrutado; sólo lo que está ausente puede ser extrañado), morales (sólo lo que se cree bueno puede ser envidiado; sólo lo que se cree malo puede ser lamentado). Pero las limitaciones que impone el objeto formal de ningún modo apuntan a qué objetos particulares pueden ser objetos de una emoción dada; como dijimos antes, sólo basta con tener la creencia adecuada para que ese objeto se convierta en objeto particular de la emoción.

¿Hay modos de garantizar la identificación adecuada del objeto material (específico)? Podemos dar un paso más y preguntarnos: ¿es acaso relevante conocer el objeto? Si hemos de adoptar un marco cognitivista, resulta que conocer el objeto será de cabal importancia, puesto que una emoción es tal dependiendo

22 Lazarus (1991) presenta los ‘temas relacionales centrales’ que identifican los significados de cada tipo emocional, así, el enojo se define a partir de la detección de una ‘ofensa degradante, contra uno y lo suyo’; la culpa por la ‘transgresión de un imperativo moral’; la vergüenza por el ‘incumplimiento de un ideal propio’; los celos por el ‘resentimiento de una tercera parte ante la amenaza o pérdida del afecto de otro’, etc.

de su objeto intencional.²³ Sin embargo, como dije más arriba, el objeto intencional de una emoción debe ser distinguido de su objeto material. Adicionalmente el objeto material debe diferenciarse de la causa original que predispone a un sujeto a sentir una emoción frente a determinado objeto. De modo tal que de acuerdo a esta perspectiva, una emoción cualquiera, para ser tal, depende en primera instancia de su causa histórica originaria (no siempre conocida), pero principalmente, en el presente, de su objeto intencional, expresable proposicionalmente. Es el objeto intencional el que determina que el episodios emocional sea de miedo, celos, u otra cosa.

3.4. Sobre la influencia la etiqueta

Estela admite que está enojada por el destrato recibido por el mozo, y al minuto recapacita y reconoce que aquello no fue intencionalmente una ofensa (como dije más arriba, lleva adelante una revaloración de la situación). Aquí podrían suceder al menos dos cosas: (a) Estela ya no se siente ofendida, su rostro pierde todo rasgo de enojo y Estela recupera la dulce mirada que la caracteriza; (b) Estela sabe que el mozo no quiso ofenderla, pero sin embargo sigue sintiéndose ‘incómoda’. Es probable que a continuación encuentre otro elemento del entorno que ‘explique’ su alteración.

Pero resulta bastante evidente que en realidad Estela se sintió sumamente insegura al ser ignorada por su esposo y no ser debidamente presentada, vio a su marido flirtear con jóvenes muchachas y se puso celosa. Pero Estela no se ve a sí misma como una mujer celosa ni insegura. De modo que, si ha de reconocer

23 La noción de objeto intencional es la clásica aristotélica, reintroducida y canonizada por Brentano: “Todo fenómeno psíquico está caracterizado por la inexistencia intencional (o mental) de un objeto [...] la dirección hacia un objeto [...] En la representación hay algo representado, en el juicio hay algo admitido o rechazado; en el amor, amado; en el odio, odiado; en el apetito, apetecido, etc.” (Brentano, 1874: 21).

Nótese cómo en esta caracterización se puede divisar el problema que de Sousa llama la “antinomía de la objetividad”: ¿queremos aquellas cosas porque son queribles o llamamos queribles a todo aquello que queremos? (De Sousa, 1987).

alguna emoción, esta habrá de ser miedo, enojo, asco, pero nunca celos (esa etiqueta no está disponible).

En caso de adoptar una posición cognitiva, quedaría establecida la *tesis de la determinación cognitiva*, y al considerar que las emociones son estados que involucran constitutivamente operaciones mentales valorativas, se estaría asumiendo, asimismo, la *tesis de la insuficiencia somática*, por la cual los cambios corporales (y los *sentires* que tenemos de ellos) no son suficientes (aun cuando puedan acompañarla). Echar mano a este conjunto de tesis, y abrazar de este modo una posición cognitivista, nos permitiría explicar por qué Estela no se *siente* celosa. Ella ha etiquetado –y así han quedado explicadas– sus sensaciones en sentido acorde. No obstante, si alternativamente optáramos por rechazar la necesidad de la intervención cognitiva (ya sea en forma de valoraciones, creencias, juicios, etc.), reivindicando el abordaje somático/fisiologista, bastaría con inferir los estados emocionales a partir de sus cualidades expresivas, es decir, de aquellos rasgos que la historia evolutiva consolidó como prototípicos, cumpliendo una función comunicativa.

Bibliografía

- Adolphs, R., et al. (2005), "Preferring one taste over another without recognizing either", *Nature Neuroscience*, 8(7), pp. 860-861.
- Arnold, M. (1960), *Emotion and Personality*, New York: Columbia University Press.
- Bearison, D. J. y H. Zimiles (1986), *Thought and Emotion: Developmental Perspectives*, Hillsdale, NJ: Erlbaum.
- Brentano, F. (1874), *Psychologie vom Empirischen Standpunkt*, Hamburg: Kraus.
- Charland, L. C. (1997), "Reconciling Cognitive and Perceptual Theories of Emotion: A Representational Proposal", *Philosophy of Science*, 64(4), 555. Disponible en: <http://doi.org/10.1086/392572> [31/10/2017].
- De Sousa, R. (1987), *The Rationality of Emotion*, MIT Press.
- Deigh, J. (1994), "Cognitivism in the Theory of Emotions", *Ethics*, 104(4), 824. Disponible en: <http://doi.org/10.1086/293657> [31/10/2017].
- Frijda, N. H. (1994), "Emotions Require Cognitions, Even If Simple Ones", en: Ekman, P. y R. J. Davidson (eds.), *The Nature of Emotion: Fundamental Questions*, New York: Oxford University Press, pp. 197-202.
- Frijda, N. H., A. S. R. Manstead y S. Bem (2000), "The influence of emotions on beliefs", en: Frijda, N. H., A. S. R. Manstead y S. Bem (eds.), *Emotions and Beliefs: How Feelings Influence Thoughts*, Cambridge: Cambridge University Press, pp. 1-9.
- Gordon, R. M. (1987), *The Structure of Emotions: Investigations in Cognitive Philosophy*, Cambridge, MA: Cambridge University Press.
- Gosling, J. C. (1965), "Emotion and Object", *The Philosophical Review*, 74(4), pp. 486-503.
- Hacker, P. M. S. (2009), "The Conceptual Framework for the Investigation of Emotions", en: Gustafsson, Y., C. Kronqvist y M. McEachrane (eds.), *Emotions and Understanding: Wittgensteinian Perspectives*, Palgrave Macmillan.
- Hutchinson, P. (2009), "Emotion-philosophy-science", en: Gustafsson, Y., C. Kronqvist y M. McEachrane (eds.), *Emotions and Understanding: Wittgensteinian Perspectives*, Palgrave Macmillan.

- Izard, C. E. (1986), "Approaches to Developmental Research on Emotion-Cognition Relationships", en: Bearison, D. J. y H. Zimiles (eds.), *Thought and Emotion. Developmental Perspectives*, Hillsdale, NJ: Erlbaum, pp. 21-37.
- James, W. (1884), "What is an emotion?", *Mind*, 9, pp. 188-205.
- . (1890), *Principles of Psychology*. Chicago: Enciclopedia Britannica, Inc.
- Kandel, E. R., J. H. Schwartz y T. M. Jessell (1997), *Neurociencia y conducta*, Prentice Hall.
- Kenny, A. (1963), *Action, Emotion, and Will*, Wiley-Blackwell.
- Lange, C. G. (1885), *Om sindsbevaegelser: Et psyko-fysiologisk studie*, Copenhagen: Jacob Lunds.
- Lazarus, R. S. (1982), "Thoughts on the relations between emotion and cognition", *American Psychologist*, 37(9), 1019-1024. Disponible en: <http://doi.org/10.1037/0003-066X.37.9.1019> [21/10/2017].
- . (1984), "On the primacy of cognition", *American Psychologist*, 39(2), pp. 124-129. Disponible en: <http://doi.org/10.1037/0003-066X.39.2.124> [31/10/2017].
- . (1991), *Emotion and Adaptation*, New York: Oxford University Press.
- . (1999), "The Cognition-Emotion Debate: A Bit of History", en: Dalglish, T. y M. J. Power (eds.), *Handbook of Cognition and Emotion*, Chichester, UK: Wiley, pp. 3-19.
- Lazarus, R. S., J. R. Averill y E. M. J. Opton (1970), "Towards a Cognitive Theory of Emotion", en: Arnold, M. (ed.), *Feelings and Emotions: The Loyola Symposium*, New York: Academic Press, pp. 207-232.
- Lyons, W. (1980). *Emotion*, Cambridge: Cambridge University Press.
- . (1999), "The Philosophy of Cognition and Emotion", en: Dalglish, T. y M. J. Power (eds.), *Handbook of Cognition and Emotion*, Chichester, UK: Wiley, pp. 21-44.
- Marañón, G. (1924), "Contribution à l'étude de l'action émotive de l'adrenaline", *Revue Française d'Endocrinologie*, 21, pp. 301-325.
- Morris, J. S. (2002), "How do you feel?", *Trends in Cognitive Sciences*, 6(8), pp. 317-319.
- Murphy, S. T. y R. B. Zajonc (1993), "Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures", *Journal of Personality and Social Psychology*, 64, pp. 723-739.
- Pérez, D. I. (2013), *Sentir, desear, creer*, Buenos Aires: Prometeo.

- Pfaffmann, C., R. Norgren y H. Grill (1977), "Sensory affect and motivation", *Annals of the New York Academy of Sciences*, 290(1), pp. 18-34. Disponible en: <http://doi.org/10.1111/j.1749-6632.1977.tb39713.x> [31/10/2017].
- Power, M. J. y T. Dalgleish (2008), *Cognition and Emotion: From Order to Disorder*, Hove: Taylor Francis.
- Prinz, J. J. (2004), *Gut Reactions: A Perceptual Theory of Emotion*, New York: Oxford University Press.
- Rorty, A. O. (1982), "From Passions to Emotions and Sentiments", *Philosophy*, 57(220), pp. 159-172.
- Schachter, S. y J. Singer (1962), "Cognitive, Social, and Physiological Determinants of Emotional State", *Psychological review*, 69(5). Disponible en: <http://doi.org/10.1037/h0021465> [31/10/2017].
- Solomon, R. C. (1973), "Emotions and Choice", *Review of Metaphysics*, 28(1).
— (1976), *The Passions: The Myth and Nature of Human Emotion*, New York: Anchor Press, Doubleday.
— (1993), "The philosophy of emotions", en: Lewis, M. y J. M. Haviland-Jones (eds.), *Handbook of Emotions*, New York: Guildford.
— (2001), "Thoughts and Feelings: What Is a 'Cognitive Theory' of the Emotions, and Does It Neglect Affectivity?", en: *Not Passion's Slave* Cambridge: Cambridge University Press.
— (2003), *Not Passion's Slave: Emotions and Choice*, New York: Oxford University Press.
- Winkielman, P., R. B. Zajonc y N. Schwarz (1997), "Subliminal Affective Priming Resists Attributional Interventions", *Cognition & Emotion*, 11(4), pp. 433-465.
- Wittgenstein, L. (1988), *Investigaciones Filosóficas*, México: UNAM.
- Zajonc, R. B. (1980), "Feeling and thinking: Preferences need no inferences", *American Psychologist*, 35(2), pp. 151-175. Disponible en: <http://doi.org/10.1037/0003-066X.35.2.151> [31/10/2017].
— (1984), "On the Primacy of Affect", *American Psychologist*, 39, pp. 117-123.
— (2001), "Mere Exposure: A Gateway to the Subliminal", *Current Directions in Psychological Science*, 10(6), pp. 224-228. Disponible en: <http://doi.org/10.1111/1467-8721.00154> [31/10/2017].

La atribución psicológica a personajes de ficción

Lucas Bucci

I

En el famoso final de *Casablanca* (1942) sus protagonistas, Ilsa y Rick, escapan de la policía y se dirigen al avión que les permitirá salir de la colonia nazi. Sin embargo, cuando el policía pregunta quiénes se subirán al avión, Rick envalentonado decide quedarse abajo y dar su lugar a Victor, el esposo de Ilsa. Este final es sorpresivo porque el espectador sabe que Ilsa quiere a Rick por sobre a Victor y, sobre todo, que Rick quiere a Ilsa. La noche anterior, ambos han dejado sus rencores de lado y han reavivado la llama de su amor. Sin embargo, a través del dialogo que Ilsa y Rick tienen al pie del avión podemos empezar a vislumbrar los motivos por los cuales el americano decide quedarse en Casablanca. Rick quiere a Ilsa pero también cree que el lugar que tiene Victor en la guerra contra los nazis es muy importante y si ella lo abandonara, lo dejaría devastado. Rick cree también que la Segunda Guerra Mundial es un evento mucho más importante que las cuestiones sentimentales que lo envuelven con Ilsa y Victor. Además, cree que Ilsa corre peligro en Casablanca y que si él no la engañara para ir al aeropuerto ella no se subiría al avión porque cree que ella está enamorada de él. Por su parte, Ilsa efectivamente lo acompaña al avión creyendo que Rick se va a subir con él, por eso, cuando él le dice que no va, la invade una tristeza profunda y sus ojos se llenan de lágrimas.

Como vemos, esta interpretación posible de los eventos que ocurren al final de la película apela a una cantidad importan-

te de estados mentales que el espectador adscribe de manera consciente o inconsciente mientras mira la película (o cuando reflexiona sobre ella). La adscripción de deseos, creencias y emociones a personajes de ficción es un fenómeno extendido en el consumo de películas, obras de teatro, lectura de ficción y cualquier tipo de obra de arte que involucre una narrativa. Este tipo de adscripciones –que suelen ser predominantemente de deseos, creencias y emociones– son centrales a la hora de entender la narrativa que estas obras de arte despliegan (Gomila, 2011). Es imposible entender una historia protagonizada por personajes si aquel que la contempla no es capaz de adscribirles ciertos estados mentales de una manera más o menos competente. Los autores construyen a sus personajes para que sus acciones sean explicadas apelando a estos estados y, tal como están construidas estas narrativas, sería imposible comprenderlas de manera cabal si aquel que las contempla no fuera capaz de adscribir diferentes tipos de estados mentales.

En el ámbito de la filosofía estética sajona este problema ha sido tratado de manera extensa, aunque la discusión se ha dado dentro de una discusión previa y más general relativa a las emociones que las distintas narrativas producen en el espectador. Este problema concerniente a la recepción o contemplación de arte narrativo tiene que ver con el fenómeno aparentemente paradójico de que los espectadores se emocionan con eventos que saben que son puramente ficticios. Este fenómeno, llamado “Paradoja de la Ficción” y señalado por Radford (1975, 1977, 1982) ha moldeado parte de la discusión sobre la recepción de obras de arte narrativo en los últimos cuarenta años (Carroll, 1987, 1990/2005, 1991; Currie 1995a, 1995b, 1997; Gaut, 1993; Gomila, 2011; Lamarque, 1981; Moran, 1994; Walton, 1978, etc.). Según la paradoja, es desconcertante que un espectador o lector de ficción se emocione con un evento ficcional, cuando efectivamente sabe que este evento es tal, dado que pareciera que necesitamos creer en la realidad de un evento para emocionarnos acerca de él. Si no, no se explica por qué nuestra emoción de pena se disipa cuando nos enteramos que, por ejemplo, la noticia de un avión caído era falsa, mientras que nuestra pena sobre el

suicidio de Ana Karenina no se disipa al saber que tal suicidio nunca tuvo lugar.

El interrogante parece centrarse, entonces, en la idea de que no resulte plausible que los espectadores experimenten emociones reales por fenómenos que saben que son ficcionales.¹ Sin embargo, y aunque no parezca que esta discusión sobre emociones en contexto de ficción tenga que ver con lo que aquí nos ocupa, el fenómeno de adscripción de estados mentales a personajes ficcionales cumple un papel importante. Los personajes son una de las herramientas más valiosas que tienen los autores para emocionar a una audiencia y la emoción que estos producen no podría darse si no fuera porque la audiencia adscribe, de manera consciente o inconsciente, estados mentales a los personajes de la ficción.

Aquí voy a revisar las posturas de Gregory Currie (1995a, 1997) y Noel Carroll (1990/2005, 2008) con respecto a estos dos temas relacionados. La elección de estos dos autores obedece a que ambos sostienen posturas contrapuestas, que por separado ejemplifican dos extremos de la cuestión que considero centrales. Por un lado, Currie sostiene que en nuestra adscripción a personajes ficticios utilizamos los mismos mecanismos psicológicos que utilizamos para la adscripción de personas en nuestro entorno cotidiano; y, por otro lado, Carroll sostiene que, en la contemplación de ficción, no existe ningún mecanismo psicológico específico de adscripción de estados sino que, más bien, los espectadores establecen relaciones con los personajes que son mediadas por la imaginación, los géneros narrativos y ciertos criterios emocionales y, de alguna manera, *manipulados* por el autor de la obra que contemplan.

Este punto de conflicto me parece importante a la hora de entender la adscripción a personajes de ficción. Me interesa indagar sobre si establecemos una diferencia clara entre personajes de ficción y personas que nos rodean, al estilo Carroll, o como

1 De hecho, la conclusión de Radford (1975) es que las emociones dirigidas a la ficción que experimentan los espectadores o lectores de ficción son en alguna medida *irracionales*.

piensa Currie, no hacemos distinción alguna y la ficción es una extensión de nuestro ámbito cotidiano y nuestros mecanismos psicológicos se activan de la misma manera que lo harían en una situación cotidiana.

II

En su artículo “The Paradox of Caring” (1997) Currie establece dos requisitos que debe cumplir una solución a la Paradoja de la Ficción. Por un lado, establece que una buena teorización acerca de la reacción emocional no debe apelar a mecanismos psicológicos nuevos o, mejor, debe hacer uso de los mecanismos psicológicos establecidos por la teorización actual de la ciencia de la psicología. Por el otro lado, Currie sostiene, siguiendo a Moran (1994), que una solución adecuada debe tratar no sólo con los casos de reacciones emocionales a la ficción sino también con aquellos casos de emoción que refieren a situaciones no actuales. En efecto, Moran sostiene que los casos de emoción dirigida a las obras representativas no son casos “desviados” dado que existen muchos casos que consideraríamos normales que dirigen la emoción a situaciones no actuales como, por ejemplo, situaciones muy remotas en el pasado, o situaciones que podrían haber pasado y no pasaron, tanto a mí como a otra persona. Estos dos requisitos se cumplen de manera natural en la teoría de Currie. Por un lado, porque Currie sostiene que la emoción ficcional se da en un contexto imaginativo, que fácilmente puede extenderse a situaciones “no canónicas” de la emoción y, por otro lado, porque dada la apelación de la simulación como imaginación Currie puede echar mano a la “teoría de la simulación” y postularla como el fenómeno central de la emoción ficcional.

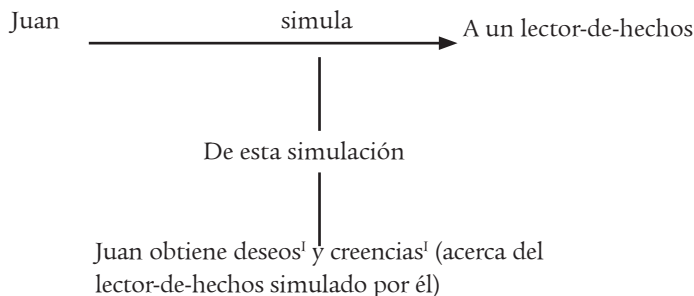
Según Currie, entonces, toda contemplación de una obra de arte representativa o narrativa es una simulación en alguna medida. Así, aquel que contempla una obra de arte simula los estados de hechos o noticias (*reader of fact*) por los que se enteraría de aquello que sucede en la obra como si realmente hubiera ocurrido en nuestra vida cotidiana y como si los hechos estuvieran

protagonizados por personas reales y no por personajes ficticios. De esta manera, el espectador *hace de cuenta* que se está enterando de los hechos de la ficción como si fueran reales. O, lo que es lo mismo, toma la perspectiva de un espectador hipotético o simulado que presencia hechos reales. De esta manera, simular este rol de un *lector* de hechos,² hace de la obra representativa una especie de instrumento o lista de instrucciones que me dice qué es lo que tengo que imaginarme.

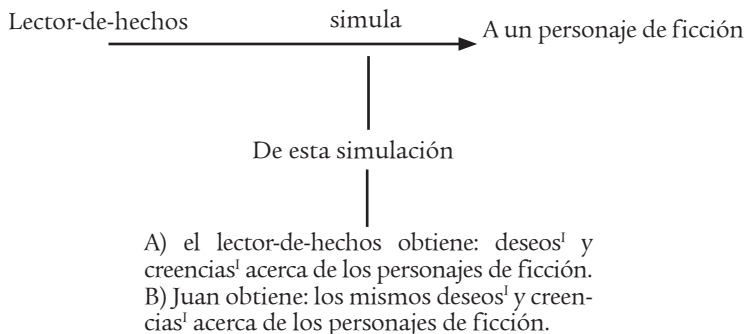
Ahora bien, queríamos comprender cómo pueden importarnos, de hecho, los destinos de los personajes de ficción, a pesar de que no creemos que ellos existan. Sabemos, por un lado, que la empatía o la lectura de mentes se hace a través de la simulación y, por el otro lado, sabemos que cuando contemplamos una obra de arte representativa simulamos a un lector hipotético de hechos. ¿Cómo habríamos de sentir algo por un personaje de la obra si, en realidad, simulamos a una persona hipotética que está de alguna manera desapegada de la obra? La razón es que este espectador hipotético, a la hora de comprender los eventos de la obra y comprender las motivaciones de los personajes para actuar, debe hacer uso de la simulación para comprender a los personajes de ficción (puesto que este es el medio por el cual comprendemos a los demás). De esta manera, un espectador, Juan, simula que es un lector que intenta saber por qué actúan los personajes de la obra y tomándolos como personajes de la vida cotidiana. En este emprendimiento, el lector simulado por Juan simula los estados mentales de los personajes de la obra. Ahora bien, simular los estados simulados de otro hace que yo tenga exactamente los mismos estados que aquel al que estoy simulando. Este es un diagrama de las simulaciones involucradas en la contemplación de una obra:

2 Currie, 1997: 68. Como vemos, el ejemplo de Currie es el de un lector. El ejemplo se puede extrapolar a otras artes narrativas sin problemas. Sin embargo, es preciso tener en cuenta que –en el caso del cine– Currie sostiene que hay un sujeto hipotético que aprende hechos reales que no hace de cuenta que ve a los personajes. El espectador ni ve a los personajes ni hace de cuenta que los ve; simplemente, se entera de los hechos por medios visuales y auditivos (Currie, 1991).

Paso 1:



Paso 2:



Según la teoría simulacionista de Currie simular estados de creencia y deseos nos proporciona “estados I”, es decir estados imaginados o simulados. La diferencia central entre los estados I y los que estos simulan es que los estados I son *offline*.³ Así, simular los estados de creencia y deseos de otros da como resultado tener creencias^I y deseos^I. Ahora bien, simular estados I, por ejemplo creencias^I y deseos^I, da como resultado tener *las*

³ Ver teoría de la simulación en Currie (1995b).

mismas creencias¹ y deseos¹. De esta manera, en la adscripción de estados mentales a personajes de ficción, las creencias y deseos de los personajes de la obra son simulados por el *lector-de-noticias* y pasan a ser creencias¹ y deseos¹; a su vez, como el espectador está simulando los estados mentales del *lector-de-noticias* obtiene exactamente los mismos estados¹ que obtiene el *lector-de-noticias* en su simulación de los personajes de ficción. Dice Currie al respecto:

¿Por qué esta diferencia entre simular una creencia y simular una creencia? Recuérdese que la simulación es el intento de replicar los estados mentales de otros en la propia mente. La simulación intenta llevarme tan cerca como sea posible al estado mental de otro siempre y cuando se constriña de darme creencias y deseos sobre los cuales serían inapropiado actuar. La solución es darme estados¹: estados como las creencias y los deseos en algunos aspectos pero diferentes a ellos en que están desconectados de la acción. Pero si el estado que busco replicar es él mismo un estado *offline* (un estado⁰), no hay peligro en que yo tenga ese mismo estado, de modo que mi replicación consista en tener ese mismo estado (1997: 69).

De esta manera, la propuesta consiste en que cuando alguien contempla una obra de arte narrativa, simula o toma el rol de un *lector-de-noticias* que aprende acerca de ciertos eventos que llevan a cabo ciertos personajes. Al simular un *lector-de-noticias*, el espectador llega a tener versiones *offline* (o estados¹) de sus estados mentales, incluyendo aquellos estados mentales *offline* que obtiene el lector por la simulación que hace de los personajes de ficción de la obra.

Esto responde entonces la pregunta acerca de cómo podemos tener emociones sin creencias. Las emociones son producto de nuestros estados simulados, de nuestras simulaciones. Además, con esta respuesta se respetan los requisitos que se habían planteado como centrales para responder la pregunta. Por un lado, no se plantea ningún mecanismo psicológico nuevo o especial para responder al problema de las reacciones emocionales a la ficción ya que es el propio mecanismo de simulación presente en nuestra interacción con los demás aquel que es el responsable de nuestras reacciones emocionales a las obras representativas. Por

otro lado, el requisito de Moran exigía que se le dé tratamiento no sólo a las respuestas emocionales a las obras de arte representativas sino a las que, de modo similar, surgían de la consideración de situaciones no actuales, o de un pasado remoto, etc. La respuesta de Currie es que todas estas reacciones emocionales responden al mismo mecanismo: la simulación.

La posición de Currie también es capaz de explicar por qué cuando contemplamos una ficción nos vemos atraídos por personajes éticamente reprobables o que en principio no tienen nada que ver con nosotros. La idea es que una obra de arte representativa no sólo me lleva a imaginar que me entero de cosas sino que, además, me lleva a compartir un punto de vista. ¿Qué punto de vista? El punto de vista del narrador o del autor implicado. Así, mientras mira *El exorcista* Juan es llevado a compartir una visión de terror acerca de los hechos que suceden, en consecuencia, al ver que Regan tuerce su cabeza siente terror y pena por su madre. Del mismo modo, Juan puede ver una película de humor negro, en la que ciertos hechos trágicos suceden pero, al compartir una mirada más bien humorística con el autor implicado, puede reírse ante estos hechos. En el mismo sentido, la serie *Breaking Bad* me invita a simular un punto de vista según el cual un traficante de drogas duras puede ser redimido, de ahí que el personaje principal me caiga simpático. Según Currie, este cambio de punto de vista se vería facilitado por cierta disposición natural a estar “en sintonía” (*in tune*) con los demás, junto con una tendencia y un gusto por probar otras perspectivas, junto con esa fuerza natural que sentimos por querer parecernos a aquellos a los que se nos presentan como atractivos.

Como vemos, la propuesta de Currie con respecto a la adscripción de estados mentales a personajes de ficción está englobada en una postura relativa a cómo es que los espectadores se emocionan con la ficción misma. La idea es utilizar los mismos mecanismos psicológicos para entender tanto a las personas como a los personajes de ficción. Esto es, según el autor, una ventaja de economía teórica puesto que para explicar cómo adscribimos estados mentales a los personajes de ficción es suficiente con apelar a nuestra mejor teoría acerca de cómo adscribimos estados mentales a las personas. Esto me da un pie para explicitar un desacuerdo que

tengo con la presentación de Currie sobre el problema. Puesto que Currie presenta a la teoría de la simulación como si fuera la única teoría o la más aceptada en el terreno de la comprensión de los otros en nuestra vida cotidiana. Sin embargo, este no es el panorama real de la discusión. Existen numerosas teorías rivales a la teoría de la simulación, i.e. la teoría de la racionalidad, la teoría de la teoría, la teoría de la segunda persona. En este sentido, hubiera sido interesante argumentativamente que el autor mostrara las ventajas de las diferentes teorías de la adscripción a personajes ficcionales a la hora de explicar el fenómeno de la contemplación de obras de arte narrativas. ¿Por qué no podemos pensar que es la teoría de la teoría aquella que explica nuestra reacción emocional para con las obras representativas? ¿No podemos acaso pensar que el espectador se imagina un lector-de-hechos que interpreta a los personajes de ficción en base a ciertos conceptos psicológicos de una teoría de la psicología *folk* y no en base a una simulación adecuada? La postura de Currie parece pasar por alto el hecho de que el debate acerca de cuáles son los mecanismos de interpretación que se utilizan para la lectura de mente está todavía vigente y que, por lo tanto, no se puede argumentar –como él argumenta– que son los mecanismos postulados por la teoría de la simulación aquellos involucrados en la reacción emocional de la contemplación de obras de arte en virtud de ser –sin disputa– los mecanismos de interpretación de las lecturas de mente.

No obstante, esto no va en desmedro de que la posición de Currie se beneficia en no tener que presentar una teoría adicional específica para la adscripción de estados mentales a personajes de ficción (ya sea esa teoría de la simulación, de la teoría, etc.). En lo que sigue, veremos la postura de Carroll, quien sí parece sostener que nuestra adscripción de estados mentales a personajes de ficción está disociada de nuestras adscripciones a personas de la vida real.

III

Carroll es un defensor de “la teoría del pensamiento” con respecto al problema de las emociones en contexto de ficción.

Esta teoría afirma que nuestras emociones ficticiales son el producto de *sostener de manera no-assertiva* en nuestro pensamiento aquello que sucede en la ficción. Esta idea propuesta por Lamarque (1981, 1983) afirma que, a diferencia de lo que sostiene la paradoja de la ficción, no es necesaria la creencia en el objeto de la emoción para que ésta se produzca. El *mero pensamiento* en un objeto, se sostiene, es capaz de causar una emoción al respecto. Así, según Carroll, mientras miro en el cine que Freddy Krueger asesina a alguien con las cuchillas de sus manos, yo sostengo en mi pensamiento la idea de que Freddy Krueger asesina a alguien y esto es causa suficiente para que yo, a pesar de que no crea que aquello que hace Freddy sea real, lo encuentre peligroso y me produzca la sensación de miedo. Según Carroll, esto es similar al miedo que me puede provocar caerme cuando estoy en un balcón. Si pienso que me voy a caer, esto puede producirme miedo. No es necesario que yo crea que me voy a caer, es suficiente con que yo piense que me voy a caer para que se produzca en mí la emoción.

Así, la idea central de la teoría del pensamiento es que la emoción es producida por los pensamientos que consideramos (*entertain*) en nuestra mente a medida que consumimos una ficción. Estos pensamientos no son simulaciones ni creencias sino que son simples *consideraciones en el pensamiento*.

Ahora bien, esto supone que para provocar una emoción en el espectador el artista debe diseñar la obra de modo que ésta ponga al espectador a sostener estos pensamientos. Carroll explica esto a través de su teoría de los criterios pre-enmarcados (*pre-focused*). Para introducir esta teoría es importante distinguir las emociones de *otros estados* mentales que nos provocan las ficciones. Por ejemplo, las películas, a través de la manipulación de la imagen y el sonido, son capaces de provocar varios de estos estados mentales no emocionales. En particular, el cine es capaz de provocar reacciones reflejas, un pico alto en el volumen de un sonido nos puede provocar un sobresalto (imaginemos una película de terror en la que, cuando aparece el asesino, suena de golpe un nota alta y fuerte), o el movimiento rápido de una cámara hacia el suelo que puede provocarnos cierta sensación en el estó-

magos, o un objeto que pasa rápidamente a través del campo de visión puede ponernos alerta o un objeto cualquiera que se acerca rápidamente a la cámara puede sorprendernos “esquivando” algo inexistente. Todo este tipo de reflejos se presentan como cognitivamente impenetrables en el sentido de que aun sabiendo que nada de lo que pase nos puede hacer daño, los estados mentales suceden, la adrenalina se descarga, etc. Los estados reflejos están diseñados por la evolución para activarse ante un estímulo y el cine es capaz de producir esos estímulos, cuyas reacciones no podemos evitar aun cuando sepamos que en este contexto no representan lo que representaban para nosotros como especie cuando vivíamos en la selva. De la misma manera, el cine o el teatro puede explotar ciertas fobias compartidas simplemente por medio de mostrar ciertos insectos o poner al espectador – mediante la puesta en escena– en una situación de fobia: como un espacio cerrado que produzca claustrofobia o un espacio muy alto que produzca acrofobia. Sin embargo, el problema aquí está centrado en un tipo de hecho mental distinto.

Las emociones son aquellos fenómenos dentro de las afectaciones mentales en los cuales ciertas valoraciones (*appraisals*) de los estímulos que son relativos a ciertos intereses dan lugar a sensaciones cualitativas viscerales que nos disponen conductualmente para actuar. Así, la percepción de ciertas cosas dentro de nuestro espectro sensorial como “peligrosas” nos producen una sensación de miedo que nos dispone a actuar, ya sea enfrentando el miedo o huyendo de él. Estas emociones son más sutiles o “inteligentes” que las reacciones reflejas ya que pueden detectar peligro no sólo en un ruido fuerte si no, también, en una sonrisa diabólica o en el suave chasquido de un revólver. A pesar de que las emociones muestran una estructura cognitiva compleja en comparación con las reacciones reflejas –por ejemplo, no son cognitivamente impenetrables, una pistola que sé que es de juguete no me provoca miedo– comparadas con la deliberación racional son bastante rápidas. Imaginemos a alguien comparando precios para comprarse zapatillas frente a alguien que al ver una navaja se asusta: la decisión y la acción será instantánea. La emoción identifica una situación determinada y produce una

sensación asociada que nos dispone inmediatamente a la acción.

Parte del misterio de la paradoja de la ficción así reformulada es: ante la sensación de la emoción ¿cómo no actuamos en consecuencia? Si la respuesta es *porque no valoramos la situación como peligrosa, porque sabemos que estamos ante ficción*, la pregunta que surge es: ¿cómo se produce la emoción si no hay valoración? De esta última pregunta surge un hecho curioso que cita Carroll (2008)⁴: ante las películas más comerciales, espectadores de distintas geografías, trasfondos culturales, idiomas, géneros, etc., demuestran las mismas reacciones emocionales, ¿a qué se debe esta coincidencia? La respuesta de Carroll refiere a que existen ciertos criterios que presentan las emociones para activarse. Así, las emociones por un lado se activan por situaciones que son gobernadas por criterios que luego, junto con las sensaciones que estas producen, organizan el estado de cosas en concordancia con ese criterio. De esta manera cada emoción tiene un criterio o criterios que especifican la situación que va a activar en el ser humano la emoción y la sensación determinada. De la misma manera, ese criterio especifica la forma en la cual el ser humano organiza la situación o el estado de cosas que tiene enfrente. Así, para que algo me provoque miedo debo considerarlo peligroso. Ante algo peligroso mi emoción de miedo se activa, la sensación recorre mi cuerpo y todo lo que percibo del mundo es catalogado en base a esta emoción reinante. De modo que, si un ladrón me amenaza con un cuchillo, percibiré más fácil hacia dónde dirige su mirada y no si su camisa es a cuadros. Del mismo modo, cuando veo que un ser querido está triste, estaré más pendiente acerca de aquello que le puede haber provocado un daño que del tamaño de su billetera. El punto central aquí es que existen criterios que definen las situaciones que activan la emoción y configuran nuestro mundo una vez que la emoción ha entrado en juego. Esos criterios son, por ejemplo, lo peligroso en el caso del miedo, el daño hecho hacia mí o uno de los míos en el caso

4 Es importante señalar que si bien Carroll utiliza como ejemplo el arte popular masivo, y en particular el cine de masas, sus argumentaciones pretenden un alcance general a todo tipo de arte narrativo.

de la bronca, que otro tenga algo que yo deseo en el caso de la envidia, etc.

La tesis de Carroll en este punto es que toda obra de cine que produzca emoción está construida para poner de relieve de manera casi grosera aquello que satisface los criterios de la emoción que se busca provocar. De esta manera, cuando vemos una película que busca provocar la emoción de miedo en nosotros, el director pondrá el acento en aquellos aspectos del film que resulten peligrosos y esto provocará en el espectador la sensación característica que acompaña al miedo. Esta forma en la cual están construidas las películas produce una disanalogía entre las emociones que tenemos en nuestra vida cotidiana y las emociones que tenemos cuando contemplamos una obra de arte representativa. Así, es posible que, ante un comentario sarcástico y dañino por parte de alguien en nuestra vida cotidiana, el comentario nos pase de largo, es decir, no seamos capaces de darnos cuenta de su carácter y sólo luego, después de reflexionar, nos demos cuenta de que se nos ha querido provocar daño y nos produzca la emoción de bronca. En cambio, cuando contemplamos una obra de arte este comentario *no puede pasarse de largo*, ya que todos los recursos están puestos en hacerle notar al espectador que tal comentario dañino ha sido dicho y en consecuencia activar la emoción pertinente. Esto explica la razón por la cual, según Carroll, diferentes personas de distintos trasfondos se emocionan igual ante las películas masivas o los libros *best-sellers*. Dice Carroll:

En contraste con la forma en que las emociones deben empezar de cero cuando se trata de gobernar nuestra atención en la vida cotidiana, cuando se trata de un caso general de la contemplación de una película, los eventos en la pantalla han sido emotivamente pre-enfocados para nosotros por parte de los creadores de la película. Los creadores han seleccionado los elementos de la escena o secuencia que piensan son emotivamente significativos y nos los han, por así decirlo, tirado en la cara. Los medios para este fin a disposición de los creadores incluyen: la posición de la cámara y la composición, la edición, iluminación, el uso del color y, por supuesto, el acompañamiento musical, la actuación, el diálogo y la misma estructura del guión o narrativa. (Carroll, 2008: 158).

Así, la forma en que está construida la película garantiza que aquello que enciende o activa las emociones sea tan evidente que sea casi imposible pasarlo por alto. Es en este sentido en el cual las obras de arte están construidas para provocar emociones. La tarea del autor consiste, por un lado, en determinar qué situación o qué aspectos de una situación son peligrosos (si quiere provocar miedo), tristes (si quiere provocar pena) o indignantes (si quiere provocar bronca). Por el otro lado, su tarea consiste en mostrar estas situaciones o aspectos de una manera que, de hecho, provoque la emoción que busca provocar.

De este modo, Carroll revela la estructura que debe tener la obra de arte para emocionar al espectador. Sin embargo, el autor sostiene que existen ciertos recursos en las obras que son más poderosos que otros. En particular, el autor sostiene que los personajes de las historias son grandes “palancas” de la emoción. La idea general del autor es que tanto los protagonistas como los antagonistas de las obras de arte narrativas generan emociones por sí mismos. Estas emociones se generan a partir de una relación *emocional central* que se establece entre los espectadores y los protagonistas (y antagonistas).

Como vimos antes, la idea de Currie es que en la contemplación de ficción se establece una relación de simulación entre los espectadores y los personajes de la ficción. En este sentido Currie no distingue personas de personajes. Sin embargo, sostiene Carroll, las obras de arte narrativas no son como nuestra vida cotidiana. Los personajes no son personas y, a diferencia de ellos, están diseñados para que su comportamiento sea comprendido. Existe, entonces, una disanalogía entre el comportamiento de los seres humanos y el comportamiento de los personajes de las obras de arte. Así, dado el comportamiento poco transparente de nuestros congéneres en la vida cotidiana, es probable que nos valgamos de la teoría de la simulación (o de la teoría de la teoría) para explicar los motivos de su comportamiento o querer predecirlo. Sin embargo, los personajes de ficción están contruidos para que los comprendamos, esto implica que las estrategias que utilizamos para comprender a los personajes pueden no ser las mismas que utilizamos para comprender a nuestros congéneres.

Un ejemplo puede ser muy ilustrativo de esto: cuando leemos una novela o vemos una película que tiene una voz en *off*, accedemos por medio del narrador *directamente* a los estados mentales del personaje. Así, una estrategia del estilo de las que usamos en nuestra vida cotidiana parece completamente redundante. La idea de Carroll es que para comprender los estados mentales de los personajes de las obras de arte representativas (cuando, por ejemplo, no hay una voz en *off* o un narrador que nos permite acceder directamente a ellas) nos valemos de ciertas guías heurísticas como esquemas, guiones preestablecidos, prototipos, guías contextuales, ejemplares, etc., para descifrar sus conductas. Así, los personajes nos dejan saber lo que están pensando, respondiendo si se quiere a cierto cliché de personalidad, o a ciertas expresiones faciales o, más evidentemente, al diálogo con otros personajes. En este sentido, los personajes de las obras de arte narrativas son menos opacos que las personas porque así han sido contruidos.

Usar la estrategia de la simulación para comprenderlos es un gasto de tiempo innecesario. La comprensión de los estados mentales de los personajes de ficción es mucho más rápida, simple y directa que el uso de una teoría *folk* de la mente o la simulación de ciertos estados mentales de otro. Esta práctica de comprensión puede, incluso, traspasar la frontera de la ficción, a veces leemos a las personas como leeríamos a personajes. Carroll argumenta que a diferencia de lo que sostiene la teoría de la simulación, en nuestra vida cotidiana somos más proclives a valernos de estas herramientas heurísticas atadas a ciertos guiones culturales compartidos que a usar la simulación para comprender a otros. Así, si nos encontramos con un compañero que ha perdido a su novia o lo han echado del trabajo inferimos en base a un guión compartido culturalmente que, en condiciones normales, nuestro compañero estará deprimido. En este sentido, no necesitamos de la simulación para establecer cuál es el estado emocional en el que se encuentra.

Con respecto a la relación fundamental que se establece con los personajes que moldean nuestras emociones, las teorías de cine han sostenido tradicionalmente que nos *identificamos* con

los personajes de una obra narrativa, sobre todo con sus protagonistas. Esta teoría ha sido adoptada incluso por el público, que suele sostener que una película o una novela les gustó porque se identificaron con el protagonista o el antagonista, o que la obra no les gustó nada porque no pudieron identificarse ni un poco con los personajes. Esta idea, incluso, parece estar detrás de ciertos actos de censura o planificación de edades para el consumo de las obras. Muchas obras pueden ser censuradas o marcadas como prohibidas para menores de edad porque se teme que los espectadores imiten el comportamiento de los personajes que ahí se retratan.

Si bien no es posible aquí determinar qué concepto de identificación específico maneja esta teoría, seguramente incluya la idea de que alguien que está identificado con otro sostiene sus mismos estados emocionales ante las mismas situaciones. Esto es lo que Carroll llama el modelo por “infección”. Así, cuando en las películas de Freddy Krueger un personaje se aterroriza por ver a Freddy, el espectador se aterroriza con él. Además, la identificación por infección de la que habla Carroll supone que el estado emocional del personaje es causa del estado emocional del espectador. De esta manera, identificarse emocionalmente con alguien es tener su mismo estado emocional *porque* ese alguien *está* en ese estado emocional. Este modelo supone que el terror del espectador al ver *El exorcista* consiste, entre otras cosas, en su identificación con el padre Karras y su susto es producto del susto del personaje.

Sin embargo, Carroll argumenta en contra de este modelo. Sostiene que no parece que el espectador se asuste porque un personaje esté asustado sino, más bien, porque Freddy le resulta peligroso. Un ejemplo podría ser el siguiente: si viéramos una película de terror mala (con una historia pobre, malos efectos y monstruos no terroríficos), en un enfrentamiento entre el monstruo y el protagonista no nos asustaríamos aun si viéramos al protagonista muerto de miedo. Y es que lo más probable es que una película de terror mal hecha nos produzca una risa burlona dirigida a lo absurdo o ridículo de la situación. Esto es así porque nuestro miedo parece estar regido no por los estados

emocionales que muestra un personaje, sino por la habilidad que tienen los artistas de mostrarnos una situación o una criatura que resulte peligrosa. Así, compartimos el mismo estado emocional que el personaje asustado pero no lo hacemos por causa de su propio estado emocional. ¿Esto implica que aquellos estados emocionales que tenemos no se deben a los estados emocionales de los personajes? Para complicar un poco más el asunto pensemos el caso de una heroína que vuelve contenta a su casa a contarle a su pareja que ha obtenido un ascenso en el trabajo y allí la espera un asesino. Aquí, nuestra emoción no es de excitación u orgullo sino de suspenso: un asesino ha matado a su pareja y la vida de la heroína corre peligro. En este caso, no sólo nuestros estados emocionales no están conectados de manera causal –como no lo estaban en el caso de *El exorcista*– sino que, como ya dijimos, ni siquiera son coincidentes, personaje y espectador no comparten un mismo estado emocional. La clave de esta asimetría de emociones reside en la asimetría de información entre el personaje y el espectador, es esta asimetría la que genera estados emocionales no coincidentes.⁵ Los estados no conectados y no coincidentes entre espectadores y personajes ficcionales son las dos grandes razones por las que Carroll sostiene que el modelo de identificación por infección no funciona.

Dicho esto, Carroll no pretende sostener que no existe ninguna relación entre espectador y el personaje que provoque emociones. Pensemos en aquellas películas en las que el protagonista finalmente consigue aquello que tanto buscaba y el espectador se alegra, o en las que dos protagonistas que se venían peleando durante toda la película finalmente se dan cuenta de que están profundamente enamorados el uno del otro y deciden convivir juntos. En este caso pareciera haber algún tipo de relación entre aquello que les pasa a los personajes y aquello que le pasa a los espectadores. Nótese que el estado emocional no tiene por qué

⁵ La asimetría de información entre espectador y personajes es un recurso central en la narración. Usualmente, aquellos que contemplan una obra narrativa suelen tener más información que los personajes pero este no siempre es el caso: Sherlock Holmes siempre sabe algo que sus lectores no saben hasta el final de las novelas.

ser coincidente: los espectadores pueden tener el estado emocional de alegría pero este difiere del de los personajes que sienten amor. Sin embargo, aquí es claro que la alegría del espectador es causa del estado emocional de los personajes. De la misma manera, Ana Karenina está triste pero el objeto de su estado emocional no es el mismo que el de los lectores: ella está triste por todas las cosas desgraciadas que ha vivido, los lectores están tristes porque ella está triste. Es más, probablemente mientras su estado emocional califique como pena, el de los lectores probablemente califique como compasión. En este sentido, Carroll sostiene que los estados emocionales de los espectadores plantean una relación de *tipo vectorial*. Así, si un espectador establece una relación con alguno de los personajes, sus estados seguirán un mismo valor.⁶ Cuando el personaje tiene emociones positivas, los espectadores experimentan emociones positivas. Si el personaje tiene emociones negativas, los espectadores experimentan emociones negativas.

Según Carroll, entonces, la relación que juega un rol central en nuestro involucramiento con los personajes de ficción es la simpatía. Esta relación emocional se entiende como una actitud de interés no pasajero o una preocupación o una actitud de aliento hacia otra persona o personaje. Así, tener simpatía por otro implica preocuparse por su bienestar y alentar –de alguna manera– para que esto ocurra. Una característica importante de la simpatía es su capacidad para generar otro tipo de emociones. Por ejemplo, si un espectador tiene simpatía por un protagonista, entonces, si alguien le hace daño, esto le provocará bronca, puesto que siente que se le ha hecho daño a uno de los suyos. De la misma manera, si el protagonista sale de una crisis personal, debido a la simpatía que le tiene el espectador, es probable que sienta alegría por él (mientras que Karras sienta alivio, por ejemplo). En este sentido, la relación emocional de simpatía tiene la capacidad de producir un conjunto amplio de *otras emociones que no son las que experimenta el personaje de ficción*. Esta capacidad,

6 Considerando que las emociones tienen valores positivos o negativos, eufóricos o disfóricos.

que Carroll llama “profundidad”, es una de las dos razones por las que el autor considera que la simpatía es la relación principal que se establece con los personajes. La segunda razón es su “alcance”. Así, el autor sostiene que la simpatía es la emoción más ubicua durante el transcurso de la obra de arte. En efecto, en el momento en el que idealmente el protagonista se gana la simpatía del espectador, esta simpatía persiste en todos los pasajes de la aventura y no se cancela ante casi nada. Del mismo modo, un sentimiento de antipatía puede recorrer la obra de arte. Un personaje que resulte odioso puede ser capaz de provocar las más variadas emociones: eufóricas cuando a este personaje le va mal o disfóricas cuando le va bien.

La simpatía y la antipatía son dos caras de una misma moneda. La simpatía puede generar antipatía en aquellos personajes que se oponen al personaje por el cual tenemos simpatía y viceversa, una antipatía profunda por un personaje puede generar cierta simpatía por alguien que se le oponga. En este sentido, el autor de la obra tiene distintas armas para convencer al espectador de la simpatía por el protagonista, por un lado puede hacer al protagonista querible o por el otro lado puede hacer el antagonista detestable. Usualmente, el autor se vale de ambos recursos aunque, a veces, si se busca una mayor profundidad, el autor puede intentar trabajar sobre esto y no hacer tan bueno al bueno o tan malo al malo. Existen combinaciones de estos casos que pueden ser interesantes: estamos del lado de Hannibal Lecter en *El silencio de los inocentes* porque es el único que puede ayudar a detener a Buffalo Bill, aquel que se opone a los intereses de Clarice Starling (la protagonista de la película y depositaria de nuestra simpatía).

Dados estos argumentos, Carroll sostiene que la simpatía es el candidato ideal para comprender qué relación se establece entre el espectador y los personajes de ficción. La identificación, por su lado, requiere que estemos en un mismo estado emocional que los personajes de la ficción y que, además, los estados estén conectados de algún modo. Y, como vimos, muchas veces el espectador no está en el mismo estado que los personajes de ficción, sin contar el hecho de que, cuando lo está, no lo hace por

una conexión particular con el estado emocional del personaje sino porque aprecia una situación dada de la misma manera en que lo hace el personaje. La simpatía, en cambio, explica por qué hay estados no coincidentes pero sí similares en sus valencias. Existe una conexión entre aquello que le pasa al protagonista y aquello que le pasa al espectador y eso tiene que ver con una actitud pro-personaje que el espectador adopta. Este hecho incluso es capaz de explicar por qué a veces ni siquiera tenemos un estado emocional con la *misma valencia* que los personajes.

Imaginemos al protagonista de una película acerca del cual guardamos la relación de simpatía. El personaje se enamora perdidamente de una chica que nosotros sabemos que tiene dobles intenciones. El estado emocional del personaje es de amor y, quizá también, alegría. La valencia de estos estados es positiva. Sin embargo, nosotros sabemos que la chica no es buena para él y nos disgustamos o nos frustramos, emociones con valencia negativa. Este hecho, en vez de ser misterioso o desconcertante, se explica perfectamente por medio de la simpatía. Queremos lo mejor para el personaje, queremos su bienestar, y sabemos que la chica no le conviene, que le va a hacer daño; por eso, nos disgustamos. Así, la simpatía es capaz de explicar estados emocionales coincidentes y no coincidentes.

Es interesante la posición de Carroll sobre cómo se logra la simpatía. La idea es que los personajes deben estar diseñados de determinada manera para provocar simpatía en el espectador. El problema consiste en que no todos los espectadores tienen los mismos intereses, trasfondo cultural o incluso creencias. La idea de Carroll es que aquello que busca el autor es encontrar un apoyo moral por parte del público. Así, los protagonistas se presentan con ciertos valores: una inclinación pro-familia o pro-social, una inclinación por la justicia, lealtad, honor, honestidad, etc. Valores morales de amplio espectro que resonarán en la audiencia como valores positivos sobre los que se debe alentar. Esto no quiere decir que todos los protagonistas sean buenos en todos los aspectos, un personaje puede tener aspectos oscuros o no simpáticos, sin embargo, el espectador debe intuir que detrás de esos aspectos oscuros se encuentra un individuo que personifica

de alguna manera uno o varios de los valores morales de los que hablamos. Incluso cuando los personajes no demuestran esa rectitud moral, existe un trasfondo o un elenco de personajes aún peores. Es el caso de *Dexter*, el asesino serial que asesina gente que es culpable de crímenes horrendos. O el caso de *El Padrino*, que es jefe de la mafia pero que tiene valores familiares mucho más altos que los valores del resto de los personajes, incluso que los policías o senadores que deberían dar el ejemplo moral en la ficción, pero que siempre terminan siendo corruptos o depravados. La idea detrás de esto es mostrar que si nadie es virtuoso, elegiremos al menos malo. O, mejor, que dentro de un ambiente sin virtud, elegiremos al que se destaque por tener alguna virtud perdida. En este sentido, el autor destaca que no es una casualidad que los protagonistas sean “los buenos” puesto que el amplio abanico de emociones que se experimentan es en base a la relación emocional que se construye con ellos.

IV

En este capítulo intenté presentar dos posiciones con respecto a la adscripción de estados mentales a personajes de ficción. La primera posición, de Gregory Currie, sostiene que la contemplación de una ficción narrativa involucra un tipo de simulación específica que garantiza que los espectadores se emocionen con algo que saben positivamente que no es el caso. Además, el autor afirma que la adscripción de estados mentales a personajes de ficción se hace también a través de una simulación parasitaria de la simulación que ocurre cuando adscribimos estados mentales en nuestro contexto cotidiano. En este sentido, la posición de Currie parece tener la ventaja de apelar a los mismos mecanismos tanto en la adscripción a personas como a personajes. Sin embargo, Currie parece tomar, sin que medie ninguna argumentación, a la teoría de la simulación como la teoría correcta de adscripción de estados mentales sin importarle que la discusión acerca de cuál es la teoría correcta siga abierta en la literatura actual.

Por otro lado, Carroll sostiene que la adscripción de estados mentales a personajes de ficción se establece por medio de la relación de simpatía entre estos y los espectadores. Así, pareciera que los espectadores sostienen ciertos pensamientos que la obra propone y a medida que los recrean forman una relación de simpatía con los personajes propuestos. Tal simpatía permite adscribir de manera correcta los estados mentales que el autor de la obra pone de manifiesto de una forma transparente. Esto contrasta con la dificultad que tenemos como humanos de adscribir estados mentales en contextos cotidianos en donde pareciera que las mentes nos son más bien opacas. Así, esta postura parece un poco en desventaja teóricamente con respecto a la anterior, puesto que debe apelar a una nueva explicación para la adscripción a personajes de ficción. Por otra parte, la postura no termina de explicitar cómo es que se desarrolla esta simpatía entre espectador y personaje y en consecuencia cuáles son los mecanismos psicológicos involucrados en ella.

Es interesante notar, entonces, que si bien el fenómeno de adscripción psicológica en contextos de contemplación de ficción es algo bastante extendido, su discusión parece bastante poco desarrollada. Así, mientras existe un consenso claro sobre la existencia del fenómeno, no existe un acuerdo acerca de si el tipo de adscripción en contextos de ficción es análoga a su contrapartida en la interacción social. Y este parece ser un punto de partida básico sobre el cual debería montarse una discusión sobre los mecanismos psicológicos involucrados en la adscripción de estados mentales a personajes de ficción.

Bibliografía

- Carroll, N. (1987), "The nature of horror", *The Journal of Aesthetics and Art Criticism*, 46(1), pp. 51-59.
- . (1990/2005), *Filosofía del terror o paradojas del corazón*, A. Machado Libros.
- . (1991), "On Kendall Walton's Mimesis as Make-Believe", *Philosophy and Phenomenological Research*, pp. 383-387.
- . (2008), *The philosophy of motion pictures*, Blackwell Pub.
- Currie, G. (1991), "Visual fictions", *The Philosophical Quarterly*, pp. 129-143.
- . (1995a), *Image and mind: Film, philosophy and cognitive science*, Cambridge University Press.
- . (1995b), "Imagination and simulation: Aesthetics meets cognitive science", en: Davies, M. y T. Stone (eds.), (1995), pp. 151-169.
- . (1997), "The paradox of caring: Fiction and the philosophy of mind", en: Hjort, M. y S. Laver S. (eds.), (1997), pp. 63-77.
- Davies, M. y T. Stone (eds.). (1995), *Mental simulation: Evaluations and applications - Reading in Mind and Language*, Wiley Blackwell.
- Gaut, B. (1993), "The paradox of horror", *The British Journal of Aesthetics*, 33(4), pp. 333-345.
- Gomila, A. (2011), "Emociones en el teatro: ¿por qué nos involucramos emocionalmente con una representación?", *Ponencia en el congreso: La representación de las pasiones*, Madrid.
- Hjort, M. y S. Laver (eds.) (1997), *Emotion and the Arts*, Oxford University Press.
- Lamarque, P. (1981), "How Can We Fear and Pity Fictions?", *British (The) Journal of Aesthetics* London, 21(4), pp. 291-304.
- . (1983), "Fiction and reality", en: Lamarque; P. (ed.), *Philosophy and Fiction: essays in Literary Aesthetics*, Aberdeen: Aberdeen University Press.
- Moran, R. (1994), "The expression of feeling in imagination", *The Philosophical Review*, pp. 75-106.
- Radford, C. (1975), "How can we be moved by the fate of Anna Karenina?", *Proceedings of the Aristotelian Society, Supplementary Volumes*, 49, pp. 67-80.
- . (1977), "Tears and fiction", *Philosophy*, 52(200), pp. 208-213.
- . (1982), "Philosophers and Their Monstrous Thoughts", *The British Journal of Aesthetics*, 22(3), pp. 261-263.
- Walton, K. (1978), "Fearing fictions", en: *The Journal of Philosophy*, 75(1), pp. 5-27.

Atribución de estados mentales conscientes: cartesianismo vs. expresivismo

Alejandro Zárate

1. Introducción

En circunstancias habituales, si alguien quiere saber lo que pienso, lo que siento o lo que pretendo, la mejor opción es que me lo pregunte a mí (por lo menos si la información que quiere obtener es acerca de mis estados mentales *conscientes*). Cuando los demás no aceptan mis auto-atribuciones mentalistas tal y como yo las presento, suele ser más porque consideran que no soy sincero que porque piensen que estoy equivocado. Después de todo, ¿puedo realmente equivocarme con respecto a mi deseo de estar ahora de vacaciones o al dolor de cabeza que siento mientras lo estoy sintiendo? Desde luego que todo esto no nos obliga a aceptar que el error esté completamente excluido en este ámbito o que otras personas que nos conozcan mucho, como nuestras parejas, nuestros amigos o nuestros familiares, no puedan realizar consideraciones más o menos confiables acerca de nuestra condición psicológica. Sin embargo, en este último caso estaríamos dispuestos a sostener que se trataría de información de “segunda mano” acerca de nuestros estados mentales, y en caso de que se suscite una discrepancia entre mi opinión acerca de ellos y la de los demás, se preferirá mi palabra.

En contraposición, la atribución de estados mentales a otras personas no goza de la misma seguridad, dado que, entre otras cosas, siempre persistiría la posibilidad de que esas personas no estén siendo honestas con nosotros o que sencillamente no demuestren sus pensamientos y sensaciones. Así, en lo que respecta a las atribuciones psicológicas dirigidas a otras personas, las cosas no siempre son lo que parecen.

Frente a este hecho (a saber, que se suele aceptar que al menos ciertas afirmaciones y pensamientos acerca de nuestros propios estados mentales son mucho más seguros y confiables que las afirmaciones que podemos hacer acerca de los estados mentales de los demás, o de cualquier otro hecho del mundo), muchos filósofos se han dado a la tarea de intentar dar cuenta de él, de determinar si nuestra confianza en dichas afirmaciones está verdaderamente justificada y, de ser así, cuál es su alcance y origen.

En el presente trabajo van a confrontarse específicamente dos maneras distintas de dar cuenta de la presunta asimetría entre las atribuciones de estados conscientes en primera persona del singular (como “Tengo un dolor”) y en tercera persona (como “Él siente dolor”). Por un lado, en la siguiente sección desplegaré una perspectiva que denominaré “cartesiana”, según la cual, la asimetría en cuestión se funda en una asimetría *epistémica*, es decir, en los diferentes tipos de conocimiento que tenemos acerca de los estados mentales conscientes propios y acerca de los estados mentales ajenos. Luego de describir algunos rasgos centrales de esta posición, en la sección 3 mostraré las dificultades a las que se enfrenta un cartesiano contemporáneo como el filósofo David Chalmers al asumir (a) que el carácter de nuestro conocimiento acerca de los estados mentales conscientes propios es, esencialmente, privado y (b) un dualismo ontológico de propiedades.

Del otro lado, en la sección 4 voy a mostrar una explicación diferente de esta asimetría desde una perspectiva “expresivista”, tal como la que presenta la filósofa Dorit Bar-On (2004, 2010). Desde esta posición, si bien se acepta que existe una asimetría entre los distintos tipos de atribuciones mencionadas, se considera que es posible dar cuenta de ella sin tener que enfrentar las consecuencias indeseables que pesan sobre la perspectiva cartesiana.

2. La idea cartesiana de la mente y el problema de la consciencia

Siguiendo la distinción de Diana Pérez (2007), una de las tesis que forma parte de una perspectiva “cartesiana” de la mente sos-

tiene que la mente humana (o al menos algunos de sus estados) es aquello sobre lo que tenemos conocimiento en el más alto grado posible. En contraposición, se considera que el mundo externo (que incluye también los estados mentales de las demás personas) es conocido de una manera más deficiente, de un modo siempre hipotético y sujeto a la posibilidad de error. Así:

Por el contrario, [se supone que] no podemos equivocarnos cuando hacemos autoatribuciones mentalistas en primera persona del presente indicativo del tipo: 'Ahora estoy teniendo un dolor'... (Pérez, 2007: 129).

Como un ejemplo paradigmático de la perspectiva cartesiana con respecto a los estados mentales conscientes y el peculiar conocimiento que tendríamos acerca de ellos, podemos citar las ideas del filósofo contemporáneo David Chalmers:¹

La experiencia consciente es, al mismo tiempo, lo más familiar del mundo y lo más misterioso. De ninguna otra cosa tenemos un conocimiento más directo que de la conciencia, pero no es claro en absoluto cómo reconciliarla con todo el resto de los que sabemos (Chalmers, 1999: 25).

En las discusiones filosóficas contemporáneas dentro de la tradición analítica, el concepto de “experiencia consciente” hace referencia a ciertos aspectos o propiedades que estarían presentes con mayor claridad en los estados perceptivos (como las experiencias visuales de rojo) y en las sensaciones (como los orgasmos y los dolores), y a los que se etiqueta con diferentes nombres como “consciencia fenoménica”, “sensaciones crudas”, “aspecto cualitativo de la experiencia”, “propiedades fenoménicas de la experiencia”, “*qualia*” o el célebre “*what it is like*” (cómo se siente)

¹ David Chalmers es sin dudas el autor que animó en gran medida el debate en torno a la conciencia a mediados de la década del 90 y principios de la década siguiente, defendiendo prácticamente en solitario una posición dualista de manera argumentada y coherente. De todos modos, el artículo que devolvió a la conciencia a la palestra de las discusiones filosóficas fue el artículo del filósofo Thomas Nagel “What Is It Like to Be a Bat?” de 1974.

tener una experiencia, introducido por el filósofo Thomas Nagel. Al mismo tiempo, la consciencia suele presentarse por parte de filósofos y científicos cognitivos de diferentes corrientes y filiaciones teóricas, como un fenómeno que constituye “el mayor obstáculo pendiente en nuestra búsqueda de una comprensión científica del universo”, un misterio, es decir, “un fenómeno para el cual todavía no hallamos una forma de pensar” (Dennett, 1995: 33).

En efecto, si bien se reconoce, por ejemplo, que la ciencia física o la biología son aún incompletas y que no todas las preguntas en esos campos cuentan en la actualidad con una respuesta acabada, se considera que tenemos de ellas una comprensión más que razonable y una buena idea de la forma que podrían tener las soluciones a esas preguntas pendientes. Por el contrario, la respuesta a la pregunta de cómo es posible que la consciencia surja, con sus características específicas, de la opaca materia gris del cerebro, aparece en principio como una cuestión cualitativamente distinta a las preguntas que se plantean en otros campos científicos, convirtiéndose así en un enigma desconcertante para la ciencia cognitiva y la neurociencia. Al parecer (si bien este modo de ver las cosas no es unánime), parte de la dificultad para dar cuenta de la consciencia podría tener que ver, como veremos enseguida, con la incompatibilidad de algunas de las propiedades que se le atribuyen más o menos implícitamente (como por ejemplo su carácter intrínseco o no relacional y el peculiar conocimiento que tendríamos acerca de ella), y las propiedades de los objetos y entidades admitidos por el resto de las disciplinas científicas.

En particular, Chalmers considera, en línea con el paradigma cartesiano, que tener una experiencia consciente es automáticamente estar en alguna relación epistémica íntima con la misma, una relación que el autor llama, siguiendo en parte a Russell, de “familiaridad” (*acquaintance*). Desde esta perspectiva, ni siquiera habría una posibilidad conceptual de que un sujeto pueda tener una experiencia de rojo sin tener ningún contacto epistémico con ella (Chalmers, 1999: 256). Sería precisamente esta “familiaridad” que el sujeto tiene con su propia experiencia consciente lo

que proporcionaría evidencia a nuestras creencias y juicios fenoménicos y lo que los transformaría en auténtico conocimiento.²

De acuerdo con Chalmers, existen razones poderosas para pensar que el conocimiento que tenemos de nuestros estados conscientes no puede estar basado simplemente en una mera relación causal o en el funcionamiento de algún mecanismo fiable. Como veremos a partir de la cita siguiente, el rechazo de la teoría causal del conocimiento no tiene que ver con que Chalmers considere que una relación natural, como lo es una relación causal, no pueda eventualmente justificar (aunque no por sí sola) una creencia o un juicio. Las razones, en cambio, son las siguientes:

Si la única fuente de justificación de una creencia acerca de X [en este caso, acerca de una propiedad fenoménica o una determinada experiencia consciente], es una conexión causal con X, entonces un sujeto no puede saber con seguridad que la conexión causal existe [ni que X existe, es decir, que haya conciencia del otro extremo]. El único modo en el que podría saber esto con seguridad sería si tuviese algún acceso independiente a X o a la cadena causal, pero esto implicaría conocimiento basado en algo más que la propia cadena causal. Siempre habrá un escenario escéptico en el cual todo le parece igual al sujeto, pero en el que la conexión causal está ausente y X no existe; de modo que el sujeto no puede saber con seguridad acerca de X [...]. En el caso del conocimiento perceptual, por ejemplo, podemos construir una situación en la cual la conexión fiable [o causal] esté ausente –por ejemplo, una situación en la cual el sujeto es un cerebro en un tanque, digamos– y todo le parecerá igual [...]. Pero nosotros sabemos con seguridad que somos conscientes: de modo que una concepción causal de este conocimiento es inapropiada (Chalmers, 1999: 253-254).

2 Cabe aclarar en este punto que Chalmers no sostiene que tener una experiencia implique automáticamente *saber* acerca de ella si el “saber” requiere creencia, porque a su juicio tenemos muchas experiencias de las que no tenemos creencias y que por lo tanto no conocemos en este sentido. Tampoco se compromete con la idea de que nuestras creencias acerca de la experiencia sean indubitables, debido a que dichas creencias podrían formarse por cualquier tipo de razones y a veces lo hacen por razones equivocadas.

Ahora bien, una consecuencia de esta concepción cartesiana es que el conocimiento de los estados conscientes de las demás personas, de las “otras mentes”, y las atribuciones mentalistas correspondientes, resultan tan poco fiables como cualquier afirmación acerca del mundo externo debido al diferente tipo de acceso epistémico que tenemos a los hechos referidos por los distintos tipos de atribuciones. En particular, Chalmers asume que el conocimiento que cada uno tiene acerca de sus propios estados mentales conscientes es *epistémicamente privado*, es decir, que sólo quien los posee puede saber si efectivamente es consciente y cuáles son las características de ese estado. En contraposición, los demás sólo pueden conjeturar la presencia y el carácter de nuestros estados mentales a partir de la observación de cierta evidencia indirecta como la conducta.³ Pero, en sentido estricto, todo lo que pueda saberse sobre los estados mentales de las personas mediante este tipo de evidencia indirecta es compatible con la ausencia de la conciencia y sus rasgos característicos.

Nuestras razones para creer en la conciencia se derivan exclusivamente de nuestra propia experiencia de ella [...]. Es verdad que ciertos hechos físicos acerca del mundo podrían proporcionar alguna evidencia indirecta en favor de la existencia de la conciencia. Por ejemplo, a partir de estos hechos podríamos averiguar que hay muchos organismos que afirman ser conscientes y dicen tener misteriosas experiencias subjetivas. Con todo, esta evidencia sería bastante inconcluyente, y podría ser muy natural extraer una conclusión eliminativa: que de hecho no hay ninguna experiencia presente en estas criaturas, sólo mucha charla (Chalmers, 1999: 141-142).

3 Desde luego, Chalmers no es el único que adhiere a al menos este aspecto del paradigma cartesiano. Por el contrario, filósofos como John Searle, que no alientan ningún tipo de dualismo ontológico, y neurocientíficos reconocidos como el premio Nobel, Francis Crick, entre otros, comparten estas ideas. Searle, por ejemplo, sostiene que “mi estado presente de conciencia es una característica de mi cerebro, pero sus aspectos conscientes son accesibles para mí de un modo en que no son accesibles para ti. Y tu estado presente de conciencia es una característica de tu cerebro y sus aspectos conscientes son accesibles para ti de un modo en que no son accesibles para mí” (Searle, 1984: 25). Crick, por su parte, afirma que “estrictamente hablando, cada individuo tiene certeza únicamente de que él mismo es consciente” (Crick, 1995: 107).

De esta manera, se desemboca rápidamente en lo que se conoce como el “problema de las otras mentes”, dado que, de acuerdo con Chalmers y la concepción cartesiana en general, “aun cuando conozcamos todas las características físicas de otras criaturas, no sabremos con seguridad que son conscientes, o cuáles son sus experiencias [...]” (Chalmers, 1999: 142). En cierto sentido, podría pensarse que este modo de ver las cosas encuentra algún apoyo en el hecho de que mientras que para atribuir estados mentales a otras personas nos basamos en la evidencia que nos brinda su conducta, en la auto-atribución de estados mentales no nos apoyamos en ese tipo de evidencia. Uno no afirma que siente dolor en virtud de percatarse de que está llorando o quejándose.

Por otra parte, esta asimetría entre los distintos tipos de accesos que tenemos a los estados mentales conscientes propios y ajenos le permite construir a Chalmers un argumento en favor de un dualismo ontológico: el argumento de la asimetría epistémica. Pero para comprenderlo, debemos primero introducir la noción de “superveniencia” que utiliza el autor para desarrollarlo y de “fiscalismo”.

El fiscalismo o materialismo, que es la posición dominante en sus distintas variantes dentro de los filósofos analíticos contemporáneos que se ocupan del “problema de la conciencia”, sostiene que todo lo que existe en el mundo es físico, o que los hechos físicos agotan en cierto sentido todos los hechos acerca del mundo, o que los hechos más básicos acerca del mundo son hechos físicos y que todos los demás hechos dependen de ellos. En este contexto, el concepto filosófico de superveniencia intenta formalizar precisamente la idea de que un conjunto de hechos puede determinar por completo otro conjunto de hechos.

Específicamente, la superveniencia es un tipo de relación que existe entre dos conjuntos de propiedades dados. Al primer *relata* le llamamos base de superveniencia, mientras que al segundo lo llamamos propiedad superveniente. La relación es tal que, si fijamos el tipo de propiedades de la base superveniente estaremos fijando las propiedades de lo superveniente. Así, de acuerdo con la definición de Chalmers (1999) de la que me voy a servir en este capítulo, “Las propiedades B *supervienen* a las propiedades A si

ningún par de situaciones posibles es idéntico respecto de sus propiedades A pero difiere en sus propiedades B”.

Establecida esta definición, debemos distinguir entre distintos tipos de superveniencia en función del tipo de noción de “posibilidad” que consideremos en juego. Para los fines que persigo, será conveniente distinguir entre dos tipos de relación de superveniencia determinados por dos tipos de posibilidad: la posibilidad lógica y la posibilidad natural.

Si hablamos de posibilidad en este mundo, con estas leyes naturales que nos rigen, hablamos de superveniencia *natural*, si hablamos de posibilidad en cualquier tipo de mundo, con cualquier tipo de leyes, hablamos de superveniencia *lógica*. Así, cuando hablemos de superveniencia natural estaremos pensando en situaciones naturalmente posibles, es decir, situaciones que no desafían las leyes de la naturaleza que rigen en nuestro mundo; y cuando hablemos de superveniencia lógica estaremos pensando en situaciones lógicamente posibles, es decir, básicamente, en situaciones no contradictorias o incoherentes pero que pueden ser incompatibles con las leyes naturales conocidas.

De acuerdo con Chalmers, la presión que ejerce un mol de gas depende sistemáticamente de su temperatura y volumen según la ley $pV = KT$. En el mundo real, cuando hay un mol de gas a una temperatura y volumen dados, su presión estará determinada: es empíricamente imposible que dos moles distintos de gas supervengan a su temperatura y volumen. De acuerdo con la definición de superveniencia que vimos, podríamos decir en este caso que la presión superviene naturalmente a la temperatura y el volumen de un mol de gas, es decir, es natural o empíricamente imposible, dadas las leyes de la naturaleza que rigen en nuestro mundo, que un mol de gas con una temperatura y volumen determinados puedan tener una presión diferente. Sin embargo, no es lógica o conceptualmente imposible que esto no sea así de la misma manera en que no lo es un universo sin gravedad.

En cambio, para Chalmers hay una relación de superveniencia lógica entre los hechos físicos de nuestro mundo y los hechos biológicos, de modo tal que no habría espacio lógico para que haya un mundo físicamente idéntico al nuestro pero que sea

biológicamente distinto porque una vez que fijemos todos los hechos físicos acerca del mundo, habremos fijado también “la forma macroscópica de todos los objetos en el mundo, el modo como se mueven y funcionan, el modo como interactúan físicamente” (Chalmers, 1999: 63). La noción de un mundo físicamente idéntico al nuestro pero biológicamente diferente es para Chalmers tan incoherente como la noción de “zorra macho”. En general, “cuando las propiedades B supervienen lógicamente a las propiedades A, podemos decir que los hechos A implican a los hechos B, donde un hecho implica lógicamente a otro si es lógicamente imposible que el primero sea verdadero y el segundo no” (Chalmers, 1999: 63).

Por el contrario, puede considerarse que parte de lo que intentaron probar muchos de los argumentos más discutidos dentro de la filosofía de la mente de los últimos años con respecto a la consciencia (como el “argumento del conocimiento” de Frank Jackson 1982 o el argumento de los zombis), es precisamente que la consciencia no superviene lógicamente a lo físico. El argumento de los zombis, en particular, sostiene como una de sus premisas que es posible concebir, sin incurrir en ningún tipo de incoherencia, un “mundo zombi”, es decir, un escenario en el cual las propiedades físicas sean idénticas a las de nuestro mundo pero en el que la experiencia consciente está completamente ausente (o es diferente). En otras palabras, para los defensores del argumento de los zombis, es posible concebir, por ejemplo, un individuo que exhiba todas las conductas verbales y no verbales asociadas al dolor en nuestro mundo, que se encuentre en el estado neurofisiológico en el que se encuentran los individuos de nuestro mundo cuando sienten dolor, pero que no sienta dolor en lo absoluto. Sin embargo, la posibilidad del escenario zombi es contraria a la verdad del fisicalismo. En efecto, si utilizáramos la noción de superveniencia para definir al fisicalismo podríamos decir ahora que el fisicalismo es verdadero si todos los hechos acerca del mundo supervienen lógicamente a los hechos físicos, o de manera alternativa, si todos los hechos físicos implican a todos los hechos positivos acerca de nuestro mundo. Ahora bien, si esto es así, es decir, si el fisicalismo es verdadero,

los hechos acerca de la experiencia consciente (en tanto parte del conjunto de los hechos verdaderos acerca de nuestro mundo) deberían estar implicados lógicamente por los hechos físicos, de manera tal que no debería ser posible pensar sin contradicción que los últimos son verdaderos y los primeros no (de la misma manera en que se supone que nos llevaría a una contradicción intentar imaginar un mundo que tuviera las mismas propiedades físicas que el nuestro pero que no instanciara al mismo tiempo las mismas propiedades biológicas de nuestro mundo).

Sin embargo, un escenario de estas características parece, en principio, perfectamente concebible, es decir, parece posible concebir un mundo zombi idéntico al nuestro en lo que respecta a las propiedades físicas pero que carezca de experiencia consciente, sin incurrir en ningún tipo de contradicción. Por lo tanto, el fisicalismo es falso. Volveré más adelante sobre este argumento. En lo inmediato, voy a reconstruir el argumento de la asimetría epistémica que había quedado pendiente.

De acuerdo con él, la única evidencia disponible para postular la existencia de la consciencia proviene de mi familiaridad de primera persona con mis propios estados conscientes, dado que aun si conociéramos todos los hechos que hay para conocer acerca de las propiedades físicas fundamentales, las funciones realizadas por ese sistema, y cómo éstas y los mecanismos neurales dan lugar a la conducta observable, nada de esto nos permitiría determinar si estas criaturas son conscientes. La presencia de todos los hechos acerca de su composición física es compatible con la suposición de que ninguno de ellas es consciente. Pero el punto del argumento no reside en demostrar que nuestra capacidad para detectar estados conscientes de otras personas tenga algún tipo de limitación natural. La idea es que no importa cuán detallada sea la información física disponible y cuán perfectas sean nuestras capacidades para aprehenderlas, nunca resultará posible establecer o deducir con certeza, basándose únicamente en esa información, la presencia y los rasgos de los estados conscientes de las demás personas.

De acuerdo con Chalmers, de esto se seguiría que la consciencia no superviene lógicamente a lo físico, dado que a partir del

conocimiento completo de los hechos físicos de bajo nivel de una criatura podemos derivar o establecer con seguridad la presencia o ausencia de todos los hechos positivos acerca de ella sin más (por ejemplo, acerca de su funcionamiento biológico, su conducta y los mecanismos cerebrales que la causan; e incluso del resto de sus estados mentales no conscientes). En estos casos, el rol del conocimiento en primera persona no es crucial. En resumen:

La asimetría epistémica en el conocimiento de la conciencia hace evidente que la conciencia no puede ser lógicamente superviniente [a lo físico]. Si lo fuese, no existiría esta asimetría epistémica; una propiedad lógicamente superviniente puede detectarse fácilmente sobre la base de la evidencia externa (Chalmers, 1999: 142).

Ahora bien, independientemente de la plausibilidad del argumento de la asimetría epistémica, lo que me interesa desatacar es que dicho argumento *supone que es posible concebir sin contradicción un escenario zombi*, puesto que sostiene que la consideración y conocimiento de todos los hechos físicos disponibles y posibles acerca de las personas y del resto del universo, descritos hasta sus más ínfimos detalles, no obligan a aceptar o a rechazar, ni hace posible deducir, que esas personas sean seres conscientes o que haya conciencia en absoluto. En otras palabras, de acuerdo con el argumento de la asimetría epistémica, no hay contradicción en aceptar la verdad de toda la información física y, al mismo tiempo, negar que haya experiencia consciente. Y esta es precisamente la definición de un “escenario zombi”.

Concluida esta breve descripción de ciertos aspectos centrales del modelo cartesiano de la mente, en la sección siguiente voy a explicitar las consecuencias perniciosas que enfrenta. En particular, voy a centrarme en las dificultades que se desprenden de aceptar (i) el carácter epistemológicamente privado del acceso a nuestros propios estados conscientes y, sobre todo, (ii) asumir la concebibilidad del escenario zombi.

3.1. Algunos dolores de cabeza para la mente cartesiana

Antes de encarar directamente estas cuestiones, voy a sostener que la caracterización que el filósofo australiano hace del co-

nocimiento por familiaridad que presuntamente tenemos acerca de nuestros estados conscientes enfrenta una suerte de dilema. En efecto, si se lo analiza detenidamente, podría concluirse que o bien no puede considerarse como un caso de conocimiento en sentido propio, o bien no es un conocimiento infalible como se pretende. En segundo lugar, tal como anticipé, voy a defender la idea de que la concebibilidad del escenario zombi que el argumento de la asimetría epistémica supone no es indiscutible a pesar de lo que puede parecer a primera vista. Ahora bien, si el argumento implica o supone algún tipo de contradicción, podemos comenzar a sospechar que algo anda mal con él.

Para desarrollar el primer argumento voy a basarme en una serie de consideraciones que realiza Dorit Bar-On (2004: 27-54), apoyándose, a su vez, en el célebre argumento de Wittgenstein del lenguaje privado, justamente con el propósito de desestimar una explicación de raigambre cartesiana para dar cuenta de la asimetría entre las atribuciones mentalistas en primera y en tercera persona.

De acuerdo con Bar-On, con respecto a las expresiones que refieren a hechos o estados del mundo, existe lo que la autora llama ‘distancia semántica’ entre la aplicación de la expresión en cuestión y el blanco putativo o referente. Una expresión de estas características tiene una extensión: es decir, un rango de objetos o cosas a las que se supone que aplica. Así, algunas cosas caerán bajo la extensión del concepto y otras no. Y debe ser siempre en principio posible que la expresión no sea adecuadamente aplicada. Esto es, debe ser posible usar la expresión para referir equivocadamente a algo que cae fuera de su extensión. Esto supone que siempre que un hablante usa una expresión para referir a algo, debe existir una distinción conceptual entre que el sujeto *piense* que aplicó el término correctamente y el hecho de si *efectivamente* aplicó el término correctamente. Por ejemplo, tomemos el término “silla”, que se usa para hacer referencia a ciertos objetos (las sillas). Decir esto es decir que se supone que el término se aplica de manera correcta únicamente a las sillas. Pero todo lo que esto puede significar es que resultará incorrecto aplicarlo a algo que no sea una silla (como cuando alguien dice, refiriéndose a una

hamaca, “¡Ey, no te sientes en mi silla!”). Y esto, a su vez, implica que es en principio *posible* aplicar el término a algo que no es una silla, incluso si nunca nadie en la historia de la humanidad pasada, presente y futura, cometiera de hecho ese error.

Ahora bien, si esta posibilidad no existiera en principio, es decir, si no existiera siquiera la posibilidad lógica o conceptual de errar el blanco a la hora de hacer referencia a un objeto, se perdería la “distancia semántica” necesaria para que podamos hablar de referencia de un término en absoluto. Para decir que un término o un concepto refiere a un objeto, y poder decir que el concepto o el término se aplican correcta o incorrectamente, los objetos referidos por dichos términos deben tener condiciones de identidad independientes del éxito de cualquier individuo en identificarlos.

Es precisamente esta posibilidad la que queda descartada en el caso del conocimiento por familiaridad del que supuestamente gozamos con respecto a nuestros propios estados conscientes, y entonces un argumento análogo puede aplicarse en este caso. En efecto, como vimos, el “huésped” de una sensación puede detectarla, o “saber” de ella, de manera infalible: de acuerdo con Chalmers, una sensación es el tipo de cosa que, si está presente, su presencia será inmediatamente ‘reconocida’ (en algún sentido) por su huésped. “Ni siquiera hay una posibilidad conceptual de que un sujeto pueda tener una experiencia de rojo [...] sin tener ningún contacto epistémico con ella: tener una experiencia es estar relacionado con ella de ese modo” (Chalmers, 1999: 256). Ahora bien, el planteo de Chalmers parece querer asegurar la infalibilidad del conocimiento acerca de nuestros propios estados conscientes descartando la “posibilidad conceptual” de que estos se instancien y el sujeto no lo “sepa” (de algún modo), es decir, “atando” esencialmente su presencia a ciertos actos de “reconocimiento” o “saber” por parte del sujeto. Sin embargo, la idea misma de reconocimiento de un objeto (sea un objeto físico o una propiedad fenoménica) parece suponer conceptualmente una cierta distancia entre la propiedad (re)conocida y el acto de reconocimiento o conocimiento de la misma por parte del sujeto, es decir, condiciones de identidad independientes para el

objeto reconocido (en este caso, las propiedades fenoménicas de la experiencia) y el acto de reconocerlas por parte del sujeto. De lo contrario, ni siquiera parece tener sentido hablar de “reconocimiento” o de “saber”.

Ahora bien, si para evitar estos inconvenientes se habilita la *posibilidad lógica* de estar equivocado con respecto a la presencia de las propiedades fenoménicas (aun cuando, de hecho, nunca lo estemos), se requiere entonces una explicación que justifique la presunta infalibilidad de dicho conocimiento, ya que se vuelve a abrir una distancia epistémica entre los sujetos y sus estados conscientes, y con esta la posibilidad de error. Como veremos más adelante, el expresivismo parece ofrecer una alternativa que, sin comprometerse con dualismo ontológico alguno, explicaría el hecho de que, en el caso de las autoatribuciones mentalistas en tiempo presente, hacemos referencia a nuestros propios estados mentales conscientes de manera segura porque somos invulnerables a caer en ciertos errores epistémicos, ya sea a la hora de identificar al sujeto que los instancia, como al momento de identificar el tipo de estado mental que se instancia.

Señalada esta primera dificultad, ahora quisiera detenerme brevemente a considerar las consecuencias que se siguen para la concepción cartesiana de la mente a partir de su aceptación de la tesis de que el conocimiento de los propios estados conscientes es “epistémicamente privado”, es decir, de la idea de que sólo podemos tener acceso directo, no inferencial y, en rigor, un verdadero conocimiento, acerca de los estados conscientes *proprios*. Acerca de los estados mentales de las demás personas únicamente tenemos, en el mejor de los casos, un acceso indirecto a través de la observación de la conducta verbal y no verbal que nos permite inferir las “causas ocultas de la conducta”. Pero, como se dijo, todo lo que puede observarse fehacientemente desde la tercera persona es compatible con la ausencia de estados conscientes.

De esta situación se deduce inmediatamente uno de los motivos por los cuales se puede pensar que la consciencia constituye, tal como sostienen muchos científicos y filósofos, un obstáculo para nuestra comprensión científica del universo. En efecto, el hecho de que ciertos aspectos o propiedades de la experiencia

consciente no puedan ser accedidos desde un punto de vista objetivo o de tercera persona, impide su estudio científico, porque el punto de vista de “tercera persona” es precisamente el punto de vista propio de la ciencia.

Siguiendo en esta dirección, podríamos comenzar a indagar de qué manera podría desarrollarse una ciencia de la conciencia desde esta perspectiva. Simplificando un poco las cosas, podríamos decir que, de acuerdo con Chalmers, existe una creencia muy extendida de que la física proporciona un catálogo completo de los elementos y las leyes fundamentales del universo. Sin embargo, a lo largo de distintos trabajos y por medio de diferentes argumentos (como el de los zombis y de la asimetría epistémica), intenta probar (junto con otros autores), que la conciencia no parece poder derivarse de leyes físicas más fundamentales. Por este motivo, considera que en lugar de intentar reducirla infructuosamente a las propiedades físicas, una teoría que pretendiera dar cuenta de “todo” debería incluir un componente extra fundamental (es decir, un componente primitivo no-físico y no derivado de otro más simple): la conciencia. Reconocido este elemento “adicional”, restaría ponerse a trabajar para establecer cuáles son las distintas “leyes puente” que rigen las relaciones entre el mundo físico y el mundo fenoménico y que mostrarán, a juicio de Chalmers, cómo depende el último del primero. En definitiva, una teoría completa tendrá dos componentes fundamentales: las leyes físicas, que nos dirán lo que les pasa a los sistemas físicos, desde los infinitesimales hasta los cosmológicos, y las que podríamos llamar leyes psicofísicas, que nos dirán cómo se relacionan los procesos físicos (o al menos algunos de ellos) con los distintos tipos y grados de experiencia consciente (o bien cómo los primeros dan lugar a las últimas). Ahora bien, admitiendo que existan tales leyes psicofísicas, ¿cómo podríamos descubrirlas? Como habíamos visto, la conciencia es privada o subjetiva; no hay, por lo tanto, manera de observarla en otros. Por lo tanto, la única salida consistiría en advertir las regularidades entre la experiencia y los estados físicos o funcionales en nuestro propio caso, y luego extenderlas a las demás personas. El problema es que, como puede verse, la formulación de estas leyes estará basa-

da, en última instancia, en una inferencia inductiva cuya base es extremadamente pobre, ya que para realizar las generalizaciones o descubrir las regularidades se basa en un único caso: la propia experiencia consciente. Sin embargo, la credibilidad de una inferencia inductiva basada en un único caso puede resultar por lo menos dudosa. Por otro lado, en caso de que se suscitara una controversia entre dos investigadores acerca de algunas de las regularidades, no habrá manera en principio de zanjar la cuestión, porque no será posible tener acceso a los datos relevantes, es decir, a los estados mentales de las otras personas.

3.2. Problemas del más allá para la mente cartesiana

En la sección precedente se dijo que alguien que, como Chalmers, aceptara el argumento de la asimetría epistémica (y el tipo de conocimiento que cada uno de nosotros tendría de sus propios estados mentales conscientes y que está a la base del argumento en cuestión), se comprometería también con la aceptación de la coherencia del escenario zombi. Sin embargo, como voy a mostrar enseguida, esta coherencia puede ser puesta en entredicho. Ahora bien, si esto es así, el argumento de la asimetría epistémica y sus supuestos también comienzan a tambalear.

Como habíamos visto, de acuerdo con Chalmers, el único motivo legítimo por el cual creemos que hay en el mundo algo tan maravilloso y misterioso como la consciencia es porque tenemos un acceso en primera persona (“privado”) a nuestros propios estados conscientes. En efecto:

Nuestras razones para creer en la consciencia se derivan exclusivamente de nuestra propia experiencia de ella. Aun si conociésemos cada detalle acerca de la física del universo –la configuración, causalidad y la evolución de todos los campos y partículas en el continuo espaciotemporal– esa información no nos llevaría a postular la existencia de la experiencia consciente. [...] Es mi experiencia de primera persona lo que me fuerza a enfrentar el problema. [...]

De todos los hechos de bajo nivel acerca de las configuraciones físicas y la causalidad, podemos en principio derivar toda clase de hechos de alto nivel acerca de los sistemas macroscópicos, su organización y la causalidad entre ellos. Podríamos determinar

todos los hechos acerca del funcionamiento biológico, la conducta humana y los mecanismos cerebrales que la causan. Sin embargo, nada en esta vasta historia causal llevaría a quien no lo haya experimentado directamente a creer que debería haber una conciencia [...] (Chalmers, 1999: 141).

En efecto, dado que, por hipótesis, todo ese tipo de cosas que menciona Chalmers pueden ser explicadas con los recursos científicos tradicionales sin residuos, no habría nada ahí que nos motive a postular la existencia de lo que parece ser un hecho adicional como la conciencia. Es más, según el autor, “el eliminativismo acerca de la experiencia consciente [es decir, la posición que niega que exista algo así como la conciencia en el sentido fenoménico] es una posición irrazonable *sólo* debido a nuestro conocimiento de ella. Si no fuese por este conocimiento directo, la conciencia podría seguir el camino del espíritu vital” (Chalmers, 1999: 142).

De todas estas declaraciones, se seguiría, en mi opinión, que (1) si nuestros gemelos zombis son realmente iguales a nosotros en todos los aspectos físicos, y (2) que la capacidad para evaluar y sopesar la evidencia con la que cuentan es también un aspecto en última instancia “físico” (tal como cabría esperar dado que, en principio, el argumento de los zombis sólo pretende mostrar la irreducibilidad de la conciencia), entonces (3) debería admitirse que estas criaturas desdichadas carecen de la única evidencia relevante para plantearse el “problema de la consciencia” y que, por este motivo, luego de considerar la evidencia relevante, es razonable suponer que en su mundo no habría publicaciones ni reuniones científicas acerca de “el problema difícil de la conciencia” (el gemelo zombi de Chalmers no tendría mucho trabajo en ese mundo alternativo). Pero si esto es así, (4) habría al menos *una* diferencia entre nuestra conducta y la de ellos que es perfectamente detectable y entonces el escenario zombi no resultaría realmente concebible después de todo. En otras palabras, no es tan claro que puedan concebirse dos criaturas que diverjan en sus propiedades fenoménicas pero que no lo hagan también en sus propiedades físicas.

En este punto, se podría intentar responder a este argumento utilizando una distinción que traza Chalmers entre la *justificación*

de los juicios fenoménicos (es decir, los juicios acerca de la experiencia consciente o propiedades fenoménicas) y los *mecanismos* (causales) por los cuales se forman estos juicios. De acuerdo con esta distinción, si bien la formación de los juicios podría en principio explicarse cabalmente en términos compatibles con el fisicalismo, la justificación de los mismos depende exclusivamente de la presencia de la experiencia relevante. Así, “por ejemplo, el propio hecho de tener una experiencia de rojo en este momento proporciona justificación a la creencia de que estoy teniendo una experiencia de rojo” (Chalmers, 1999: 255). En contraposición, si bien un zombi podría tener la misma creencia acerca de su experiencia y realizar los mismos juicios que su gemelo consciente, dado que, por definición, carece de experiencias, sus juicios y creencias fenoménicas no estarán justificados y serán falsos. Si aplicamos ahora esta distinción al argumento anterior, podríamos decir que si bien los zombis podrían plantearse “el problema de la consciencia”, sus creencias acerca de la existencia del fenómeno que generaría el problema (la consciencia) resultarían falsas y estarían absolutamente injustificadas.

Esta réplica, sin embargo, trae otras dificultades al partidario de los zombis. En efecto, dado que se supone que los juicios, creencias y pensamientos son el tipo de cosa que en principio resultan físicamente reducibles, debería ser posible explicar en términos compatibles con el fisicalismo por qué *creemos*, o podemos ser llevados a *creer*, que la consciencia no superviene lógicamente a lo físico sin tener que reconocer su existencia (o, mejor dicho, sin reconocer la existencia de algo que no superviene a lo físico).

De acuerdo con Chalmers, esta explicación sería análoga a aquella que busca explicar la creencia en la existencia de Dios o de los extraterrestres, sin postular la existencia de una divinidad o de seres de otros planetas, en términos de fenómenos sociales y psicológicos. Sin embargo, continúa Chalmers, la analogía falla, porque la consciencia, a diferencia de Dios, es ella misma un *explanandum*, y no algo que postulamos para explicar otra cosa, como por ejemplo nuestros juicios y creencias acerca de ella.

Es posible que la existencia de Dios se haya formulado como hipótesis para explicar principalmente todo tipo de hechos eviden-

tes acerca del mundo, tales como su orden y su aparente diseño. Cuando resulta que una hipótesis alternativa puede explicar igualmente bien la evidencia, entonces no hay necesidad de la hipótesis de Dios. No existe ningún fenómeno independiente de Dios al que podamos señalar y decir: eso necesita explicación. [...].

De modo similar, suele postularse que la existencia de los OVNIS explica sucesos extraños en el firmamento, marcas en el suelo, desapariciones en el Triángulo de las Bermudas, las aseveraciones de los ‘sobrevivientes’ de los OVNIS, etc. Si resulta que esta evidencia puede explicarse sin postular la existencia de los OVNIS, entonces nuestras razones para creer en ellos desaparecen (Chalmers, 1999: 245).

En contraposición, adoptar un enfoque que sostenga que lo único que hay que explicar son nuestros juicios y creencias acerca de la conciencia es, para Chalmers, una posición eliminativista que desconoce que la “conciencia no es una construcción explicativa postulada para ayudar a explicar la conducta o los sucesos en el mundo. Más bien, es un *explanandum* primitivo, un fenómeno por derecho propio que necesita explicación” (Chalmers, 1999: 245). De hecho, para Chalmers *todo* lo que probaría la posibilidad de explicar en términos compatibles con el fisicalismo los juicios y creencias acerca de la experiencia es precisamente que los juicios y creencias pueden explicarse de esta manera. Pero lo que realmente querríamos explicar no es esto, sino la propia conciencia.

Con todo, no creo que esta potencial respuesta a mi argumento en contra de la concebibilidad del escenario zombi funcione. En particular, considero que hay en Chalmers una cierta ambigüedad en el uso de la idea de “explicar los juicios fenoménicos en términos compatibles con el fisicalismo”. Por un lado, puede interpretarse, como hace Chalmers, como una posición eliminativista que sostiene que lo único que hay que explicar son los juicios y creencias fenoménicos y nada más, porque de hecho no hay nada más que explicar. Sin embargo, la idea puede también interpretarse como sosteniendo que es posible explicar los juicios y creencias fenoménicas en términos compatibles con el fisicalismo sin necesidad de negar que, además, exista un fenómeno adicional (el *explanandum*) al que estos juicios refieren. En

todo caso, lo que se podría estar negando es cierta *caracterización* del fenómeno en cuestión (en particular, una caracterización que presente a la consciencia como un hecho que no superviene lógicamente a lo físico) mostrando, por ejemplo, qué rasgos de nuestra experiencia podrían conducirnos a conclusiones dualistas como las de Chalmers.

En efecto, tal como señalan Carruthers y Schier “hay un número de personas que dan explicaciones estructurales/funcionales de la consciencia fenoménica. Al hacer eso no niegan que la consciencia fenoménica exista, en lugar de eso niegan la caracterización de Chalmers de ella [...], i.e. como siendo necesariamente inexplicable en términos de estructura y función de estados mentales” (Carruthers y Schier, 2014: 5). Pretender que aceptemos sin reparos que la consciencia es físicamente irreducible con el pretexto de evitar ser acusados de eliminativistas constituiría, en el contexto de esta discusión, una petición de principio.

4. Una alternativa expresivista para la asimetría entre la primera persona y la tercera y en contra del carácter “epistémicamente privado” de la experiencia

Como vimos, para la idea cartesiana de la mente que delineamos en este trabajo, existen ciertos hechos acerca de los cuales tengo un acceso epistémico privilegiado: mis estados conscientes presentes. De acuerdo con esta perspectiva, se supone que conozco estos estados de manera directa, inmediata, no inferencial (e, incluso, infalible). Los conozco de un modo en que nadie más puede conocerlos y de una manera en que no puedo conocer otra cosa. Como contrapartida, se asume que la atribución de estados mentales conscientes en tercera persona está basada en la observación de la conducta de las personas que, en el mejor de los casos, nos permite inferir o conjeturar el estado “escondido” detrás de los actos.

La filósofa Dorit Bar-On (2004) se aparta de esta presuposición (compartida también por posiciones materialistas) según la cual tenemos un acceso epistémico privilegiado a nuestros es-

tados mentales presentes, y parte en cambio de la premisa menos controvertida de la peculiar seguridad de la que gozan las atribuciones psicológicas en primera persona del singular en tiempo presente o, en inglés, *avowals*. Esta seguridad especial se refleja en el hecho de que si alguien afirma de manera espontánea y sincera “Siento un dolor muy fuerte en este momento”, no consideramos que sea adecuado preguntarle qué fundamentos o evidencias tiene para hacer esa afirmación, ni cuestionamos normalmente su verdad. En este sentido, Bar-On sostiene que sobre estas afirmaciones existe una presunción de verdad. En este escenario, parte de la tarea que asume la autora es explicar esta seguridad (que daría lugar a la asimetría mencionada arriba), sin recurrir a ningún acceso epistémico privilegiado ni favorecer algún tipo de dualismo ontológico.

De acuerdo con esta perspectiva, la seguridad de la que gozan este tipo de auto-atribuciones está justificada en parte por la inmunidad o invulnerabilidad a cometer ciertos errores que tenemos al realizarlos. En efecto, si bien desde el punto de vista semántico informan acerca de estados de cosas contingentes, específicamente, identifican a un individuo –el que los emite– y le adscriben cierto estado mental, cuando nos atribuimos estados mentales de esta manera no podemos equivocarnos ni en la identificación del sujeto al que se le atribuyen (quien realiza la atribución en este caso), ni en la identificación del tipo de estado que se adscribe, ni en el carácter del mismo. Sin embargo, esta imposibilidad de cometer ciertos errores no supone un funcionamiento seguro o exitoso de algún proceso cognitivo sumamente fiable o infalible. Por el contrario, la idea de Bar-On es que la imposibilidad de cometer estos errores de identificación se basa en que no llevamos adelante ninguna tarea de identificación a la hora de realizar las adscripciones mentalistas en primera persona en tiempo presente.

En particular, no se trata de que, al realizar estas auto-atribuciones, primero me percató de que alguien, por ejemplo, siente un dolor, y después, en un segundo momento, reconozco que ese alguien soy yo (como sucedería a alguien que dijera: “Hay alguien que siente dolor ¿pero soy yo el que lo siente?”). Tampoco sucede

que primero identifico que siento algo y luego reconozco que eso que siento es un dolor de cabeza (como si dijera, “siento algo, ¿pero es un dolor de cabeza o de panza?”).

De todas maneras, todo esto no implica que para Bar-On la verdad de estas atribuciones mentalistas esté necesariamente garantizada o que crea que son infalibles. Para la autora, existe la posibilidad de que sean falsas. Lo que sí podríamos afirmar desde esta perspectiva es que si hay algún tipo de error, ese error no puede deberse a un error por identificación o algún tipo de error de corte epistémico.

Al mismo tiempo, esta doble invulnerabilidad al error se explica, según Bar-On, por el carácter *expresivo* de estas atribuciones en tanto actos. En efecto, de acuerdo con la propuesta de la autora, la asimetría entre los dos tipos de atribuciones psicológicas que venimos considerando descansaría en el hecho de que las auto-atribuciones en primera persona en tiempo presente constituyen la forma lingüística que tenemos de *expresar* nuestros estados mentales, de “dar voz” a nuestra mente (*speaking our mind*).⁴

Para desarrollar su posición, Bar-On hace una distinción entre tres sentidos distintos de “expresar”:

EXP1 en el sentido de la acción: una persona expresa un estado propio haciendo algo intencionalmente. Por ejemplo cuando intencionalmente doy un abrazo o digo “¡Qué bueno verte!”, intencionalmente hago algo y doy expresión a mi alegría de verte.

EXP2 el sentido causal: una emisión o conducta expresa un estado subyacente por ser la culminación de un proceso causal que empieza con dicho estado. Por ejemplo, una mueca involuntaria o las manos temblorosas pueden expresar, en el sentido causal, el dolor o el nerviosismo, respectivamente.

EXP3 el sentido semántico: e.g. una afirmación expresa una proposición abstracta, un pensamiento o un juicio por ser una representación (convencional) de la misma. Por ejemplo, la frase

4 Existen otras versiones (a mi juicio menos elaboradas) de esta perspectiva expresivista en la literatura filosófica contemporánea en los trabajos de Hacker, 2005, y Finkelstein, 2010. Por lo demás, la perspectiva expresivista tiene su origen en cierta lectura de los aportes de Wittgenstein acerca de los *avowals* y de la relación de la conducta con los estados mentales.

“Está lloviendo afuera”, expresa en el sentido semántico la proposición de que está lloviendo afuera en el momento de su emisión en la vecindad del emisor (Bar-On, 2004: 248).

De esta manera, Bar-On sostiene que las atribuciones mentalistas en primera persona en tiempo presente se asemejan, en tanto *acciones*, a otras conductas expresivas que incluyen expresiones naturales no lingüísticas (como los abrazos); pero en tanto *productos*, son oraciones de un lenguaje que expresan³ proposiciones auto-adscriptivas. Así, por ejemplo, supongamos que me siento contento de ver a un amigo después de mucho tiempo sin hacerlo y digo (o pienso) “¡Estoy tan contento de verte!”. En este caso, mi auto-adscripción expresa³ una proposición que es verdadera si efectivamente es cierto que estoy contento de ver a mi amigo. Así, en la medida en que mi atribución proviene, según Bar-On, “directamente” desde mi estado mental de estar contento de ver a mi amigo, podemos decir que esta auto-atribución expresa² el sentimiento de alegría, como podría hacerlo una expresión facial (una sonrisa amplia) o un gesto. Pero la auto-atribución mentalista, así como la afirmación “¡Qué bueno verte!” (que no es auto-adscriptiva), y como un abrazo que puedo darle a mi amigo, no son un simple acto reflejo de mi cuerpo. Es algo que *yo hago*; un acto intencional de mi parte. De hecho, el abrazo, la afirmación no auto-adscriptiva y la auto-adscriptiva son todos actos intencionales genuinos que sirven para expresar mi alegría pero no meramente en un sentido causal.

Ahora bien, el producto de estos actos sí es diferente. Sólo las auto-atribuciones mentalistas como las que venimos considerando, y sólo ellas, tiene propiedades semánticas que expresan³ una *proposición* acerca de mi estado (emocional en este caso). Dicho de otra manera, en tanto *actos*, los auto-atribuciones mentalistas expresan¹ los estados mentales del individuo que los emite. Y en tanto actos *lingüísticos*, expresan³ en qué estado mental se encuentra el individuo que los realiza. De esta manera, el expresivismo rechaza la idea cartesiana de que (al menos algunas de) las auto-atribuciones psicológicas sean una suerte de reporte de mis estados mentales basado precisamente en la

“observación” de los objetos y sucesos internos que pasan por nuestra mente.

Además, y tal como enfatiza Pérez (2014), el planteo de Bar-On la lleva a “cuestionar una de las consecuencias tradicionales más ampliamente aceptadas de las explicaciones epistémicas clásicas de la asimetría entre la primera y la tercera persona: la idea de que sólo tenemos acceso a nuestros propios estados psicológicos, y que a los estados psicológicos ajenos sólo accedemos por inferencia” (Pérez, 2014: 13). En efecto, para Bar-On,

[m]ientras que el comportamiento expresivo puede considerarse, en cierto sentido igualmente abierto a la observación tanto por parte del observador como del sujeto, para el sujeto el comportamiento servirá para dar voz a, o dar aire a, su actual estado mental, mientras que para el observador es la manifestación del estado psicológico de otra persona. El sujeto está en el estado mental y se está expresando a través de su comportamiento, el observador percibe ese estado. No hay que asumir que sólo aquel que está en un estado lo puede ver con un ‘ojo interno’, mientras que los otros que observan su comportamiento a lo sumo pueden conjeturar al respecto. Todo lo contrario [...]. De todos modos, si seguimos al sentido común, deberíamos considerar a la conducta natural expresiva como permitiéndonos ver la furia, el temor, la vergüenza, etc., de los sujetos. [...] [Pese a] la perspectiva tradicional introspeccionista (y sus herederos contemporáneos), no percibimos meramente la conducta e inferimos la presencia y el carácter de los estados internos del sujeto como la mejor explicación de la conducta que percibimos; en lugar de eso, percibimos sus estados mentales presentes. [...]

En este sentido, la conducta expresiva es transparente-a-la-condición-mental del sujeto (Bar-On, 2004: 278-279).

De acuerdo con este esquema, los estados mentales son estados expresables de los diferentes individuos que muy frecuentemente se hacen perceptibles para sus congéneres mediante diferentes tipos de conducta expresiva. Esto significa que no se piensa a los estados mentales como condiciones o estados esencialmente privados y ocultos dentro de la mente de los sujetos, que son los únicos que en principio podrían captarlos, y cuya presencia y carácter serían conjeturados por los observadores a

partir de la observación de los efectos conductuales típicos. Más específicamente, desde la perspectiva expresivista, las diferentes manifestaciones conductuales no son meros signos o síntomas que sirven como indicadores más o menos fiables de sus causas (i.e. los estados mentales de los sujetos). Por el contrario: para Bar-On la conducta expresiva es suficiente para *mostrar* o, como dijimos, para “hacer perceptibles” los estados mentales de los sujetos que la realizan. De hecho

si [...] pensamos a las expresiones conductuales como meros efectos de condiciones internas, seguiría siendo desconcertante por qué deberíamos tratar a [las auto-adscripciones psicológicas] *avowals* (entendidos como expresiones) como proveyendo más que pistas contingentemente seguras para realizar inferencias causales por parte de los observadores (Bar-On, 2004: 418).

Ahora bien, dicho esto, debemos agregar que, como anticipamos, Bar-On no supone que las auto-atribuciones en cuestión sean infalibles o inevitablemente verdaderas. Lo único que parecía asegurarnos el carácter expresivo era la invulnerabilidad a cometer cierto tipo de errores epistémicos. Sin embargo, esto no nos evade de la posibilidad de cometer errores de otra clase. De hecho, la autora considera que hay tres tipos de situaciones en las cuales las atribuciones psicológicas en primera persona en tiempo presente resultarían erróneas o desacertadas y por lo tanto falsas.

A. Casos de expresión silenciosa o ausente, como cuando el sujeto se encuentra en cierta condición mental expresable pero sin expresarla. En este caso ninguno de sus potenciales observadores podrá determinar su condición.

B. Casos que involucran la simulación, como las actuaciones que ejecuta un actor en un escenario o alguien que nos miente. En estos casos, Bar-On considera que el sujeto expresa un estado o condición mental pero no la suya propia. Aquí parecerá que el individuo está en cierto estado cuando en realidad no lo está.

C. Por último, están los casos de expresión “fuera de lugar”, en los que la conducta expresiva no concuerda con el estado en que se encuentra el sujeto, como el caso de la persona que va al

dentista y, precavida de todo lo que son capaces estos profesionales, comienza a gritar antes de que el médico utilice el torno. En este caso el sujeto estaría expresando una condición (de dolor, por ejemplo) que no es la que de hecho atraviesa.

Ahora bien, una vez que se reconoce que la conducta expresiva no es una condición *necesaria* para considerar que efectivamente estamos en cierto estado mental (como mostraría el caso de la simulación o de la expresión ausente), ni tampoco *suficiente* (como probaría el caso del mentiroso y de aquel que expresa “mal” el estado en que se encuentra), ¿no se abre la posibilidad para plantear ciertos escenarios escépticos en los cuales las conductas expresivas resultan sistemáticamente inadecuadas con respecto a los estados que se supone que deberían expresar y en donde es imposible determinar por observación de la conducta en qué condición se encuentran nuestros semejantes? ¿Podrían los propios agentes determinar en su propio caso la no concordancia de sus actos con sus estados si no poseen algún tipo de conocimiento “interno” acerca de estos últimos que les permita comparar?

Por otra parte, creo que un adversario como Chalmers podría llegar a admitir de buena gana la explicación de la seguridad de la que gozan ciertas auto-adcripciones psicológicas de acuerdo con el esquema que ofrece Bar-On. Sin embargo, podría insistir también con la idea de que, además de esta asimetría entre las *atribuciones* en primera y en tercera persona que Bar-On explica sin verse obligada a postular la presencia de algún tipo de acceso epistémico privilegiado, existe un conocimiento por familiaridad acerca de nuestra propia experiencia como el que describimos arriba, que no es conceptual o que es pre-conceptual y que, por este motivo, no es abarcado por la explicación expresivista.

Frente a esta posible objeción, creo que alguien como Bar-On podría ensayar una respuesta “terapéutica”. En particular, podría afirmar que el hecho de que Chalmers piense que gozamos de un conocimiento de esas características tiene su origen precisamente en una interpretación inadecuada de un hecho menos controvertido: la especial seguridad que poseen las atribuciones mentalistas en primera persona en tiempo presente. En otras pa-

labras, lo que sería cierto –podría sostener un expresivista– pero que aparece completamente distorsionado en la imagen cartesiana del auto-conocimiento y de la auto-percatación es que mi palabra acerca de algunos de mis estados mentales tiene un estatus “privilegiado”.

Sin embargo, no es necesario aceptar que esto se debe a que tengo un acceso *sui generis* a un espectáculo privado que me permite luego describir o reportar lo que veo en él mientras que los demás tienen el acceso vedado a las funciones que allí se presentan. En lugar de eso, lo que afirmamos acerca de algunos de nuestros estados mentales es especialmente seguro porque constituye una expresión o manifestación de nuestra experiencia, mientras que lo que otros dicen no lo es.⁵ De esta manera, es posible mantener la intuición de que hay algún tipo de diferencia importante entre las afirmaciones que hacemos acerca de nuestros propios estados mentales y las que hacemos acerca de los estados mentales de los demás (y del resto de los hechos del mundo en general), que justifica la confianza de la que gozan, pero que no precisa comprometerse, ni con las dificultades que entraña el dualismo, ni con la existencia de un conocimiento de carácter privado y misterioso inaccesible a la ciencia.

5. Conclusiones

En este trabajo me ocupé de mostrar dos maneras diferentes de abordar la asimetría entre las atribuciones de estados conscientes en primera y en tercera persona. De acuerdo con una de ellas, la perspectiva cartesiana, la asimetría en cuestión se explica por la existencia de una asimetría epistémica entre el tipo de conocimiento que tenemos acerca de nuestros propios estados

5 En este punto de la esgrima dialéctica habitual, Chalmers podría insistir con la idea de que lo que queremos explicar no son los juicios, atribuciones o creencias acerca de la conciencia y sus características, sino el fenómeno mismo, a saber: la propia experiencia consciente. Más arriba indiqué de qué manera podemos responder a este tipo de declaraciones sin ser acusados de eliminativistas ni de dejar de tomarnos en serio la conciencia.

mentales conscientes y acerca de los estados mentales de las demás personas. Sin embargo, como se mostró, el carácter privado del conocimiento de los propios estados mentales conscientes impide el abordaje científico de estos fenómenos. Además, junto con este punto, el compromiso de la mente cartesiana con la dudosa coherencia de la noción de zombi la convierten definitivamente en una opción poco atractiva.

Como una alternativa, en la sección 4 reconstruí algunos aspectos de una interpretación expresivista de determinadas autoatribuciones psicológicas, tal como la desarrolla Dorit Bar-On, que promete poder dar cuenta de las asimetrías mencionadas sin que se susciten los problemas que enfrentan los partidarios de la mente cartesiana. En efecto, de acuerdo con esta posición, la seguridad especial que identifica a las mencionadas atribuciones se podría explicar por el carácter expresivo de los actos que los producen en lugar de hacerlo postulando algún tipo de acceso epistémico privilegiado o infalible al espectáculo “privado” que se proyecta en nuestra mente.

Al mismo tiempo, el enfoque expresivista se destaca como una opción que, aunque su presencia no resulta nada frecuente en las discusiones en torno a la consciencia, merece ser explorada en la medida en que representa una ruptura con la idea de toda una larga tradición en la filosofía y la psicología, de acuerdo con la cual tenemos un acceso epistémico peculiar a nuestros propios estados mentales, mientras que los estados psicológicos de las otras personas permanecen ocultos y conjeturados detrás de su conducta pública. De acuerdo con el expresivismo, en cambio, la idea es que así como aquellos que están adecuadamente situados y poseen el conocimiento necesario *ven* el significado de las palabras en una página, si tenemos las sensibilidades adecuadas y el conocimiento relevante, *vemos* los fenómenos psicológicos en el comportamiento de las personas.

Bibliografía

- Bar-On, D. (2004), *Speaking my Mind. Expression and Self-Knowledge*, Oxford: Clarendon Press.
- . (2010), “Avowals: Expression, Security, and Knowledge: Reply to Matthew Boyle, David Rosenthal and Maura Tumulty”, en: *Acta Anal*, 25, pp. 47-63.
- Carruthers., G. y E. Schier (2014), “Why are We Still Being Hornswoggled? Dissolving the Hard Problem of Consciousness”, *Topoi. An International Review of Philosophy*, pp. 1-13.
- Chalmers, D. (1999), *La mente consciente. En busca de una teoría fundamental*, Barcelona: Gedisa.
- . (2010), *The Character of Consciousness*, Oxford University Press.
- Crick, F. (1995), *The Astonishing Hypothesis*, Londres: Touchstone.
- Dennett, D. (1995), *La conciencia explicada. Una teoría interdisciplinar*, Barcelona: Paidós.
- Ezcurdia, M. y O. Hansberg (2003), *La naturaleza de la experiencia. Volumen I: Sensaciones*, México: UNAM.
- Finkelstein, D. (2010), *La expresión y lo interno*, KRK ediciones.
- Hacker, P. M. S. (2005), “Of knowledge and of knowing that someone is in pain”, en: Pichler, A. y S. Saatela (eds.), *Wittgenstein: The Philosopher and His Works*, The Wittgenstein Archives at the University of Bergen, pp. 203-235.
- Jackson, F. (1982/2003), “Qualia epifenómicos”, en: Ezcurdia, M. y O. Hansberg (2003), *La naturaleza de la experiencia. Volumen I: Sensaciones*, México: UNAM.
- Pérez, D. (2007), “¿La conciencia?, ¿Qué es eso?”, en: *Estudios de Psicología: Studies in Psychology*, 28:2, pp. 127-140.
- . (2014), “Asimetría y conceptos psicológicos”, en: *Páginas de Filosofía*, año XV, Nº 18, Departamento de Filosofía, Universidad Nacional del Comahue, pp. 5-26.
- Searle, J. (1984), *Mind, Brain and Science- the 1984 Reith Lectures*, Londres: BBC.

La atribución psicológica: entre la primera y la tercera persona

Diego Lawler

1. Introducción

Cada uno de nosotros espera decir de sí mismo cosas como las siguientes: qué está haciendo ahora y por qué, si está o no de acuerdo con un comentario realizado por otra persona, si desea ir al próximo recital de rock de su grupo favorito, si está enamorado de su pareja, si la temperatura del día es agradable, etcétera. Todas estas expresiones forman parte de un amplio trasfondo de cosas que cada uno de nosotros puede expresar inmediatamente, sin realizar ninguna de las averiguaciones en las que debería embarcarse si quisiera saber qué está haciendo ahora su amigo Javier o si Julieta está enamorada o si Andrés irá a la fiesta. Resultaría muy extraño que alguien nos preguntara: “¿Tenés calor?”, “¿Creés que existe el paraíso extraterrenal?” o “¿Tenés ganas de tomar un helado?”, y que nosotros respondiéramos: “¿Cómo podría saber esas cosas?” o “En un par de días, cuando nos veamos de nuevo, te respondo”. Si nos expresáramos así, concitaríamos miradas suspicaces; nuestras compañeras ocasionales de conversación intuirían que algo no anda bien.

Me interesa destacar que estos episodios de la vida diaria indican que el conocimiento de nuestras propias actitudes proposicionales –por ejemplo, deseos, intenciones y creencias– difiere del conocimiento que tenemos de esas mismas actitudes en otras personas.¹ Sé inmediatamente si estimo que el desempleo au-

¹ Los predicados mentales, esto es, las palabras con las cuales describimos las mentes de las personas, se dividen en dos grandes clases. Una de ellas es la clase de las actitudes proposicionales. Éstas son las actitudes atribuibles

mentó, mientras que no sé inmediatamente si Francisco estima lo mismo; por el contrario, tengo que preguntarle a Francisco o deducirlo de sus opiniones sobre la situación socioeconómica. Hay una asimetría de hecho entre estos casos. En situaciones normales, disfruto de un acceso inmediato y no inferencial a mis creencias, deseos e intenciones; sin embargo, cuando atribuyo actitudes a los otros, me embarco en un esfuerzo interpretativo, agrupo evidencia, realizo inferencias a partir de sus palabras, compromisos y toda clase de conductas expresivas. No obstante, esto no descarta que en algunas ocasiones no ocurra lo mismo conmigo mismo, esto es, a veces debo recurrir a este último trabajo interpretativo para, por ejemplo, saber lo que quiero o llegar a saber qué es lo que opino sobre cierto asunto, es decir, puedo adoptar sin más sobre mí la perspectiva interpretativa que adopto frente a otros; pero en la mayoría de los contextos cotidianos respondo qué es lo que opino, qué persigo con mis acciones o qué deseos tengo sin iniciar una investigación sobre mi universo de opiniones, deseos, conductas, etcétera.

En este capítulo quiero narrar dos historias. Una de ellas se enmarca en una tradición cuyas raíces se remontan a Kant. El punto de partida de esta narración es que los seres humanos se autodeterminan cognitiva y prácticamente como criaturas racionales. Esta historia enlaza el autoconocimiento con nuestra capacidad de juzgar.² Según esta narración, nuestra capacidad para

con una cláusula relativa ‘que’, por ejemplo, desear, creer, conocer, querer, intentar, pensar, afirmar, decir, etc. Por consiguiente, las actitudes proposicionales son aquellas actitudes indicadas de forma característica por medio del uso de, para decirlo con palabras de Russell, “Esas palabras como ‘creer’, ‘desear’, ‘dudar’, las cuales, cuando se presentan en una oración, deben seguirse de una oración subordinada que diga qué es lo que se cree o se desea o se duda” (1940: 65). La segunda clase de predicados mentales es la clase que incluye sensaciones, *qualia* (propiedades fenoménicas de la experiencia), sentimientos, ganas y humores, etc. Se trata de una clase que reúne todos esos atributos mentales que no pueden ser adscriptos con una construcción que suponga una cláusula relativa.

- 2 En su forma de concebir la mentalidad, Kant enlazó la capacidad de juzgar con el autoconocimiento. Este enlace reunió la conjunción de dos líneas de razonamiento. Contra sus predecesores empiristas, Kant consideró que había una diferencia radical entre, por ejemplo, tener una impresión de un

conocer de manera inmediata nuestra mente es parte de nuestra capacidad para determinarla cognitiva y prácticamente. El autoconocimiento está enlazado con la agencia racional. Elucidar el autoconocimiento conlleva desarrollar un guion donde el fenómeno de la autodeterminación cognitiva y práctica desempeña el papel del personaje principal.

Quienes cuentan esta historia otorgan una importancia filosófica especial a la presencia de esa asimetría de hecho, antes mencionada, entre la primera persona y la tercera persona. Consideran que no sólo refleja un modo de conocimiento fundamentalmente diferente (una asimetría epistémica), sino que, además, muestra que hay una manera distintiva en que los estados mentales se presentan a los sujetos que los tienen y que su descripción es necesaria para entender la mente. El razonamiento que subyace a este enfoque tiene esta forma: para entender la mente se debe entender la subjetividad, que se expresa primariamente en el modo especial en que conocemos o devenimos conscientes de ciertos estados mentales, a saber, en virtud de conocerlos o ser conscientes de ellos por disfrutar de una posición especial, precisamente la posición de estar en esos estados mentales o de tenerlos. Denominaré a este punto de vista “la perspectiva de primera persona”.

triángulo azul y el pensamiento de que el triángulo es azul. La impresión es un evento causado por el mundo. El pensamiento es una actitud de creencia que el sujeto asume sobre el mundo. Cuando percibo un triángulo azul, estoy en un cierto estado perceptivo. Al menos que la percepción vaya acompañada de un juicio, no se plantea la cuestión de su verdad o falsedad. Hay solamente un hecho: tengo una percepción de un triángulo azul. Por el contrario, cuando juzgo que un triángulo es azul, mi actitud está abierta a evaluación, mi creencia puede ser verdadera o falsa, racional o irracional. Para Kant, una teoría de la experiencia cognitiva es una teoría del juicio; una sugerencia que resulta del modo en que Kant desplaza de la percepción el análisis de la cognición hacia la estructural proposicional del pensamiento (véase Rorty, 1989: cap. III). La segunda línea de razonamiento afirma que el sujeto debe estar en una cierta relación consigo mismo como prerequisite de su capacidad de juzgar. El conocimiento comienza con el autoconocimiento. Este es el principio de apercepción kantiano: “El *Yo pienso* tiene que poder acompañar todas mis representaciones” (Kant, 1978, B132: 153). Para una discusión de esta genealogía, véase Kitcher (2011: cap. 15).

En la segunda narración no se cuenta ninguna historia sobre esta dependencia entre el autoconocimiento y nuestra capacidad para autodeterminarnos. Esta narración no le otorga ninguna importancia filosófica a la asimetría presente en la vida ordinaria cuando informamos sobre nuestra mente y la de los otros. Para quienes cuentan esta historia, nuestros modos para saber cosas sobre nuestros deseos, creencias, intenciones, etc., son los mismos que empleamos para saber estas mismas cosas de otras personas. No hay una diferencia de clase, sino meramente de grado entre el autoconocimiento y el conocimiento de los otros. Denominaré a este punto de vista “la perspectiva reductiva o de la tercera persona”.

A continuación describiré cada uno de estos puntos de vista, comenzando con el “punto de vista reductivo o de la tercera persona” y siguiendo con el “punto de vista de la primera persona”. Me apoyaré en las elaboraciones del filósofo Daniel Dennett (1991) para desarrollar la primera sección. Dennett, siguiendo las sugerencias de Ryle (2005), construye una imagen que ha sido muy influyente a favor de este punto de vista. Recurriré a los trabajos de Shoemaker (1996), Moran (2001, 2003, 2012) y Boyle (2009, 2011a, 2011b) para describir el enfoque de la primera persona. Finalmente, contrastaré estos dos puntos de vista con el propósito de mostrar la especificidad de la perspectiva de la primera persona.

2. La perspectiva de la tercera persona

En primer lugar quisiera introducir con cierta libertad el contexto filosófico al que pertenece lo que he denominado “el punto de vista de la tercera persona”. En términos generales, la posición de Dennett en filosofía de la mente se encuadra dentro de una corriente que podría denominarse “conductismo filosófico”.³

³ Es muy habitual que el término inglés *behaviorism* se traduzca sin más al español por “conductismo”. Sin embargo, aquí emplearé la etiqueta de “conductismo filosófico”, entre otras cosas para evitar la asociación del conductismo con sus versiones mecanicistas.

Esta es una etiqueta que requiere ser precisada en un número importante de aspectos. Comenzaré esta sección presentando escuetamente las afirmaciones básicas del conductismo filosófico. Luego ubicaré la posición de Dennett dentro de ese marco.

El conductismo filosófico puede encuadrarse dentro de dos formulaciones. La primera enfatiza lo que el conductismo filosófico afirma; la segunda subraya lo que particularmente niega. A continuación examinaré rápidamente el significado de cada una de ellas.

Veamos la formulación afirmativa. El behaviorismo está comprometido con la siguiente tesis: la interpretación interpersonal es en un sentido o en otro previa a la introspección intrapersonal. Esta formulación rechaza un estilo de pensamiento sobre la mente muy familiar y extendida, un estilo que impregna nuestro modo cotidiano de aproximarnos a ella. Se trata de un estilo que concibe la mente bajo la imagen de un teatro interno, que Dennett (1995) denominó con precisión “el teatro cartesiano”.⁴ Esta imagen contiene la siguiente descripción. En el escenario o pantalla de ese teatro ocurren (o mejor, aparecen) las representaciones del mundo. El público de esa puesta en escena se compone de una sola persona, el propietario de esas representaciones. Esa única persona (i.e. el público en su totalidad) tiene un conocimiento inmediato de sus pensamientos. Por consiguiente, el conocimiento que cualquier otra persona ajena al público, llamémosla “intérprete”, pueda tener de lo que allí sucede será conocimiento mediato, esto es, conocimiento inferido de la observación de la conducta de la única asistente a la función y del conocimiento inmediato que ella misma posea de sus propios contenidos mentales. El conductismo filosófico se rebela contra esta imagen, rechazándola porque representa un modo de acceso a la vida mental que se reduce en primera y última instancia a mirar dentro de uno mismo (introspección), independientemente de la interpretación que lleven a cabo los otros. En resumen, en esta formulación la tesis conductista filosófica concluye des-

4 Rorty (1989), en cambio, se refirió a esta misma imagen con la ya famosa expresión “el espejo de la naturaleza”.

cartando una imagen de la mente de acuerdo con la cual el sujeto posee un acceso privado y original (en el sentido de primero) a sus propios pensamientos, un acceso que es de alguna manera previo a cualquier relación interpersonal.

La formulación negativa expresa lo que el conductismo filosófico naturalmente niega: dos tesis que están presupuestas en la imagen de la mente como un teatro privado: “la tesis representacionalista” y “la tesis del acceso especial”.⁵ Brevemente, la

⁵ Es habitual encontrar desarrollos que superponen descuidadamente estas dos tesis; en particular, la tesis representacionalista se confunde a menudo con la tesis del acceso especial. Sin embargo, son lógicamente independientes. Por ejemplo, Wittgenstein (2003: 5.541-5.5423) sugiere en el *Tractatus* que las actitudes proposicionales deben ser entendidas como oraciones complejas al modo de estructuras en el sujeto. La creencia de que p es un estado de cosas en el creyente que figura (*pictures*) p. Esta teoría es representacionalista: las actitudes proposicionales están constituidas por estados de cosas internos. No obstante, no hay en el *Tractatus* ninguna sugerencia de que el sujeto deba tener un acceso especial a sus propias representaciones. Los ítems internos son estados de cosas como cualesquiera otros y, como tales, pueden ser representados por otros sujetos (el subjetivismo entra desde otro ángulo y posteriormente en el *Tractatus*, a saber, en su discusión del solipsismo y lo místico). Por consiguiente, la tesis representacionalista no implica a la tesis del acceso especial. Tampoco la tesis del acceso especial implica la tesis representacionalista. El argumento de Thomas Nagel (1981) ilustra lúcidamente esta cuestión puesto que muestra que hay un acceso especial desde la primera persona a los aspectos subjetivos de la experiencia: lo que efectivamente es el experimentar, que no puede ser capturado desde el punto de vista de la tercera persona. Nagel apela a aquello en lo que debe consistir ser un murciélago; a nuestra intuición de que percibir vía ecolocación debe de ser radicalmente diferente de aquello en lo que efectivamente consiste meramente oír o ver. La perspectiva subjetiva es, de acuerdo con Nagel, *sui generis*, no descansa en otras perspectivas; por el contrario, flota independientemente de ellas. Sin embargo, el argumento a favor de esta tesis no recurre a una premisa específicamente representacionalista. Nagel no afirma en ningún lugar que lo que él denomina “el punto de vista subjetivo” esté correlacionado con ítems internos o eventos. La tesis del acceso especial no implica la tesis representacionalista. Sin embargo, se podría argüir que se requiere un ámbito interno de representaciones para hablar con sentido del acceso especial. ¿Dónde podría estar localizado el punto de vista subjetivo si no es en el interior? Todo lo que está situado fuera es motivo de inspección pública. Pero este no es un argumento muy convincente. Descansa en una premisa que dice que un hecho disponible debe estar en algún lugar: si no está dentro debe estar fuera. Esta premisa podría impugnarse. Por ejemplo,

tesis representacionalista afirma que hay ítems internos, eventos o estados, que explican y justifican que se tengan actitudes proposicionales o que son constitutivos del hecho de tener actitudes proposicionales. La tesis del acceso especial asevera que el sujeto tiene una clase de acceso a hechos sobre sí mismo que (1) sólo está disponible desde el punto de vista de la primera persona y que (2) tiene lugar de manera *sui generis*, esto es, su constitución no descansa en ningún otro punto de vista.⁶

La importancia de distinguir entre estas dos tesis radica en que si bien ambas están presupuestas por la imagen de la mente como un teatro interno, el conductismo filosófico sólo se compromete a rechazar tajantemente una de ellas, a saber, la tesis del acceso especial. Por el contrario, no niega *in toto* la tesis representacionalista. Sólo rechaza una versión de ella. Para decirlo con precisión: el conductismo filosófico es indiferente a la tesis representacionalista siempre y cuando las representaciones no sean concebidas como representaciones con significado independiente de (o previo a) las relaciones interpretativas interpersonales. En caso contrario, el conductismo filosófico se compromete con su rechazo categórico.

Enriquecida la tesis del conductismo filosófico se compromete con el rechazo de la tesis de que hay una clase de acceso que el sujeto tiene a los hechos sobre sí mismo que (1) está sólo disponible desde la perspectiva de la primera persona y (2) está dada de manera *sui generis*, esto es, para su constitución no depende de ninguna otra perspectiva.⁷

podríamos preguntar, con espíritu ryleano, ¿dónde están localizadas las habilidades o los rasgos de carácter de las personas? Si las habilidades y los rasgos de carácter no necesitan una localización, ¿por qué habría de necesitarla una perspectiva subjetiva?

6 La agudeza observacional y claridad expresiva de John Locke retratan esta forma de pensar: “Estas precisas y nudas apariencias en la mente, el entendimiento las erige [...], sin considerar cómo, de dónde y con qué otras ideas fueron recibidas en la mente, como patrones para ordenar en clases las existencias reales, según se acomoden a esos patrones, y para denominarlas de acuerdo con ellos (1956: Libro II, capítulo XI, § 9).

7 En este punto se pueden realizar otras dos distinciones en relación con la tesis del acceso especial. En primer lugar, la tesis del acceso especial

En resumen, la formulación positiva del conductismo filosófico afirma que la interpretación interpersonal es anterior a la introspección intrapersonal. La formulación negativa asevera que no hay acceso especial *sui generis*.

Presentado de manera general el marco del conductismo filosófico, pasaré a describir la posición de Dennett. No explicitaré, sin embargo, uno a uno los compromisos filosóficos de la perspectiva dennettiana. Me ceñiré a su estrategia relacionada con la adscripción y autoadscripción de actitudes proposicionales. Mi propósito será mostrar que el predominio en su estrategia filosófica de un interés meramente descriptivo-predictivo constituye un obstáculo para la comprensión de lo que está genuinamente en juego en la perspectiva de la primera persona.

debe distinguirse de la tesis que podríamos denominar 'tesis del acceso privilegiado'. Esta última afirmaría algo como lo siguiente: hay una clase de acceso que el sujeto tiene a los hechos acerca de sí mismo y que sólo está disponible desde la perspectiva de la primera persona. En esta formulación, esta tesis es la tesis del acceso especial sin la segunda cláusula, sin la afirmación de que el acceso es *sui generis*. El behaviorismo en general no tendría en principio nada en contra de la tesis del acceso privilegiado *per se*. El acceso privilegiado es compatible con el behaviorismo mientras que se respete la prioridad de la dimensión interpersonal, esto es, mientras no se conciba que el acceso especial es una habilidad que una persona puede tener independientemente de su participación en el juego de la interpretación interpersonal. Obviamente, quien se comprometa con la tesis del acceso privilegiado deberá explicar convenientemente cómo la participación en prácticas interpretativas interpersonales da lugar al desarrollo de una habilidad puramente intrapersonal. Y aunque esta explicación no parece estar fácilmente disponible, no sería *a priori* imposible. En consecuencia, se debe reconocer la existencia de esta posibilidad. En segundo lugar, la tesis del acceso especial tiene que distinguirse de la tesis de la privacidad. Esta última tesis afirmaría algo así como lo siguiente: solamente el sujeto que tiene determinadas representaciones posee una relación de familiaridad inmediata con esas representaciones y, por tanto, sólo él puede tener algún conocimiento sobre qué son sus representaciones. La tesis del acceso especial no implica la tesis de la privacidad. De hecho, la idea de que hay una facultad especial, desde el punto de vista de la primera persona, para conocer el contenido de nuestras propias representaciones, no excluye la idea de que pueda haber formas igualmente fiables de acceder a las representaciones internas desde, por ejemplo, la perspectiva de la tercera persona.

Dentro del conductismo filosófico, Dennett suscribe un punto de vista particular. Este puede reunirse en una tesis: la tesis interpretativista. La tesis interpretativista afirma que tener ciertas actitudes proposicionales consiste en ser adecuadamente interpretado o interpretable como teniendo esas actitudes proposicionales. En líneas generales, si uno acepta esta tesis, la consecuencia inmediata es la siguiente: tener actitudes proposicionales depende en gran parte de la práctica interpersonal (léase interpretativa) de atribuir actitudes proposicionales.

No obstante, la perspectiva interpretativista admite diferentes versiones. Hay que descender a un nivel de particularización para recoger la versión dennettiana. En su versión de la tesis interpretativista, los humanos atribuimos actitudes proposicionales a las otras personas para describir y predecir sus conductas. Para decirlo con un ejemplo, si yo sé que Valentina cree que lloverá y que desea permanecer seca, y si supongo que ella es una ser racional que tiene un buen número de creencias verdaderas sobre el tiempo y los medios para bregar con él, entonces puedo predecir que antes de salir de su casa tomará un paraguas. Mi atribución de creencias y deseos a Valentina tiene éxito cuando puedo utilizarla para describir y predecir exitosamente su conducta. De hecho, el significado de la atribución, su contenido empírico, está dado por el complejo descriptivo-explicativo de sus disposiciones conductuales. Desde este punto de vista, atribuir a otras personas actitudes proposicionales es una estrategia predictiva, un medio para prever lo que efectivamente harán.⁸ Para decirlo de otro modo, desde el punto de vista de Dennett, la fuerza pragmática de nuestras atribuciones de actitudes proposicionales –el uso que hacemos de ellas– es predictivo-descriptivo: atribuimos actitudes proposicionales para describir la conducta pasada y presente, y predecir la conducta futura.⁹

8 Véase Dennett (1991: 26-27).

9 Dennett está elaborando su posición en la estela de Ryle (1984). Ahora bien, mientras que Ryle y Dennett comparten la actitud de poner el acento en el uso descriptivo-predictivo que tiene nuestra forma de hablar de la mente, Ryle se propone meramente describir la geografía existente de nuestros conceptos mentales; en cambio, Dennett pretende racionalizarla y mejorarla.

La teoría resultante es una versión fuerte de la tesis interpretativista que se conforma al marco del conductismo filosófico. Tener una creencia, una intención, etc., es presentar esas disposiciones conductuales que permiten que uno sea interpretable como teniendo esa creencia, intención, etcétera.¹⁰ Las actitudes proposicionales en general son entidades de cálculo al servicio de una teoría que se propone predecir la conducta de los otros desde la plataforma de la actitud intencional. Así como los centros de gravedad ‘existen’ en el contexto de una teoría sobre el comportamiento mecánico de los objetos físicos, las actitudes proposicionales ‘existen’ en el contexto de las prácticas interpretativas.

Si la práctica de la interpretación interpersonal es la condición de posibilidad de las actitudes proposicionales en general, se sigue que la interpretación interpersonal es previa a las actitudes proposicionales, incluido el autoconocimiento (la autoatribución de actitudes proposicionales) o la introspección personal. La siguiente cita de Dennett ilustra acabadamente que el conocimiento de uno mismo no puede ser previo o independiente de las prácticas interpretativas interpersonales. La autoadscripción de actitudes proposicionales tiene el mismo significado que las adscripciones realizadas desde la tercera persona.

Postulamos todas esas actividades y procesos mentales aparentes para encontrarle sentido a la conducta que observamos, para, en realidad, encontrarle el mayor sentido posible a la conducta, especialmente cuando la conducta que observamos es la nuestra propia. Los filósofos de la mente solían desviarse de su camino para insistir en que el acceso que uno tiene a su propio caso en esos temas es muy diferente al acceso que uno tiene a los de los

Por otro lado, Ryle no presta prácticamente atención a la estructura holística de las actitudes proposicionales, cuestión que se sitúa en el foco de la mirada dennettiana.

- 10 Dennett señala que una creencia “puede ser discernida solamente desde el punto de alguien que adopta cierta *estrategia predictiva*, y cuya existencia puede ser confirmada sólo por una evaluación del éxito de esa estrategia” (1991: 26). Esto es, “Lo que es ser un verdadero creyente es ser un *sistema intencional*, un sistema cuyo comportamiento se puede predecir en forma confiable y amplia por medio de la estrategia intencional” (1991: 27).

otros, pero a medida que aprendemos más acerca de las distintas formas de psicopatología y hasta de las debilidades de personas aparentemente normales se vuelve más plausible suponer que todavía hay algunos rincones de privilegio indisputable, algunos temas en los cuales nuestra autoridad es invencible, cada uno de nosotros es, en muchos aspectos una especie de autopsicólogo inveterado, que inventa sin esfuerzo interpretaciones intencionales de nuestras propias acciones en una mezcla inseparable de confabulación, autojustificación retrospectiva y (a veces, sin duda) buena teorización (1991: 89).¹¹

En algunas circunstancias puede tener sentido adscribir a una persona actitudes proposicionales de segundo orden sobre sí misma para describir o predecir su conducta. Por ejemplo, adscribir a una persona la habilidad de tener creencias sobre sus propias actitudes proposicionales es una forma útil de describir y predecir su conducta lingüística. Podría correlacionar la creencia de Valentina de que ella intenta tomar un paraguas con una disposición a proferir “intento tomar un paraguas”. Esto implica tratar al autoconocimiento como una clase de actitud proposicional entre otras.

Desde el punto de vista de Dennett no podría ser de otra manera. Dado que su enfoque niega la tesis del acceso privilegiado, socava al mismo tiempo la base para cualquier idea de que hay algo especial en el autoconocimiento. Cuando me pregunto si Valentina tomará o no un paraguas, lo que necesito saber es si ella cree que está lloviendo, si ella tiene la intención de no mojarse, si ella cree que tomar un paraguas es un medio efectivo para no mojarse, etc. En ningún sentido parece ayudarme o realizar una contribución suponer que si tiene actitudes proposicionales, Valentina debe ser capaz de pensar que las tiene. Estoy en condiciones de predecir si tomará o no un paraguas sin asumir esta suposición. En definitiva, según el enfoque dennettiano cualquier elaboración de una idea que suponga que el sujeto tiene una relación especial con su vida mental sería empíricamente falsa o superflua. Para darnos cuenta de que sería empíricamente falsa

11 Otra vez, Dennett está siguiendo la huella de Ryle (1984).

basta con recordar que alcanzamos un éxito descriptivo y predictivo aceptable atribuyendo actitudes proposicionales a cosas que son incapaces de tener creencias de segundo orden –por ejemplo, bebés, computadoras, mascotas, etc. Para cotejar que podría ser superflua basta con señalar que el propósito de las atribuciones psicológicas se alcanza sin presuponer el autoconocimiento.

La posición de Dennett sobre las atribuciones de la primera persona no tiene nada de particular. Las atribuciones desde la tercera persona son correctas o incorrectas sobre la base de la conducta observable. Lo mismo vale para el caso de la primera persona. Las atribuciones desde la tercera persona son utilizadas para describir y predecir la conducta. Lo mismo vale para el caso de la primera persona. En definitiva, dado que el autoconocimiento es una actitud proposicional más, en tanto que tal presupone las prácticas interpretativas. De este modo se confirma el rechazo de la tesis del acceso especial, tesis que anteponeía el acceso al yo desde la primera persona en desmedro del acceso interpretativo desde la tercera persona. En cualquier caso, si quisiéramos una conclusión más moderada sería esta: las diferencias entre la aprehensión desde la primera persona o desde la tercera de lo que uno está, por ejemplo, haciendo o pensando, por ejemplo, es meramente una diferencia de grado. Por consiguiente, no habría una brecha sustantiva entre el autoconocimiento y el conocimiento de los estados mentales de otros, sino que, además, en ambos casos emplearíamos el mismo método para determinar nuestros estados mentales o los estados mentales de los otros.¹²

3. La perspectiva de la primera persona

Quienes son partidarios de contar una historia en la que se le otorga un lugar especial a la perspectiva de primera persona

12 Véase Ryle (1984: cap. IV), quien señala que siempre está en juego la simple pregunta metodológica de cómo establecemos y atribuimos ciertas actitudes proposicionales para capturar el sentido de las conductas públicas y silenciosas de las personas; luego si se trata de nosotros o de los otros es una cuestión contingente y sin mayor importancia.

suponen, entre otras cosas, que hay una relación entre el autoconocimiento de ciertos estados mentales (estados con contenido proposicional) y nuestra capacidad agencial, manifestada, por ejemplo, la racionalidad de nuestros juicios. Richard Moran (2001) lo expresó de esta forma:

[l]a idea básica que cobra expresión en la idea de “autoridad de la primera persona” puede no ser que la persona misma siempre “conoce mejor” lo que piensa sobre un asunto dado, sino, por el contrario, que es efectivamente su asunto qué es lo que piensa sobre eso, que su pensamiento depende de ella. Al declarar su creencia no se expresa como una testigo experta respecto de un ámbito de hechos psicológicos, sino que expresa su autoridad racional sobre ese ámbito (124). [La traducción es mía]

La afirmación de Moran condensa una intuición que puede ser formulada así:

“Un sujeto *S* está en condiciones de conocer de manera inmediata o transparente sus propias actitudes proposicionales (por ejemplo *C* creencias, *D* deseos o *I* intenciones) porque esas mismas actitudes (*C*, *D* e *I*) son expresiones de su agencia racional”.

La clave de esta intuición está en entender por qué las autoadscripciones son inmediatas o transparentes.¹³ El entendimiento de este fenómeno contribuiría a volver inteligible nuestra capacidad ordinaria de autoconocernos. El fenómeno de la transparencia, que explota una sugerencia de Evans (1982) sobre la epistemología de las creencias de segundo orden, señala que las respuestas a preguntas sobre nuestros estados mentales, formuladas desde la primera persona, están correlacionadas sistemáticamente con nuestras respuestas a preguntas sobre el mundo. Así, nuestro punto de vista, desde la primera persona, sobre nuestras vidas mentales es transparente a nuestro punto de vista sobre el mundo. Para saber qué pensamos, deseamos o intenta-

13 El hecho de que nuestras autoadscripciones sean inmediatas no está relacionado con el acierto epistémico sobre el contenido de esas autoadscripciones. Son inmediatas porque no son inferidas o no descansan en algo más básico.

mos, por ejemplo, debemos atender correctamente a ciertos aspectos del mundo y no a nuestros contenidos psicológicos internos.¹⁴ A continuación describo, siguiendo a Lawler y Vega (2016: 37), el fenómeno de la transparencia para el caso de la creencia, pero podría extenderse para otras actitudes proposicionales:

(a) ¿Creo que mi profesor no me dijo toda la verdad? (esta es una pregunta sobre uno de mis estados mentales, a saber, qué es lo que creo).

(b) ¿Me dijo mi profesor toda la verdad? (esta pregunta se refiere a qué es lo que ocurrió en el mundo, esto es, si mi profesor dijo o no toda la verdad).

(c) ¿He de creer o no que mi profesor no me dijo toda la verdad? (esta pregunta versa sobre qué es lo que debo creer, y en tanto tal tiene un aire normativo cuando se la formula en un contexto deliberativo).

La transparencia hace que la pregunta (a) se responda resolviendo la pregunta (b); cuando esto ocurre, entonces (a) se convierte en (c).¹⁵ Adviértase que para que la transparencia se satisfaga, el sujeto adopta una actitud deliberativa respecto del hecho de si su pareja le dijo o no toda la verdad y de su relación con la determina-

14 Este es el párrafo de Evans (1982: 225): “Si alguien me preguntara: ‘¿Crees que tendrá lugar la tercera guerra mundial?’, para responderle debo prestar atención a los mismos fenómenos externos a los que prestaría atención si estuviese respondiendo la pregunta ‘¿Habrá una tercera guerra mundial?’. Me encuentro en posición de responder la pregunta sobre qué es lo que creo poniendo en operación el procedimiento que sea que tengo para responder la pregunta sobre qué es lo que ocurrirá”.

15 Por supuesto, no todas las cosas que una persona llega a creer, las cree de esta manera. Agustina, por ejemplo, podría llegar a tener creencias sobre su madre, y el amor que esta le prodigó, a través de la terapia que realiza, donde ella adopta un punto de vista empírico, de tercera persona, respecto de su propia vida mental, como si fuera cualquier otra persona. Sin embargo, si no pudiese acceder a esas creencias reflexionando sobre los hechos relacionados con su madre y su vínculo con ella, entonces esas creencias no serían *prima facie* completamente racionales.

ción de ese hecho.¹⁶ Teniendo en cuenta este fenómeno, dos asuntos devienen inmediatamente evidentes. Por una parte, es inteligible para un sujeto que la pregunta sobre qué creer es transparente respecto de la pregunta sobre qué es o no, en el mundo, el caso, puesto que al reflexionar sobre las razones que tiene para aceptar lo que es el caso en el mundo, está simultáneamente determinando a sabiendas qué es lo que cree respecto de ese asunto. Por otra parte, un sujeto solamente puede encontrarse en esa situación si adopta la instancia deliberativa, dado que en ella el sujeto asume que su creencia depende de él, en el sentido de que su reflexión sobre los hechos determina qué es lo que habrá de creer.

La estructura que subyace a la transparencia, siguiendo el caso de la creencia, tiene esta forma:

- (1) Un sujeto *S* sabe qué es lo que cree sobre un asunto (digamos *p*), si sabe transparentemente o inmediatamente que cree que *p* es el caso.
- (2) Un sujeto *S* conoce transparentemente su creencia *p* reflexionando sobre los hechos (razones) que tiene para juzgar que *p* es el caso.
- (3) Un sujeto *S* juzga que *p* es el caso si tiene y se compromete con hechos (razones) para juzgar que *p* es el caso. Juzgar que *p* es el caso supone de parte de *S* un ejercicio de su agencia racional, donde *S*, al juzgar que *p*, brinda su asentimiento interno y se compromete con el mundo siendo de una manera y no de otra. Este compromiso trae aparejado una responsabilidad: si alguien le preguntara a *S* por qué cree que *p*, *S* respondería mencionando sus razones para creer que *p* es el caso.
- (4) Juzgar que *p* es creer *p*. La deliberación y su resolución, el juicio, es la determinación de la creencia misma. El acto mental de la resolución, el juicio, constituye el estado mental: creer que *p*.

16 Véase Moran (2001: 62-63): “el vehículo de la transparencia reside en el requerimiento de que aborde la pregunta por mi estado mental con un espíritu deliberativo, decidiendo y declarándome sobre el asunto en cuestión”.

(5) Un sujeto *S* llega a saber que cree que *p* porque conocer (o devenir consciente de) su creencia supone el ejercicio de las mismas habilidades empleadas para determinarse a creer que *p*; o dicho de otra manera, al juzgar que *p* es el caso, *S* suscribe hechos (razones) a favor de *p*; resolverse racionalmente a creer que *p* es lo que vuelve posible que *S* sepa que cree que *p*. *S* conoce su creencia puesto que esta es el resultado de su involucrarse activamente en el mundo.

(6) Juzgar que *p* para un sujeto *S* no es distinto de saber que cree que *p*. Un sujeto que juzga que *p* está en posición de hablar a favor de que *p* es el caso (saber que cree *p*) desde su convencimiento a favor de sostener a *p* como verdadero.¹⁷

Hay algo peculiar en esta transición, que vuelve a la transparencia filosóficamente relevante para elucidar el punto de vista de la primera persona. No hay ningún juicio ni experiencia que aparentemente desempeñe un papel mediador, que cumpla un papel de intermediario que habilite al sujeto a pasar de su resolver deliberativamente si *p* es o no el caso a sostener activamente y a sabiendas que *p* (Lawler y Vega, 2016: 37-38). La clave reside en advertir que no hay dos condiciones psicológicas diferentes, esto es, creer que *p* y creer que uno cree que *p*; por consiguiente, no habría que explicar ninguna transición entre dos condiciones psicológicas diferentes, donde la segunda se vuelve conocimiento de la primera –esto es, la transición vinculada a cómo se relaciona una condición psicológica en la que el sujeto está con una creencia sobre esa condición psicológica. Este giro práctico para entender la perspectiva de la primera persona no ofrece una explicación estrictamente epistémica del autoconocimiento.

17 Adviértase que cuando le preguntamos a alguien “¿Por qué deseás, creés o intentás *p*?” no estamos presuponiendo meramente que ese alguien conoce una cierta evidencia y su relación con *p*; nuestro preguntar adquiere su sentido del hecho de que lo que desea, intenta o cree depende él o ella. Y esto está presupuesto en la direccionalidad de la pregunta, una cierta relación del sujeto con su estado. Al determinar su relación con *p* sabe de lo que lo convence a afirmar *p*.

to; por el contrario, apunta en otra dirección, en la dirección de lo que significa estar en una condición psicológica de creer, desear o intentar.¹⁸ La transparencia sugiere que un sujeto que se encuentra en un estado mental de creer, intentar o desear, se halla en una condición psicológica tal que tácitamente conoce que está en esa condición. Desde el punto de vista de la primera persona, creer, desear o intentar es, en los casos significativos de satisfacción de la condición de transparencia, creer, desear o intentar a sabiendas. Al mismo tiempo, son los mismos poderes cognitivos bajo aspectos diferentes los que están en juego cuando creo, intento o deseo p y cuando sé que creo, intento o deseo p .

Sin embargo, este creer, desear o intentar a sabiendas es compatible con creer erróneamente que uno tiene una creencia, deseo o intención determinada –por consiguiente no sería el caso de que el sujeto estuviese anclado en una posición de infalibilidad; tampoco el sujeto estaría en una posición omnisciente respecto de sus actitudes proposicionales, puesto que podría darse el caso de que un sujeto tuviese una creencia o un deseo, donde estos dominen su comportamiento, pero no los suscribiese conscientemente. Así, sucedería con el conocimiento de la creencia, del deseo o de la intención, algo análogo a lo que podría suceder con la creencia, el deseo o la intención de primer orden: S sabe que p pero no es consciente de que sabe que p –su conocimiento, así como su creencia, podría estar reprimido, impedido, etcétera. En estas ocasiones la transparencia no se satisfaría, puesto que el sujeto conocería algo, pero no sería consciente de conocer ese algo. Una cosa es que algo sea conocido y otra cosa es que eso conocido sea accesible a la reflexión consciente (Boyle, 2011b). En condiciones anormales, por tanto, se rompería lo que la transparencia nos enseña: que lo que se conoce o es conocido es accesible a la reflexión consciente.

18 Boyle (2011a) explora la naturaleza de los distintos estados mentales para analizar qué condiciones deben darse para que satisfagan el criterio de que su existencia implica que el sujeto que los tiene cuenta con un conocimiento tácito de ellos. La discusión sobre la naturaleza de los diferentes estados mentales excede este trabajo.

Se puede estar muy impresionado por el fenómeno de la transparencia o poco impresionado por él. Cuando se está impresionado se supone que dicho fenómeno elucida la perspectiva de la primera persona sobre las atribuciones psicológicas y por qué estas tienen rasgos particulares. Sin embargo, quienes están poco impresionados por la transparencia sugieren, entre otras cosas, que este fenómeno promueve una imagen de una persona hiperracional, como si estuviéramos frente a un *homo philosophicus* y no frente a un ser humano común y corriente (Cassam, 2014). Si este fuera el caso, se habría extraviado esta manera de entender las atribuciones psicológicas desde la primera persona, puesto que tendría en mente una persona que consiste en realidad en una abstracción. Los poco impresionados por el fenómeno de la transparencia sugieren que las actitudes proposicionales que se autoatribuyen terminan siempre siendo las que racionalmente deben ser, y esto promueve una imagen irreal de la persona.

No obstante, los defensores de la transparencia, como *via regia* para el análisis de la perspectiva de la primera persona, estiman que aunque sostuviéramos que la condición de transparencia afirma que un sujeto sostiene cierta actitud proposicional como transparente a la cuestión de si debe o no racionalmente sostener esa actitud, no se seguiría todavía que la vigencia de la condición de transparencia dependa de que mis actitudes sean en general las que racionalmente deben ser.

Imaginemos el siguiente caso: mi creencia sobre una proposición p (por ejemplo, María es una buena tenista) no es como racionalmente debe ser, y esto con independencia de lo cuidadoso que haya sido en mi proceso deliberativo al considerar la evidencia –i.e. a pesar de la evidencia presente a favor de que p no es el caso (de que María es una mala tenista), yo sostengo que p es el caso (que María es una buena tenista). Si la condición de transparencia se aplicase, me conduciría a creer que creo que p dado que soy insensible a la evidencia que afectaría mi resolución al deliberar si debo o no racionalmente creer que p . Lo que este hecho significa es que la condición de transparencia es un vehículo para determinar mi conocimiento de la creencia con independencia de si mi creencia no es como racionalmente debería

ser. Por consiguiente, la condición de transparencia no sobreactuaría mi condición de criatura racional.

Sin embargo, el hecho de que la condición de transparencia dice que mi creencia respecto de p se determina a partir de mi consideración deliberativa (reflexión) sobre las razones a favor de p , ¿no estaría suponiendo un lazo íntimo entre la transparencia y la racionalidad? Sí, por supuesto, pero esto solamente señalaría lo siguiente: que hay una conexión entre una pregunta por uno de mis estados psicológicos (si creo o no que p) y una pregunta dirigida al mundo (si p es o no el caso). Esta última pregunta no supone una investigación de mi propio estado psicológico, sino una inspección del mundo, esto es, si es verdadera la proposición que concierne al mundo (me encuentro mirando el mundo, no introspeccionando).

Si p es o no el caso demanda un ejercicio de mi capacidad de juzgar si p es o no verdadera. En algunas ocasiones no necesito deliberar, ya estoy resuelto; por ejemplo, cuando considero en qué ciudad vive mi hermano. En otras ocasiones considero la evidencia, pero la evalúo de una manera en la que me desvío de las normas de la racionalidad, quizás mis inclinaciones me llevan hacia una dirección que se aparta de la correcta evaluación de la evidencia; por ejemplo, cuando mi deseo de que Poxipol, el caballo, gane el derby me lleva a sobreestimar sus posibilidades. Sin embargo, en cualquier caso, el poder que ejercito al responder la pregunta sobre si p es o no el caso, es mi poder de juzgar si p es el caso. No puedo tener este poder sin tener la capacidad de considerar la evidencia a favor o en contra de la verdad de p y afirmar o rechazar una proposición sobre la base de esa consideración. Implicarme en esta actividad es deliberar si aceptar o no una proposición. Aceptar o no p supone considerar mi relación con p como un asunto que me corresponde a mí resolver, no como un hecho con el que me encuentro. Esto es así con independencia de si resuelvo si p es o no el caso, reafirmando un punto de vista que ya tengo o juzgando de una manera que es racionalmente defectuosa.

En definitiva, lo que estas consideraciones muestran es que el fenómeno de la transparencia no promueve por sí mismo

una posición hiperracionalista sobre las autoatribuciones psicológicas. Que las atribuciones psicológicas “dependan de mí” no significa que deba atribuirme las que racionalmente deben ser en función de cómo está hecho el mundo; por el contrario, puedo equivocarme. No obstante, esto no es un obstáculo para dar cuenta de cómo me resuelvo a creer, intentar, desear y a conocer lo que intento, resuelvo o deseo, y cómo esta situación es diferente del modo en que atribuyo a otros creencias, deseos e intenciones.

A modo de conclusión

Quiero llamar la atención sobre la siguiente cita de Shoemaker (1996):

Es esencial para la comprensión filosófica de lo mental que apreciemos que hay una perspectiva de la primera persona sobre lo mental, una manera distintiva en que los estados mentales se presentan a los sujetos que los tienen, y que una parte esencial de la tarea filosófica es ofrecer una explicación de la mente que vuelva inteligible la perspectiva que los sujetos tienen sobre sus propias vidas mentales (157).

Shoemaker no nos está sugiriendo que orientemos nuestras investigaciones a los aspectos fenoménicos de nuestra vida mental consciente; por el contrario, sugiere que la relevancia de la primera persona reside en que expresa nuestra subjetividad. El fenómeno de la transparencia, analizado en la sección anterior, captura esta expresión, vuelve “inteligible la perspectiva que los sujetos tienen sobre sus propias vidas mentales”, una perspectiva activa del sujeto en la determinación de las actitudes proposicionales autoatribuidas. El punto de vista de la primera persona es diferente del punto de vista de la tercera persona, en el que el sujeto conoce sus actitudes como si conociera cualesquiera hechos sobre sí mismo, por ejemplo, que su nariz está torcida al mirarse al espejo, o que su corazón no funciona bien al leer su electrocardiograma junto con el informe del médico.

La atribución psicológica desde la primera persona conlleva un rasgo distintivo: en estas atribuciones está imbuido el entendimiento de parte del sujeto de que adopta, a través de ellas, una relación con el mundo con la que está comprometido. Al atribuirme *transparentemente* la creencia de que la tercera guerra mundial tendrá lugar, me comprometo con mi estado psicológico, donde eso significa que me comprometo con el mundo como siendo de una manera y, al mismo, estoy abierto, si alguien me lo solicitara, a expresar mis razones para exponer este compromiso. Y esto es parte de la comprensión que los sujetos tienen de su propia vida mental, y del concepto de vida mental en general; algo que se expresa únicamente desde el punto de vista de la primera persona.

Sin embargo, esta línea argumentativa no significa que una persona siempre habrá de conocer su mente desde esta posición. Por supuesto, este no es el caso. Cualquiera de nosotros puede obtener información sobre sí mismo desde la perspectiva de tercera persona, por ejemplo, nuestro mejor amigo nos dice que en realidad, en virtud de lo que hacemos y decimos, creemos que vivir en el campo es más edificante que vivir en la ciudad más allá de nuestra opinión sobre este asunto. La línea argumentativa a favor de la perspectiva de la primera persona señala que hay una conexión importante entre el conocimiento inmediato de algunos de nuestros propios estados mentales y el asunto de lo que es pensar en uno mismo desde la posición de la primera persona, i.e. nuestra subjetividad. Por consiguiente, la diferencia que está en juego es la diferencia entre conocer mis estados mentales como hechos de mi vida psicológica, como quien conoce la temperatura del día empleando un termómetro, o verme a mí mismo en mis estados mentales. Entonces, ¿qué está supuesto en estos episodios? ¿Qué supone asir proposiciones que tienen esta forma “Creo que creo que p”; “soy F”, etc.? La intuición que aquí está en juego sugiere que responder estas preguntas supone establecer una relación entre el pensamiento desde la primera persona y el autoconocimiento inmediato, una relación que hace al núcleo de lo que denominamos subjetividad. En definitiva, el fenómeno de la transparencia nos pone sobre la pista de ese nú-

cleo. Así, conocer inmediatamente nuestra mente es una marca de la posición de la primera persona, que disfruta de un involucramiento con el mundo diferente de la posición de la tercera persona. Quienes piensan que la perspectiva de la primera persona se reduce a la perspectiva de tercera persona estiman que no hay tal marca; o lo que es lo mismo, no hay nada especial en la perspectiva de la primera persona que deba ser elucidado como camino para volver inteligible lo que es tener una mente o una vida psicológica.

Las perspectivas de la primera y de la tercera persona son dos formas muy extendidas de considerar las atribuciones psicológicas, dos formas de caracterizar cómo conocemos cosas sobre nosotros mismos. Quienes son partidarios de la perspectiva de la tercera persona consideran que la otra perspectiva no tiene nada de especial; es más, a veces apelan a nuestras ignorancias respecto de nosotros mismos, por ejemplo, respecto de nuestras verdaderas motivaciones o nuestro verdadero carácter en muchas circunstancias vitales, para mostrar no sólo que la primera persona tiene serias limitaciones epistémicas, sino que además puede llegar a ser una empresa vana y superflua. Sin embargo, quienes son partidarios de la perspectiva de la primera persona consideran que la sugerencia anterior choca contra el trasfondo amplio de conocimientos sobre nosotros mismos que nosotros, y aquellos con los que interactuamos, asumimos normalmente. El autoconocimiento ordinario, a saber, qué estoy haciendo, qué deseo u opino, son formas básicas de la experiencia humana, formas que expresan lo que es disfrutar de una condición subjetiva, lo que es tener una vida. Cuando este autoconocimiento ordinario está bloqueado nuestra comprensión de lo que es tener una vida mental se desfigura o se vuelve ininteligible. Sería, por tanto, un error no tomar en serio las asimetrías que nos muestran las interacciones cotidianas respecto de la comprensión de los otros y de nosotros mismos.

Bibliografía

- Boyle, M. (2009), “Active Belief”, en: Hunter, D. (2009), *Belief and Agency*, Calgary: University of Calgary Press.
- . (2011a), “‘Making up Your Mind’ and the Activity of Reason”, *Philosophers’ Imprint*, Vol. 11, N° 17.
- . (2011b), “Transparent Self-Knowledge”, *Proceedings of the Aristotelian Society, Supplementary Volume*, 85:1, 2011, pp. 223-241.
- Cassam, Q. (2014), *Self-Knowledge for Humans*, Oxford: Oxford University Press.
- Dennett, D. (1991), *La actitud intencional*, Barcelona: Gedisa.
- . (1995), *La conciencia explicada*, Barcelona: Paidós.
- Evans, G. (1982), *The Varieties of Reference*, Oxford: Oxford University Press.
- Kant, I. (1978), *Crítica de la Razón Pura*, Madrid: Alfaguara.
- Kitcher, P. (2011), *Kant’s Thinker*, Oxford: Oxford University Press.
- Lawler, D. y J. Vega (2016), “La autoridad de la primera persona y la estrategia de la autoría”, en: Pérez, D. y R. Ibarlucía (comps.) (2016), *Hechos y valores en filosofía teórica, filosofía práctica y filosofía del arte*, Buenos Aires: CIF/SADAF.
- Locke, J. (1956), *Ensayo sobre el entendimiento humano*, México: Fondo de Cultura Económica.
- Moran, R. (2001), *Authority and Estrangement*, Princeton: Princeton University Press.
- . (2003), “Responses to O’Brien and Shoemaker”, en: *European Journal of Philosophy*, 11 (3), pp. 402-419.
- . (2012), “Self-knowledge, ‘Transparency’, and the Forms of Activity”, en: Smithies, D. y D. Stoljar (2012), *Introspection and Consciousness*, Oxford: Oxford University Press.
- Nagel, T. (1981), “¿Qué se siente ser murciélago?”, en: Nagel, T. (1981), *La muerte en cuestión. Ensayos sobre la vida humana*, México: Fondo de Cultura Económica.
- Rorty, R. (1989), *La filosofía y el espejo de la naturaleza*, Madrid: Cátedra.
- Russell, B. A. W. (1940), *An Enquiry into Meaning and Truth*, London: George Allen and Unwin.

Ryle, G. (2005), *El concepto de lo mental*, Barcelona: Paidós.

Shoemaker, S. (1996), *The First-Person Perspective and Other Essays*,
Cambridge: Cambridge University Press.

Wittgenstein, L. (2003), *Tractatus logico-philosophicus*, Madrid: Tecnos.

La atribución psicológica y la normatividad de lo mental

Karina Pedace

Introducción: el encuentro de la normatividad y lo mental

La idea de que la normatividad guarda alguna relación con la naturaleza de lo mental surgió explícitamente hacia la mitad del siglo XX, especialmente con las contribuciones de Ludwig Wittgenstein (1953), Donald Davidson (1980) y Saul Kripke (1982). Y, probablemente, puede retrotraerse hasta la obra de Immanuel Kant.

En efecto, en la medida en que Kant entiende los conceptos como *reglas*, un aspecto central de su legado reside en el tratamiento de la intencionalidad como una actividad estructurada conceptualmente, que se distingue por su carácter *normativo*. De este modo, lo que distingue al juzgar y al hacer intencional no es que involucren algún tipo especial de procesos mentales, sino que son cosas de las que los sujetos cognoscentes y los agentes somos, de un modo característico, responsables.

Desde esta perspectiva, Kant ofreció una caracterización de lo mental según la cual estar en un estado intencional tiene una significación normativa, en virtud de que cuenta como asumir una obligación o compromiso. El contenido del compromiso está determinado por las reglas que son los conceptos en términos de los cuales está articulado el estado. Así, una de las preguntas más urgentes (e inquietantes) de Kant es cómo entender la autoridad de los conceptos.

El intento de comprender la fuente, naturaleza y significación de las normas ínsitas en nuestros conceptos, tanto las que gobiernan su empleo teórico en la investigación y el conocimiento, como las que gobiernan su empleo práctico en la deliberación

y la acción parecen estar, pues, en el centro de la empresa filosófica kantiana.

Hundiendo sus raíces en este legado, hacia la mitad del siglo XX surgió explícitamente la idea de que la normatividad tiene algo que ver con la naturaleza de lo mental, especialmente, con las contribuciones de Wittgenstein (1953), Davidson (1980) y Kripke (1982).

Me interesa subrayar que, pese a sus contundentes (e imprescindibles) antecedentes modernos y decimonónicos (en especial a partir de la querella metodológica que se suscitó con la irrupción de las “Ciencias del Espíritu”) es durante el siglo XX que tuvo lugar un giro en la concepción de lo mental, que denominaré “giro interpretativo”, a partir del cual la normatividad asoma explícitamente como un rasgo de la atribución psicológica. En efecto, una vez que lo mental deja de concebirse como un espectáculo privado que se vislumbra desde un ojo interior y se lo inscribe en la esfera pública a la que ha de acceder el intérprete, la normatividad entra decisivamente en escena como una de las restricciones fundamentales del juego de la atribución psicológica.

El giro de la mirada hacia la interpretación, hacia la comprensión mutua, nos invitará entonces a que pensemos la normatividad más allá de la esfera deóntica (propia de las reglas), hacia el horizonte evaluativo (i.e. inherente a valores) y desde un diálogo necesario de la subjetividad con la intersubjetividad.

En efecto, a partir de su raíz común en el legado kantiano, las derivas contemporáneas de la *normatividad* de lo mental suponen su encuentro con las nociones de intencionalidad y racionalidad. De este modo, desde la normatividad propia de la comprensión, i.e. de la atribución mutua que nos hacemos de estados mentales intencionales a los efectos de explicar, predecir y comprender nuestra conducta, se concibe la racionalidad como un rasgo que nos hace “especiales”. Se ha sostenido así que la racionalidad guarda una íntima conexión con la agencia y se ha consolidado una singular concepción relacional de la intencionalidad y el lenguaje, según la cual ser quienes somos es algo que viene en compañía de otros.

Entre los filósofos contemporáneos que han abordado la normatividad de lo mental desde su encuentro con la racionalidad y la intencionalidad me interesa subrayar que desde los aportes canónicos de Donald Davidson (1980), hay varios que han “recogido el guante” y mantienen la discusión sumamente vigente. Tales son los casos paradigmáticos, *inter alia*, de Daniel Dennett (1987) y Robert Brandom (1994), en cuyos interesantes aportes no me voy a detener aquí.

Simplemente me interesa exaltar que, por ejemplo, para Brandom somos los únicos seres que estamos sujetos a la peculiar fuerza de la “mejor razón”. Esta fuerza es, a su juicio, un tipo peculiar de fuerza normativa: un “deber” racional. Ser racional es estar constreñido por esta norma, esto es, estar sujeto a la autoridad de las razones. Decir “nosotros” es, en este sentido, ubicarnos a nosotros mismos y a los demás en el espacio de las razones, dando y pidiendo razones para nuestras actitudes y actuaciones. Según Brandom (1994) hacernos explícito a nosotros mismos quiénes somos requiere una explicación teórica de qué es en la práctica tratar a otro como a uno de nosotros. Esto abre, en mi opinión, una remisión expresa hacia la empresa de Davidson, en el sentido de que su teoría de la interpretación radical podría leerse como su explicitación de lo que hacemos al atribuir intencionalidad a otro, a saber: tratarlo como persona.

En efecto, desde la perspectiva fundacional de Davidson lo que nos hace singulares es ser racionales. Ser racional es, a su juicio, tener actitudes proposicionales; esto supone participar del juego de la interpretación gobernado por principios constitutivos entre los cuales la racionalidad resulta esencial en tanto abre la ruta a la normatividad de lo mental.

Dado, pues, el modo en que el aporte de Davidson sentó un precedente canónico en torno a dicho rasgo, voy a centrarme en su obra como vertebradora de este capítulo, sin que ello implique que no sea consciente de su impacto en la de otros autores, para quienes la normatividad de la atribución psicológica resulta también de tratamiento acuciante, toda vez que nos disponemos a pensar en nosotros mismos como “animales racionales”.

Sin embargo, creo que es importante destacar que a pesar de la profusa cantidad de trabajos publicados recientemente sobre este tópico, no hay un acuerdo razonable entre los autores acerca de cómo entender la normatividad. Es por ello que uno de los objetivos de este capítulo es clarificar de qué hablamos cuando hablamos de “normas de lo mental”, sus variedades y alcance, así como mostrar algunos puentes entre la filosofía de la mente y la filosofía práctica, que son posibilitados –según argumentaré– por la normatividad propia de la atribución psicológica.

1. El giro “interpretativo” en filosofía de la mente y la normatividad de lo mental

Como es sabido, con el surgimiento de las así llamadas ‘Ciencias del Espíritu’ en el siglo XIX el tópico de la *interpretación* pasó a ocupar un lugar central en la “querella metodológica”. En efecto, la interpretación se constituyó en una de las claves para defender la peculiaridad epistemológica de dichas ciencias. Se trataba de hacer lugar al “dualismo metodológico” frente a las pretensiones monistas de quienes consideraban que todo saber (a los efectos de considerarse “científico”) debía emular al método propio de las ciencias naturales.¹

No obstante, fue recién en el siglo XX cuando asistimos a un profundo desplazamiento en torno a la interpretación. En efecto, se produce el tránsito desde su consideración estrictamente epistemológica, esto es, como un mero método de acceso a la mutua comprensión, hacia el tratamiento de su dimensión antropológica, es decir: como un rasgo constitutivo del existente humano.

Ahora bien, la manera en que este “giro interpretativo” impactó en la filosofía de la mente es, en mi opinión, un paso conceptualmente indispensable para comprender de modo cabal el derrotero que lleva al íntimo encuentro de la normatividad con lo mental. En este sentido, cabe recordar que la inflexión en cues-

1 En este sentido resulta muy iluminador el capítulo 1 de Von Wright (1971).

tión está enraizada en la elección de la perspectiva del intérprete como aquella desde la cual se ha de investigar el pensamiento, el lenguaje y su relación. Esta adopción estuvo motivada por la idea de que el pensamiento y el lenguaje son, esencialmente, fenómenos *sociales* que deben ser indagados desde una dimensión *pública*. Aquello que, en mi opinión, liga a la empresa interpretativa con la filosofía de la mente es la convicción de que el pensamiento y el lenguaje son inescindibles y puesto que este último es, por su naturaleza, un instrumento para la comunicación, entonces, el significado y la creencia deben ser accesibles desde un punto de vista público. Los datos relevantes se restringen, pues, a lo disponible para un intérprete radical, quien debe construir su teoría acerca de los pensamientos y palabras de los demás y su conexión con la realidad a partir de los recursos públicos disponibles para cualquier sujeto cognoscente.

Con esta suerte de “giro copernicano” en filosofía de la mente, se deja de lado la clásica concepción cartesiana de la mente como un espectáculo privado y accesible sólo al “ojo interior” de cada sujeto.² Como veremos, esa concepción de lo mental resulta desacreditada por el análisis de los supuestos necesarios de la interpretación y la comunicación intersubjetiva, entre los cuales la normatividad juega un papel central. Asistimos, pues, desde mediados del siglo XX a un movimiento que posa la mirada en la figura del intérprete y conduce a una transformación profunda de los supuestos en que se ha movido la reflexión filosófica desde los tiempos de Descartes. Las importantes consecuencias que se desprenden de este giro dependen, pues, crucialmente de la peculiar concepción de la mente humana desde la perspectiva de la *interpretación* del lenguaje y de la conducta, es decir, desde la perspectiva del proceso por el que un sujeto trata de hallar sentido en la conducta y las emisiones lingüísticas de otro.

A la luz de este giro, la mente deviene un producto de la interpretación y de la comunicación intersubjetiva. Como señala Carlos Moya (1992), la mente resulta aquello que atribuimos

2 Para una profundización de esta concepción de la mente pueden verse, especialmente, los capítulos de Burdman y Zárate en esta compilación.

a los demás para hacernos inteligible su conducta (lingüística y no lingüística). La mente se revela, pues, como interactiva y social. Esta renovada investigación de la comunicación y de la interpretación lingüística que, según veremos, es heredera del aporte fundamental de Wittgenstein y del análisis de Quine de la traducción radical, se consolida en el proyecto de Davidson (1973/1984) de la *interpretación radical*.

No se trata de un mero cambio terminológico respecto de la contribución de Quine (1960): en principio, podemos saber que una oración traduce a otra, sin saber qué significa ninguna de las dos; en cambio, la interpretación de una oración tiene que darnos su significado. El intérprete radical pretende construir una teoría del significado de las emisiones (presuntamente lingüísticas) de un sujeto cuyo lenguaje le es completamente desconocido.

Como veremos, situar el punto de partida del análisis de la interpretación en esta situación radical, extrema, es un artificio metodológico destinado a poner de manifiesto los aspectos implicados en la comunicación cotidiana entre los seres humanos. Como en “La carta robada” de Edgar Allan Poe, a veces de tan familiar que nos resulta, dejamos de ver lo que tenemos ante nosotros. Así, la ventaja que reviste es que nos permite evitar que pasen inadvertidos presupuestos importantes de la comunicación, cosa que puede suceder fácilmente si analizamos la comunicación en el caso de sujetos que comparten un lenguaje y una cultura.

Ahora bien, cabe advertir que es posible discernir al menos dos aspectos en la teoría de la interpretación radical. Por un lado, (i) encontramos el intento de construir una teoría composicional del significado para un lenguaje natural, que ha de tener la forma de una teoría tarskiana de la verdad para demostrar cómo podemos comprender expresiones semánticamente complejas sobre la base de que comprendemos las semánticamente primitivas y sus combinaciones. Por otro lado, (ii) encontramos el rol que tiene la adopción de la *perspectiva del intérprete* respecto al hablante; un intérprete cuya evidencia disponible excluye toda información acerca del significado de las expresiones lingüísti-

cas del hablante y todo conocimiento acerca de los contenidos de sus creencias, deseos y demás actitudes proposicionales (cf. Moretti, 2008).

No obstante, a los efectos de echar luz sobre el encuentro de la normatividad y lo mental, no voy a detenerme en su inscripción en el programa semántico que hace del concepto de condiciones veritativas de una oración la clave para desarrollar las ideas de significado y comprensión, sino en una de las restricciones surgidas a la luz de la consideración de la situación del *intérprete radical*: el rol del principio de caridad.

Lo que Davidson denomina “interpretación radical” es la interpretación de las emisiones de un hablante de un lenguaje totalmente desconocido para nosotros. El intérprete radical supone, entonces, que está frente a una actividad lingüística que le es imposible, por mera observación, entender (cf. Davidson, 1973). En tal instancia, su único elemento de juicio van a ser grupos de emisiones, *prima facie* oracionales, que sean sostenidas como verdaderas [*hold true*] por los presuntos hablantes. Inicialmente, entonces, el intérprete debe descubrir la actitud del hablante hacia sus propias emisiones, esto es, si sostiene como verdadera una oración –o no– en circunstancias particulares. De este modo, si no asumimos al menos provisionalmente que al emitir un sonido el sujeto está profiriendo una oración y que al hacerlo pretende enunciar algo que sostiene como verdadero en las circunstancias en las que realiza dicha emisión, no podemos siquiera intentar una tarea de interpretación.

Ahora bien, tal como señalan Amoretti y Vassallo (2008), sostener como verdadera una oración es ya una actitud semántica, pero de acuerdo con Davidson se trata de una actitud que puede preceder a la interpretación. Por ejemplo, el intérprete puede saber que un sujeto sostiene como verdadera la oración “Sta piovendo” sin haber reconocido qué verdad específica entraña. Así, aún cuando el intérprete sabe que el hablante sostiene una oración como verdadera, todavía no es capaz de determinar lo que el hablante cree ni lo que significa su emisión efectiva. Imaginemos que durante una tormenta el hablante profiere: “Sta piovendo”. A los efectos de interpretar esta emisión y averiguar que signifi-

ca “Está lloviendo”, el intérprete debe ser capaz de adscribirle al hablante la creencia de que está lloviendo; sin embargo, para atribuirle dicha creencia, el intérprete debe conocer el significado de sus palabras. En otros términos: supongamos que el intérprete sabe que el hablante sostiene una cierta oración como verdadera; si supiera el significado de la oración, podría establecer lo que el hablante cree y, conversamente, si supiera lo que el hablante cree, podría determinar el significado de la oración proferida. De esta suerte, la interdependencia de creencia y significado es inextricable. En consecuencia, resulta crucial encontrar una manera de romper el círculo que los liga.

Según Davidson, podemos lograrlo definiendo una teoría general de la interpretación que pueda ofrecernos “simultáneamente una teoría de la creencia y una teoría del significado” (1974: 144) y que nos suministre “un método para mantener constante un factor mientras estudiamos el otro” (1975: 167). La teoría en cuestión es, precisamente, la de la interpretación radical, en tanto que el método aludido es suministrado por un principio que Davidson toma de Quine: el *principio de caridad*.

Davidson introduce, entonces, el principio de caridad para disipar el problema que suscita la interdependencia de la *creencia* y el *significado*. Dado que la tarea de despejar ambas incógnitas es excesiva a la luz de la magra evidencia provista por la actitud de sostener una oración como verdadera, sugiere que el intérprete mantenga fijo un factor mientras se estudia el otro. La caridad, precisamente, proporciona un método para mantener constante el factor *creencia* mientras se resuelve el factor *significado*. En virtud de este principio el intérprete presupone que la oración que el hablante sostiene como verdadera generalmente expresa una creencia que es de hecho verdadera, y que el hablante no tiene creencias distintas de las que él mismo tendría si estuviera en su lugar.

Si suponemos que las creencias del hablante están de acuerdo con las nuestras y son de hecho verdaderas –al menos en los casos más simples y básicos– podemos emplear nuestras creencias acerca del mundo como guía hacia las creencias del hablante.

Ahora bien, el principio de caridad comprende tesis muy distintas y ha sido formulado de maneras muy diversas. En la con-

tribución específica de Davidson lo encontramos caracterizado, recientemente, a partir de dos vertientes que denominó: el *principio de coherencia* y el *principio de correspondencia* (cf. 1991).

El principio de *coherencia* lleva al intérprete a descubrir un cierto grado de consistencia lógica en el pensamiento del hablante, mientras que el principio de correspondencia lleva al intérprete a considerar que el hablante está respondiendo a los mismos rasgos del mundo a los que el intérprete respondería en circunstancias similares. De este modo, “un principio dota al hablante de una módica cantidad de lógica y el otro lo dota en un cierto grado con lo que el intérprete considera que son creencias verdaderas acerca del mundo” (1991: 211).

En otras palabras, según el principio de coherencia la *caridad* es entendida como *racionalidad*: desde esta perspectiva, el intérprete está obligado a presuponer que el hablante es –más o menos– racional de acuerdo *con sus propios estándares de racionalidad*. Adoptar la caridad no es, pues, una opción.

En consecuencia, el intérprete no puede evitar *proyectar su propia lógica* sobre el discurso, el pensamiento y la acción ajenos; la consistencia entre las creencias y entre estas y la acción resulta un requisito esencial de la interpretación.

A su vez, según el principio de correspondencia, la caridad es asociada con la verdad: el intérprete está obligado a presuponer, asimismo, que la mayor parte de las creencias del hablante son verdaderas de acuerdo con sus *propios estándares*.

En suma, dada esta doble vertiente, el principio de caridad implica presuponer que las creencias son comunes a hablantes e intérpretes y que la mayor parte de esas creencias compartidas son verdaderas.

De este modo, si el supuesto hablante no es inteligible y mayoritariamente correcto, no es interpretable, es decir: no es realmente un hablante para el intérprete. En efecto, Davidson sostiene que en nuestra necesidad de hallar comprensible lo que alguien dice, trataremos de “construir una teoría que la haga consistente, creyente de verdades y amante de lo bueno” (1970: 222). De acuerdo con esto, las diferencias de tipo y grado entre los sistemas de creencias del hablante y el intérprete sólo pueden

ser comprendidas a partir del acuerdo general. De ello se sigue a juicio de Davidson que: “en la medida en que no logramos descubrir un patrón coherente y plausible en las actitudes y acciones de los otros, simplemente renunciamos a la posibilidad de tratarlos como *personas*” (1970: 221-222; mis *itálicas*).

A la vera de esta transformación, vamos a asistir a una filosofía de la mente post-ontológica en la que estados de la mente como creencias, deseos e intenciones no son entidades, sino modificaciones de una persona. Describimos estas modificaciones relacionando a la persona con una oración, pero no se trata de que haya objetos abstractos que estén “ante la mente” de la persona a la que se le atribuye el pensamiento (Davidson, 1989). Los supuestos de la interpretación obligan a concebir el contenido de nuestras creencias básicas como un evento u objeto público (y no como una entidad intermedia entre el sujeto y el mundo). Los contenidos de nuestras creencias básicas son, pues, parte del mundo público e intersubjetivo, y no objetos entre este y nosotros. En suma, la concepción representacionista de la mente como un espectáculo privado y accesible sólo al “ojo interior” de cada sujeto resulta, entonces, socavada por el análisis de los supuestos necesarios de la interpretación y la comunicación intersubjetiva. Asistimos, de este modo, a un movimiento que conduce a una transformación profunda de los supuestos en que se ha movido la reflexión filosófica desde los tiempos de Descartes.

En suma: las importantes consecuencias que se desprenden del “giro interpretativo” en cuestión dependen de una peculiar concepción de la mente humana desde la perspectiva de la interpretación del lenguaje y de la conducta, es decir, desde la perspectiva del proceso por el que un sujeto trata de hallar sentido en la conducta y las emisiones de otro. De ello, según vimos, resulta la concepción de la mente bajo el principio constitutivo de la racionalidad y bajo el supuesto de la veracidad de las creencias: las dos caras del célebre “principio de caridad”.

La normatividad, en clave del principio de caridad, se exhibe entonces como un rasgo constitutivo de nuestras prácticas cotidianas de atribución psicológica.

2. ¿De qué hablamos cuando hablamos de “normas” de lo mental?

A la luz de la lectura que propuse del “giro interpretativo”, la caridad asoma, pues, como el rostro clave de la normatividad de lo mental. Ahora bien, es importante insistir en que a pesar de la cantidad de trabajos publicados recientemente sobre este tópico, no hay un consenso razonable entre los autores acerca de cómo entenderlo. Es por ello que una cuestión crucial que pretendo indagar aquí es: en qué sentido la normatividad es un rasgo constitutivo de nuestra práctica cotidiana de hacernos atribuciones psicológicas y en qué medida responder a esta pregunta involucra responder a la cuestión respecto de qué nos hace ser quienes somos.

Antes de seguir avanzando sobre estas cuestiones se impone, entonces, distinguir tipos y sentidos de *normatividad*. A tal efecto, a continuación describo algunas elucidaciones que hice previamente al respecto (Pedace, 2012). Tradicionalmente, se considera que la esfera de la normatividad puede descomponerse en las dos siguientes dimensiones: la deóntica (recogida por nociones tales como “correcto”, “incorrecto”, “deber”, “obligación”, “permiso”, etc.) y la evaluativa (recogida por nociones tales como “bueno”, “malo”, etc.)³. De este modo, se considera que la normatividad “es un rasgo común a ambos lados de la distinción deóntico/evaluativo” (Dancy, 2000: vii).

Comienzo, entonces, por la dimensión deóntica de la normatividad; para ello voy a valerme del ya célebre trabajo de Georg H. von Wright (1963). En *Norm and Action*, él distingue entre tres tipos de normas, a saber: las prescripciones, las directivas y las reglas.

En primera instancia, Von Wright caracteriza las *prescripciones* del modo siguiente. Tienen su fuente en una autoridad que demanda que los sujetos adopten una cierta conducta; a los efectos de que su voluntad sea conocida, la autoridad promulga la nor-

3 Esta presentación clásica puede encontrarse, entre otros, en Dancy (2000) y Engel (2001).

ma y añade, en vistas de que sea efectiva, una sanción o amenaza de castigo. En términos generales, entonces, las prescripciones son órdenes o permisos dados por alguien en posición de autoridad a alguien en posición de sujeto (Von Wright alude a las leyes del estado como ejemplo paradigmático de prescripciones).

En segundo lugar, encontramos las *directivas o normas técnicas*. Este tipo de norma concierne a los medios que han de emplearse para la consecución de un fin determinado. De este modo, resultan ilustraciones típicas de las directivas las así llamadas “instrucciones de uso”, en las que se presupone que la persona que las sigue apunta a la obtención de un fin o resultado. Su formulación estándar se da mediante oraciones condicionales en cuyo antecedente se menciona algo que se quiere y en cuyo consecuente figura algo que debe (o no debe) hacerse a tal efecto.

Finalmente, encontramos el tercer tipo de norma: las *reglas*. El ejemplo prototípico que nos da Von Wright son las reglas de un juego. Jugar un juego es una actividad humana que se lleva a cabo bajo patrones estandarizados que podemos llamar “movidas” del juego. La idea es que las reglas del juego determinan estas movidas y, mediante ellas, el juego mismo y la actividad de jugarlo. Desde la perspectiva del juego mismo, las reglas determinan cuáles son las movidas correctas y, desde la perspectiva de la actividad, cuáles son las movidas permitidas (de este modo, las movidas que no son correctas están prohibidas para los jugadores del juego y aquella movida que es la única correcta en una cierta situación, es la obligada cuando se está jugando el juego).

Ahora bien, dado este marco general ¿bajo qué sentido de normatividad resulta adecuado elucidar el tipo de normatividad que nos hace especiales, que resulta distintiva de la atribución psicológica?

Comencemos por el primer tipo de normas: las prescripciones. Estos candidatos no parecen ofrecer una clave apropiada para comprender la normatividad de lo mental al interior de la propuesta de Davidson. Como es sabido, él rechaza de manera explícita aquellos abordajes de la comunicación que reposan en la idea de que para nuestra comprensión mutua debemos seguir prescripciones o convenciones particulares ligadas a las

palabras o a los conceptos que determinen su uso correcto.⁴ Así, por ejemplo, en “Communication and Convention”, Davidson intenta hacernos ver –contra la idea de David Lewis– que no es, en absoluto, una trivialidad o un lugar común que haya (o deba haber) convenciones en el lenguaje, sino que se trata de una tesis filosófica cargada de contenido teórico. Él aborda esta cuestión argumentando que “la convención no ayuda a explicar qué es básico para la comunicación” (Davidson, 1984/1984: 280).

Examinemos, entonces, si el segundo tipo de norma, las directivas, permite echar luz sobre la peculiar normatividad de lo mental que Davidson propone. Como se advierte en Von Wright (1963: 9 y ss.), las directivas deben su validez, su carácter de guías para acción, a relaciones previas e independientes de cualesquiera de los imperativos hipotéticos derivados de ellas, esto es: deben su validez a las leyes naturales. Ahora bien, Davidson nos recuerda la lección de Wittgenstein respecto de que debe haber un modo de interpretar que no es cuestión de seguir directivas o instrucciones, en el sentido de que el tipo de conocimiento que describe una teoría de la creencia/significado no es irrelevante, pero nunca puede instruirnos acerca de cuándo aplicarla. No hay ninguna directiva en sentido estricto, como opuesta a meras máximas y generalidades metodológicas. De este modo, tampoco obtenemos una clarificación de la normatividad de lo mental por esta vía.

Pasemos, entonces, a considerar a nuestro tercer candidato: veamos si puede resultarnos de ayuda concebir la normatividad de lo mental en términos de reglas. Como ya señalé, los juegos son los casos paradigmáticos de actividades esencialmente constituidas por reglas. Sus reglas no dependen, como en el caso de

4 Soy consciente de que la asimilación que estoy haciendo entre prescripciones y *convenciones* amerita, por lo menos, el siguiente par de aclaraciones. Las convenciones o hábitos, nos dice Von Wright, se comportan como prescripciones en tanto influyen en la conducta ejerciendo “presión normativa” sobre los miembros de la comunidad. No obstante, en el caso de las convenciones sólo podemos hablar de “autoridad” en tanto nos referimos a la comunidad misma y, de esta suerte, las convenciones pueden caracterizarse como prescripciones “anónimas” (cf. 1963: 9).

las directivas, de leyes naturales previas e independientes; por el contrario, las reglas de los juegos son claramente “creaciones humanas, hechas, no encontradas” (Baker y Hacker, 1985: 63). En este sentido, son arbitrarias. No hay algo tal como el juego correcto o el conjunto correcto de reglas; podemos crear incontables conjuntos de reglas, incontables juegos. La segunda característica central de las actividades gobernadas por reglas, como jugar vóley, es que no sólo es imposible jugar al vóley sin reglas, sino que es imposible jugar vóley sin las reglas del vóley. Tales actividades están definidas o constituidas por sus reglas: si cambiamos las reglas, automáticamente cambiamos la actividad.⁵ Ahora bien, consideradas conjuntamente estas dos características, la arbitrariedad y la constitutividad, tornan complejo considerar la normatividad de lo mental en términos del principio de racionalidad entendido como regla de juego de la práctica de nuestra atribución psicológica.

En efecto, la analogía entre las reglas de los juegos y las relaciones básicas constitutivas de nuestros conceptos parece colapsar en un aspecto decisivo: la arbitrariedad. Recordemos que, a juicio de Davidson: “la caridad no es una opción” (1974/1984: 197). Hay diferentes juegos que podemos jugar, pero hay un único sistema de conceptos. Tener conceptos es básicamente tener nuestros conceptos. Son estos mismos conceptos los que son una condición para tener creencias, deseos, intenciones y los medios para expresarlos a través del lenguaje.

Puesto en términos de “juegos”, aquí ya no parece tener sentido decir que podríamos jugar un juego distinto al intencional; este es el único “juego” que no podemos dejar de jugar, no importa qué hagamos. Su validez es condición para la posibilidad

5 En este punto, cabe recordar la clásica distinción trazada en *Speech Acts*. Allí Searle intenta clarificar la distinción entre las reglas regulativas y las reglas constitutivas. Su idea es que las primeras regulan formas de conducta previamente existentes (ejemplo de ellas son las reglas de etiqueta que gobiernan relaciones interpersonales que existen independientemente de tales reglas), en tanto que las segundas no regulan meramente, sino que constituyen o definen nuevas formas de conducta (en este sentido, las reglas del fútbol, por ejemplo, crean la posibilidad misma de jugar ese juego).

de la intencionalidad y el significado y, como tal, es *sui generis*. En consecuencia, si se la piensa en términos de la analogía con los juegos, parece haber dificultades a los efectos de considerar la normatividad de lo mental como las reglas del juego de nuestra mutua comprensión.⁶ No obstante, si acotamos nuestro tratamiento sólo a la dimensión constitutiva del principio de racionalidad (con prescindencia de su carácter no opcional, no arbitrario), quizás podamos arrojar algo de luz sobre su normatividad. Tal vez aquello que Davidson está subrayando es que se trata de una regla constitutiva de un tipo especial, en el sentido de que su peculiaridad consiste en que sea esta regla. Veamos más detenidamente si podemos esclarecer este punto.

Cabe señalar que es el holismo de lo mental el rasgo que conduce al reconocimiento de su normatividad. En mi opinión, para apreciar la relevancia de esta observación, debemos reparar en que “holismo” se dice de distintas maneras en la teoría propuesta por Davidson. A mi juicio, es posible deslindar los dos tipos siguientes: por un lado, un holismo metodológico, que constriñe la teoría y, por otro, un holismo constitutivo, que constriñe aquello acerca de lo cual versa la teoría. Lo que, en mi opinión, amerita que concibamos a Davidson como comprometido con un holismo metodológico es que, a los efectos de la aceptabilidad de la teoría, lo que debe buscarse no son interpretaciones correctas de muestras, sino la corrección material de la teoría como totalidad.

No obstante, creo que es en el holismo constitutivo donde se exhibe la peculiaridad de su aporte, y a partir de donde podemos comenzar a elucidar el tipo especial de normatividad que tiene lo mental. En efecto, en respuesta a las objeciones de Fodor y Lepore (1992), Davidson distingue, a su vez, este último tipo de holismo en dos variedades, igualmente importantes para su concepción de lo mental: el holismo intra-actitud y el holismo inter-actitud

6 En un espíritu similar, pero a propósito de la normatividad semántica, Kathrin Glüer (2001), emplea la analogía wittgensteiniana con los juegos para sugerir que el uso significativo de las expresiones lingüísticas carece de los rasgos característicos requeridos por las actividades esencialmente gobernadas por reglas.

(variedades que prefiero denominar, como holismo semántico o del contenido (lingüístico/mental) y holismo psicológico o de las actitudes proposicionales, respectivamente). Según su tesis holista del contenido, la identidad de un pensamiento dado depende, en parte, de sus relaciones con otros pensamientos. En virtud de que esta variedad de holismo concierne a las relaciones entre varias creencias dentro de la categoría de creencia, Davidson la denomina “holismo intra-actitud”. La idea es que la individuación de lo creído depende de otras cosas creídas, esto es, no podemos tener un pensamiento aislado porque cada pensamiento supone un contenido y ese contenido está determinado por su relación con los contenidos de otras creencias atribuidas. A su vez, dichos contenidos poseen condiciones de corrección que se juegan en su dirección de ajuste al mundo. Así, para considerar un ejemplo que Davidson (1982) proporciona en “Rational Animals”: si creo que hay una moneda en mi bolsillo, el contenido de mi creencia puede ser verdadero o falso, según se ajuste o no al mundo y, de manera correlativa, mi creencia será correcta o incorrecta.

Por su parte, el holismo inter-actitud (u holismo psicológico o de las actitudes proposicionales) concierne a las relaciones entre una categoría de pensamiento o juicio y otra (de allí, que sea “interactitud”): esto es, concierne, por ejemplo, a las relaciones entre los actos de creencia, deseo e intención.

Tras este rodeo en torno al compromiso davidsoniano con el holismo de lo mental, creo que es posible comenzar a avizorar en qué sentido el principio de racionalidad es una regla constitutiva de tipo especial. La normatividad es un rasgo esencial de la atribución psicológica porque lo mental –a diferencia de lo físico– se constituye en un marco de actitudes que tienen contenidos proposicionales que guardan entre sí relaciones lógicas.

En este sentido, no podemos tener cualquier colección arbitraria de pensamientos: nos restringe la lógica. Esto es, nuestros pensamientos están constreñidos por las relaciones que imponen la consistencia y la coherencia lógica propias del principio de racionalidad, que resulta una regla de tipo especial en tanto no tiene eco en la teoría física.

Ahora bien, aún bajo la presunción (optimista) de que hemos

logrado clarificar el tipo peculiar de normatividad de lo mental a partir de su sentido deóntica, cabe preguntarnos si ello es suficiente para capturar la profundidad que la torna aquel rasgo que nos hace ser, distintivamente, quienes somos. En otros términos: ¿es posible escindir esta acepción de la normatividad de aquella otra que completa su tratamiento tradicional, esto es, su acepción evaluativa?

Davidson (2005) nos dice cuál es su objeción central a la idea de concebir la normatividad en términos de meras reglas: aquellas que gobiernan nuestra práctica comunicativa son dependientes de “valores ulteriores” (*further values*). ¿Tenemos que internarnos, entonces, en la dimensión evaluativa de la normatividad para terminar de explicitar la dimensión deóntica peculiar de las reglas que constituyen la práctica de nuestra mutua comprensión?⁷ ¿Cuáles son, entonces, esos valores ulteriores de los que depende la normatividad que advertimos en la “superficie” de dicha práctica? Según acabamos de ver, las condiciones de coherencia y consistencia racional parecen crear la posibilidad misma de la interpretación. No podemos atribuir inteligiblemente creencias y deseos, o describir movimientos como conducta, “a menos que estemos comprometidos a encontrar, en el patrón de conducta, creencia y deseo un amplio grado de racionalidad y consistencia, y si fallamos en descubrir un patrón coherente en las actitudes y acciones de los demás perdemos la oportunidad de *tratarlos como personas*” (Davidson 1970/80: 222, mis itálicas).

En este punto, mi invitación es a que consideremos la posibilidad de interpretar la noción de “persona” desde una connota-

7 Además de las razones que encuentro al interior del planteo de Davidson, creo que es posible argumentar a favor de una “imbricación” entre la dimensión deóntica y la dimensión evaluativa a partir de razones externas a él. En este sentido, puede resultar iluminador el debate en torno a normas y valores que han mantenido Putnam y Habermas (2008). Si bien está precedido por un acercamiento progresivo entre las filosofías de ambos autores desde mediados de los años ochenta, el debate se abre propiamente cuando Putnam pronuncia en 1999 la conferencia “Valores y normas”. Allí expone su rechazo de la dicotomía que, a su juicio, Habermas postula entre los valores y las normas, para defender, en cambio, una tesis de “imbricación” (*entanglement*) entre ambos.

ción evaluativa: la normatividad de las atribuciones psicológicas que se exhibe en la interpretación resulta especial en la medida en que depende del siguiente “valor ulterior”: el de tratarnos mutuamente como personas.⁸

Así, leemos en Davidson:

[N]o podemos empezar a descifrar lo que dice una persona si no comprendemos sus actitudes en torno a sus oraciones, tales como sostener, desear o querer que sean verdaderas. Empezando por estas actitudes, debemos construir una teoría de lo que la persona quiere decir, dando así, simultáneamente, contenido a sus actitudes y a sus palabras. En nuestra necesidad de hallarla comprensible, trataremos de construir una teoría que la haga consistente, creyente de verdades y amante de lo bueno (1970/1980: 222).

Es, precisamente, esta noción de *persona* que el propio Davidson pone en juego en el corazón de su filosofía de la mente aquella que, en mi opinión, opera como un gozne, como una bisagra, para tender puentes hacia la filosofía práctica.

3. La normatividad de la atribución psicológica: un puente entre la filosofía de la mente y la filosofía práctica

En un espíritu consonante con el que propongo aquí, B. Ramberg nos ayuda a ver que Davidson está tratando de quebrar la

⁸ Parece razonable preguntarnos en qué sentido la noción de persona entrañaría un “valor”. Como argumenté en Pedace (2012), a los efectos de elaborar mi sugerencia, creo que puede resultarnos de ayuda la reflexión que Korsgaard (1996) hace en torno a la filosofía kantiana. Ella distingue entre el valor de los fines y el valor de la agencia: a su juicio, Kant advierte que consideramos cosas como importantes (*value of ends*) porque son importantes para nosotros y concluye, por tanto, que debemos tomarnos a nosotros mismos como importantes (*value of agency*). En esta línea, recordemos la segunda formulación del imperativo categórico, en la que se pone en juego el reconocimiento de la humanidad en la propia persona y en la de los otros como fin en sí mismo.

distinción entre la mente como espectadora y como participante responsable en las prácticas sociales.⁹ En efecto, al combinar una teoría del significado con una teoría de la acción, Davidson entreteje la distinción mente-cuerpo con la distinción persona-cosa. Mientras que la noción de mente se vincula con una noción de racionalidad teórica, entendida como posesión de creencias y deseos, la noción de persona se vincula con una noción de racionalidad práctica, que involucra la idea de ser uno de nosotros, esto es, ser un miembro de nuestra comunidad, en la que estamos ligados por nuestras responsabilidades recíprocas. En este sentido, Davidson afirma que “cada mente se revela como parte de una comunidad de *yoes libres*” (Davidson, 2001: 9).

Recordemos que la empresa de más largo aliento que persigue Davidson es la de proporcionar una teoría unificada del pensamiento, el significado y la acción (cf. “A Unified Theory of Thought, Meaning and Action”, 1980/2004). Cabe subrayar que la teoría de la interpretación radical no es una mera teoría de condiciones de verdad, sino que a los efectos de su aplicabilidad empírica hay que complementarla con elementos ajenos a ella, a saber: las restricciones que surgen de la adopción de la perspectiva del intérprete radical. Su teoría unificada nos invita, a su vez, a precisar dichas restricciones mediante los aportes de los teóricos de la decisión. En este sentido, Lepore y Ludwig (2006: 257) sostienen que la teoría unificada, lejos de reemplazar el procedimiento de la interpretación radical presentado inicialmente en (1973), constituye un refinamiento del mismo. Lo que me interesa enfatizar aquí es que el carácter unificado de su teoría del pensamiento/significado y de la acción parece apuntar en dirección a un entrelazamiento de la racionalidad teórica (tener una mente/ser interpretable) y la racionalidad práctica (ser persona).

En su respuesta a Bill Martin (1999), el propio Davidson sostiene que hay cuatro áreas en las cuales considera haber tocado cuestiones relevantes para la ética, a saber: la naturaleza del razo-

9 Cf. Ramberg (2000). Allí afirma que en lugar de leer a Davidson bajo el signo de Brentano, esto es, en términos de la irreducibilidad de la intencionalidad, deberíamos pasar a leerlo en la clave hegeliana de nuestro mutuo reconocimiento (*Anerkennung*) como personas.

namiento práctico, la cuestión de la objetividad de los juicios de valor, la autonomía de lo intencional y los aspectos normativos de la interpretación. En el presente capítulo, he propuesto una elucidación de estos aspectos a partir del tratamiento tradicional de la normatividad, esto es: su dimensión deóntica y su dimensión evaluativa. La idea que intenté presentar y sostener es que la normatividad propia de la atribución psicológica, aquel rasgo que a juicio de Davidson porta la clave de lo humano, debe comprenderse en términos del principio de racionalidad como regla constitutiva de la práctica atributiva especial por la que nos tratamos mutuamente como *personas*.

Ahora bien, un punto controvertido en el que me interesa detenerme en esta intersección entre la filosofía de la mente y la filosofía práctica es si la atribución psicológica, tal como ha sido caracterizada desde la interpretación radical, nos compromete con una posición etnocéntrica.

En efecto, tal como ha señalado Pablo Quintanilla (2001), podría afirmarse que al abrazar el principio de caridad no se está dejando resquicio alguno para aceptar una diferencia radical, en virtud de que dicho principio, ropaje clave de la normatividad, no permitiría que el otro al que ha de interpretarse pueda ser radicalmente diferente de uno. Así leemos:

[a]l proyectar nuestros estados mentales en el otro, podría decirse, lo estamos convirtiendo en un alter ego de nosotros mismos y, entonces, no estamos capturando lo que él realmente es sino sólo una fabricación de nuestra propia imaginación. Así, la intérprete se estaría engañando a sí misma, pensando estar comprendiendo al agente cuando en realidad sólo estaría imponiendo sus propias creencias en él. Mientras que en la interpretación individual esto es una forma de ego-centrismo, en la interpretación intercultural es una forma de etnocentrismo; en ambos casos no es comprensión sino imposición. Si estos cargos están justificados, ciertamente constituirían argumentos devastadores contra la teoría de la interpretación basada en el principio de caridad (2001: 7).

En lo que resta, deseo explorar si la normatividad de la atribución psicológica aquí considerada conduce, en el plano práctico,

a la imposición etnocéntrica mentada. A tal efecto, mi estrategia será examinar si la noción de una alteridad “radical” introduce (o no) un dilema genuino para quien quiera abrazar la teoría de la interpretación radical. Cabe, entonces, preguntarnos qué concepción de la alteridad subyace a la noción de subjetividad pensada al interior de la teoría de la interpretación radical.

En otros términos: la invitación es a que pensemos cuán otro puede ser el otro con el que interactúa el intérprete radical, es decir, cuán radical ha de ser esta alteridad que se cierne frente al intérprete radical. ¿Hay un genuino lugar conceptual para esta noción?

El principio de caridad parece comprometernos con un espacio conceptual en el que el otro debe ser suficientemente parecido a nosotros, so pena de no poder interpretarlo. En este sentido, *prima facie* parece quedar atrapado por el paradigma asimilacionista en que el otro es concebido como un mero alter ego. En efecto, sin el supuesto de un conjunto crítico de creencias verdaderas compartidas y de valores conjuntamente asumidos, no hay comunicación posible desde el marco de la teoría en cuestión.

El principio de caridad pareciera, en principio, ponernos pues ante el siguiente dilema: (i) o se interpreta al otro y, en consecuencia, se lo reconoce como *persona*, a expensas de “fagocitarlo”, esto es, de asimilarlo a la mismidad del intérprete o (ii) se retiene su radical alteridad, a costa de no poder reconocerlo siquiera como otro, si es que vamos a tomarnos en serio la idea de su incompreensión profunda, su carácter inapresable. En una formulación más condensada, las opciones desesperadas parecieran ser: fagocitación (antropofagia) o incompreensión (antropoe-mia), y transpuestas en una clave política, pues las alternativas oscilarían entre: el etnocentrismo radical (que se fagocitaría al otro) o el relativismo cultural extremo (que no permitiría siquiera avizorarlo como otro).

Ahora bien, la pregunta que creo que cabe formularnos es si, en efecto, estamos ante un dilema genuino que podría conducirnos al eventual abandono de la teoría de la interpretación radical por sus consecuencias prácticas. En sintonía con esta inquietud, el corazón de lo que me interesa advertir a modo de mero esbozo

es que: la formulabilidad misma del dilema en cuestión depende, a mi juicio, de cuán dispuestos estemos a abrazar la idea de una *alteridad radical*.

Desde dicha noción, se alude –entre otras versiones– a que la apertura incondicionada a una otredad resulta, por definición, opaca a todo intento de asirla. La idea es que “el otro no puede ser introyectado: he allí su singularidad irrenunciable, su imposibilidad de ser sometido a una lógica identificatoria” (Cragnolini, 2007: 55). El espacio entre nosotros y los otros se cifraría así en un “entre” que no asumiría ninguna dialectización posible, sino un ámbito de incertidumbre que no podría ser saldado por ninguna síntesis.

En la idea misma de *alteridad radical* parece anidar, pues, la siguiente tensión. Por un lado, (1) si se trata de una *alteridad radical*, no puedo reconocer siquiera al otro como otro y, por tanto, resta explicar por qué en ese contexto me sentiría interpelada éticamente hacia él/ella. Parece, pues, estallar la pretensión de estar frente a una “alteridad”. Por otro lado, (2) si trata de una *alteridad radical*: i.e. de un otro al que puedo reconocer como otro y comprenderlo/a como *persona*, entonces estalla la pretensión de su profunda e inapresable “radicalidad”.

A la luz de esta elucidación sostengo, pues, que la carga de la prueba de que la teoría de la interpretación radical nos sitúe ante un genuino dilema de opciones prácticas desesperadas reposa en quien abraza la noción de una “alteridad radical”. En efecto, a mi juicio la formulación misma de las opciones en pugna depende de la posibilidad de dotar de inteligibilidad al supuesto de una *alteridad radical*.

En suma, a la luz del puente que tracé desde la filosofía de la mente a la filosofía práctica, tomando como *via regia* la noción de *persona* como rasgo central de la normatividad de nuestras prácticas de atribución psicológica, he intentado argumentar lo siguiente. No tenemos por qué vernos conducidos al abandono inexorable de la teoría de la interpretación radical por las presuntas consecuencias prácticas del principio de caridad. El corazón de la normatividad de la atribución psicológica sólo nos dejaría en esa posición en la medida en que se conceda, según argumen-

té, una problemática noción de alteridad radical. Transfiero, entonces, la carga de la prueba.

No obstante, soy plenamente consciente de que el desafío de pensar el “giro interpretativo” en filosofía de la mente a la luz de las consecuencias prácticas que conlleva la normatividad, sin dudas, todavía persiste. Aunque su profundización será tarea para otra ocasión.

Bibliografía

- Amoretti, M. C. y N. Vassallo (2008), *Knowledge, Language, and Interpretation. On the Philosophy of Donald Davidson*, Frankfurt: Ontos Verlag.
- Baker G. P. y P. M. S. Hacker (1985), *Wittgenstein, Rules, Grammar, and Necessity*, Vol. 2 de *An Analytical Commentary on the "Philosophical Investigations"*, Oxford: Basil Blackwell.
- Brandom, R. (1994), *Making it explicit: reasoning, representing and discursive commitment*, Cambridge, MA: Harvard University Press.
- Brandom, R. (ed.) (2000), *Rorty and His Critics*, Oxford: Blackwell Publishers.
- Cragolini, M. (2007), *Derrida, un pensador del resto*, Buenos Aires: La Cebra.
- Dancy, J. (2000), *Normativity*, Oxford: Blackwell Publishers.
- Davidson, D. (1970/1980), "Mental Events", en: Davidson (1980), pp. 207-228.
- . (1973/1984), "Radical Interpretation", en: Davidson (1984), pp. 125-140.
- . (1974/1984), "On the very idea of a conceptual scheme", en: Davidson (1984), pp. 183-198.
- . (1975/1984), "Thought and Talk", en: Davidson (1984), pp. 155-170.
- . (1980), *Essays on Actions and Events*, Oxford: Oxford University Press. Trad. esp.: *Ensayos sobre acciones y sucesos*, Barcelona: Ed. Crítica, 1995.
- . (1980/2004), "A Unified Theory of Thought, Meaning and Action", en: Davidson (2004), pp. 151-166.
- . (1982/2001), "Rational Animals", en: Davidson (2001), pp. 95-106.
- . (1984), *Inquiries into truth and interpretation*, Oxford: Oxford University Press.
- . (1984/1984), "Communication and Convention", en: Davidson (1984), pp. 265-280.
- . (1989/2001), "What is present to the mind?", en: Davidson (2001), pp. 53-68.
- . (1991/2001), "Three varieties of knowledge", en: Davidson (2001), pp. 205-220.
- . (1992), *Mente, mundo y acción*, Barcelona: Paidós.
- . (2001), *Subjective, Intersubjective, Objective*, Oxford: Oxford University Press.

- (2001b), “Comments on Karlovy Papers”, en: Kotatko, P., P. Pagin y G. Segal (eds.), *Interpreting Davidson*, Stanford: CSLI, pp. 285-308.
 - (2004), *Problems of Rationality*, Oxford: Oxford University Press.
 - (2005) *Truth, Language and History*, Oxford: Oxford University Press.
- Dennett, D. (1987), *The intentional Stance*, Cambridge MA: MIT Press.
- Engel, P. (2001), “Is Truth a Norm?”, en: Kotatko, P., P. Pagin y G. Segal (eds.) (2001), *Interpreting Davidson*, Stanford: CSLI, pp. 37-52.
- Fodor, J. y E. Lepore (1992), *Holism: a shopper´s guide*, Cambridge, Mass.: Basil Blackwell.
- Glüer, K. (2001), “Dreams and Nightmares: Conventions, Norms and Meaning in Davidson’s Philosophy of Language”, en: Kotatko, P., P. Pagin y G. Segal (2001) (eds.), pp. 53-74.
- Habermas, J. y H. Putnam (2008), *Normas y valores*, Madrid: Trotta.
- Hahn, L. E. (1999) (ed.), *The Philosophy of Donald Davidson*, Library of Living Philosophers, vol. 27, Chicago: Open Court.
- Korsgaard, Ch. (1996), *Creating the Kingdom of Ends*, New York: Cambridge University Press.
- Kripke, S. (1982), *Wittgenstein on Rules and Private Language*, Cambridge: Harvard University Press.
- Lepore E. y K. Ludwig (eds.) (2006), *The Essential Davidson/Donald Davidson; with an introduction by Ernie Lepore and Kirk Ludwig*, Oxford: Oxford University Press.
- Martin, B. (1999), “Interpretation and Responsibility: Excavating Davidson’s Ethical Theory”, en: Hahn, L. E. (ed.) (1999), pp. 345-356.
- Moretti, A. (2008), *Interpretar y referir (Ejercicios de análisis filosófico)*, Buenos Aires: Grama.
- Moya, C. (1992), “Introducción a la filosofía de Davidson: mente, mundo y acción”, en: Davidson (1992), pp. 9-45.
- Pedace, K. (2012), “La normatividad de lo mental y el rol de la segunda persona”, en: *Areté, Revista de Filosofía*, Pontificia Universidad Católica del Perú, Vol. XXIV, N° 1, 2012, pp. 109-152.
- Quine, W. V. O. (1960), *Word and Object*, Cambridge: MIT Press.
- Quintanilla, P. (2001), “El lugar de la racionalidad en la comprensión del otro”, en: López Maguiña, S. et al. (eds.), *Estudios Culturales. Discursos*,

poderes, pulsiones, Lima: Red Para el Desarrollo de las Ciencias Sociales en el Perú.

Ramberg, B. (2000), "Post-Ontological Philosophy of Mind: Rorty versus Davidson", en: Brandom, R. (ed.), *Rorty and His Critics*, Oxford: Blackwell Publishers.

Von Wright, G. (1963), *Norm and action, A logical Enquiry*, London: Routledge & Kegan Paul.

—. (1971), *Explanation and understanding*, Cornell University Press.

Wittgenstein, L. (1953), *Philosophical Investigations*, Oxford: Basil Blackwell.

Sobre los autores

Tomás Balmaceda es doctor en filosofía por la Universidad de Buenos Aires. Se dedica a la filosofía de la mente, interesado por los cruces de la psicología del sentido común con los abordajes corporizados y enactivos. En la actualidad es docente en la materia Fundamentos de Filosofía en la Universidad de Buenos Aires, en el programa de educación a distancia UBA XXI y es profesor asociado de la materia Filosofía de la Mente en UCES. Publicó el libro *Psicología de sentido común. Pasado, presente y futuros* (Título, 2014) y participó de numerosas revistas y compilaciones, como *La segunda persona y las emociones* (SADAF, 2017). Es miembro de SADAF. Actualmente participa de los proyectos de investigación “La atribución psicológica. Problemas y perspectivas”, dirigido por Diana Pérez, y “Las definiciones del arte en la estética contemporánea”, dirigido por Ricardo Ibarlucía.

Lucas Buccì es doctor en filosofía por la Universidad de Buenos Aires. Se dedica a la filosofía de la ficción y el cine. Da clases de Introducción al Pensamiento Científico en el Ciclo Básico Común (UBA) y de Metafísica en la Facultad de Filosofía y Letras (UBA). Es miembro de SADAF y sus últimas publicaciones incluyen “La metafísica del Quijote. Un problema para la teoría davidsoniana de la metáfora”, en *Páginas de Filosofía*, y “El enigma de la reacción emocional en el cine. Un análisis crítico de la postura de Kendall Walton sobre la paradoja de la ficción”, en *Boletín de Estética*. Ha presentado trabajos en la Sociedad filosófica del Uruguay y en el Congreso Iberoamericano de Filosofía, en Chile. Actualmente participa de los proyectos de investigación “La atribución psicológica. Problemas y perspectivas”, dirigido por Diana Pérez, y “Las definiciones del arte en la estética contemporánea”, dirigido por Ricardo Ibarlucía.

Federico Burdman es doctor en filosofía por la Universidad de Buenos Aires. Especialista en temas de filosofía de la mente y filosofía de la psicología, se desempeña actualmente como docente de Fundamentos de Filosofía en la Facultad de Filosofía y Letras de la Universidad de Buenos Aires y en las carreras de Filosofía y de Psicología de otras universidades de la región metropolitana. Becario posdoctoral del CONICET, desarrollando su investigación en el Instituto de Investigaciones Filosóficas de SADAFA, ha publicado diversos artículos en revistas nacionales e internacionales, así como el libro *Razón y sinsentido. Apuntes para una lectura de Sobre la certeza de Wittgenstein* (Teseopress, 2016). Beneficiario del Premio Eduardo Rabossi otorgado por SADAFA en 2009, ha participado de diversos grupos de investigación en la Universidad de Buenos Aires, el CONICET y la Agencia para la Promoción de la Ciencia y la Tecnología.

Antoni Gomila es doctor en filosofía por la Universidad de las Islas Baleares (España). Especializado en filosofía de la psicología, filosofía de la mente y psicología teórica. Se desempeña como docente en Psicología del Pensamiento y del Lenguaje y como investigador en la Universidad de las Islas Baleares. Recibió una beca posdoctoral Fulbright en 1990-91 para trabajar en la Rutgers University con Jerry Fodor, y pasó un año sabático en 2009 en The School of Psychology, University of St. Andrews. Editor, junto a Paco Calvo, del *Handbook of Embodied Cognitive Science* (Elsevier, 2009) y autor de *Verbal Minds: Language and the Architecture of the Cognition* (Elsevier, 2013), además de múltiples artículos que pueden leerse en: antonigomila.wordpress.com.

Diego Lawler es doctor en filosofía por la Universidad de Salamanca (España). Especializado en filosofía de la tecnología, filosofía de la acción y filosofía de la mente. Es profesor de grado y posgrado en diferentes universidades nacionales y extranjeras, públicas y privadas. Investigador independiente del CONICET. Publicó el libro *Las acciones técnicas y sus valores* (2007) y numerosas compilaciones y artículos sobre temas de filosofía de la tecnología, filosofía de la acción y filosofía de la mente en revistas

especializadas del país y el extranjero. Presentó trabajos y conferencias en Argentina, Brasil, Chile, Colombia, España, Grecia, Holanda, Portugal, México, Reino Unido y Uruguay. Dirigió y forma parte de proyectos de investigación financiados por instituciones locales (CONICET, ANPCyT) y extranjeras.

Andrea F. Melamed es doctora en filosofía por la Universidad de Buenos Aires. Especialista en filosofía de la mente y filosofía de la ciencia cognitiva. Se desempeña actualmente como docente en Introducción al Pensamiento Científico del Ciclo Básico Común (UBA). Ha publicado artículos en revistas especializadas, como “Las emociones como metaprogramas: lo que la psicología evolucionista nos puede enseñar sobre las emociones”, en *Ludus Vitalis*, y “Las teorías de las emociones y su relación con la cognición: un análisis desde la filosofía de la mente”, en *Revista Cuadernos FHyCS-UNJu*. Ha presentado numerosos trabajos en congresos y conferencias en Chile, Brasil, España y Argentina. Desde 2005, ha participado en proyectos de investigación financiados por CONICET, UBA y ANPCyT.

Karina Pedace es doctora en filosofía por Universidad de Buenos Aires. Especializada en metafísica, filosofía de la mente y metodología de la investigación. Como docente de grado e investigadora se desempeña como profesora adjunta a cargo de Metodología de la Investigación II (Universidad Nacional de La Matanza) y como jefa de trabajos prácticos en Metafísica (UBA). Como docente de posgrado es profesora titular en la Maestría de Estudios Interdisciplinarios de la Subjetividad (UBA), en el Doctorado en Derecho (Universidad de Palermo) y en el Doctorado en Psicología (UCES). Publicó el libro *Mente y lenguaje. La filosofía de Donald Davidson, modelo para armar* (SADAF, 2017) y fue coeditora de *Cuestiones epistemológicas. Una introducción a la problemática científica* (EdUNLu, 2010) y de *Olimpiadas de Oratoria: una década de expresiones. Argumentos y estilos* (2017). Presentó numerosos trabajos y conferencias en Argentina, Brasil, Chile, España y Uruguay. Fue galardonada con el Premio Estímulo a la Investigación SADAF (2006). Dirigió, continúa dirigiendo y participa de

numerosos proyectos de investigación financiados por diversas instituciones (CONICET, UBA, UNLaM y ANPCyT, entre otras).

Diana Pérez es doctora en filosofía por la Universidad de Buenos Aires. Trabaja en temas de filosofía de la psicología, metafísica de la mente y filosofía del arte. Profesora asociada regular de Fundamentos de Filosofía y Metafísica (UBA). Investigadora Principal del CONICET. Ha publicado dos libros: *La mente como eslabón causal* (Catálogos, 1999) y *Sentir, desear creer. Una aproximación filosófica a los conceptos psicológicos* (Prometeo, 2013), además de varias compilaciones y numerosos artículos en revistas especializadas. Ha participado en congresos y dictado conferencias en Brasil, Chile, México, Perú, Uruguay, España, Francia, Grecia y Estados Unidos. Ha recibido el Premio Konex 2006 y el Premio Houssay 2003. Ha dirigido proyectos financiados por la UBA, CONICET y ANPCyT.

Alejandro Zárate es doctor en filosofía por la Universidad de Buenos Aires. Su tesis de doctorado es sobre el llamado “problema difícil” de la conciencia. Entre sus publicaciones relacionadas con estos temas se encuentra el capítulo “¿Qué hay de nuevo, viejo? Acerca de la respuesta de Dennett al argumento del conocimiento contra el fisicalismo”, publicado en el libro *Poderes causales, tropos y otras criaturas extrañas. Ensayos de metafísica analítica* (Título, 2017). Recibió una beca doctoral UBACyT (2013-2017) y participó en diferentes proyectos de investigación, como “La atribución psicológica. Perspectivas y problemas” (FONCyT PICT 2013-1419) y “Conceptos psicológicos: conciencia, intencionalidad, emoción” (UBACyT), entre otros. Se desempeñó como docente asistente de Lógica y Argumentación y Filosofía de la Ciencia en la Universidad Nacional General Sarmiento. Actualmente dicta la materia Introducción al Pensamiento Científico en el Ciclo Básico Común (UBA) y el Taller de Ciencias en la Universidad Nacional de Moreno. Participó en diversos congresos y jornadas.

Impreso en julio de 2018
en los talleres gráficos de Bibliografika
Buenos Aires, Argentina.