

# DAR FORMA A LA MENTE<sup>1</sup>

SHAPING THE MIND

Víctor Fernández Castro<sup>2</sup>  
Universidad de Granada (España)

Recibido: 28-01-2015

Aceptado: 15-02-2015

---

**Resumen:** El objetivo de este trabajo es exponer y analizar críticamente el libro de Tadeusz Zawidzki *Mindshaping*. Este libro presenta una visión alternativa a la visión estándar de la cognición social de acuerdo con la cual el eje central de nuestras capacidades socio-cognitivas es el mecanismo psicológico conocido como lectura de mentes. Por el contrario, el libro defiende que ciertos mecanismos de mindshaping dan forma a nuestra mente de manera que nuestra conducta es más cooperativa y fácilmente interpretable para los demás. De este modo, algunos de los rasgos sociales característicos del ser humano dependieron evolutivamente de los mecanismos de mindshaping

**Palabras-clave:** Mindshaping, cooperación, lectura de mentes, evolución, lenguaje, cognición social.

**Abstract:** The aim of this paper is to expound and critically analyse TadeuszZawidzki's book *Mindshaping*. The book presents an alternative account of social cognition according to which the linchpin of our socio-cognitive capacities is the psychological mechanism known as mindreading. On the contrary, the book defends that certain psychological mechanisms shape our minds in order to make our behaviour more cooperative and easily interpretable. Therefore, unique human social capacities depend evolutionarily on *mindshaping* mechanisms.

**Key-words:** Mindshaping, cooperation; mindreading; evolution; language; social cognition;

---

[1] Quiero agradecer los comentarios y consejos de Fernando Martínez Manrique y la financiación del proyecto FFI2011-30074- C02-01 del Gobierno de España

[2] (vfernandezcastro@gmail.com) Máster en Lógica y Filosofía de la ciencia (USC), Máster en Ciencia Cognitiva y Lenguaje (UB), actualmente es estudiante de Doctorado (FPI) en la Universidad de Granada. Principales áreas de investigación son la filosofía de la mente, el lenguaje y la psicología.

La habilidad del ser humano para interactuar y desenvolverse en entornos sociales no tiene parangón dentro del reino animal. El alto grado de cooperación, el aprendizaje, la sobre-imitación o la comunicación compleja son algunos de los rasgos distintivos del *síndrome socio-cognitivo humano*. La cuestión es elucidar cómo emergieron estos rasgos distintivos en nuestra historia evolutiva y cuáles de esas características hicieron de trampolín para la adquisición de las demás.

## 1. La perspectiva Mindshaping

El síndrome socio-cognitivo humano tiene cuatro pilares básicos: la cooperación, el lenguaje, la lectura de mentes y la capacidad para aprender y enseñar patrones de conducta y raciocinio (*mindshaping*)<sup>3</sup>. De acuerdo con *el punto de vista estándar*, la lectura de mentes es el eje central de estas capacidades, es decir, es la habilidad sobre la que descansan las otras tres. Sin ser capaces de atribuir estados mentales a otros agentes para explicar su comportamiento, no podríamos cooperar, comunicarnos o aprender ciertas normas y patrones culturales. Por el contrario, *Mindshaping: A New Framework for Understanding Human Social Cognition* de Tadeusz Zawidzki pretende mostrar que la importancia de la lectura de mentes ha sido sobrestimada. La lectura de mentes no sólo no es el eje central de nuestra cognición social sino que, además, depende ontogenética y filogenéticamente de contextos sociales donde una serie de mecanismos dan forma a la mente (*mindshaping mechanisms*) para mejorar nuestra capacidad cooperativa y hacernos más fácilmente interpretables. O dicho de otro modo, los seres humanos están biológica y ecológicamente predispuestos a aprender y enseñar normas y patrones de conducta racionales que facilitan la cooperación y los proyectos complejos conjuntos, además de hacer nuestro comportamiento más transparente para los demás agentes.

La obra de Zawidzki se divide en siete capítulos en los que se desarrollan distintos argumentos a favor de la idea de que los mecanismos de *mindshaping* son el eje central de nuestras habilidades socio-cognitivas, además de una gran cantidad de evidencia empírica de diferentes ámbitos como la psicología del desarrollo o la antropología. En los tres primeros capítulos se nos presenta la visión del *mindshaping* como eje, una taxonomía de los diferentes mecanismos de *mindshaping* y los principales argumentos contra la visión estándar. En el resto del libro, Zawidzki explica cómo los mecanismos de *mindshaping* jugaron un papel fundamental en la adquisición de nuestra

---

[3] El término *Mindshaping* puede traducirse como dar forma a la mente o moldear la mente, sin embargo, a falta de una expresión más compacta para referirme a esta capacidad he preferido mantener el término en inglés.

conducta cooperativa, el lenguaje y la lectura de mentes sofisticada. La teoría de Zawidzki podría enmarcarse dentro de una corriente de visiones sobre la cognición social que hacen un especial énfasis en el carácter normativo de las interacciones sociales. Esta visión, que podría rastrearse hasta Davidson o Dennett<sup>4</sup>, ha tenido una nueva ola de defensores con autores como Mamelli, McGeer o el propio Zawidzki<sup>5</sup>.

El principal aporte de la teoría de Zawidzki con respecto a las posiciones clásicas de Dennett o Davidson es su énfasis en el aspecto de la adquisición y regulación de patrones de comportamiento y normas de conducta. Esto hace que la generación de expectativas de conducta a partir de la presunción de racionalidad no sea una mera ficción útil, sino que dichas expectativas funcionan porque los seres humanos tenemos mecanismos psicológicos que nos ayudan a seguir esos patrones y normas de conducta que concuerdan con dichas expectativas. Por otra parte, aunque aprender y seguir dichas normas exija en cierto sentido entender la conducta, esto no implica que necesitemos recurrir a la lectura de mentes. Zawidzki defiende que es suficiente con una habilidad de bajo nivel que se caracteriza en términos de lo que se denomina *estrategia intencional*. La estrategia intencional es la capacidad de percibir la conducta como patrones abstractos que se atienen a pautas de racionalidad. Esta capacidad para entender la conducta contrasta con la lectura de mentes sofisticada que Zawidzki liga a la adquisición del lenguaje y que define como la capacidad de atribuir actitudes proposicionales, esto es, estados mentales internos comprendidos como nexos causales no observables y duraderos que postulamos para explicar el comportamiento<sup>6</sup>.

La teoría de Zawidzki no sólo aporta una perspectiva diferente a la hora de entender la evolución de nuestras capacidades sociales, sino que presenta algunas ventajas importantes. Por un lado, recupera las ventajas de las teorías normativas clásicas como la de Dennett o Davidson: explicar las restricciones normativas de nuestras prácticas sociales. Pero además, al cambiar el énfasis desde la predicción hacia la perspectiva del aprendizaje y regulación, está bien posicionada para evitar algunos de los problemas que estas teorías presentaban. Por poner un ejemplo, la teoría de Dennett puede ser acusada de ser demasiado racionalista y no poder dar cuenta de casos donde predecimos la

---

[4] La exposición clásica de este punto de vista puede verse en Davidson, D.: *Inquiries into truth and interpretation*, Oxford, Clarendon, 1984. Y Dennett, D.: *The intentional stance*, Cambridge, MIT Press, 1987.

[5] Por ejemplo, McGeer, V.: "The Regulative Dimension of Folk-Psychology" en Hutto, D. Y Ratcliffe, M.: *Folk-psychology Reassessed*, Springer, 2007, p.137-56. Y Mamelli, M.: "Mindreading, mindshaping, and evolution" en *Biology and Philosophy*, 16, 2001, p. 597-628.

[6] Aunque Zawidzki no menciona cuál de las dos capacidades daría cuenta de los estados no proposicionales como los estados perceptivos y emocionales, parece que estos estarían ligados a la estrategia intencional

conducta de los demás aunque estos no sigan patrones de racionalidad. Sin embargo, como explica McGeer esto se soluciona cuando atendemos a estos casos y nos damos cuenta de que reaccionamos a ellos con conductas de regulación. Es decir, en esos casos, instamos a los demás a que se comporten con respecto a las normas.

## 2. Cognición Social y Lenguaje

En esta nota crítica me ocuparé principalmente de dos aspectos centrales para la tesis del libro. Por un lado, las consideraciones de Zawidzki sobre la relación entre cognición social y lenguaje. Por otro lado, el rol de lo que Zawidzki denomina la lectura sofisticada de mentes. En los capítulos 5 y 6 Zawidzki aborda la emergencia de la comunicación compleja en el marco de su teoría. Antes de presentar su visión, se consideran diversos modos de evitar los argumentos a favor de las posiciones neo-griceanas del lenguaje que consideran que la comunicación lingüística necesita de la capacidad de atribuir intenciones comunicativas y por tanto, de lectura de mentes. En primer lugar, se señala la implausibilidad debido a los costes cognitivos. Además, la capacidad de un hablante para hacer manifiestas sus intenciones comunicativas e informativas puede ser entendida desde la estrategia intencional. Cuando un hablante escoge ciertas oraciones o palabras para hacer manifiestas sus intenciones comunicativas no está guiado por hipótesis sobre los estados mentales de su audiencia sino por expectativas basadas en la presunción de racionalidad sobre cómo su audiencia *debería* reaccionar en dicho contexto. Por otro lado, Zawidzki intenta reinterpretar la evidencia empírica de que niños con trece meses de edad pueden pasar versiones no verbales del test de falsa creencia, lo que implicaría que los niños pre-lingüísticos poseen el concepto de creencia. De acuerdo con Zawidzki, esta interpretación presupone dos cosas. Primero, que el ser capaz de atribuir creencias falsas es condición suficiente para poseer el concepto de creencia. Segundo, que dicha evidencia empírica no es interpretable de otro modo. Estas presuposiciones tienen sus problemas. Por un lado, el concepto de creencia parece exigir algo más que ser capaz de atribuir creencias falsas, por ejemplo, ser capaz de dar cuenta de la *intensionalidad*. Por otro lado, existen diferentes modos de interpretar la evidencia empírica. Podría argumentarse que la razón por la que los niños pasan más tiempo observando el escenario de falsa creencia es porque viola sus expectativas de racionalidad.

Los argumentos de Zawidzki parecen convincentes contra la posición neo-griceana estándar. Parece poco plausible que agentes sin capacidad lingüística puedan tener lectura de mentes, especialmente si tenemos en cuenta la evidencia empírica que muestra la relación entre ciertas capacidades lin-

güísticas y el éxito para pasar el test de falsa creencia<sup>7</sup>. Sin embargo, no creo que de estos argumentos se siga necesariamente la posición de *mindshaping*. En principio se podría optar por defender una co-evolución entre lenguaje y lectura de mentes sin que eso suponga abandonar el marco completo de la posición estándar. Por ejemplo, podría defenderse que para que el lenguaje pudiera emerger basta con que el agente posea un mecanismo de lectura de mentes mínima que ayudara a rastrear el significado del hablante sin necesidad de una lectura de mente sofisticada. Después, el lenguaje enriquecería a su vez la lectura de mentes mínima para hacer emerger una atribución de estados mentales de nivel alto gracias a las capacidades representacionales del propio lenguaje. En esta línea, Bermúdez<sup>8</sup> defiende que podemos distinguir en el reino animal entre estas posibilidades, siendo el ser humano como animal lingüístico el único capaz de atribuir actitudes proposicionales gracias a las capacidades estructurales del lenguaje. Por tanto, parece que para hacer volcar la balanza a favor de la visión de *mindshaping* frente a la posición co-evolutiva, deberíamos analizar si la estrategia intencional explica mejor nuestra capacidad para rastrear las intenciones comunicativas de los hablantes que otro tipo de mecanismo de predicción como una lectura de mentes mínima. Esta último mecanismo es interesante porque uno de los principales argumentos de Zawidzki contra la visión neo-griceana de Sperber y Wilson<sup>9</sup> es que fallaría para dar cuenta de la automaticidad de nuestras conductas lingüísticas. Y es concretamente la automaticidad de algunas de nuestras prácticas de lectura de mentes lo que motivó la posición de lectura de mentes mínima de Butterfill y Apperly.

Veamos el caso con más detenimiento. De acuerdo con Sperber y Wilson, la razón por la que la lectura de mentes es necesaria para dar cuenta de la comunicación es que para que esta sea exitosa se necesita que el oyente reconozca la intención comunicativa e informativa del hablante. Es decir, la intención de hacer manifiesta para la audiencia un conjunto de asunciones (informativa) y la intención de hacer mutuamente manifiesta entre el hablante y la audiencia que el hablante tiene dicha intención informativa (comunicativa). Sperber y Wilson consideran que dado el principio cognitivo de relevancia, un ítem de información es relevante cuando minimiza la distancia entre el estado del oyente y su objetivo epistémico o pragmático. Cuanto más ayude el ítem

---

[7] De Villiers, J.G. y De Villiers, P.A.: "Language for Thought: coming to understand False Beliefs" en Gentner, D. y Goldin-Meadow, s.: *Language in Mind: Advances in the Study of Language and Cognition*, MIT press, 2003, p. 335-284.

[8] Bermúdez, J.L.: "Mindreading in the animal kingdom" en Lurz, W.R.: *The Philosophy of Animal minds*, MIT Press, 2009, p. 145-164.

[9] Para una exposición de la teoría ver Sperber D. Y Wilson, D.: *Relevance: Communication and Cognition*, Oxford, Blackwell. 1986/1995. Para ver cómo la comprensión lingüística puede verse otros en la misma lde un agente si necesidad de representarlos explícita está relacionada con la cognición social y la hipótesis de la modularidad ver Sperber, D. Y Wilson, D.: "Pragmatics, modularity and mindreading", *Mind and Language*, 17" 2002, p. 3-23.

a llevar a cabo el objetivo, más relevante es. Suponiendo que todos tenemos como objetivo comunicarnos de modo exitoso, el hablante solo necesita emitir una señal relevante para dicho objetivo, es decir, ayudar a la audiencia a comprender las intenciones del hablante. La razón por la que Sperber y Wilson consideran que una lectura de mentes es necesaria se debe a que el hablante solo puede escoger los ítems adecuados para una audiencia si tiene en cuenta los objetivos, creencias, motivos o trasfondo cognitivos de esta audiencia. Contra esto, Zawidzki motiva su propuesta con dos argumentos. En primer lugar, la visión de Sperber y Wilson implica un gran coste cognitivo para llevar a cabo una práctica, la comunicación, que en principio parece muy automatizada. En segundo lugar, la estrategia intencional podría dar cuenta de la comunicación. Como hemos dicho antes, el hablante podría escoger sus señales en base a expectativas guiadas racionalmente, sin necesidad de incluir estados meta-representacionales. Además, los mecanismos de *mindshaping* mitigarían el problema de contextos en los que la audiencia es reacia a la comunicación o que tienen objetivos diferentes. Si estamos socializado para ser motivados antes los mismos contextos, cooperar y usar heurísticas similares, el esfuerzo para involucrarnos en la comunicación sería mínimo.

El problema es que estos dos argumentos pueden ser usados en principio para defender que la comunicación está guiada por un mecanismo de lectura de mente mínima. Esto no sólo serviría para dar cuenta de los dos motivaciones de Zawidzki contra la posición de Sperber y Wilson, sino que nos permitiría mantener las ventajas explicativas de la versión estándar. Un mecanismo de lectura de mentes mínimo es un mecanismo que puede registrar ciertos estados mentales de un sujeto sin necesidad de representar su contenido. Estos mecanismos pueden dar cuenta de la automatización que guía la comunicación. Además, posee la capacidad explicativa necesaria para dar cuenta de algunos de los test de falsa creencia en niños pre-lingüísticos que considera Zawidzki. Butterfill y Apperly consideran que existen varios mecanismos de lectura de mentes mínimos que dan cuenta de diferentes comportamientos sociales. Por otro lado, podríamos pensar que hay varios mecanismos involucrados a la hora de tener en cuenta componentes motivacionales y perceptivos de nuestra audiencia, lo que daría cuenta de los problemas de indeterminación que tanto Sperber y Wilson como Zawidzki consideran.

Otro posible modo que tiene Zawidzki de atacar la posiciones neo-griecanas con respecto al lenguaje supone alguna ventaja explicativa. Su alternativa es básicamente una teoría pragmatista al estilo de Brandom<sup>10</sup>. De acuerdo con su punto de vista, el lenguaje apareció como un mecanismo para señalar *compromisos*. Los primeros usos del lenguaje eran parecidos a promesas, es decir, señalaban compromisos con cursos de acción que involucran ob-

[10] Brandom, R.: *Making it Explicit. Reasoning, Representing, and Discursive Commitment*, Cambridge, Harvard University Press, 1994.

jetos salientes y propiedades. El señalar compromisos de modo honesto es importante para la cooperación y la coordinación, ya que sólo los individuos que realmente están comprometidos con el grupo serían capaces de señalar dichos compromisos a pesar de los costes. Este punto de vista tiene la ventaja de explicar la emergencia del lenguaje de una manera análoga a la emergencia del canto de los pájaros. Del mismo modo que los cantos de los pájaros se hicieron cada vez más complejos por la presión evolutiva de la selección sexual, el lenguaje humano se hizo cada vez más complejo por la presión de grupo para que las señales de compromisos fueran cada vez más sofisticadas. Además, la misma señal de compromiso funcionaría como un mecanismo de *mindshaping*, al obligar al agente a ser coherente con los patrones de acción legítimos que son compatibles con la señal.

Este argumento no es del todo concluyente si tenemos en cuenta que una visión del lenguaje de corte neo-griceana también está bien posicionada a la hora de dar una explicación evolutiva. En un entorno social como el que proponen los defensores de la *inteligencia maquiavélica*<sup>11</sup>, no es de extrañar que aparecieran individuos con mecanismos de lectura de mentes mínimas. Así, el lenguaje emergería con la función de facilitar la lectura de mentes mínima facilitando el reconocimiento de intenciones y la comunicación. Por tanto, si Zawidzki quiere convencernos de que su versión naturalizada de la teoría de Brandom es mejor que una posición como la de Sperber y Wilson, entonces necesita desarrollar más argumentos. Quizás una posible vía a explorar es intentar hacer suyos algunos de los argumentos contra las teorías basadas en intenciones.

En el debate sobre la naturaleza de los actos de habla existen varias posiciones. En la visión basada en intenciones como las que estamos comentando, los actos de habla se entienden en términos de intenciones comunicativas e informativas. Por otro lado, la visión normativa de los actos de habla defiende que nuestros actos comunicativos se fundan en convenciones sociales que se traducen en obligaciones, compromisos y derechos que adquieren los interlocutores. Los defensores de la teoría normativa han venido señalando varios problemas de la propuesta intencionalista. Por un lado, no parece correcto afirmar que el significado de nuestros actos de habla esté especificado por nuestras intenciones informativas y comunicativas. En muchos contextos comunicativos el significado de lo que queremos comunicar es más difuso. Esto se hace patente en fenómenos como la deferencia, las semi-citas o las constricciones sociales de nuestras interacciones comunicativas. Otro problema es que la visión intencionalista parece demasiado amplia, lo que implica tratar fenómenos como las implicaturas o las

---

[11] La exposición clásica de esta hipótesis se encuentra en Byrne, R. W. Y Whiten, A.: *Machiavellian intelligence: social expertise and the evolution of intellect in monkeys, apes and humans*, Oxford: Oxford University Press, 1988.

presuposiciones como casos de lo que se dice<sup>12</sup>. Según la visión intencionalista, cuando asevero que Pedro tiene una excelente caligrafía en una carta de recomendación estaría expresando la creencia de que Pedro es un mal investigador ya que es parte de mi intención al comunicar. Sin embargo, parece que esto no es parte de lo que digo sino de lo que queda implicaturado. Por último, como los defensores de la visión normativa<sup>13</sup> han venido defendiendo, nuestros actos comunicativos implican la instanciación de ciertos compromisos, derechos y obligaciones de los que no se puede dar cuenta sólo en términos intencionales. Por ejemplo, para dar una orden a alguien necesito estar en cierta posición social para que mi orden sea exitosa, si no, la orden no cumplirá su cometido aunque yo emita una oración con la intención de dar dicha orden.

Aunque este es un debate abierto, este tipo de argumentos fortalecerían el enfoque Zawidzki al mostrar porqué una teoría del compromiso está mejor posicionada para dar cuenta de algunos de estos problemas. En cualquier caso, lo que parece inevitable es que Zawidzki necesita hacer frente a estos dos problemas para convencernos de su visión con respecto al lenguaje y la necesidad de los mecanismos de *mindshaping* a la hora de abordar la emergencia del lenguaje. Por una lado, debe tener en cuenta que la visión estándar puede rescatarse atendiendo a una posición co-evolutiva, donde mecanismos de lectura de mentes mínima produjeran la emergencia del lenguaje, y este sirviera como trampolín para una lectura de mentes más sofisticada. Por otro lado, si quiere favorecer su posición en términos de compromisos, debe dar argumentos más potentes que el de la evolución mediante selección de grupo.

### 3. La aparición de la lectura de mentes sofisticada

Finalmente, voy a analizar otra de las partes que considero centrales del libro de Zawidzki, y es su posición sobre la lectura de mentes sofisticada. En el último capítulo, se nos presenta cómo las habilidades *mindshaping* explican la adquisición de esta lectura de mentes sofisticada. Según Zawidzki, las actitudes proposicionales son estados mentales comprendidos como nexos causales no observables y duraderos que postulamos para explicar el comportamiento. Este punto de vista contrasta con la estrategia intencional, la cual no necesita postular ninguna realidad mental oculta tras la apariencia del comportamiento. Sin embargo, los seres humanos, frente a otros seres sociales, parecen ser esencialistas y categorizar ciertos eventos como siendo causados por entidades no observables. Por tanto, Zawidzki debe explicar por qué esta

---

[12] Este argumento y otros en la misma línea puede verse en Mcfarlane, J.: “What Is Assertion?”, en Brown, J. Y Cappelen, J.: *Assertion*, Oxford, Oxford University Press, 2011, p. 79-96.

[13] La versión normativa de los actos de habla puede verse defendida en Sbisà, M.: “Communicating citizenship in verbal interaction”, en Hausendorf, H. y Bora, H.: *Analysing Citizenship Talk: Social positioning in political and legal decision-making processes*, 2006, p. 151–180.



visión esencialista apareció a pesar de la eficiencia de la estrategia intencional. Su argumento es que la atribución de actitudes proposicionales apareció con una función social (no epistémica) asociada al lenguaje. Como ya he señalado, los primeros usos del lenguaje eran parecidos a promesas, es decir, señalaban compromisos con cursos de acción que involucran objetos salientes y propiedades. Esto es lo que Brandom denomina compromisos doxásticos. Esto tiene una ventaja social clara, podemos informarnos sobre hechos de manera indirecta a otros miembros del grupo que no tienen acceso a ellos. Por otro lado, alguien que expresa compromisos doxásticos no solo está comprometido con ciertos patrones de acción sino que está abierto a sanción. Aseverar que hay una manada de presas al norte del río habilita a la audiencia a verificar la aseveración y sancionar al hablante en caso de que sea falsa. Este tipo de prácticas básicas facilitarían los proyectos cooperativos y se irían refinando hasta crear señales de compromisos más complejas.

Una vez que ha aparecido un sistema comunicativo complejo en términos de señalar compromisos, es posible entender las actitudes proposicionales de alto nivel en estos términos: Creer que P no es más que exhibir un compromiso doxástico con P. Sin embargo, en el momento en que tenemos una población capaz de expresar compromisos y racionalizar cursos de acción de otros, es normal que aparezcan conflictos. Entonces es cuando los intérpretes deben postular una distinción entre apariencia de la conducta y realidad mental. En el momento en que dos interpretaciones de un comportamiento entran en conflicto se necesita de esta distinción para mitigar el conflicto. En este sentido, la atribución de actitudes proposicionales de alto nivel (como nexos causales y holísticamente restringidos) tiene una función social, la de justificar un comportamiento que de otro modo estaría abierto a sanción por ser percibido como anómalo.

Una de las cosas que más me llaman la atención de este último capítulo del libro es la defensa de la idea de que los seres humanos necesitan postular estados internos con poderes causales para dar cuenta de la adscripción de actitudes proposicionales. En este punto parece que se puede optar por dos posiciones. Una primera posición podría ser defender que esta visión realista de los agentes no es más que un modo de hablar. Siguiendo a Wittgenstein podríamos defender que cuando los agentes hablan como si las creencias y los deseos fueran estados internos de un sujeto, solamente están proyectando sombras metafísicas al confundir algunos de nuestros usos de estos conceptos. Desde este punto de vista, y en línea con la posición brandomiana, podría defenderse que adscribir un creencia es adscribir un conjunto de compromisos y que la tendencia a hablar en términos esencialistas de los agentes no es más que una metáfora que no añade nada a nuestra explicación de la conducta de los demás. La segunda posición consiste en defender que, aunque nuestra adquisición del vocabulario mental se hace en términos de compromisos discursivos, la función de nuestros conceptos mentales es en última instancia es referirnos a estados

internos del sujeto. Ese esencialismo es algo más que una metáfora y tiene una función explicativa que no es sustituible.

Esta última opción es la que parece defender Zawidzki cuando defiende que las tensiones entre las atribuciones de primera y tercera persona nos exigen que haya una distinción entre apariencia de la conducta y realidad mental. El argumento parece defender que el realismo intencional es necesario para mitigar las tensiones entre la explicación en primera y tercera persona. Por ejemplo, cuando damos una explicación en tercera persona de un curso de comportamiento, digamos “Juan corre porque cree que el edificio está en llamas”. Y luego Juan nos dice en primera persona. “corría porque había olvidado un papel importante”, tenemos una tensión entre ambas explicaciones. Esta tensión parece ser la que motiva el realismo intencional que exhiben las atribuciones de los interpretes, ya que sin él, no se podría distinguir entre la causa del comportamiento aparente y la real. Sin embargo, este movimiento no me parece del todo convincente. Las tensiones entre las atribuciones de primera y tercera persona no exigen que haya una distinción entre apariencia de la conducta y realidad mental. Si miramos a los contextos cotidianos, solemos privilegiar la explicación de primera persona, a menos que esta sea incoherente o pueda ponerse en duda. Por tanto, siempre y cuando podamos dar cuenta de por qué dotamos de autoridad a la primera persona sin apelar a estados internos a los que accedemos, no necesitamos los postulados de los que habla Zawidzki. Es decir, si mantenemos que nuestra atribución de primera persona tiene cierta autoridad y explicamos dicha autoridad sin apelar al realismo intencional, entonces no tenemos que comprometernos con la idea de que nuestras prácticas sociales necesitan de un postulado de estados internos con poderes causales. En el mercado filosófico podemos encontrar varias teorías en esta línea. Por poner un ejemplo, de acuerdo con McGeer<sup>14</sup>, la autoridad de primera persona se basa en la responsabilidad que el agente adquiere con sus propias emisiones al ser estas expuestas a sanción pública. En el momento en que un hablante tiene que ser coherente con lo que se sigue de sus emisiones en primera persona, entonces regula su propio comportamiento de acuerdo con ellas. Es por esto por lo que las explicaciones en primera persona están dotadas de autoridad, porque se presupone que regulamos nuestro comportamiento de acuerdo a las expectativas que dichas explicaciones nos imponen. En este sentido, no tenemos por qué postular una realidad interna que mitigue las tensiones entre primera y tercera persona.

A lo largo de esta exposición he intentado mostrar algunos de los puntos débiles del libro de Zawidzki. A pesar de sus problemas, el libro muestra de manera convincente una hipótesis arriesgada que es, incluso para sus detractores, una herramienta interesante para repensar algunos problemas clásicos

---

[14] McGeer, V.: “The moral development of First-Person Authority”, *European Journal of Philosophy*, 16(1), 2007, p. 81-108.