TESIS DOCTORAL

La Importancia de la Segunda Persona en la Constitución del Autoconocimiento: Un Enfoque Expresivista

The Significance of the Second Person in the Constitution of Self-Knowledge: An Expressivist Approach

Doctorando

José Ferrer de Luna

Director

Dr. Manuel de Pinedo García

Programa de Doctorado en Filosofía

Departamento de Filosofía I

Universidad de Granada

Editor: Universidad de Granada. Tesis Doctorales
Autor: José Ferrer de Luna
ISBN: 978-84-1306-236-5
URI: http://hdl.handle.net/10481/56222

A Clarita

Agradecimientos

"Persons are *second* persons,
who grow up with other persons"
—Annette C. Baier.

Todos tenemos, o deberíamos tener, una segunda persona, una persona que forma parte de nosotros, una persona que llevamos y llevaremos dentro a lo largo de nuestra vida. En mi caso es Clarita, mi madre, a quien debo no solo la vida sino también los principios por los que en ella me guio, así como el valor y la responsabilidad necesarios para no dejar de ser yo mismo. A ella va dedicada esta tesis, a Clarita, mi segunda persona, la persona que nunca dejaré de llevar en mi interior.

Esta tesis tampoco hubiera sido posible sin la ayuda de otras muchas personas que a lo largo de mi carrera universitaria han estado siempre a mi lado, apoyándome, confiando en mí, formando parte de ella. En primer lugar quiero agradecer a mi director, Manuel de Pinedo, no solo el haber confiado en mí a la hora de llevar a cabo este proyecto sino también por haberlo hecho en todos y en cada uno de los que he decidido llevar a cabo desde que, en tercero de carrera, tuve la suerte de tenerlo como profesor. Gracias Manolo por tu amistad y por el apoyo que siempre me has brindado tanto en lo profesional como en lo personal. Sin tu ayuda, consejos e indicaciones no hubiera siquiera comenzado a escribir esta tesis. Tu envidiable rapidez mental, tu entusiasmo por la filosofía, tu capacidad para situarte a cualquier nivel, han sido siempre una guía y un ejemplo para mi. Gracias también por las interminables horas que has debido pasar leyendo y corrigiendo cada versión de esta tesis que, ciertamente, no han sido pocas. Otra de las personas fundamentales durante todo mi recorrido universitario ha sido Cristina Borgoni, la mejor profesora de la que uno pueda disfrutar. Gracias Cristina por creer en mí y ayudarme a saber desarrollar mis ideas, a saberlas expresar, refinar, escribir, y por recordarme que son esas y no otras las que he de defender, aunque eso vaya en contra de lo establecido. Gracias también por tus maravillosas clases y tu confianza a la hora de recibirme como visitante en la Universidad de Graz. Gracias a tu dedicación, a tus comentarios, a tus sugerencias, en definitiva, gracias a ti allí comencé a escribir esta tesis. I would also like to thank my fantastic supervisor Naomi Eilan for her dedication and kindness while I stayed in Warwick under her supervision. Naomi, your philosophical orientation was not only illuminating but also decisive in my development within Philosophy of Mind. Thanks for your invaluable comments and

trust in my ideas and for the countless conversation hours that I was fortunate enough to share with you. Y cómo no, gracias también a mi gran referente intelectual y académico, Juan José Acero. Gracias profesor Acero, sin su apoyo, enseñanzas y confianza esta tesis tampoco hubiera sido posible. Gracias por enseñarme a enfrentarme a cualquier reto, por tener la capacidad de hacer fácil lo difícil y conseguir que, ya en tercero de carrera, Frege y Wittgenstein dejarán de ser infranqueables para mí.

Además de estas grandes maestras y maestros también debo un agradecimiento especial a mi hermanito pequeño durante mis años de carrera, máster y doctorado, Nemesio García-Carril Puy, a quien debo mucho desde el punto de vista personal y profesional; a Noé Expósito, por ser mi compañero de andanzas tanto en lo académico como en lo laboral y lo lúdico; a Alejandro Armenteros, por sus acertadas sugerencias sobre la gramática inglesa y por su vasto conocimiento enológico, al que he recurrido en más de una ocasión durante el desarrollo de esta tesis; a José Ramón Torices, por las interminables horas al teléfono aguantando mis elucubraciones; a Pedro García, por los fantásticos ratos de charlas personales y filosóficas y tu introducción a la gastronomía granadina; a Manuel Heras, por su valioso apoyo en los momentos clave; a Víctor Fernández, por su envidiable actitud filosófica y capacidad cervecera; a Mirco Sambrotta, por los buenos ratos juntos; a todos los miembros del Gang: David, Andrés, Liñán, Palma, Francesco, Dani, Manu, Xavi, Llanos, Amalia, Eduardo, Alba, es un placer teneros como compañeros.

Quiero también expresar mi agradecimiento, asimismo, a los demás profesores que, de una u otra manera, han tenido que ver con la realización de esta tesis. A Esther Romero, Fernando Martínez, Mª José Encinas, Henrik Zinkernagel, Mª José Frápolli y Neftalí Villanueva, gracias a todos por formar parte de mi vida académica como profesores durante mis años de carrera y como compañeros en los de doctorado. También quiero agradecer a Mª José Alcaraz, de la Universidad de Murcia, su amabilidad y disponibilidad a la hora de comentar algunos de mis trabajos. Gracias Mariajo por tus acertados comentarios.

Ni esta tesis ni nada en mi vida hubiera sido tampoco posible sin mis hermanos. Alicia, Gerardo, Eduardo, sois los mejores hermanos que una persona puede tener. Edu, sin tu altruista ayuda no hubiera siquiera acabado la carrera; Alicia, sin tu cariño hubiera sido mucho más difícil seguir adelante; Gerardo, ¿qué hubiera sido de mí sin tus enseñanzas farmacognósticas? Gracias también por diseñar la portada de esta tesis, siempre has sido el artista de la familia. Gracias a todos por confiar siempre en mí, por

estar a mi lado, por respetarme y quererme como siempre lo habéis hecho. Tampoco nada hubiera sido posible sin mis otros hermanos. Oscar, Luis, sin vosotros nada tendría sentido. Miguelito, gracias por haber sido y seguir siendo mi maestro. Nando, la distancia de las moradas no separa el cariño de los corazones.

Por último, debo un agradecimiento a los proyectos de investigación *Naturalismo, expresivismo y normatividad: FFI2013-44836-P* y *Expresivismos contemporáneos y la indispensabilidad del vocabulario normativo: alcance y límites de la hipótesis expresivista: FFI2016-80088-P,* por posibilitar mi asistencia a varios congresos nacionales e internacionales en los que he defendido, criticado y mejorado mis ideas, así como al Ministerio de Educación, Cultura y Deporte por financiar un contrato y dos estancias de investigación para la realización de esta tesis.

Publicaciones

Ferrer, J. (forthcoming). Intersubjectivity in infancy: a second-person approach to ontogenetic development. *Philosophical Psychology*

Ferrer, J. (2018). ¿Pueden los bebés comunicar y conectar sus experiencias con los adultos? *Ciencia Cognitiva 12*(2), 33–35. ISSN 1988-7884

Ferrer, J. (2016). Expressivism and Self-Knowledge: Second Person Relations and the Spokesperson. Actas del VIII Congreso de la Sociedad Española de Filosofía Analítica. Oviedo. KRK Ediciones, pp. 66-69. ISBN: 978-84-8367-547-2

Ferrer, J., Pinedo, M. (2015). Dos tipos de segunda persona. interiorizada. Actas del VIII Congreso de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España. Barcelona. Universidad de Barcelona, pp. 202-206. ISBN: 78-84-606-9303-1

Ferrer, J. (2014). El papel de la segunda persona en la constitución del autoconocimiento. *Daimon. Revista Internacional de Filos*ofía 62, 71–86. ISSN: 1130-0507

Índice
Contents

BLOQUE III

Resumen

El principal objetivo de esta tesis es ofrecer una explicación del autoconocimiento que tome como punto de partida la perspectiva de segunda persona y que recoja las intuiciones tanto del expresivismo clásico, como del neo-expresivismo y del expresivismo semántico.

La principal motivación de esta investigación surge del descontento con un presupuesto común a todos los enfoques tradicionales acerca de este fenómeno. Adoptar este presupuesto —que denomino "presupuesto individualista"— implica afrontar la investigación sobre el autoconocimiento partiendo de una dicotomía exclusiva y excluyente entre las perspectivas de primera y tercera persona, obviando, por tanto, la posibilidad de la existencia de otra perspectiva genuina y diferente desde la cual realizar el estudio. Partir de esta dicotomía no solo conduce a una visión parcial del fenómeno sino también a una serie de problemas que, como muestro en este trabajo, se dirimen al adoptar un enfoque que toma como punto de partida la perspectiva de segunda persona.

Según el enfoque expresivista de segunda persona objeto de esta investigación, el estudio del fenómeno del autoconocimiento no debe realizarse de manera aislada, sino que debe contar asimismo con una explicación tanto del desarrollo ontogenético de la mente como de nuestra capacidad para la auto-atribución de estados mentales y la atribución de estados mentales a los demás. Con objeto de profundizar en cada uno de estos temas, divido la presente investigación en tres grandes bloques; sin que ello implique una desconexión entre ellos. Dado que el presupuesto individualista se encuentra presente no solo en los estudios acerca del autoconocimiento, sino también en las investigaciones realizadas en los otros dos ámbitos, discuto las carencias y/o problemas que conlleva dicho presupuesto en cada uno de ellos.

El primer bloque tiene como objetivo mostrar la importancia de la perspectiva de segunda persona en el estudio del desarrollo ontogenético de los bebés, así como en la constitución de la conciencia de sí mismos y de los demás, es decir, de las perspectivas de primera y tercera persona. En este bloque sostengo que el enfoque de segunda persona ofrece una explicación más plausible del desarrollo de los bebés en este período que los enfoques alternativos que parten de la dicotomía entre la primera y la tercera persona. En primer lugar, defiendo la existencia de una comunicación expresiva mediante la cual los bebés son capaces de conectar y comunicar sus

experiencias subjetivas con los demás, es decir, de participar en interacciones intersubjetivas. Para fundamentar esta comunicación expresiva, ofrezco una explicación de la noción de expresión compatible con la misma, así como con la disolución de la dicotomía entre primera y tercera persona. Según esta noción, las expresiones son constitutivas de los estados mentales, en concreto, son su condición comunicativa.

De otra parte, sostengo que la perspectiva de segunda persona no solo es anterior en el desarrollo ontogenético, sino que juega un papel constitutivo en la emergencia de las otras dos perspectivas. Finalmente, presento un modelo de comunicación reguladora mutua, según el cual los bebés conciben a la persona que está a su cuidado como una autoridad —que denomino autoridad de segunda persona. Defiendo que mediante esta autoridad los bebés extienden la regulación emocional que tiene lugar en sus interacciones diádicas, a una regulación normativo-social en sus interacciones triádicas a través de los triángulos de atención conjunta.

El segundo bloque está dedicado a la presentación y crítica de las diferentes propuestas que se han llevado a cabo sobre el fenómeno del autoconocimiento. De acuerdo las similitudes inherentes a estas propuestas, distingo tres tipos de enfoques: el introspeccionista, el agencialista y el neo-expresivista. En el primer apartado, dedicado al enfoque introspeccionista, señalo que la explicación que ofrecen estos autores resulta diametralmente opuesta a la ofrecida por el enfoque de segunda persona. Según los introspeccionistas, las interacciones con los demás no tienen relevancia alguna en el estudio del fenómeno del autoconocimiento. El autoconocimiento proviene, a su entender, de un acceso privilegiado del individuo a sus propios estados mentales. Tras la presentación de sus propuestas, expongo las principales críticas a las mismas destacando entre ellas la dificultad para ofrecer una explicación de la existencia de estados mentales en los demás.

En el apartado dedicado a los agencialistas sostengo que aunque estos autores sí incluyen las interacciones en sus propuestas, lo hacen desde una perspectiva teórica y centrada en el individuo. Este hecho conduce a una comprensión parcial del fenómeno del autoconocimiento y supone un problema a la hora de dar cuenta del comportamiento que, de hecho, tenemos los seres humanos. La comprensión parcial de este fenómeno conlleva la ausencia del elemento relacional en su explicación de las nociones de razón, responsabilidad, compromiso y autoridad, así como la ausencia del tratamiento de la normatividad social en sus enfoques. Por ello, presento una explicación relacional de estas nociones que elimina las carencias a las que se enfrenta la explicación parcial de

las mismas. Asimismo, muestro la necesidad de la inclusión de la normatividad social en la explicación del fenómeno del autoconocimiento.

Finalmente, en el apartado dedicado a los neo-expresivitas sostengo que el presupuesto individualista conduce, en su caso, a una visión parcial del proceso de adquisición de las (auto-)atribuciones de estados mentales. Estos autores centran su explicación exclusivamente en la capacidad para la auto-atribución de estados mentales, obviando su relación con la capacidad para la atribución de estados mentales a los demás. Por el contrario, de acuerdo con el enfoque expresivista de segunda persona, ambos capacidades se adquieren simultáneamente. Este hecho, sostengo, resulta crucial para una correcta comprensión de dicho proceso. Asimismo, su explicación sobre el autoconocimiento se extiende a todos los estados mentales sin distinción. Defiendo que la no diferenciación de los distintos tipos de estados mentales conduce a una visión errónea del mismo. Finalmente, suscribo la validez de algunos de los argumentos sobre la transparencia de los estados mentales, en concreto los referentes a la noción de la transparencia-a-la-condición-mental-del-sujeto.

En el tercer y último bloque analizo las propuestas de los enfoques tradicionales acerca de la Folk Psychology, así como las propuestas actuales que rechazan la explicación de estos enfoques. En este bloque ofrezco, asimismo, una explicación del autoconocimiento desde el enfoque expresivista de segunda persona. Al igual que en los estudios acerca del desarrollo ontogenético y del autoconocimiento, los enfoques tradicionales sobre la Folk Psychology parten de la dicotomía entre primera y tercera persona. En su caso, este presupuesto ha separado la investigación en dos tipos de teorías contrapuestas, cada una de ellas fundamentada en uno de los dos polos de la dicotomía. De una parte, los defensores de la Teoría de la Teoría desarrollan su enfoque partiendo de la perspectiva de tercera persona. De otra parte, los defensores de la Teoría de la Simulación lo hacen desde la perspectiva de primera persona. En el apartado dedicado a estas teorías, sostengo que el mayor problema al que se enfrentan radica en el presupuesto individualista que ambas comparten. Según este presupuesto, los demás son seres a comprender, por lo que la principal función de la Folk Psychology es, para estos autores, la de predecir y explicar la conducta de los demás.

En el siguiente apartado, analizo los distintos argumentos que contra este supuesto se han llevado a cabo desde algunos de los enfoques actuales, suscribiendo la validez de los mismos y ofreciendo una explicación de la Folk Psychology desde el enfoque de segunda persona. Este enfoque entiende la Folk Psychology no como una

teoría, sino como una práctica normativo-regulativa cuya principal función no es la de predecir y explicar la conducta sino una función regulativa consistente en aprender, enseñar y exhortar a otros a comportarse de acuerdo a las normas sociales compartidas que rigen las interacciones de los miembros de la comunidad.

Para concluir, ofrezco una explicación del autoconocimiento que recoge lo defendido tanto acerca del desarrollo ontogenético, como acerca de la Folk Psychology. Sostengo que la transparencia de lo mental no lo es solo hacia los propios estados mentales sino también hacia los estados mentales que los demás nos comunican de manera expresiva durante las interacciones, es decir, mediante la comunicación expresiva que tiene lugar a través de la perspectiva de segunda persona. De acuerdo a esta definición de la transparencia, argumento la existencia de dos tipos de conocimiento: el conocimiento epistémico, en sentido tradicional, y el conocimiento pre-conceptual, entendido este último como el conocimiento con el que contamos ya desde nuestras primeras semanas de vida. Sostengo que este tipo de conocimiento, lejos de desaparecer una vez introducidos en el lenguaje y, por tanto, una vez poseemos los conceptos necesarios para ello, continúa presente a lo largo de nuestra vida. En cuanto a la autoridad de primera persona, defiendo una concepción relacional de la misma según la cual esta está vinculada tanto con el conocimiento de nuestros estados mentales, como con la responsabilidad de hacer que nuestras auto-atribuciones de estados mentales coincidan con nuestros actos (con lo que, según las normas sociales y lingüísticas se espera de alguien que se auto-atribuye el estado mental en cuestión), así como con el reconocimiento por parte de los demás miembros de la comunidad. Finalmente, sostengo que el autoconocimiento incluye no sólo la perspectiva de primera o de tercera persona sino la perspectiva de primera, la de segunda y la de tercera persona.

Summary

The primary aim of this dissertation is to propose an explanation of self-knowledge that takes the second-person perspective as a starting point and that gathers the intuitions from classic expresivism, neo-expressivism and semantic expressivism.

The main motivation for this research stems from the discontent with an assumption shared by all traditional approaches to this phenomenon. To adopt this assumption–that I call 'individualistic assumption'–implies facing the research on self-knowledge beginning at an exhaustive and excluding dichotomy between the first and third person that ignores the possibility of the existence of another genuine and different perspective from which to conduct the study. Beginning at this dichotomy not only leads us to a partial vision of the phenomenon, but also to a series of problems that, as I show with my work, are settled by adopting an approach that takes as its starting point the second-person perspective.

Under the second-person expressivist approach–the object of this study–, the analysis of the phenomenon of self-knowledge should not be carried out in an isolated manner. It should be accompanied by an explanation of both the ontogenetic development of mind and the human capacity for self-attribution of mental states and the attribution of mental states to others. In order to examine each of these topics in depth, I divide the present inquiry within three large blocks; in no way implying a discontinuity between them. Given that the individualist assumption is present, not only in the studies on self-knowledge, but also in the research carried out in the other two areas, I discuss the shortcomings and/or problems that this assumption entails.

The first block is devoted to the relevance of the second-person perspective both in the ontogenetic development of infants and in the constitution of their self-awareness and the awareness of others, i.e., of the first- and the second-person perspectives. In this block I argue that the second-person approach allows for a more plausible understanding of infants' development during this period than the alternative approaches that assume the dichotomy between the first and the third person. First, I sustain the existence of an expressive communication through which infants are able to connect and communicate their subjective experiences with others, i.e., of participating in intersubjective interactions. I bring forward and explanation of the notion of expression compatible with this expressive communication as well as with solving the

dichotomy between the first and the third person. According to this conception, the expressions are constitutive of the very mental states, specifically, they are their communicative condition. Moreover, I argue that the second-person perspective not only precedes the ontogenetic development of the other two perspectives, but that it also plays a constitutive role in their emergence.

Finally, I present a mutual communicative regulation model according to which infants conceive their caregivers as an authority–which I call second-person authority. I sustain that it is through this authority that infants extend the emotional regulation that takes place within their dyadic interactions over a socio-normative regulation in their triadic interactions by manner of the joint-attention triangles.

In the second block I present and criticize the different proposals that have been brought forward regarding the self-knowledge phenomenon. Following the inherent similarity within these proposals, I distinguish three different approaches: introspectionist, agencialist and neo-expressivist. In the first section, dealing with introspectionism, I point out that the explanation that these authors present opposes radically the one offered by the second-person approach. Introspectionists affirm that interactions with others have no relevance whatsoever in the study of self-knowledge. Self-knowledge arises, they argue, from a privileged and exclusive access to their inner mental states that individuals have. Once their proposals are unfolded, I detail the primary critiques highlighting among them the difficulty to produce an explanation of the existence of others' mental states.

In the section dealing with agentialism, I argue that, although these authors do include interactions in their proposals, they do it from a theoretical perspective and centred on the individual. This leads to a partial understanding of the phenomenon and it has serious problems when it comes to accounting for the behaviour that, in fact, we humans have. This partial understanding of the phenomenon implies the absence of the relational element in their explanation of the notions of reason, responsibility, commitment and authority, as well as the lack of a treatment of social normativity in their approaches. Because of this, I present a relational explanation of these notions that resolves the shortcomings that the partial explanation of them tackles. Furthermore, I show the necessity of including social normativity within the explanation of self-knowledge.

Finally, in the section dealing with neo-expressivism, I maintain that the individualist assumption leads, in their case, to a partial vision of both the process of

learning the self-attributions of mental states and the notion of expression. Regarding the first question, these authors focus their explanation exclusively on the capacity for self-attribution of mental states, not taking into consideration its connection with the capacity for attributing mental states to others. In contrast, according to the second-person expressivist approach, both capacities are learned simultaneously. This fact, I contend, happens to be crucial for a proper understanding of such a process. Regarding the notion of expression, in the neo-expressivist approach, it appears severed from the notion of mental state, which leads to its understanding as a means through which we know mental states. Moreover, their explanation of self-knowledge comprises all mental states without distinction. I argue that both the separation of the notions of expression and mental state and the lack of differentiation between different mental states, results in an inadequate understanding of them. Lastly, I agree with some of the arguments about transparency of mental states, particularly about transparency-to-the-subject's-mental-condition.

I analyse the proposal of the traditional approaches on Folk Psychology in the third and last block, as well as the recent proposals that reject the explanation of these approaches. In this third block I also provide an explanation of self-knowledge from the second-person expressivist approach. Just as the studies on ontogenetic development and self-knowledge, the traditional approaches on Folk Psychology are based on the dichotomy between the first and the third person. This assumption has split the research into two types of opposing theories, each one grounded in one of the poles of the dichotomy. The supporters of the Theory Theory construct their approach from the third-person perspective, while the supporters of the Simulation Theory embrace the first-person perspective. In the section dealing with these theories, I argue that the main problem they faces arises from the individualist assumption that both share. According to this assumption, the others are beings to be understood. For these authors, then, the primary function of Folk Psychology becomes predicting and explaining the behaviour of others.

In the next section I analyse the diverse arguments held by some current approaches, subscribing their validity and providing a second-person approach to Folk Psychology. This approach understands Folk Psychology, not as a theory, but as normative-regulative practice whose primary function is not predicting and explaining behaviour, but a regulative function consisting of learning, teaching and exhorting

others to behave according to the shared social norms that rule the interactions of the community members.

To conclude, I provide an explanation of self-knowledge that gathers what has been defended both on the ontogenetic development and Folk Psychology. I maintain that the transparency of the mental is not only towards the one's own mental states, but also towards the mental states that others communicate in an expressive manner during interactions, i.e., through the expressive communication that occurs by means of the second-person perspective. Following this definition of transparency, I argue for the existence of two kinds of knowledge: epistemic knowledge, in a traditional sense, and the pre-conceptual knowledge, this last one understood as the knowledge we have from the first weeks of life. I defend that this type of knowledge, far from disappearing once introduced within language and, therefore, once we possess the necessary concepts for it, stays present through life. Regarding first-person authority, I argue for a relational conception of it, according to which it is linked both with the knowledge of our mental states and the responsibility to put in line our self-attributions with our deeds (with what is expected from someone who self-attributes the mental state in question, according to social and linguistic norms), as well as with the recognition by other members of the community. Finally, I maintain that self-knowledge comprises not only first- or third-person perspective but first-person, second-person and third-person perspectives.

0. Introducción

Los dos paradigmas históricos más reconocibles respecto al autoconocimiento están representados en los trabajos de Descartes (1641) y Ryle (1949). En el esquema cartesiano se entiende que el autoconocimiento deriva de un sentido interno por el cual tenemos un acceso directo, completo e infalible a nuestros estados mentales. Desde este punto de vista, la perspectiva subjetiva, la perspectiva de primera persona, es la única mediante la cual podemos conocernos, es decir, es la única mediante la cual podemos tener acceso a nuestros estados mentales sin riesgo de error. Para Ryle, en cambio, el autoconocimiento deriva de una serie de auto-adscripciones de estados mentales hechas a partir de evidencia externa, inferencias, análisis o auto-interpretaciones. Desde esta perspectiva, la perspectiva de tercera persona, la observación del comportamiento, tanto lingüístico como no lingüístico, es la principal base sobre la que se obtiene el conocimiento de nuestros estados mentales. Ambos enfoques son, por tanto, excluyentes. Descartes anula la perspectiva de tercera persona con su dualismo epistemológico: el único conocimiento indubitable que podemos tener de nuestros estados mentales es el que se da mediante el acceso privilegiado que tenemos en primera persona a través de la introspección. Ryle, por su parte, niega la existencia de un acceso privilegiado y, con ello, la existencia de una perspectiva de primera persona, ya que para él no existe una diferencia, sino de grado, entre la perspectiva de tercera y la de primera persona, reduciendo, de esta manera, la segunda a la primera (Ferrer, 2014).

Esta dicotomía entre primera y tercera persona no solo ha marcado la investigación sobre autoconocimiento en las últimas décadas, sino también las investigaciones tanto en psicología del desarrollo como en filosofía de la mente. En lo referente al autoconocimiento, esta dicotomía ha conducido la investigación hacia el esclarecimiento de las peculiaridades de las auto-atribuciones en primera persona *frente* a las atribuciones de estados mentales a los demás. En los estudios en psicología del desarrollo ha llevado a centrar la investigación en el desarrollo de las capacidades cognitivas de los bebés para *llegar a entender* a los demás. Finalmente, en los estudios en filosofía de la mente, la investigación ha estado centrada en la *predicción y explicación* del comportamiento y marcada por el enfrentamiento entre dos teorías contrapuestas, la Teoría de la Teoría y la Teoría de la Simulación, cada una de ellas

situada en uno de los dos polos de la dicotomía, la primera en la perspectiva de tercera persona y la última en la de primera persona.

Sin embargo, en los últimos años, autores como Vasudevi Reddy (2003, 2005, 2008), Naomi Eilan (2005, 2017), Antoni Gomila (2001, 2002), Daniel Hutto (2004, 2008), entre otros, han mostrado un interés creciente en la inclusión de las relaciones interpersonales en el estudio de estos tres ámbitos, señalando el olvido filosófico al que la perspectiva de segunda persona ha sido sometida y subrayando su importancia para la comprensión de los mismos. De otra parte, ha habido un crecimiento exponencial en el interés acerca de esta perspectiva en otros campos de estudio: en filosofía de la acción encontramos a J. David Velleman, cuyo trabajo, aun cuando no se desarrolle en términos de segunda persona, gira en torno a la intersubjetividad, es decir, en torno a la interacciones con los demás. Velleman (2009) describe su posición como un "término medio entre Williams y Kant" (Velleman 2009, p. 148), pues, para él, la naturaleza de la razón práctica no hace irracional la inmoralidad, como sostiene Kant, ni tampoco hace que la moral sea opcional, como sostiene Williams. Los seres humanos somos, según Velleman, 'improvisadores conjuntos' que buscamos 'auto-aprobación', es decir, improvisamos en nuestras interrelaciones con los demás con la intención tanto de auto-comprendernos y corroborarnos a nosotros mismos, como de ser entendidos por otros.

En teoría de la racionalidad y en metaética Stephen L. Darwall (2006) desarrolla una teoría moral basada en las actitudes prácticas de las interacciones en segunda persona. Para Darwall, al contrario que en las posturas tradicionales que tratan la moralidad en términos de primera persona, el concepto de obligación moral posee un carácter interpersonal. Las personas tenemos una responsabilidad de unos a otros como miembros de la comunidad moral, la cual presupone una autoridad para exigir, para reclamar, al otro dicha responsabilidad. Darwall sostiene, por tanto, la existencia de una autoridad ética de segunda persona.

Finalmente, Schilbach y su grupo (Schilbach et al., 2013) defienden la necesidad del estudio en neurociencia de los encuentros sociales interactivos en tiempo real. Para Schilbach, la cognición social es esencialmente diferente cuando nos encontramos en una interacción 'cara a cara', que cuando nos limitamos a observar. Schilbach, basándose en pruebas de neuroimagen y estudios fisiológicos, desarrolla una teoría basada en un enfoque de segunda persona hacia las otras mentes resaltando su importancia para la creación de una neurociencia realmente social.

El objetivo general de esta tesis consiste en ofrecer una explicación del autoconocimiento desde un enfoque de segunda persona que recoja las intuiciones tanto del expresivismo clásico de Wittgenstein (1953, 1958, 1981), como del neo-expresivismo de Bar-On y Finkelstein (Bar-On, 2004, 2009, 2011, 2015; Finkelstein, 1999, 2003, 2010, 2012) y del expresivismo semántico de Frápolli y Villanueva (2012). Según este enfoque de segunda persona, existe una continuidad entre la capacidad para la intersubjetividad en el desarrollo ontogenético, el autoconocimiento y la psicología del sentido común (*Folk Psychology*), por lo que estos tres ámbitos que tradicionalmente han sido tratados de manera independiente pueden, y deben, tratarse de manera conjunta. La tesis se enmarca, por tanto, dentro de dos discusiones de gran relevancia en la filosofía contemporánea: la discusión sobre el papel de la segunda persona en el autoconocimiento y en la constitución de la racionalidad, por una parte, y los análisis expresivistas de enunciados que contienen términos mentales, por otra.

El objetivo específico de esta tesis consiste en la esclarecimiento de una noción de expresión que sea compatible con la disolución de la dicotomía entre primera y tercera persona. Para ello, desarrollaré una explicación de la constitución de dichas perspectivas basada en lo que denominaré "comunicación expresiva". Mostraré que la perspectiva de segunda persona juega un papel constitutivo en la emergencia de las perspectivas de primera y tercera persona y situaré esta tesis en el contexto de un tratamiento expresivista de los avowals, i.e., de las declaraciones sinceras e inmediatas que el sujeto hace con respecto a sus propios estados mentales.

0.1 Por qué un enfoque de segunda persona

Como he señalado en el apartado anterior, los enfoques tradicionales tanto en psicología del desarrollo como en filosofía de la mente y en los estudios acerca del autoconocimiento, han desarrollado sus propuestas partiendo de la dicotomía entre la primera y la tercera persona. Partir de esta dicotomía conlleva obviar la posibilidad de que esta contraposición no sea en realidad exclusiva y excluyente, y de que sea posible, por tanto, distinguir una perspectiva diferente, genuina, acerca de la mente y de su desarrollo ontogenético. Es por ello que el objetivo principal de este trabajo sea presentar una propuesta que parta de la perspectiva de segunda persona, destacando su importancia para la compresión tanto del desarrollo ontogenético de la mente, como de

nuestra capacidad para el autoconocimiento y el conocimiento de los demás. Partir de una perspectiva de segunda persona no solo modifica el objetivo de la investigación sino que evita una serie de problemas derivados de la adopción de dicha dicotomía. Veámoslo con más detalle.

0.1.1 Desarrollo ontogenético e intersubjetividad

La intersubjetividad, entendida como la capacidad de compartir y conectar nuestras experiencias subjetivas con las de los demás, es un tema que sigue generando una gran controversia en psicología del desarrollo. La discusión está centrada en las capacidades cognitivas y comunicativas que pueden atribuirse a un bebé desde el nacimiento hasta, aproximadamente, el primer año de vida. De una parte, los teóricos que parten de la dicotomía entre primera y tercera persona niegan la existencia de capacidades intersubjetivas en los bebés hasta que alcanzan los nueve meses de edad (Baron-Cohen, 1995; Barresi & Moore, 1996; Tomasello, 1999). De otra parte, los teóricos que parten de la segunda persona afirman la existencia de intersubjetividad desde las primeras semanas de vida (Bråten, 2007; Bråten & Trevarthen, 2007; Murray & Trevarthen, 1985; Reddy, 2003, 2005, 2008; Trevarthen, 1979, 1982, 1993b, 1998).

La diferencia fundamental entre ambos radica en los presupuestos que conllevan cada una de las opciones. Estos presupuestos incluyen una comprensión de los estados mentales radicalmente diferente que, a su vez, conlleva una comprensión de la interacción completamente distinta. Para los primeros, los estados mentales son objetos invisibles que el bebé solo puede conocer mediante la realización de inferencias. De acuerdo con este supuesto, se asume que los bebés perciben las características físicas de la persona con la que interactúan pero no tienen conciencia de su condición psicológica. El acceso a lo mental está basado en el dominio de representaciones abstractas y relaciones inferenciales que los bebés no alcanzan hasta la edad de, aproximadamente, nueve meses. Consecuentemente, a la hora de enfocar el estudio acerca de la intersubjetividad, estos teóricos entienden a los bebés y a los adultos con los que interactúan como seres independientes y separados entre los que existe una brecha psicológica, un espacio que han de cruzar para *llegar a comprender* la condición psicológica del otro. Por tanto, para que pueda darse la intersubjetividad es necesario trazar un puente de unión que permita cerrar dicho espacio y permitir la comprensión.

Un puente de unión que, según estos teóricos, se constituye mediante la formación de representaciones mentales que se infieren de la observación del otro y de las relaciones de las mismas con los estados mentales propios.

Para los teóricos que parten de la segunda persona, en cambio, el bebé es capaz de captar la condición psicológica la persona con la que esta interaccionando sin necesidad de este "puente representacional" ni de la realización de inferencias. Para estos teóricos, no es necesario que el bebé alcance la edad de nueve meses para que la interacción pueda ser considerada como intersubjetiva puesto que el bebé nace con la capacidad de conectar sus experiencias subjetivas con las experiencias subjetivas de los demás. Bajo este enfoque se asume que los bebés reconocen directamente el estado mental expresado por el rostro de la persona con la que interactúan ya desde las primeras semanas de vida. De esta manera, los demás no son entendidos por los bebés como seres a comprender sino como seres con los que interaccionar y regular sus emociones mediante el intercambio de los estados mentales que comunican sus expresiones. Ahora bien, ¿qué concepción de los estados mentales requiere esta comprensión de la comunicación intersubjetiva?

En este trabajo sostendré una concepción de los estados mentales según la cual estos no deben ser entendidos como entidades ocultas e inobservables que los bebés infieren *a través* de la expresión. Según esta concepción las expresiones y los estados mentales no son dos entidades separadas pertenecientes a reinos ontológicos distintos. Las expresiones no son un medio a través del cual el bebé infiere un estado mental oculto sino que son constitutivas de los mismos estados mentales. Ser constitutivo en este contexto no significa que el estado mental constituya la expresión ni, al contrario, que la expresión constituya el estado mental. Ser constitutivo significa que no hay una brecha ontológica entre los estados mentales y las expresiones. Los gestos faciales, los movimientos corporales, el tono de la voz, son una condición o aspecto de los estados mentales mismos, a saber, su *condición comunicativa*. Los bebés ven, oyen, sienten los estados mentales de la persona con la que están interaccionando de manera directa sin necesidad de realizar inferencias o interpretaciones. De acuerdo con esta concepción de los estados mentales, la intersubjetividad tiene lugar a través del intercambio de expresiones (de estados mentales en su condición comunicativa) entre los bebés y las personas con las que interactúan, intercambio que en este trabajo denominaré *comunicación expresiva*. Esta comunicación expresiva mediante la cual tiene lugar la intersubjetividad conlleva, por tanto, una comprensión de la interacción que va más allá

de la contraposición exclusiva y excluyente entre la perspectivas de primera y tercera persona. Según esta comprensión la perspectiva mediante la cual se llevan a cabo las interacciones no es ni la perspectiva de primera ni la de tercera, sino la de segunda persona.

En este trabajo argumentaré que la perspectiva de segunda persona no solo es, como acabo de señalar, la perspectiva genuina mediante la cual se llevan a cabo las interacciones intersubjetivas sino que es, asimismo anterior a, y constitutiva de, las perspectivas de primera y tercera persona.

0.1.2 Perspectivas de 1ª, 2ª y 3ª persona

Los enfoques tradicionales que parten de la dicotomía entre primera y tercera persona ofrecen una explicación del desarrollo ontogenético de la autoconciencia del bebé como separada de la conciencia de los demás. De esta manera, el "yo" es concebido en oposición a un "él" o "ella", por lo que el desarrollo de la comprensión del "yo" por parte del bebé es entendido como separado de la comprensión de los demás. De una parte, hay autores que sostienen que los bebés desarrollan la capacidad de verse a sí mismos como agentes a través de la perspectiva de primera persona, es decir, a través de la comprensión de sus propias acciones, emociones, sentimientos y experiencias que luego proyectan en los demás mediante la simulación de la situación en la que estos se encuentran. De esta manera, se asume que la auto-comprensión precede a la comprensión de los demás (Goldman, 1993, 2006; Tomasello, 1993, 1999, 2008). De otra parte, otros autores parten de la perspectiva de tercera persona. Estos autores sostienen que los bebés desarrollan la capacidad de verse a sí mismo como agentes de manera inversa, es decir, aplicándose a sí mismos la "teoría de la agencia" que desarrollan a partir de la observación y la comprensión del comportamiento de los demás (Gopnik, 1993,2003; Gopnik & Meltzoff, 1997; Gopnik, Meltzoff, & Kuhl, 1999; Gopnik & Wellman, 1992; Meltzoff, 2002).

Ambos enfoques, por tanto, parten de un presupuesto individualista, en el sentido de que la hora de enfocar el estudio del desarrollo de la autoconciencia del bebé presuponen un escenario en el que esta se desarrolla de manera individual y separada de su conciencia de los demás. En el enfoque de segunda persona que sostendré en este trabajo, en cambio, el estudio del desarrollo de la autoconciencia del bebé no se

considera de manera separada al desarrollo de la comprensión de los demás. Según este enfoque, la interacción que tiene lugar entre los bebés y los adultos no es una interacción entre dos seres separados cuyas mentes infieren la existencia de la otra por medio de representaciones mentales provenientes de la observación y la realización de inferencias, o de la simulación del otro. Por el contrario, durante la comunicación expresiva (señalada en el apartado anterior) tiene lugar una *interacción mental* mediante la que el bebé regula sus emociones con las de la persona con la que está interaccionando. En este enfoque, los bebés y los adultos con los que interaccionan no son tratados como entidades separadas, sino como una sola unidad de estudio dentro un modelo de *comunicación reguladora mutua*. El "yo" no es tratado como opuesto a un "él" o "ella", sino como un "yo-tu" interactivo. Según este enfoque, la perspectiva por la cual se lleva a cabo el desarrollo de la autoconciencia del bebé y, por tanto, la perspectiva que aparece primero en el desarrollo ontogenético, no es ni la de primera ni la de tercera persona, sino la perspectiva de segunda persona, la *perspectiva de la interacción mental*.

En este trabajo desarrollaré una explicación del surgimiento de las perspectivas de primera y tercera persona desde el enfoque de segunda persona. Según esta explicación, las perspectivas de primera y tercera persona aparecen en el desarrollo mediante la perspectiva de segunda persona, cuando los bebés alcanzan la edad de entre 12-14 meses. A esta edad los bebés comienzan a involucrarse en triángulos de atención conjunta en los que la interacción con el otro y, por tanto, la comunicación expresiva, incluye un tercer elemento. Con la inclusión del tercer elemento los bebés ya no solo interactúan de forma directa con los objetos externos, sino también de forma indirecta, a través de la persona con la que están interaccionando. Interaccionar de forma indirecta implica un distanciamiento por parte del bebé tanto de la persona con la que están interaccionando como del objeto en cuestión, es decir, del tercer elemento. Asimismo, conlleva que el bebé comience a percibir que las expresiones del otro hacia el objeto pueden no coincidir con las suyas propias. Este proceso, que tiene lugar a través de la perspectiva de segunda persona, permite que los bebés comiencen a entender el mundo como algo independiente tanto de su propia perspectiva como de la perspectiva del otro, es decir, permite el surgimiento de una comprensión básica tanto de la perspectiva de primera persona como de la de tercera persona.

Sin embargo, no solo el surgimiento de esta comprensión básica de las perspectivas de primera y de tercera persona está basada en la perspectiva de segunda

persona. En este trabajo argumentaré que, asimismo, tanto el desarrollo de una conciencia básica de la normatividad como del sentido de autoridad están basados en, o provienen de, la perspectiva de segunda persona.

0.1.3 Normatividad y autoridad

Como he señalado en el apartado anterior, cuando a la edad de entre de 12 a 14 meses el bebé comienza a involucrarse en triángulos de atención conjunta, la comunicación expresiva entre él y el adulto que, hasta el momento, se había desarrollado entre ellos de manera diádica, se torna en una comunicación expresiva triádica al incluir un tercer elemento. Esta inclusión del tercer elemento, es decir, la comunicación expresiva triádica, permite el surgimiento de la conciencia de la existencia de perspectivas distintas acerca de los objetos externos. En esta etapa del desarrollo, los bebés comienzan a ver en su cuidador o cuidadora (madre, padre, etc.) un referente social, alguien con quien regular no sólo las emociones que sienten durante la interacción sino, asimismo, las emociones que sienten hacia los objetos externos con los que interaccionan de manera triádica (Brink, 2008; Franco, 2005; Reddy, 2008; Source et al., 1985; Werner & Kaplan, 1962). La confianza emocional que los bebés sienten hacia la persona que cuida de ellos hace que esta sea la que determine cómo afrontar una situación desconocida para la cual no tienen respuesta o en la que no están seguros de cómo han de comportarse.

Dado que a partir de ese momento, cuando los bebés se enfrenten a una situación similar, responderán de la misma forma que han aprendido a responder en base a la respuesta que obtuvieron de su referente social, en este trabajo sostendré que la explicación de esta conducta se halla en que los bebés asimilan el valor de la información que obtienen de la persona que está a su cuidado, del referente social, no solo en el sentido descriptivo sino también en el sentido normativo. En base a esa explicación y en relación a las propuestas de Rochat (2001) y Roessler (2005), sostendré, asimismo, que los bebés conciben a la persona que cuida de ellos, al referente social, como una autoridad, autoridad que denominaré *autoridad de segunda persona.*

0.1.4 Autoconocimiento

Al igual que en los estudios acerca del desarrollo ontogenético, en los estudios acerca del autoconocimiento también se halla presente la dicotomía entre la primera y la tercera persona. Como he señalado anteriormente, esta dicotomía ha conducido la investigación sobre el autoconocimiento hacia el esclarecimiento de las peculiaridades de las auto-atribuciones en primera persona *frente* a las atribuciones de estados mentales a los demás. El objetivo de estos autores es explicar lo distintivo de los avowals (las auto-atribuciones de estados mentales en primera persona del presente de indicativo) *frente* a las atribuciones de estados mentales a los demás, es decir, explicar la autoridad de primera persona (el hecho de que nuestros avowals se presupongan verdaderos por defecto) y la transparencia de lo mental (el hecho de que no necesitemos inferir nuestros estados mentales presentes de nuestro comportamiento) en base a la asimetría entre primera y tercera persona obviando, por tanto, las relaciones interpersonales y, con ello, la perspectiva de segunda persona.

En este trabajo presentaré una crítica a este presupuesto individualista presente en todos los enfoques actuales acerca del fenómeno del autoconocimiento. Argumentaré que obviar la importancia de las relaciones interpersonales conduce a una visión sesgada y parcial de dicho fenómeno. Con objeto de llevar a cabo dicha crítica, dividiré los enfoques en tres grandes grupos: los introspeccionistas, que defienden que la autoridad proviene de acceso epistémico especial de primera persona a los propios contenidos mentales (Armstrong, 1968, 1981, Lycan, 1987, 1996); los agencialistas, para quienes la autoridad no es una cuestión de acceso privilegiado a los contenidos mentales sino que tiene que ver con lo que hacemos como agentes responsables (Bilgrami, 1998, 2006, 2012; Moran 2001, 2003; Wright, 1984, 1986 1987, 1989a, 1989b, 1991, 1996a, 1996b, 2001b, 2015; Wright, Smith, & Macdonald, 1998); finalmente, los neo-expresivistas, para quienes la autoridad de primera persona proviene de nuestras capacidades expresivas (Bar-On, 2004, 2009, 2011, 2013, 2015; Finkelstein, 1999, 2003, 2010, 2012). De esta manera, señalaré las críticas generales de las que estos enfoques han sido objeto, así como las carencias y/o errores que, según el enfoque de segunda persona, adolecen dichos enfoques, como también lo acertado de algunos de sus argumentos.

Entre los aciertos de estos enfoques destacaré la inclusión de las relaciones interpersonales (aunque de manera parcial, como mostraré a lo largo de este trabajo) en

el caso de los agencialistas y los neo-expresivistas. Para los agencialistas, la autoridad de primera persona está relacionada con nuestro estatus como agentes, el cual conlleva la inclusión de nuestra relación con los demás miembros de la comunidad. Esta inclusión de los demás miembros de la comunidad revela la importancia de nuestras relaciones interpersonales a la hora de explicar tanto la autoridad como la transparencia de los estados mentales. Asimismo, introduce en la explicación del autoconocimiento las nociones de compromiso, responsabilidad y normatividad, nociones que, según el enfoque de segunda persona que defenderé en este trabajo, resultan básicas y fundamentales para una correcta comprensión del fenómeno del autoconocimiento. Finalmente, en el caso de los neo-expresivistas, la inclusión de las interacciones resulta en una explicación acertada, aunque parcial, de la adquisición del lenguaje de las auto-atribuciones de estados mentales. Su explicación de la transparencia, en cambio, escapa a dicha parcialidad, coincidiendo con la definición relacional que ofreceré de acuerdo al enfoque de segunda persona objeto de este trabajo.

Entre las críticas más relevantes señalaré, en el caso del introspeccionismo, el denominado "problema de las otras mentes", según el cual, dado que el acceso a los estados mentales queda restringido, según estos autores, al conocimiento introspectivo de los estados mentales propios, no podemos tener la seguridad de que los otros también tengan estados mentales similares a los nuestros. En el caso del agencialismo, argumentaré que, que aun cuando estos sí incluyen las interacciones personales en su enfoque, estas son consideradas exclusivamente desde una perspectiva teórica y en relación al individuo. Por un lado, la transparencia es vista únicamente como una relación del individuo respecto a sus propios estados mentales, obviando la posibilidad de otro tipo de transparencia relacionada con nuestra interacción con los demás miembros de la comunidad. Por otro, la autoridad es vista como una característica inalienable del individuo, obviando la posibilidad de que la relación con los demás miembros de la comunidad pueda influir en ella, como mostraré que ocurre en los casos de injusticia epistémica (Borgoni, 2018; Fricker, 2007). En el caso del neo-expresivismo, el presupuesto individualista conduce, como he señalado en el párrafo anterior, a una visión parcial de la adquisición del lenguaje de las (auto)-atribuciones de estados mentales. En el análisis de la adquisición del lenguaje de las (auto)-atribuciones de estados mentales, el neo-expresivismo obvia, o no tiene en cuenta, el aspecto relacional de este proceso, el cual es entendido por estos autores en un sentido

unidireccional, es decir, solo tiene en cuenta el aprendizaje de las auto-atribuciones de estados mentales.

Según el enfoque de segunda persona que defenderé en este trabajo, el análisis del proceso de aprendizaje de las auto-atribuciones de estados mentales no está separado del aprendizaje de las atribuciones de estados mentales a los demás. Sostendré que al igual que los bebés aprenden a auto-atribuirse estados mentales mediante la interacción con la persona que está a su cuidado, aprenden, asimismo, a reconocerlos en las expresiones de esta. De otra parte, argumentaré que existe una diferencia fundamental entre los estados mentales que forman pare de la comunicación expresiva y las actitudes proposicionales de creencia y deseo. Esta diferencia influye tanto en la manera en la que conocemos ambos tipos de estados mentales como en la comprensión de los mismos y la función que cumplen respecto a los compromisos y la responsabilidad que tenemos en cuanto agentes pertenecientes a una comunidad.

Este análisis de los compromisos y la responsabilidad del agente está directamente relacionado tanto con el análisis de los agencialistas sobre el autoconocimiento como con el análisis del expresivismo semántico acerca de los predicados de segundo orden. Estos análisis, a su vez, están relacionados con la interpretación que ofrece el enfoque de segunda persona que defenderé en este trabajo acerca de la adquisición, comprensión y función de la Folk Psychology. A lo largo de este trabajo, por tanto, mostraré que el análisis del autoconocimiento no debe realizarse de manera separada ni del análisis del desarrollo ontogenético ni del análisis de la adquisición de las normas sociales de las que está compuesta la Folk Psychology.

0.1.5 Desarrollo ontogenético, autoconocimiento y Folk Psychology

Como señalé al principio de esta introducción, al igual que en el caso del estudio del desarrollo ontogenético y del fenómeno del autoconocimiento, los enfoques tradicionales en filosofía de la mente también asumen la dicotomía entre la primera y la tercera persona como presupuesto de partida. Esta dicotomía ha llevado a una separación entre dos grandes grupos de teóricos respecto al estudio de la comprensión de la conducta de los demás, como he señalado en el apartado dedicado a las perspectivas de 1ª, 2ª y 3ª persona. De una parte, están los defensores la denominada

Teoría de la Teoría, que sostienen que la comprensión de la conducta (y de la mente) de los demás se realiza desde la perspectiva de tercera persona. Según estos teóricos, comprendemos tanto nuestra conducta como la conducta de los demás en base a una "teoría de la agencia" que formamos mediante las inferencias provenientes de la observación de la conducta de los demás (Gopnik, 1993, 2003; Gopnik & Meltzoff, 1997; Gopnik et al., 1999; Gopnik & Wellman, 1992, 1994; Meltzoff, 2002; Wellman, 1990). De otra parte, están los teóricos que defienden la denominada Teoría de la Simulación. Según esta teoría, la comprensión de la conducta (y de la mente) de los demás se lleva a cabo mediante la simulación mental de la situación en la que el otro se encuentra. El proceso de simulación se lleva a cabo, según estos autores, desde la perspectiva de primer persona (Goldman, 1989, 1993, 2000, 2006; Gordon, 1986, 1995, 2007, 2009). El principal problema al que estas teorías se enfrentan proviene del supuesto individualista del que parten, según el cual los demás son *seres a comprender* (como he señalado que ocurre en el desarrollo ontogenético a la hora de explicar la intersubjetividad), por lo que la función de la Folk Psychology es, según estos autores, la de *explicar y predecir* la conducta de los demás.

Sin embargo, en los últimos años autores como Andrews (2009, 2012, 2015), Hutto (2004, 2008), McGeer (2001, 2007, 2015) y Zawidzki (2013), entre otros, rechazan que la función de la Folk Psychology sea la de predecir y explicar el comportamiento de los demás. Según estos autores, la principal función de la Folk Psychology es una función regulativa que consiste en aprender, enseñar y exhortar a otros a comportarse de acuerdo con las normas sociales compartidas que rigen nuestras interacciones. Siguiendo las aportaciones de estos autores, sostendré una concepción de la Folk Psychology como una práctica normativo-regulativa en la que la perspectiva de segunda persona juega un papel fundamental. Según esta concepción, el modelo adecuado para describir y comprender la adquisición de la Folk Psychology no es el observacional ni el de la simulación sino el modelo de *comunicación reguladora mutua* que defenderé en el desarrollo ontogenético. A lo largo de la explicación sobre la adquisición de la Folk Psychology, se mostrará, asimismo, la relación de esta concepción normativo-reguladora con la comprensión del fenómeno del autoconocimiento, tanto en lo que respecta a los enfoques agencialistas como a los neo-expresivistas y al enfoque de segunda persona objeto de este trabajo.

Finalmente, presentaré una concepción del autoconocimiento basada en todo lo defendido en los tres ámbitos analizados. Según esta concepción, la transparencia de los

estados mentales no tiene un sentido unívoco puesto que engloba tanto la transparencia hacia los propios estados mentales como la transparencia a los estados mentales de los demás que tiene lugar durante la comunicación expresiva. Sostendré, por tanto, que la comunicación expresiva continúa presente a lo largo de nuestra vida en nuestras interacciones con los demás. Esta afirmación conlleva la asunción de dos nociones de conocimiento. De una parte, está el conocimiento en sentido tradicional, epistémico, que requiere la posesión de conceptos y de lenguaje proposicional. De otra parte, la explicación de la comunicación intersubjetiva que defenderé en este trabajo presupone que los bebés "conocen" los estados mentales de la persona con la que interactúan, así como los suyos propios. Esta forma de "conocimiento", sostendré, no requiere de elementos conceptuales ni de lenguaje proposicional, por lo que puede ser considerado como un conocimiento *pre-conceptual* (*pre-reflexivo*).

En cuanto a la autoridad de primera persona, siguiendo lo defendido a lo largo de todo el trabajo, sostendré que, al contrario de lo que afirman los enfoques acerca del autoconocimiento, la autoridad de primera persona no tiene solo que ver con cómo conocemos nuestros estados mentales, sino también con la responsabilidad de hacer coincidir nuestras auto-atribuciones con nuestros actos, así como con el reconocimiento de la misma por parte de los demás miembros de la comunidad.

Por último, argumentaré que el autoconocimiento incluye tanto la perspectiva de primera, como la de segunda y la de tercera persona. Esta definición de autoconocimiento engloba lo defendido tanto en el desarrollo ontogenético como en lo referente a la adquisición, comprensión y principal función de la Folk Psychology. De esta manera, quedará mostrada la pertinencia de la continuidad en el estudio de estos tres ámbitos así como la importancia de la segunda persona para la comprensión de los mismos.

0.2 Plan de la tesis

Como he afirmado, el objetivo general de esta tesis es ofrecer una explicación del autoconocimiento desde un enfoque de segunda persona que recoja las intuiciones del expresivismo clásico, así como del neo-expresivismo y del expresivismo semántico. La principal motivación tras esta propuesta reside en la necesidad de acabar con uno de los sesgos que han formado parte en los estudios tanto acerca del autoconocimiento, como acerca del desarrollo ontogenético y de la Folk Psychology, a saber, la dicotomía entre

la primera y la tercera persona. En esta tesis se ofrecerá una alternativa a esta dicotomía que no toma como punto de partida ni la perspectiva de primera ni la de tercera persona sino la perspectiva de segunda persona.

Según este enfoque, la explicación del autoconocimiento no debe (o no puede) analizarse de forma aislada, sino que debe incluir tanto la manera en la que llegamos poseer la capacidad de (auto)-atribuir estados mentales, como la manera en la que comprendemos a los demás. En otras palabras, una explicación del fenómeno del autoconocimiento debe contar tanto con una explicación del desarrollo ontogenético como con una explicación de la comprensión y el dominio de las normas sociales que rigen las interacciones de los miembros de la comunidad. Dados estos tres ámbitos, el desarrollo ontogenético, el autoconocimiento y la Folk Psychology, la tesis se estructurará en tres grandes bloques, cada uno de ellos dedicado a uno de estos ámbitos. Los bloques irán precedidos por una introducción al tema en cuestión y contarán con unas breves conclusiones al final de los mismos. Sin embargo, a lo largo de esta tesis se mostrará tanto la continuidad de estos tres ámbitos como la necesidad de su tratamiento conjunto.

El primer bloque estará dedicado al los análisis que en psicología del desarrollo se han llevado a cabo respecto a la cuestión de la intersubjetividad durante el desarrollo ontogenético. Este bloqué contará, por tanto, con la evidencia empírica propia de este campo. Se defenderá que la intersubjetividad está presente en las interacciones de los bebés con los adultos ya desde las primeras semanas de vida. Este bloque se dividirá en cuatro apartados principales. El primero de estos apartados constará de una introducción general al tema en cuestión, mostrando los diversos enfoques respecto al mismo. Se adoptará la clasificación de Trevarthen (Bråten & Trevarthen, 2007; Trevarthen, 1979; Trevarthen & Hubley, 1978) según la cual el desarrollo ontogenético puede dividirse en tres etapas diferenciadas: la intersubjetividad primaria, la secundaria y la terciaria. El segundo apartado estará dedicado a la primera de estas etapas, la intersubjetividad primaria, que comprende desde el nacimiento del bebé hasta, aproximadamente, los nueve meses de edad. En este apartado se ofrecerá una explicación de la imitación neonatal y de las protoconversaciones contrastando los enfoques individualistas con el enfoque de segunda persona. Se argumentará que el enfoque de segunda persona ofrece una explicación más plausible de estos fenómenos que la propuesta por los enfoques individualistas. Se defenderá una noción de expresión que fundamenta la existencia de una comunicación expresiva mediante la cual los bebés participan en interacciones

intersubjetivas.[1] El tercer apartado ofrecerá una explicación de la segunda etapa, la intersubjetividad secundaria, que comprende desde, aproximadamente, los nueve meses de edad, momento en el que aparecen en el desarrollo ontogenético los triángulos de atención conjunta, hasta los 20-24 meses. Se sostendrá, en contraste con las explicaciones individualistas, que existe una continuidad entre las interacciones diádicas y las triádicas, es decir, entre las interacciones cara a cara y las interacciones que, asimismo, incluyen un tercer elemento. De otra parte, se defenderá que la perspectiva de segunda persona no solo es anterior ontogenéticamente sino que es la base de la cual surge la comprensión de las otras dos perspectivas. Finalmente, en el último apartado se sostendrá que a través de la perspectiva de segunda persona surgen en el bebé tanto una concepción básica de la normatividad como del sentido de autoridad.

El segundo bloque estará dedicado a la presentación y crítica de los diversos enfoques acerca del fenómeno del autoconocimiento. Para ello, dividiré estos enfoques en tres grandes grupos, centrándome en los autores más representativos de cada uno de ellos. Se sostendrá que todos estos enfoques parten de un presupuesto individualista que obvia la importancia de las interacciones, hecho que conlleva una visión sesgada y parcial de este fenómeno. El bloque se estructurará de acuerdo a esta división, constando de tres apartados para el desarrollo y crítica de cada uno de estos tres grupos. El primer apartado estará dedicado al introspeccionismo. En este apartado analizaré las propuestas de Armstrong (1968, 1981) y Lycan (1987, 1996), señalando las principales críticas a las mismas. Se argumentará que estos enfoques son diametralmente opuestos al enfoque de segunda persona objeto de esta tesis. El segundo apartado estará dedicado a los agencialistas. En este apartado se analizarán las propuestas de Moran (2001, 2003), Wright (1984, 1986, 1987, 1989a, 1989b, 1991, 1996a, 1996b, 2001b, 2015) y Bilgrami (1998, 2006, 2012). Se argumentará que aun cuando estos sí incluyen las interacciones en su enfoque, lo hacen desde una perspectiva teórica y centrada en el individuo. Sin embargo, se sostendrá la validez de algunos de sus argumentos. El último apartado estará dedicado a los neo-expresivistas. En este apartado se analizarán los enfoques de Bar-On (2004, 2009, 2011, 2013, 2015) y Finkelstein (1999, 2003, 2010, 2012). Se sostendrá que, al igual que los otros enfoques, estos autores parten de un presupuesto individualista que, en su caso, conduce a una interpretación parcial tanto de

[1] Salvo algunos cambios menores para adaptarlo a la tesis, este apartado se corresponde con el artículo "Intersubjectivity in infancy: a second-person approach to ontogenetic development", aceptado para su publicación en la revista *Philosophical Psychology* [SJR Q1].

la introducción al lenguaje de las (auto-)atribuciones de estados mentales como de la noción de expresión.

El tercer y último bloque estará centrado en el análisis de los enfoques que en filosofía de la mente se han llevado a cabo acerca de la explicación de la Folk Psychology. Asimismo, en este bloque se ofrecerá la explicación que el enfoque de segunda persona sostiene acerca del fenómeno del autoconocimiento. El bloque estará dividido en tres apartados principales, el primero de ellos dedicado al análisis de los enfoques individualistas. En este apartado se analizarán la Teoría de la Teoría, defendida por Gopnik, Meltzoff y Wellman (Gopnik, 1993, 2003; Gopnik & Meltzoff, 1997; Gopnik & Wellman, 1994) y la Teoría de la Simulación, defendida por Goldman, 1989, 1993, 2000, 2006) y Gordon (1986, 1995, 2007, 2009). Se argumentará que, al igual que los enfoques acerca del desarrollo ontogenético, estos autores parten de la dicotomía entre primera y tercera persona, lo que conduce, en este caso, a una explicación de la Folk Psychology según la cual su principal función es la predecir y la explicar la conducta de los demás. En el segundo apartado se presentarán los enfoques que en los últimos años han surgido en oposición a los enfoques individualistas (Andrews, 2009, 2012, 2015; Hutto, 2004, 2008; McGeer 2001, 2007, 2015; Zawidzki, 2013). Según estos autores, la principal función de la Folk Psychology no es la predicción y la explicación de la conducta de los demás sino una función regulativa consistente en aprender, enseñar y exhortar a otros a comportarse de acuerdo a las normas sociales compartidas que rigen nuestras interacciones. Se argumentará a favor de estas propuestas y se ofrecerá una explicación sobre la adquisición de la Folk Psychology desde un enfoque de segunda persona. Según este enfoque, la Folk Psychology es una practica normativo-regulativa en cuyo eje se encuentran las interacciones y, por tanto, la perspectiva de segunda persona. En el último apartado se presentará, finalmente, la explicación del autoconocimiento que ofrece el enfoque de segunda persona objeto de esta tesis.

BLOCK I
The Ontogenetic Development of Infants

1. Introduction

As humans, we have subjective experiences of the world and the people around us which have both phenomenal and psychological aspects: we feel fear, sadness, joy, anger; we are cold, tired, hungry, dizzy; and we have hopes, desires, beliefs, and preferences. Although most theorists assume intersubjectivity in the interactions between adults, much controversy persists among researchers of ontogenetic development concerning the phenomenon of intersubjectivity in earliest stages of infancy. The debate focuses mainly on infants' capacity to share their subjective experiences and to understand the subjective experiences of others. It is clear that adults can use conventional language to share and understand their subjective experiences when interacting, but it is not a common assumption that the early interactions between infants and their caregivers can be classified as intersubjective interactions. On the one hand, there are researchers who deny the existence of intersubjective capacities in infants until they reach the age of 9-12 months (Baron-Cohen, 1995; Barresi & Moore, 1996; Tomasello, 1999). Discussing early interactions, Tomasello states: "they cannot be intersubjective until infants understand others as subjects of experience –which they will not until nine months of age" (Tomasello, 1999, p. 59). Other researchers argue that from the first months of life intersubjectivity is present at the interactions between the infant and the adult (Bråten, 2007; Bråten & Trevarthen, 2007; Murray & Trevarthen, 1985; Reddy, 2003, 2005, 2008; Trevarthen, 1979, 1982, 1993b, 1998). Trevarthen asserts that "the idea of infant intersubjectivity is no less than a theory of how human minds, in human bodies, can recognize one another's impulses, intuitively, with or without cognitive or symbolic elaborations" (1998, p. 17). In his view, it is possible to affirm the existence of an innate capacity related to intersubjective exchanges, which he calls *innate intersubjectivity* (Trevarthen, 1998). In the first few weeks of life, infants engage in direct face-to-face exchanges, coordinating vocal and gestural expressions

because “the infant is born with awareness specifically receptive to subjective states in other persons” (Trevarthen & Aitken, 2001, p. 4).

The core question is to explain the kind of awareness of the self and of the other that infants have as well as the kind of personal interaction in which they engage in their interactions with adults. In order for intersubjectivity to occur, the subjective experiences of both participants in the interaction have to be shared, and, for subjective experiences to be shared, they need to be *communicated*. This issue raises another set of questions: How does communication begin and develop in infancy? What kind of communication is required for intersubjectivity to occur? What capacities can we plausibly attribute to infants, which enable them to participate in communication? What role do interactions play in the development of these capacities?

I propose a non-individualist framework to answer these questions based on both philosophical and psychological research into ontogenetic development, distinguishing it from an individualist framework (i.e., one which considers separate entities rather than relations). Following Reddy (2008), I will pursue a second-person approach rather than one of the traditional approaches based on the first or third person. Contrary to the individualist idea that there is a parallel development between infants’ self-awareness and their awareness of others’ self—which thereby assumes a contrast between the first and third person—I will argue for an intersubjective story of the development by assuming a link between the first and the second person.

My aim is to show the importance of the second-person perspective in both the ontogenetic development of infants and in the constitution of the awareness of themselves and of the others’ perspective. I will argue that the second-person perspective is prior and constitutive of both the first- and the third-person perspective. I will defend that it is through second-person relationships (second-person perspective) that infants develop themselves as having an independent perspective of the person with whom they interact (first-person perspective) and to discover the existence of others’ perspective (third-person perspective) when they engage in triangles of joint attention.

In the next section, I will briefly sketch infants’ ontogenetic development from birth up until they begin to be considered as conceptual beings. In Section 3, I will concentrate on the dyadic interactions contrasting an individualist approach, based on the first and the third person, with a second-person approach. I will argue that understanding protoconversations as what I will call ‘expressive communication’ allows understanding them as entailing intersubjectivity. Section 4 focuses on triadic

interactions, the so-called joint attention. I will argue that joint attention has its roots in second-person relations, and that it enables infant's understanding of the first- and the third-person perspective. In Section 5 I will develop the idea of a second-person authority in the ontogenetic development. Finally, I will draw some conclusions from the framework presented.

2. Infant Development: A Preliminary Sketch

From birth, infants are social creatures. From very early on, infants and mothers—or other caregivers[2]—engage in mutual interactions. During the first weeks of life, the interactions of the infant with the mother which have been researchers' main focus of study are the ones they have in face-to-face emotional exchanges. It is traditionally assumed that in these interactions infants are spontaneously and instinctively reacting without understanding that they are responding to or influenced by something external to them. Infants do not yet feel differentiated from their environment or from their mother. In Mahler's words, the relationship of the infant with the mother is one of *symbiotic fusion* (Mahler, Pine, & Bergman, 1975; Mahler, 1952)[3]. Today, however, most theorists hold a different opinion regarding newborns' capacities (Kokkinaki & Kugiumutzakis, 2000; Kugiumutzakis, 1985, 1996, 1998; Meltzoff, 2007; Meltzoff & Moore, 1983, 1989, 1998; Reddy, 2003, 2005, 2008; Trevarthen, 1979, 1982, 1993b). Advances in research and new experiments allow them to attribute some more capacities to infants from the beginning. Newborns, they argue, are not fused with the mother and the environment. There is a social connectedness present from birth, but it is not an undifferentiated fusion.

Shortly afterwards, between two and four months, with the development of the infants' senses, their attitudes begin to be established as basic answers to the environment, in this case, as *responses* to the mother. Trevarthen calls this period *primary intersubjectivity* (Trevarthen, 1979). At this stage, through the use of facial

[2] I use the relationship of the infant with the mother as the most common case. Henceforth, the term 'mother', in this context, it is to be understood as referring to the person who is taking care of the infant, to the caregiver, or to the adult with whom the infant interacts, whether it is the mother or not.

[3] However, it should be noted that Mahler and their colleagues did not claim that the symbiotic period included the first weeks of life. "This corresponds to the entry into that period which we have named the symbiotic phase. While primary narcissism still prevails, in the symbiotic phase it is not so absolute as it was in the autistic phase (the first few weeks of life)" (Mahler et al., 1975, p. 46).

expressions, babbling, and movements of the body, infants seek the attention of their mother (Trevarthen, Delafield-Butt, & Schögler, 2011). Once they grasp her attention they begin to interact with her, understanding her gestures as a response to something that they themselves may have caused (Reddy, 2003). The way in which infants behave in these interactions indicates that what they are doing are not merely spontaneous, instinctive reactions to external stimuli but communication. These interactions are what Bateson has called *protoconversations*: nonverbal discourses regulated by dynamic relational affects and alternating shifts of vocalizations and gestures (Bateson, 1975, 1979).

A little later, between three and five months, infants begin engaging in a new activity: shifting their attention to others. In their interactions with their mother, their gaze shifts to other people who are within their visual field and who become the object of their attention (Reddy, 2005, 2008). During the period from five to seven months, the infants' attention, as well as their emotional responses, begin to be related to the physical objects surrounding them and the objects to which others attend. Infants follow other people's gaze to objects that are in front, nearby, or in someone else's hand (Reddy, 2008). Objects become appealing to them when they are linked to human activity, whether it's their own activity or that of other people. Infants exhibit different behaviour toward others than toward physical objects, based on the different experiences they have in dealing with the former. Infants "'identify' with other people" but not with objects (Hobson, 2002, p. 271).

Although the attentive contact between the infant and the mother occurs at an early age (e.g., in the early face-to-face contact centred on the gaze), it is towards the first year of life, after nine months old, that we can see them attending together, and in a lasting way, to the environment (Brinck, 2004; Eilan, 2005). At the age of about eight or nine months, infants are able to coordinate interpersonal actions and attitudes with their mother, including physical objects in the mother's hands. Over the age of nine months, in addition to having become familiar with their mother's face and gestures, infants have also learned the different intonations of their mother's voice. At this stage, infants already have a level of familiarity with their mother that allows them to capture her emotions, both towards themselves and towards the objects of the shared external world with which she interacts. Infants have already discovered that the mother's emotional expressions, her different tones of voice, her facial and body gestures, can be directed at—or be responses to—an external object, an object which is not part of their

self and with which they could also interact. The beginning of these triadic interactions is a sign of the emergence of *secondary intersubjectivity* (Trevarthen & Hubley, 1978).

During the following three months, infants acquire a set of new capabilities, such as following the other person's gaze determining the point at which they target (Franco, 2005), or indicating, showing or requesting objects to others (often looking into the other person's eyes to see if they are attending, that is, "checking that accompan(ies) pointing" (Eilan, 2005, p. 15). These pointings are the so-called *proto-declaratives* and *proto-imperatives* (Bates, Camaioni, & Volterra, 1976; Bretherton, McNew, & Beeghly-Smith, 1981). It is then, about the age of twelve months, that we can affirm that *joint attention* begins to manifest itself. The triangle formed by joint attention (infant-adult-environment) is not a mere contact of casual or sporadic attention, but rather a relatively durable one in which the infants activity is structured according to that state of shared attention. At the age at which the triangles of joint attention begin to develop, infants begin to appreciate a difference between their actions and the effects they have on their mother, responding with rejection, satisfaction or ambivalence to the attention it provides to her actions.

From then on, with the infants' participation in more sophisticated triangles of joint attention, they begin to acquire the linguistic abilities that will lead them to be considered as adult conceptual beings. Some theorists have suggested that experiencing joint attention is a fundamental requirement in the earliest stages of vocabulary development (Morales et al., 2000; Tomasello, 1993, 1999; Tomasello & Farrar, 1986). According to Tomasello "to acquire a new word—to learn to comprehend and produce it in conventionally appropriate contexts […] the child must enter into a joint attentional focus with an adult" (1993, p. 117). Infants are yet in a third level of intersubjective development called by Bråten and Trevarthen (2007) *tertiary intersubjectivity*. In tertiary intersubjectivity arises a high-order development of intersubjectivity that is strongly language-mediated and in which the infant understands nuanced dialog, narrative imagination, emotional empathy, and other phenomena we associate with adult conceptual beings.

In this section, I have provided a succinct explanation of infants' ontogenetic development. Following Trevarthen's classification, which is based on intersubjectivity, I have distinguished three stages: *primary intersubjectivity*, which goes from the first weeks of life of the infant until approximately nine months of age, stage in which joint-

attention appears; *secondary intersubjectivity*, the period from nine months to 20-24 months of age; and *tertiary intersubjectivity,* the point at which infants begin to get involved in complex linguistic and emotional exchanges.

Once this framework is established, I will analyse it in more detail in the following sections, contrasting the intersubjective story, the second-person approach that I advocate, with the individualist story that is based on the opposition between first and third person. I will show that an intersubjective story fits and explains, in a more appropriate and natural way than an individualistic one, both the development of infants during primary intersubjectivity and the subsequent behaviour and development in secondary intersubjectivity, the episodes of joint attention. I will start the next section by discussing the first stage of infants' development in *primary intersubjectivity*, in which I will distinguish between *neonatal imitation* and *protoconversations*.

3. Relations in Early Development: Primary Intersubjectivity

As I briefly explained in the introduction, theorists assume that conventional language could make intersubjectivity possible. Language radically transforms a child's ways of communicating, but this does not imply that intersubjectivity should be relegated to this stage of development. It is possible to assume, as other theorists do (Kokkinaki & Kugiumutzakis, 2000; Kugiumutzakis, 1985, 1996, 1998; Reddy, 2003, 2005, 2008, Trevarthen, 1979, 1982, 1993b), that intersubjectivity is an important aspect of psychology from the outset of development, long before children learn to speak. Not only do basic, innate biological mechanisms (such as mirroring and imitation mechanisms related to intersubjectivity) suggest this; it also seems plausible because of how infants interact socially.

In what follows, I will focus on primary intersubjectivity, explaining what occurs in the dyadic interactions between the infant and the mother from birth until secondary intersubjectivity. I will divide it in two sections: the first section focuses on the first weeks of life and the researches on neonatal imitation, where I will answer some of the questions raised above. In the second section, I shall concentrate on the so-called *protoconversations,* from the age of two to four months. I will argue that

protoconversations can be understood as an expressive *communication* between infants and their mother that entails intersubjectivity.

3.1 Neonatal Imitation

It was traditionally assumed, perhaps due to the influence of Freud's and Piaget's works in the psychology of development (Freud, 1911; Piaget, 1952, 1954), that, after birth and during the first weeks of life, until the age of two months, the responses of infants to their mother in their interactions are no more than spontaneous and instinctive reactions. It is assumed that infants lack the distinction between self and non-self: "this primitive relation between subject and object is a relation of undifferentiation, corresponding to the protoplasmic consciousness of the first weeks of life when no distinction is made between the self and the non-self" (Piaget, 1954, p. 355). Infants are assumed to be in a *symbiotic phase* in which the environment and their mother are not recognized as external to them. Life begins, according to these theorists, in a state of profound confusion between the self and the world. Infants are somehow fused so they cannot tell themselves apart from the world around them. As they put it:

> The normal symbiotic phase is marked by the infant's increased perceptual and affective investment in stimuli that *we* (the adult observers) recognize as coming from the world outside, but that (we postulate) the infant does not recognize as having a clearly outside origin. (Mahler et al., 1975, p. 48)

Currently, however, several psychologists and philosophers, such as Kugiumutzakis (1985, 1993, 1996, 1998) Trevarthen (1979, 1982, 1993b), Meltzoff (Meltzoff, 2007; Meltzoff & Moore, 1983, 1989, 1998) and Reddy (2003, 2005, 2008, 2010), among others, have challenged this idea. This series of experiments, based on the behaviour of infants, seem to show that the real story is a markedly different one. The results of these experiments, the infants' responses to the interactions with adults, the *neonatal imitation*, lead them to abandon the assumption of the symbiotic fusion or, as it is also called, the *adualism* (Kugiumutzakis, 1998).

Contrary to the idea that the movements made by infants are unintentional purposeless reflexes, results from several studies indicate that neonatal imitation is more complex than previously assumed and that it helps infants to socially engage with adults. Newborns are perfectly capable of responding to their surrounding events but not immediately, for imitation can occur after some delay (Meltzoff & Moore, 1989). An

experiment conducted by Van der Meer, Van der Weel, and Lee, "indicate(s) that newborns can purposely control their arm movements in the face of external forces and that visual control of arm movement is underway soon after birth" (1995, p. 693). In research carried out by Von Hofsten (1982), when an object is placed in the so-called "reach-space", infants look at it and try to reach it with their arms. Delafield-Butt and Gangopadhyay (2013) show that there is evidence of primary sensorimotor intentionality that demonstrates agency in neonatal movement. Moreover, newborns are observed to prompt responses from the attentive adult, implying that the infant's movements are not mere unintentional reflexes (Nagy & Molnar, 1994, 2004). It can be argued, contrary to the Freudian and Piagetian tradition, that there is no confusion in infants between self and non-self.

On the one hand, this assumption leads us to answer some of the questions raised above: newborns have some capacity to engage in "intentional", purposeful interactions, as we have seen in the way that neonatal imitations develop. They are, therefore, somehow aware of themselves at birth as differentiated from the mother, as someone who can interact with her. However, this awareness is revealed *only* in interactions. Interactions, therefore, play a key role in an infant's development of self-awareness. On the other hand, this same assumption leads us to some new questions: What sort of "self" can we attribute to infants? What kind of awareness could infants have of themselves and of others? What role do interactions and others play in an infant's self-awareness?

The way theorists answer these questions depends on the approach they adopt. Martin Buber's (1958) states that there are two forms of relating and knowing, namely, the *I-Thou* and the *I-It*. In the latter, the *I* "experiences" a detached thing, the *It*, while, in the former, there are no bounds—the *I* "participates" by engaging directly whit the *Thou*. The way in which we understand the subject determines the way the study will be conducted. "For the *I* of the primary word *I-Thou* is a different *I* from that of the primary word *I-It*" (Buber, 1958, p. 3).

In an individualist approach—or in Buber's terms an *I-It* approach—the behaviour of infants is studied in a unidirectional way, in which the concept of correspondence is static, centred on the infant. The self is understood as detached from the other, that is to say, the self is an "I" as opposed to a "he" or "she". This individualist approach is based on a parallel story about the self-awareness of infants and their awareness of others, in which infants' development is understood in a first- or

third-person way. On the one hand, first-person understanding of infants' development signifies that infants develop capacities to see themselves as agents by understanding the effects of their own actions, feelings, emotions, and experiences, and they apply this understanding of themselves to others by simulation, namely, by "putting herself in the other shoes". This first-person approach presupposes that infants possess the materials for self-knowledge regardless of their knowledge of others (Tomasello, 1993, 1999, 2008). On the other hand, a third-person understanding of infant development reverses this explanation: infants develop the capacity to see themselves as an agent by applying to themselves the "agency theory" that they develop through observing and understanding the behaviour of others[4] (Gopnik, 2003; Gopnik & Meltzoff, 1997; Gopnik, Meltzoff, & Kuhl, 1999; Gopnik & Wellman, 1992; Meltzoff, 2002). In either case, both first- and third-person approaches presuppose a scenario in which infants' self-awareness and their awareness of others develop in parallel.

By contrast, in the intersubjective approach—the second person understanding of development that I am proposing—the study is done in a bidirectional way. Infants and adults are treated as a single unit of study within a *mutual regulatory communication model*. The self is not understood as an "I" opposed to a "he" or "she" but as an interactive "I-You" (*I-Thou*). This is not a parallel development but a development based on shared-interactive experiences and regulation of emotions. A bidirectional development gives rises to efficiencies that are not be possible merely in terms of a parallel development, which may explain its evolution. In Reddy and Trevarthen's words:

> The most powerful meaning of a smile or gaze or a frown emerges in the infant's engagement with these human events, not through an abstracted observation nor simply as a predetermined given. If we didn't engage with infants, they wouldn't learn very much at all about us, just as we wouldn't learn very much about them. We draw their knowledge into existence and they draw ours. That is how infants, and we too, 'learn how to mean' from each other. (2004, p. 14)

Infants' understanding of their own emotions is linked with the emotions of the mother. Through interactions with the mother, infants share and regulate their emotions with hers. This shared interactive relation and regulation is carried out by an "expressive communication" that *comes together with* the capacity to be self aware and aware of

[4] In Block III I will develop these approaches more extensively.

others. Allow me to explain this proposal by comparing it with another one based on a parallel development.

Although Andy Meltzoff, a key figure in the studies on infant development, certainly recognizes the importance of the interpersonal exchanges between infants and adults, he adopts an individualist approach based on the contrast between first person and third person. According to Meltzoff, infants can relate themselves with the other's self because their bodily actions can be compared: I can act like my mother, and she can act like me (Meltzoff & Moore, 1998). In this way, infants are somehow reflected in the other. They reproduce actions of the other as a way of testing their identity or of retrieving interactive personal "games". Meltzoff's theory provides a way of conceptualizing how both the infant and the mother can feel the other's state through the perception of correspondences. According to Meltzoff, infants have a proprioceptive mechanism called AIM (active intermodal mapping), which works through a transmodal coincidence: infants associate what they see with what they feel proprioceptively in their face. As he puts it:

> Infants' self-produced movements provide proprioceptive feedback that can be compared to the *representation* of the observed act. AIM proposes that such comparison is possible because the observation and execution of human acts are coded within a common framework. We call it a 'supramodal act space'. (Meltzoff, 2002, p. 24, italics added)

By detecting coincidences, the infant can, from the beginning of her life, translate environmental stimuli into internal states. Meltzoff contends that this ability produces in the infant the "you like me" feeling, this being the origin of pre-symbolic intersubjectivity (Meltzoff, 1985, 1990). The apprehension that the other is similar to oneself constitutes the origin of a theory of mind: other people have similar states to one's own.

One of the main problems with the individualist approach is the way in which it explains infants' access to the other. Infants understand that the other and they are similar to the other through comparing the representations of the observed act with the proprioception of their own act. Infants need to *learn* this relation, and, thus, they need to link their own self with the others' self or link the others' self with their own self by using the comparison shown above.

On the contrary, adopting a second-person approach resolves the problem. In a second-person approach, there is no need for a representational bridge to account for

how infants recognize the other as a subjective being similar to themselves, and there is no need for a parallel story of the development of the infant self and the self of others. Rather, in a second person approach, infants' self-awareness is linked with the awareness of others (I will return later to the implications of this assertion).

Vasudevi Reddy shares with Meltzoff the interest in neonatal imitation, but, contrary to Meltzoff, she develops a second-person approach based on an affective-engagement view. She shares with Meltzoff the rejection of unity—the symbiotic fusion in which no distinction is made between the self and the non-self—but she believes that a second-person approach more accurately explains the infant capacities, behaviour, and development to which the results of neonatal imitation experiments attest. On the one hand, it explains in a more parsimonious way how infants recognize others, since does not need a bridge between self and other. On the other hand, it accounts for the imitating behaviours of the infant in a more appropriate way, since it explains not only *how* but also *why* infants imitate. The two issues are related to each other. I will begin with the latter, and then I will come back to the former and its implications.

According to Reddy (2008), the explanations of imitation based on the AIM hypothesis presented above may describe in neurological terms how infants imitate, but something is missing in Meltzoff's explanation of these interactions. Neonatal imitation is not just an automatic, involuntary reaction of "matching". It seems to be more of a *response to* than a *mimic of* the mother's gesture (Reddy, 2008). "When interacting with people infants don't just imitate, they respond" (Reddy, 2008, pp. 58–59). Kugiumutzakis' experiments with newborns demonstrate that infants try to imitate and succeed: "Neonatal imitation has been shown to be highly dependent on motivational state. It has the characteristics of selective and effortful behaviour" (Kugiumutzakis, 1993, p. 25). Moreover, newborns do not interrupt the mother until she has finished her communicative act. Their actions usually do not overlap with those of the mother. Newborn behaviour is more about turn-taking[5] than mere reaction (Gratier, Devouche, Guellai, Infanti, Yilmaz & Parlato Oliveira, 2015). In one of Kugiumutzakis' experiments, the act to be imitated was presented to the newborn a maximum of five times with an interval of three seconds between each, but the act stopped being presented as soon as the newborn started reproducing it, regardless of the number of

[5] There is even evidence of turn-taking vocalizations in pre-term infants A study carried out with pre-term infants, still admitted to the NICU, showed that, at 32 weeks of gestation, infants produce reciprocal vocalizations, supporting the hypothesis that taking turns is an innate human ability (Caskey, Stephens, Tucker, & Vohr, 2011).

presentations already made. Although they were able to interrupt the presentations at any time, more than three-quarters of newborns (77%) reacted after the fifth presentation, when the presentations were finished. These results show that in "natural mother-infant interactions co-actions are the exception rather than the rule" (Kugiumutzakis, 1998, p. 68).

The AIM hypothesis may be correct in explaining *how* newborns imitate, as it offers what may be an adequate explanation of the equivalence in neurological terms between the newborn and the mother. However, this explanation does not answer a fundamental question of the imitation process. The key question is not to answer *how* infants imitate (although it is undoubtedly significant) but *why* they do so; it is to answer the *motivation question*, the *relevance* of imitation in the interactions between the infant and the mother (Reddy, 2008).

The question is, then, why do infants imitate? As we have seen, infants do not just mimic the mother's gestures, they respond to them. However, infants do more than respond to the actions of their mother. As indicated above, newborns can also prompt responses from the attentive mother. Nagy and Molnar (1994, 2004) showed that the heart rate of the newborn accelerates just before they imitate a movement and strongly decelerates when they are about to prompt a movement in the mother. According to a second-person approach, these findings of provocation and turn-taking interactions suggest that neonatal imitation is the first dialogue in which newborns become involved. This is why imitation is relevant to them; this is the reason why they imitate. In imitating, newborns are engaging in a simple form of communication with their mother (Kugiumutzakis, 1993; Reddy, 2008; Trevarthen & Reddy, 2007). Newborns imitate *because* they are responding to the "communicative attempt by the mother" (Reddy, 2008, p. 61). They are completing the mother's as of yet incomplete interpersonal actions, and, furthermore, they expect a response from their mother to their own interpersonal communicative attempt. In the words of Kugiumutzakis, when interacting with the mother, the newborn knows "when *you* have stopped *your* action and when *I* have to start *my own*" (1998, p. 77).

To state that infants communicate with the mother requires an explanation of the nature of the communication that is involved in their interactions. I will explain what I call "expressive communication" in the next section, but, before I do so, it is necessary to examine the implications of both approaches to explaining the bridge or link between self and other.

An individualist approach such as Meltzoff's assumes that infants cannot observe the psychological features of their mother. Intentions are not linked to movements in this model. Infants have access only to the physical properties of their mother, namely, her visible movements. Infants, therefore, need to understand that the physical features which they see in their mother and that they imitate are related to her psychological state. According to Meltzoff, infants *learn* this relation through the process of imitation and by making *inferences*. In his model, infants need extensive experience, filled with observations of similarities, in order to understand by "an inferential process" (2002, p. 35) that the other "is like me", that the movements of the mother are related to unobservable properties, and that she has intentions (Meltzoff, 1990, 2002; Meltzoff & Moore, 1989, 1998). An individualist approach thus requires demanding cognitive abilities from the infant.

On the contrary, in a second-person approach, infants are assumed to *directly* recognize the expressed state in the face of the mother—or, as I will argue, to directly see *the mental state itself*. This is not to say that what it is recognized by infants is the expression as such. What infants recognize in interactions is the *occurrence* of the expression within the *ongoing* context. Infants understand the expression as *directed at them*; they grasp that there is a motivation in the mother's gestures, an intention to communicate, and this intention *moves them to respond*. This view demonstrates that a "conversational or communicative answer to the motivation question may be cognitively far less demanding than an individualist 'identity testing' one" (Reddy, 2008, p. 61). Results of experiments on neonatal imitation, carried out by Kugiumutzakis, show that there is no need for infants *to learn* to recognize the similarity of themselves to their mother by an inferential process, because they already "know" that they are like the mother, and they "know" how to relate to her. There is no need for the infant to make inferences to discover that the mother "is like me". In Kugiumutzakis' studies, 75% of newborns reproduced their mother's mouth opening correctly in their *first* imitative effort. The recognition must take place before the onset of imitation; otherwise, it is not explained how newborns are able to make their first imitative movement correctly. As Kugiumutzakis maintains, and contrary to what Meltzoff assumes, "the recognition of the isomorphism is a precondition, not a result of the active intermodal *matching*" (Kugiumutzakis, 1998, p. 77).

In this section on neonatal imitation, we have seen that a second-person approach does not require such cognitively demanding abilities as other accounts. According to this approach, infants do not need to bridge the gap between physical and psychological properties, since these properties, in fact, arise together for them. Moreover, newborns do not need *to learn* to recognize similarities between themselves and their mother. There is no need for infants to recognize that "you are like me"; there is no need to make such inferences to bridge a gap made by representations because *there is no gap to bridge*. Infants do not experience the other as something to understand but as someone to *tune in with*, someone to *communicate* with by sharing expressions and regulating emotions. Infants are born into the world actively seeking interpersonal experiences; they are ready to engage in face-to-face interactions. The dialogic nature of neonatal imitation and the innateness of intersubjectivity is shown through an infant's direct recognition of the expressions their mother directs at them, their ability to "provoke" her responses, and their ability to know when to take turns in these interactions.

Regarding the question raised above about the kind of awareness of the self one can attribute to newborns, we can argue that, according to a second-person approach, this awareness is one of *self in relation with the other*, a sort of interpersonal unit in which the infant's "self" is separate from the other but existing within the same *interactive interpersonal unit*. Thus, to answer another question I raised above, I propose that interactions and the other play a constitutive role in infants' self-awareness. Interpersonal interactions, face-to-face interactions, are a constitutive feature of newborns' inner life.

In the following section, I will continue my descriptions of child development with the stage from two to nine months of age, during which protoconversations take place. I will show that a second-person approach can also explain the behaviour of the infant in this later period, taking place after neonatal imitation. I will argue that the communication that infants have with their mother in their interactions, which I call "expressive communication", entails intersubjectivity. To show this, I will describe some psychological experiments on ontogenetic development and put forward a philosophical explanation of how one must understand expressions in order to understand how such communication is possible; more specifically, I will explain how expressions are the communicative condition of mental states.

3.2 Protoconversations

As we have seen, from the second-person approach on neonatal imitation, the interactions between the newborn and the mother can be seen not as mere reactions but as a kind of communication—a communication in which the newborn not only *responds* to the mother's gestures but also prompts them. The dialogic nature of neonatal imitation shows that newborns are born ready to engage in face-to-face interactions. Indeed, newborns can see and distinguish faces in the first days after birth (Trevarthen, 2002). It is worth noting that, when people are present, newborns move to seek experiences with the people while ignoring other events or objects. There is a great deal of experimental evidence since the late 1960s which shows that infants, even at an early age, engage in forms of interpersonal relationships that substantially diverge from the way they relate to inanimate objects. Brazelton, Koslowski, and Main (1974) studied the differences between the way that four-week-old newborns relate to an object (a small monkey suspended by a string) brought toward their "reach space" and the way that they interact face-to-face with their mother. Brazelton and his colleagues commented:

> We had felt that we could look at any segment of the infant's body and detect whether he was watching an object or interacting with his mother –so different was his attention, vocalizations, smiles, and motor behaviours with the inanimate stimulus as opposed to the mother. (Brazelton et al., 1974, p. 35)

Face-to-face interactions of two-month-old infants with adults show evidence of engagement. After studying mother-infant vocal exchanges in films made by Margaret Bullowa, Mary Catherine Bateson (1979) introduced the term "protoconversations" for the spontaneous face-to-face interactions between an infant of two to three months and her mother. Bateson described the phenomenon as follows:

> [T]he mother and infant were collaborating in a pattern of more or less alternating, non-overlapping vocalisation, the mother speaking brief sentences and the infant responding with coos and murmurs, together producing a brief joint performance similar to conversation, which I called 'protoconversation'. (Bateson, 1979, p. 65)

However, most developmental psychologists are sceptical about whether protoconversations can be considered to be truly communicative. Infants are not supposed to be sensitive to the purpose of another person's signs until they acquire the cognitive skills that they need to learn throughout their first year of life. On the one

hand, the cognitivist tradition (the supporters of the Piagetian individualism) argues that infants learn through the development of mental representations. According to these theorists, it is not possible for infants to have genuine communication before the last quarter of their first year, at about the age of nine months. To have genuine communication, infants need to be endowed with the requisite communicative skills. They need to have an awareness of intentionality which they gain only at that age. On the other hand, social constructionism moves away from Piaget's individualism. Derived from the update of Lev Vygotsky's theory of social learning (Vygotsky, 1962, 1977), social constructionists place more emphasis on social factors that contribute to cognitive development. According to these theorists, higher mental processes in the individual have their roots in social processes. To understand individual development, we must refer to the social and cultural context within which it is embedded. Nonetheless, according to constructionists, protoconversations between infants and their mother do not really qualify as genuine communication. The perceived genuineness of communication in protoconversations is just an illusion, a fiction created by the mother. Infants do not yet have the required awareness of what the adult is intending or understanding. Mothers just act "as-if" the infant could understand them. And it is precisely because mothers play out this fiction that it eventually turns out to be true (Kaye, 1982). Social interaction plays a fundamental role in the development of cognition. It is through these interactions that the awareness of intentionality and understanding emerge in infants at the end of the first year.

There is, however, another way to interpret protoconversations. Reddy (2008, 2010) argues that the social constructionists' conclusion, namely, that mothers play out the "as if" fiction, is a consequence of their interpretation of the Ragnar Rommetveit's famous dictum: "Intersubjectivity has to be taken for granted in order to be achieved" (1974, p. 56). According to Reddy, social constructionists applied this dictum to infants by indicating that, in order for infants to actually come to understand, their mothers have to act "as if" they can understand them. However, Reddy argues that we can also interpret this dictum in another way: "that the infant initiation of communication is evidence of the infant's recognition of another's intersubjectivity" (2010, p. 9).

The core question is how we interpret the necessities of genuine communication, more specifically, the necessity of intersubjectivity. A basic requirement for a communication to be genuine is that the participants have to be aware of being involved in it. In other words, communication has to be *mutually manifest* (Eilan, 2017).

Otherwise, communication is just an unidirectional act—that is, infant's crying, for instance, would be a form of non-genuine communication in which the infant communicates a state (hunger, irritation, fatigue, etc.) but does not engage in a genuine communication). If mutual awareness implies *conceptual* awareness of intentionality by both partners in the relation, as cognitivists and constructionists hold, then there would be no genuine communication between the infant and the mother. Infants do not have concepts at that age. However, this assumption, as I see it, relies on an erroneous view, namely, the existence of a mind-body (or mind-behaviour) dualism. According to this assumption, it is assumed that infants can perceive physical features but cannot have an awareness of their mother's psychological state. As seen above, these theorists assume that the psychological features are invisible objects only knowable through inferences. These invisible mental objects cannot be perceived; they must be *theoretically conceived of* or *understood* by the infant through an understanding of the relations between objects, which, if sufficient, enables inference-making. The access to the mental realm is based on the mastery of abstracted representations and inferential relations which infants lack at that age and will not possess until the end of their first year of life.

However, is this necessarily so? Must we inevitably distinguish between the physical features that infants are assumed to perceive and the psychological features that they are assumed not to detect? It is so that infants cannot be aware of their mother's psychological orientation, except through making inferences?

As we have seen in neonatal imitation, second-person explanations about infant's awareness of their mother's expressions allow us to provide a different interpretation. This also applies to the case of protoconversations. There is a great deal of empirical as well as theoretical evidence that shows that this does not have to be the case. On the one hand, there are explanations based on the results of a substantial body of empirical research carried out by developmental psychologists that show that protoconversations can be understood as truly communicative by demonstrating the existence of genuine *expressive communication* between infants and their mother which is *mutually manifest* for both. These explanations are based on the infant's ability to exchange *significant expressive information* with the mother, rather than on their theoretical knowledge of other's minds. On the other hand, there are second-person interpretations about the infant's awareness of the attention that allow us to give a different explanation not only of protoconversations but also of both primary and

secondary intersubjectivity, when the triangles of joint attention take place in the development of the infant. Let's start with the former in this section and conclude with the latter in the fourth section, which will pave the way for the explanation of secondary intersubjectivity, the triangles of joint attention.

Several studies by developmental psychologists show the existence of *tune-in* emotional patterns in the interpersonal face-to-face exchanges between the infant and their mother in the first months of life. Jeffrey Cohn and Edward Tronick (1983) conducted an experiment in which mothers were asked to interact with their three-month-old infants using normal and sad expressions, each one over a period of three minutes. The results showed that when the mothers' expression was sad, infants became more and more negative, even to the point of getting upset and protesting. Infants continued behaving in this manner for a short time after the mothers had interacted with them normally again. Another experiment, carried out by Jeanette Haviland and Mary Lelwica (1987), with ten-week-old infants revealed the different affective states that infants manifest in relation to their mothers' emotional expressions in face-to-face exchanges. Faced with the expression of joy, infants responded by intensifying their own expressions of joy and interest; accompanied by a decline in oral responses. Oral responses, however, increased when the mother showed an expression of sadness. In addition, when the mother presented an expression of anger, infants reduced their movements and showed an increase of anger. Colwyn Trevarthen argues that what the infants are doing in the interactions with the mother is engaging in a *communication of emotions*. Even in the first weeks of life, newborns apprehend the intrinsic motives in their mother's attempts at communication. "It is communication, made with emotions of pleasure, interest, surprise, etc., as the baby intently watches and listens" (Trevarthen, 2002, p. 167). A short time later, in protoconversations, infants share timing and narratives of emotion. Trevarthen (1993a) has used the term "emotional narratives" to refer to the spontaneous fluctuations of energy and emphasis that mother and infant share. In emotional narratives infants reveal their emotional state and perceive that of their mother. They take an active role in engaging in chatter with their mother, chatter in which the tone expressed by each of them is not identical, but *reciprocal*.

There are more features in the expressive communication between infants and the mother than those of shared information, attunement, and reciprocity. There is a temporal organisation and quality of vocalisation in mother–infant interaction that has been described by Malloch (1999) as a "communicative musicality". Gratier and

Trevarthen (2008) state that "(a)n intrinsic timing or pulse of expression […] drives the emotional energy of the moving subject" (p. 131). Infants show a precocious sense of rhythm and an interest in the qualities of expression and their mothers' sounds and gestures (Trevarthen, 1974). There is conclusive evidence that infants of two months of age are able to *predict the timing* and emotion of the mother's expressions when they engage in expressive communication. This evidence comes from a number of experiments conducted by Trevarthen (1974, 1977) as well as by other researchers, such as Malloch (1999; Malloch & Trevarthen, 2009), Murray (1980; Murray & Trevarthen, 1985), Tronick (1989; Tronick, Als, Adamson, Wise, & Brazelton, 1978) and Nadel (Nadel, Carchon, Kervella, Marcelli, & Reserbat-Plantey, 1999), among others.

Contrary to the idea defended by constructionists such as Kenneth Kaye (1982) about the "as if" illusion explained above, experiments such as the "still" or "blank-face test" or "perturbation experiments" demonstrate that infants actually respond to their mother's interaction and expect a proper answer from her. In still-face experiments, mothers are asked to start a happy communication with her child and after a while stop their expressions and maintain an inexpressive face looking at the infant and not talking or responding in any way. In response, the infant first asks for reactions by smiling, gesturing, and vocalising, then becomes sober, withdrawing eye contact and returning to seek it, and finally gets distressed, sometimes crying. The results of these experiments show that what the infant detects are not mere contingencies but the emotional relations of interpersonal acts. Moreover, a lack of response is not only noticed but *matters* to the infant. An experiment by Murray (1980) to test the effect on the infant of inappropriate timing of the mother's behaviour shows that timing also matters to the infant. Murray gets infants from two to three months of age to interact with their mothers through television monitors. Mother and infant are placed in different rooms and the images of the two partners are relayed via video recorders to the monitor. They can see each other in the monitor full-face and life-size, as in a direct interaction, with perfect eye-to-eye contact being possible. They start then to interact as usual. Despite the strangeness of the technological mediation the interaction develops in a similar way as in a direct condition, with happy vocalizations and facial expressions. Once this happy interaction has been established, the images of the monitor are covertly rewound and, without the mother and infant realizing it, replayed to the infant. The reactions of infants in the replayed condition (where the mother's behaviour was the same as in the live condition but not longer responsive and appropriately adjusted to the infant's on-going activity)

were very similar as those of the still-face experiments: infants look away, become sober and try to regain an adjusted interaction showing indications of protest or distress. These results demonstrate that it is not just an odd behaviour that the infant detects but also the *inappropriateness* of the responses (Reddy, 2008, 2010). According to Hobson (1993) it is not only a question of coordination of mother and infant behaviour; there is also a psychological bond present in the face-to-face interactions between infants and mothers "that when established—or when broken—has psychological consequences for both participants. In a sense, the infant (or some "mechanism" in her) seems to expect the appropriate forms of expressive, bodily, and dynamic responses from the other person" (Hobson, 1993, p. 37). Indeed, the mothers, who were unaware that the infant was no longer viewing their live image, also reacted with perplexity, becoming less engaged when detecting that the infant behaved oddly and was no longer engaging with them. Kokkinaki (2010) conducted an experiment in which the fluctuations in the expressions of children aged two to six months and their parents were micro-analysed through recording the emotional reactions in free interactions between them. Kokkinaki installed cameras in the homes and proceeded to the recordings in the habitual environment of the baby in the afternoon schedule, when infants were likely to be relaxed and alert. The results of the experiment proved the existence of emotional coordination, the coincidence of dyadic facial expressions and/or attunement of dyadic emotional intensity, between infants and parents during protoconversations and preceding or following pauses in spontaneous dyadic interactions. An experiment by Malloch (1999) shows that mother-infant communication manifests rhythmic patterns of engagement that could be represented as 'musical' or 'dance-like'. This "communicative musicality is vital for companionable parent/infant communication" (p. 29). There is "pulse", a regular succession of expressive events through time, and "quality" which, when combined, give shape to narratives of expression and intention. The narrative form shows a modulation of rhythms and expression composing *introduction, development, climax*, and *resolution*, with rhyming vowels at key points, and infants engage with all of it with anticipation (Delafield-Butt & Trevarthen, 2015; Malloch, 1999; Trevarthen, 1999, 2008). Moreover, "these 'musical' narratives allow adult and infant, and adult and adult, to share a sense of sympathy and situated meaning in a shared sense of passing time" (Malloch & Trevarthen, 2009, p. 4).

According to the second-person interpretation, these findings imply that infants are born seeking harmony in their connections with adults and that protoconversations

can be understood to entail intersubjectivity. It can be argued that infants have an ability to detect and exchange significant expressive information. Two-month-old infants have no need for a conceptual awareness of emotions and intentionality, but rather an interactive awareness. It is in the context of interactions, of shared expressions, that infants can be aware of the meaning of the expressions and of the intention of the mother to communicate. The emotional detection by infants in face-to-face interactions and their attempts to influence and maintain the mother's affective state show that they have a non-conceptual understanding of the psychological significance; that they perceive expressions meaningfully, as directed at them; that infants recognize the mother's intention to communicate; and that they are involved in genuine communication with her, namely, in mutually manifest communication. The infant is "draw into the mental life of their caregiver *through* her perception of and responsiveness to the bodily expressed attitudes of the adult" (Hobson, 2002, p. 271).

How can we conceive of expression to understand how this kind of communication is possible? As shown above, individualists accounts based on the first or the third person assume that 'mutual awareness' implies that both partners in the interaction have conceptual awareness of intentionality and that 'psychological features' are invisible objects only knowable through inferences. According to this view, mental states and expressions exist in different ontological realms, and, therefore, it is not possible to have direct access to other's mind. There is, however, another way to interpret this that leads to a different explanation of the phenomenon.

On the one hand, Delafield-Butt and Gangopadhyay (2013) claim the existence of a *pre-conceptual and pre-reflective bodily intentionality* in newborns (and even in the foetus) that they call "sensorimotor intentionality", understood "in terms of prospectivity—the idea that control of movement is necessarily geared to the future—in an embodied agent's interactions with the world" (p. 401). According to their view, there is a structural continuity between the emergence of sensorimotor intentionality and the structure of mental states as intentional or, in more advanced forms, content-directed.

On the other hand, to say that infants (and adults) have a direct access to their mother's mind does not mean that her mental states are available *by means of her* expressions. What it means is that expressions are *constitutive* of mental states. To be constitutive, in this sense, means that expressions are not *related to* mental states but

that *they are a condition of them*[6]. Expressions are the *communicative conditions* of mental states, the communicative mode of being of mental states. There is no gap between mental states and expressions. Tears are not a sign or symptom of sadness but a condition of it. Mental states are not in another ontological realm[7]. Facial gestures, body movements, tone of the voice, are the communicative conditions of mental states. Infants see, hear[8], and feel the mental states of their mother as well as perceive her intention when interacting with her. It is this exchange of expressions between the infant and the mother that makes the communication mutually manifest.

In my view, understanding expressions as the communicative condition of mental states and, by extension, understanding protoconversations as expressive communications allows understanding the latter as entailing intersubjectivity.

In this section I analysed infants' ontogenetic development during their first months of life by contrasting a second-person approach with an individualistic one, showing that the former fits and explains both neonatal imitation and protoconversations in a better and more parsimonious way than the latter.

The second-person approach does not place excessive demands on infants' cognitive abilities. In explaining neonatal imitation, the second-person approach shows that newborns do not need *to learn* to recognize the similarity of themselves to their mother and that there is no gap that needs to be bridged with the making of inferences. Infants are not just detached observers who see their mother as someone to understand. They are born *ready for* and *actively seeking* interpersonal experiences. Infants are aware of themselves in relation to their mother, namely, as detached from her but within the same interactive interpersonal unit. Thus, it can be argued that interpersonal, face-to-face interactions are a *constitutive* feature of newborns inner life. Additionally, the

---

[6] I choose to use the term 'condition' rather than 'aspect' because, although they have the same meaning in this context, the term 'aspect' has a different sense in Wittgenstein, which I want to avoid. According to Wittgenstein, 'noticing an aspect' is related with seeing something *through* seeing another thing: "I contemplate a face, and then suddenly notice its likeness to another. I see that it has not changed; and yet I see it differently. I call this experience 'noticing an aspect'" (1953, part II, section xi, p. 193).

[7] As Wittgenstein highlights: "The epithet 'sad', as applied for example to the outline face, characterizes the grouping of lines in a circle. Applied to a human being it has a different (though related) meaning. (But this does *not* mean that a sad expression is *like* the feeling of sadness!". (1953, part II, section xi, p. 209)
There is not relation of similarity, resemblance or of any other kind. Feeling of sadness and expression of sadness are in the same ontological realm because they are two conditions of one and the same thing. The former is the internally felt condition while the latter is the communicative condition of sadness.
It is worth noting that this does not imply that the expression has to be always there. For instance, sometimes we do not feel fatigue but it is expressed in our face, and conversely sometimes we feel pain without it being expressed.

[8] Regarding the possibility to hear a mental state Wittgenstein states: "Think of this too: I can only see, not hear, red and green, —but sadness I can hear as much as I can see it." (1953, part II, section xi, p. 209).

second-person approach explains not only how but also *why* infants imitate. As I have shown, the reason why newborns imitate is that they are *responding* to the mother's attempt to communicate, as well as expecting a response from their mother to their own interpersonal, communicative attempt.

Furthermore, the second-person approach explains the communication between infants and mothers without appealing to capacities for conceptualization. The characteristics of protoconversations, the way in which infants behave during their course, show that infants are aware of communicating and that communication is *mutually manifest* for both the infant and the mother. By understanding expressions as the *communicative condition* of mental states, we can argue that, in protoconversations, infants and mothers engage in a *meaningful expressive communication* which reveals that infants recognize the psychological qualities in their mother without making any inferences or having any mastery of concepts. The mechanisms that sustain mutuality in protoconversations are personal-level, conscious ones, and they can operate in the absence of reflective concepts (Eilan, 2005). It is the conscious exchange of expressions between the infant and the mother which makes the communication mutually manifest. Infants can already grasp whether the mental state they see in their mother is the same mental state that they feel because seeing and feeling a mental state are two ways of grasping one and the same thing. When infants see (or hear) their mother's mental state while interacting with her, they are aware of it as communicated to them. Hence, in their interactions with their mother, infants directly recognize not only the mental state but also that it is directed at them. They grasp that there is a motivation, an intention to communicate, because they are aware of the mental state in its communicative condition, which moves them to respond. Moreover, in protoconversations, infants and mothers share informational content: they are involved in interactional and emotional attunement-timing and mutual affect regulation; they manifests rhythmic patterns of engagement; and they show coherency and reciprocity, appropriateness of the emotional responses, and repertoire of expressive communicative acts. Therefore, contrary to what cognitivists and constructionists argue, adopting a second-person approach allows for a plausible understanding of protoconversations as instances of genuine communication involving *intersubjectivity*.

In the next section I will describe the period from protoconversations to the emergence of the triangles of joint attention and the beginning of secondary intersubjectivity. I will focus on the concept of attention and the consequences the way

we interpret it have for the understanding of both the phenomenon of joint attention and the emergence of the first- and third-person perspectives.

4. Secondary Intersubjectivity: Joint Attention

Just as in the case of communication, the phenomenon of attention is also a controversial topic among researchers of infant's development. Infants' understanding of mother's attention is generally considered to arise when they begin to engage in triadic attentional engagements, i.e., when infants redirect attention to match the mother's focus of attention to another object or try deliberately to direct the mother's attention towards a target. However, some researchers argue that understanding of attention begins much earlier in the infant's development, namely, in the context of dyadic interactions (Reddy, 2003, 2005, 2008). Adamson and Bakeman (1991) differentiated two kinds of shared attention between the infant and the mother. First, they call *mutual attention* the attention that takes place between the infant and the mother when both are attending to each other. Second, they call *joint attention* the shared attention of the infant and the mother to an object, what Bates calls the "third element" (Bates et al., 1976). The debate runs then between those who place the infant's understanding of attention in joint attention and those that place it in mutual attention.

In this section, I will contrast Tomasello's approach, which is based on a *cognitive* revolution in infants at the age of nine months that allows them to rationally understand others as intentional agents, with the second-person approach I advocate, to which there is not a *cognitive* revolution in infants at the age of nine months, but an improvement in the development of their visual perception that allows them to grasp the mother's expressions as directed to an external object. I will argue that the second-person approach explains this phenomenon in a more natural and appropriate way because it advocates for the continuity between mutual and joint attention and do not need to postulate the existence of a *cognitive* revolution.

I will also argue that the problem of children with autism in engaging in triangles of joint attention is not a question of a lack of the capacity for triadic representations but a problem they also have in dyadic interactions, namely, a deficient one-to-one interpersonal engagement. It is because of this that they may need to build a theory to understand the other's behaviour.

Finally, I will defend that it is through second-person interactions (through the second-person perspective) that infants become able to understand both the first- and the third-person perspectives.

In the next section I will begin with a description of the experimental data about the behaviour of infants from the beginning until joint attention and, then, I will examine the understanding of attention on the period of secondary intersubjectivity, the point at which joint attention takes place in the infant's development.

4.1 The Development of Attention in Infants: Some Empirical Data

In the first weeks of life newborns only share attention in a dyadic way with people or with objects. However, they clearly show preference for looking at people, in particular, faces and face-like stimuli (Batki, Baron-Cohen, Wheelwright, Connellan, & Ahluwalia, 2000; Johnson & Morton, 1991). Moreover, they prefer watching their mothers' faces rather than the faces of strangers (Bushnell, Sai, & Mullin, 1989). Experiments by Farroni and her colleagues (Farroni, Csibra, Simion, & Johnson, 2002) show that newborns prefer to look at faces that engage them in mutual gaze. They seek to find and keep eye-to-eye contact with people.

At the age of three months infants smile in response to eye contact and decrease smiling when the other's gaze is averted (Hains & Muir, 1996; Wolff, 1987). Indeed, the onset of mutual gazes is a compelling elicitor of smiles (Reddy, 2008). From three to four months of age infants begin not just to seek for the gaze but to 'call' for attention. Through different vocal tone and intensity—as those that they use when attention is already received—they call their parents when they are not present or present but inactive, stopping this 'calling' when they recover attention (Reddy, 2008). From three to five months of age, infants begin to engage in a new and significant activity. They start following the gaze of the mother when interacting with her, sifting their gaze to other people who are in their field of vision. However, this displacement is only oriented towards people[9]. At this age infants are not able to coordinate triadic

[9] As noted in Section I, infants' behaviour differs in a singular way in their relation with people. Infants' identifying with people "involves *feelings and attitudes*" (Hobson, 2002, p. 105), something that does not occur with objects.

interactions, they just shift the attention to others and engage in dyadic interactions with them. They are still engaging only in mutual attention.

At the age of six months, direct gaze works as a form of ostensive communication and boosts the frequency of successive gaze following (Senju & Csibra, 2008). Infants' attention begins, then, to be related to the physical objects surrounding them and to the objects to which the mother attends, which they are now able to reach and explore with their hands (Fogel, 2011). Infants follow their mothers' gaze to objects in front of them, nearby or in their mother's hands. Some researches suggest that, already at the age of three to four months, the object in a mother's hand becomes somewhere a three-person triangle in which the third 'object' of attention is the hand-with-object (Amano, Kezuka, & Yamamoto, 2004). However, they do not consider this kind of interaction as joint attention but as the precursor of it. There are two main reasons why they are not considered as joint attention. First, infants continue looking at the adverted mother for a few seconds before they turn to look at the object held by the mother. So, there is not a following gaze as it appears typically in cases of joint attention. Second, infants do not care about the object if the mother's gaze remains on them and they do not look at the hand if there is no object in it or was doing nothing interesting (Reddy 2008). So, there is not a deliberative intent by infants to engage in joint attention.

From seven to nine moths of age infants are able not only to follow mother's gaze towards objects but also to coordinate interpersonal actions and attitudes with her. However, they still remain engaging in dyadic interactions, i.e., there is only mutual attention between them. During the following three months infants improve the skills that they already have and develop a new and significant one. On the one hand, infants become more skilled at following gaze, becoming able to locate the object of the mother's gaze among several situated in the same direction (Woodward, 2005). On the other hand, they begin pointing to seek to manipulate the attention of the mother checking to see whether their actions have been successful (Bates, Benigni, Bretherton, Camaioni, & Volterra, 1979; Eilan, 2005; Franco, 2005). These kinds of pointing are the so-called *proto-declaratives* and *proto-imperatives* (Bates et al., 1976; Bretherton et al., 1981). It is said that infants make a proto-declarative pointing when they want to show things to the mother and are satisfied when she engages with their attention to the object. In the case of proto-imperative, infants demand an object and they are satisfied only when they get it. Although we can find infants pointing as young as three months,

this kind of pointing is not considered a pointing gesture but a 'pointing hand posture'. The difference is that three-months-olds do not point to any significant aspect of the environment. The gesture of pointing as a deliberative intent and with a specific target emerges in infants at the age of about 12 months (Franco, 2005). It is then, when proto-declaratives and proto-imperatives come into play, that we can affirm that *joint attention* begins to manifest itself, that is, when the fact that both the infant and the mother are attending to the same object is mutually manifest and done deliberatively and with an specific target (Eilan, 2005).

After this description of the empirical data of infants' development of attention I will explain the development of triangles of joint attention. I will show how the way in which we interpret infants' understanding of attention leads to different explanations of this phenomenon.

4.2 The Idea of Joint Attention

The beginnings of the study of the phenomenon of joint attention date back to the middle of the 70's on the basis of Jerome Bruner's works, who pointed up the fact that "[l]ittle is know about how visual attention of the mother-infant pair is directed jointly to objects and events in the visual surround during the first year of life" (Bruner & Scaife, 1975, p. 265). Although this phenomenon was initially underestimated—taken for granted—, thanks to Bruner's work (Bruner, 1977, 1995) it is nowadays considered by many philosophers and psychologists to play an indispensable role in the study of both linguistic and conceptual development of toddlers. Bruner drew attention to the developmental psychologists' disinterest on epistemological questions. Questions such as "How infants move from mutual to joint attention?" or "How they come to understand other minds?" "never entered in the discussion" (Bruner, 1995, p. 1). According to Bruner, we must face these questions in order to highlight the understanding of the ontogenetic and the conceptual development of infants. We have to find out whether young children understand that they are involved in episodes of joint attention and, if so, the way they arrive at the understanding of sharing a common experience. In Bruner's words:

> [D]oes the young child naturally 'understand' (after a while, anyway) that she and her mother are looking at something together, sharing a common experience? […] But could any infant, or anybody for that matter, ever learn from scratch, from experience

> alone, that somebody was looking at something, and that it was the same thing the infant was looking at? (Bruner 1995, p. 2)

The way we answer these questions yields different interpretations of the understanding of the phenomenon. One would understand the child's developing the mental capacities involved in the awareness of attention—and hence in the understanding of others' minds—as a matter of theorizing about other's observed behaviour by assuming a contrast between the first and the third person. Another interpretation would see it as emerging in the context of social interaction, assuming a second-person approach.

In what follows I will consider and contrast two approaches to this phenomenon. I will begin with a description of joint attention to frame the debate, followed by a description of Tomasello's approach, which is based on the contrast between the first and the third person and for which the capacity to engage in joint attention implies the occurrence of a mental revolution in infant's development. Finally, I will describe the second-person approach I advocate for, in which there is continuity between dyadic and triadic modes of intersubjective engagement, i.e., continuity between mutual and joint attention.

The most comprehensive of joint attention's definitions given so far we can find in Elian's work:

> As I will understand the term 'joint attention', to say of an event that it is an event of two subjects (or more) jointly attending to the same object is to be committed, at least, to the truth of the following four claims about the event.
>
> a. There is an object that each subject is attending to, where this implies: (i) a causal connection between the object and each subject, and (ii) awareness of the object by each subject.
> b. There is a causal connection of some kind between the two subjects' acts of attending to the object.
> c. The two subjects' experiences exploit their understanding of the concept of attention.
> d. Each subject is aware, in some sense, of the object *as* an object that is present to both subjects. There is, in this respect, a 'meeting of minds' between both subjects, such that the fact that both are attending to the same object is open or mutually manifest.
>
> Let us say that when these conditions are met, we have in play a 'joint attention triangle'. (2005, p. 5)

Not all theorists of the phenomenon of joint attention assume that this definition can be applied in the case of childhood. The question is how much of the conditions cited above we can attribute to a child between one and two years of age. Theorists might assume that at that age joint attention involves only causal factors, accepting only points (a) and (b). This would be a radically 'lean' theory of joint attention in which there is no

understanding by infants of the concept of attention. Other theorists would be committed to a definition encompassing the (a)-(c) points accepting that, in addition to causal relationships, there is some understanding of attention but not sufficient to make it mutuality manifest. In these cases we would have a modest rich account of joint attention. Finally, theorists would admit that infants between the ages of one and two meet all these conditions. In these cases we would have a rich theory[10] of joint attention in which both infants and adults were aware of each other as intentional agents whose attention is directed towards the same object (Eilan, 2005).

For the matter at issue, the two accounts I will describe can be considered as rich theories of joint attention. In both the first- and the second-person accounts infants are assumed to be aware of the object *as* the focus of the other's attention and of the sharing of attention towards the object in question. These accounts "hold that the kind of casual co-ordinations of attention we find in among 1–2–year–olds provide for mutual awareness" (Eilan, 2005, p. 6). Let's see how these theories explain these capacities of infants to engage in joint attention.

4.2.1 From Mutual to Joint Attention

According to Tomasello (1993, 1999) between the ages of 9 and 12 months a cognitive revolution occurs in infants that allows them to understand others as intentional agents. It is then when they start engaging in joint attentional interactions. In Tomasello's words: "infants begin to engage in joint attentional interactions when they begin to understand others as intentional agents like the self" (1999, p. 68). Tomasello's explanation of this phenomenon relies on the attribution to infants of two abilities. One is what Tomasello, borrowing Gopnik and Meltzoff's label, describes as the 'like me' stance: in attempting to understand others infants apply what they already experience of themselves. Thus, around nine months of age infants begin to understand themselves as intentional agents and they apply this understanding to others

The main difference between Tomasello's and Gopnik and Meltzoff's explanation of the 'like me' stance relies on the way they describe infants'

[10] To avoid any kind of ambiguity, the sense in which I will use the notion of 'rich theory' implies that in these theories causal co-ordinations of attention of infants between one and two years olds provide for mutual awareness (see Eilan, 2005, p. 6).

understanding of themselves and of the others' self. Although Tomasello's and Gopnik and Meltzoff's approaches are what I described above as individualistic approaches, Tomasello argues for a first-person understanding of infants' development rather than for the third-person understanding defended by Gopnik and Meltzoff. According to the latter (1997), infants come to understand that other persons are 'like me' from birth through the imitation of others and the proprioception of their own acts –although the precise age at which it emerges and how many personal experiences are needed remains unclear. Once they 'know' the others are similar to them they begin to develop an 'agency theory' based on direct observation of and inferences about the others' behaviour that they apply both to others and to themselves (see Section II.1 above). However, in Gopnik and Meltzoff's account the 'like me' stance does not play a real role in the process of joint attention. On the contrary, in Tomasello's account the 'like me' instance does play a central role because it is a key element together with the second ability (the distinction between means and ends).

The second ability that comes into play and Tomasello attributes to infants to explain the phenomenon of joint attention is the ability to distinguish means from ends. At the age of nine months and due to a cognitive revolution—what he calls "the Nine-Month Revolution" (1999, pp. 61 ff.)—infants become able to differentiate goals from behavioural means in their own sensory actions. They come to understand that they can use different behavioural means to the same goal, to choose a given method—including intermediaries—to achieve a specific goal. It is by combining these two abilities that infants are able to understand others as intentional beings and to begin to engage in joint attention. The idea is that since infants have a sense of the other as being "like me" they use their new acquired ability—the understanding of themselves as intentional beings—in simulating others' intentional actions. Tomasello's approach "may thus be thought of as one version of simulation model in which individuals understand other persons in some sense by analogy with the self" (Tomasello, 1999, p. 70).

According to the second-person approach, the infant's capacity to engage in joint attention begins and develops over the earliest months of life. The kind of psychological engagement that infants need to participate in joint attention is already present in the early dyadic interactions of mutual attention, in the attention that takes place between the infant and the mother when both are attending to each other. As showed above (see Section 3) the engagement between infant and mother, the intention to communicate, is mutually manifest. Infants' understanding of the mother's

expression is neither the result of an intellectual deduction nor an empathic recognition but something that they grasp directly as *directed at them*, something that *moves them to respond*. Infants do not just grasp that the mother is attending to them and expressing an emotion, they grasp the directedness of the mother's expression as an attempt to communicate—since the expression is the communicative condition of her mental state. Infants feel engaged in *expressive communication* and, consequently, they respond through the expression of their emotions. Now the question is: why infants come to engage in joint attention? How do they move from mutual to joint attention? Instead of a *cognitive* revolution I suggest that at the age of nine months, since infants have become more skilled at following gaze, being able to locate the object of the mother's gaze among several situated in the same direction because of the improvement of their visual perception (Franco, 2005; Woodward, 2005), and they already have a familiarity with the mother, as all infants have at that age, they become able to capture her expressions both towards them and towards the objects of the shared external world with which she interacts. By following the mother' gaze infants grasp that her expressions are related to something external to them and that somehow the external object plays a role between them. They are now involved not just in a direct interaction but also in an interaction mediated by a third element, i.e., they engage not only in dyadic interactions but also in triadic ones.

The main question now is: which is the particularity of joint attention? What comes into play in joint attention that is not present in mutual attention? In order to clarify this question it will be helpful to refer to the new activity that infants develop at the age of nine months and that arises in line with joint attention: the intentional pointing.

4.2.2 Intentional Pointing: Proto-Imperatives and Proto-Declaratives

Characterizing the difference between the two kinds of pointing in terms of means-ends relations, proto-imperatives can be described as infants using the adult in order to get the object, while in proto-declaratives infants use the object to get the adult's attention/response (Bates et al., 1976). Some theorists have interpreted this difference as the difference between understanding of action and understanding of mind (Baron-

Cohen, 1995; Franco, 2005; Tomasello, 1999). As Franco puts it "the social partner is attributed physical agency (agent of her action) in proto-imperative acts, but mental agency in proto-declarative acts (e.g. agent of her own attention)" (2005, p. 132). This interpretation relies on the assumption that mental states are invisible objects that cannot be perceived but are instead theoretically conceived or understood by the infant (see Section 3.2 above). Infants are supposed to form mentalistic interpretations of others' actions according with the theory of mind they are supposed to have rather than directly grasping the mental state and being moved for the directedness of the expression as the second-person approach sustains.

This is also the reason why Tomasello and other theorists interpret the lacking of the development of proto-declaratives in children with autism. While non-autistic children engage in both proto-declaratives and proto-imperatives, autistics develop only proto-imperatives (Gómez, 2005). Indeed, children with autism appear to have significant difficulties engaging in joint attention (Franco, 2005). These theorists assume that the absence of proto-declaratives and of joint attention behaviours in children with autism indicates damage to the cognitive 'mindreading' system that impairs the ability to form mentalistic interpretations of other's actions. Children with autism are supposed not to be capable of constructing triadic representations that specify joint attention–i.e., self and other attending together to the same object (Baron-Cohen, 1995). In Tomasello's words: "the difficulties of children with autism in understanding other persons as intentional agents leads to deficits in their symbolic skills, which then may create difficulties in representing situations perspectivally" (1999, p. 133).

In contrast, the second-person approach views the absence of joint attention behaviours in children with autism as a problem of the shared nature of experience, the 'jointness' of joint attention rather than as a lack of the capacity for triadic representations. The difference between children with autism and other children lies not only in triadic interactions but also in dyadic ones—not only in joint attention but also in face-to-face interactions (Leekam, 2005; Moore, Hobson, & Lee, 1997; Reddy, 2005). Children with autism do not manifest basic expressive communicative acts such as greeting the parents and waving by raising the arms to be picked, enjoying and participating in lap games or sociability in play. They fail "to display not only signs of joint attention, but also signs of one-to-one interpersonal engagement" (Hobson, 2005, p. 195). In accordance with the hypothesis that I am defending, I suggest that a basic

requirement for infants to be capable of engaging in joint attention is the capacity to engage in expressive communication. Infants go from mutual to joint attention by following the mother's gaze and grasping the directedness of her expressions. Children with autism are no capable of doing that. Although they grasp the mental state expressed they are not aware of the mother's attempt to communicate. The directedness of the expressions of the mother does not move them to respond because they do not join in the psychological engagement necessary for expressive communication. Hence, they cannot be aware of the directedness of her expressions towards the objects of the shared external world with which she interacts.

One of the basic characteristics of interpersonal engagement that is absent in children with autism is the mutual affect regulation that takes place during expressive communications. Children with autism are reported not to relate their emotions with the emotions of the mother neither to relate their emotions with the behaviour of others. An experiment carried out by Moore, Hobson and Lee (1997) shows that children with autism do not interpret other persons in terms of emotions, they do not refer to emotions to explain the behaviour of others but to movements and actions. Children with autism do not see the mother as someone with whom to *tune-in* but someone to understand. Sue Leekam (2005) sustains that children with autism start out in basic impairment in orienting to others without meaning that they are unable to develop any joint attention, although the quality of sharing might still be missing. According to Leekam it is possible that "they make another route to the development of linguistic symbol use and to the acquisition of joint attention" (2005, p. 225). Contrary to the standard way of understanding the issue, I suggest that it is in children with autism rather than in other children where the third-person approach *could* make sense[11]. Due to the absence of the capacity to engage emotionally with the mother—of the capacity to see her as someone to *tune-in*—children with autism are not moved by her expressions. They may find necessary to build a theory to understand *why* the mother expresses what she expresses––the mental state expressed—and use this theory to make inferences of the observed behaviour. They may need to build a theory to understand the reasons why others behave as they do.

[11] According to Hobson (2005), children with autism "do not appropriate another's person mode of action *through identifying with the other person*. […] (E)ven when the child with autism copies the strategy by which a goal is achieved (something that appears to be beyond non-human primates), this is accomplished not by moving to the position of the other and appropriating the person's style and mode of dealing with the problem, but rather by 'watching from outside' and adopting the requisite strategies to achieve a goal" (2005, p. 198).

Returning to the main question, in Tomasello's account the particularity of joint attention lies in the infants' cognitive revolution that allows them to *rationally* understand others as intentional agents. According to Tomasello, in order to share attention infants have to *interpret* other's behaviour in terms of shared intentions. Sharing attention implies, then, the idea that infants have a primitive theory of mind. Joint attention is understood as a form of rational cooperation in which the infant's agentive conception of the mind does all the work. Tomasello sustains that the understanding of a communicative intention must have the same structure as a Gricean communicative intention: to understand the adult's communicative intention when pointing, the infant must understand that "You intend for [me to share attention to (X)]" (1999, p. 102).

On the face of it, it seems that this account of 'sharing' faces a significant problem. Although Tomasello recognizes that "infants' understanding of all of this is still not fully adult-like" (2008, p. 132), assuming the Gricean structure in infants of 12 months of age implies some rational requirements that are quite implausible for infants to have them at that age. As Roessler says, "to interpret proto-declarative joint attention as a matter of expressing, recognizing, and acting on communicative intentions is to over-mentalize the phenomenon" (2005, p. 242). Roessler points out that in assuming the Gricean structure Tomasello shares a crucial feature with Grice's account of assertion. Roughly, this feature consists of the gap between recognizing that the other intends for the infant to share attention to X and the infant forming the intention to attend to X. Tomasello says that "by 12 months of age they also understand that actors choose to attend intentionally*, for some reason*, to some subset of things they perceive" (Tomasello, 2008, p. 139, italics added). This claim implies that infants form the intention of attending to the other *for some reason*, which means they have acquired not only the beginnings of means-end reasoning but also "the ability to weigh up potentially competing practical projects—for example, to decide, in the light of one's own desires and plans, whether to accept an invitation to share attention with someone else" (Roessler, 2005, p. 240). Infants are, thus, supposed to be able of this kind of practical reasoning. This seems to endow infants with rational abilities that far exceed those of an infant of 12 months of age.

On the contrary, to say that the infant's understanding of the mother intention is a part of *what* is for her to be aware of the directedness of the expression does not commit to the idea that such awareness implies the capability for this kind of practical

reasoning. As I have argued, infants go from mutual to joint attention by following the mother's gaze and grasping the directedness of her expressions towards an external object. The mechanism that allows infants to engage in joint attention is the capacity to engage in expressive communication, which is already present in mutual attention. Infants begin to relate in a triadic interaction without being necessary to postulate a *cognitive* revolution in them. By following the mother's gaze they grasp that the expressions that used to be exchanged in dyadic expressive communication now include a third element.

The question is: what comes into play in joint attention that is not present in mutual attention? As we have seen, to understand the difference between proto-imperatives and proto-declaratives in terms of the difference between understanding of action and understanding of mind leads, in Tomasello's account, to an over-mentalization of the phenomenon of joint attention. There is, however, another way to understand the difference between proto-imperatives and proto-declaratives that results in a different explanation of this phenomenon. It seems clear that when infants point to an object in a declarative sense (proto-declarative) they do not do this in order to obtain or manipulate the object as they do when they point in an imperative way (proto-imperative). In accordance with the second-person approach that I am advocating, I suggest that what infants are doing with proto-declaratives is treating the object as an element of contemplation as well as including the object in the expressive communication. In dyadic interactions the infant and the mother engage in an expressive communication in which both participants are aware of their mutual attention. Both the infant and the mother are aware of being attending to each other. They exchange vocalizations and expressions with one another and engage in mutual affect regulation. In triadic interactions, when the third element comes into play, the expressive communication goes from being a mutual exchange of attention and affective regulation to a sharing of an appreciation of what the object—to which they are both attending—is like. While in mutual attention infants are moved to respond to the mother's attempt to communicate, in joint attention infants are moved to share the mother's and their own experience towards the object. The difference between dyadic and triadic interaction is the difference between *to respond to* and *to contemplate and comment on.* The object—the third element—comes into play in an indirect way becoming an object of contemplation as well as an object to comment on.

Roughly, the process develops as follows: the infant and the mother are engaged in expressive communication and, then, the third element comes into play as I have described—the infant follows the mother's gaze and grasps the directedness of her expressions toward the object. The infant engages with the object through this process, not in a direct but in an indirect way, what means that she *distances* herself from both the mother and the object. The object *being mediated by the mother* becomes not only something to engage with directly but something to *contemplate* together as well as something to *comment on*. Once the infant becomes familiar with this process, from then on she will use proto-declaratives to engage with it: to 'ask' about the world, to comment her impressions about it and to regulate her own impressions with those of her mother.

The idea of a contemplative stance goes back to Werner and Kaplan's discussion of "the primordial 'sharing' situation" (1963, pp. 42ff.). According to them "there arises *reference* in its initial, nonrepresentational form: child and mother are now beginning to contemplate objects together […] a fundamental transformation in the relation of person and object occurs with the shift from ego-bound things-of-action to ego-distant objects-of-contemplation" (1963, p. 43–44). It is worth noting that Werner and Kaplan argue that the contemplative instance takes place in the context of joint attention, "the act of reference emerges not as an individual act, but as a social one: by exchanging things with the Other, by touching things and looking at them with the Other" (1963, p. 43). The contemplative stance contrasts with a purely practical one. Infants perform proto-declaratives not in order to manipulate or obtain the object but in order to contemplate it together with the mother. By means of proto-declaratives infants want the mother to show that there is something outside, a third element that somehow has modified their relation: a *distancing* has occurred between the infant and the mother and between the infant and the object.

With regards to the idea that the declarative pointing should be seen as the beginning of commentary on the world, results of experiments made by Franco (2005) show that the infant typically produces declarative pointing in order to share attention with her mother about the object and *exchange comments about it*. As she puts it: "with proto-declarative pointing, the main goal is to bring someone else's attention to the referent object or event that the infant finds interesting. Once this is realized, some exchange of information concerning the referent may take place—for instance, identity

('This is a X'), properties ('Big', 'Red', 'Fast'), actions ('Moved', 'Stopped'), or the self (['I FIND THIS'] 'Interesting', 'Funny', 'Scary', 'Novel', etc.)" (2005, p. 142).

There are, thus, two new activities in which infants engage in joint attention: sharing a contemplation of an object with the mother and commenting on the object– i.e., including the object in expressive communication. The inclusion of the third element generates a progressive distance between the infant and the mother as well as between the infant and the object that modifies their relation. The question now is: what are the consequences of this on the psychological development of infants? Naomi Elian proposes a suggestive answer to this question. According to Eilan (2005):

> Introducing a third element requires treating her (the mother) as someone who can have a take on the world, can have a perspective on it—where the differentiation comes, in the more the child needs to take into account the difference of perspective. Taking the latter into account simultaneously strengthens the grip on the idea of a mind-independent world. Awareness of others as having a different perspective and awareness of the world as being independent of one's perspective come together. (p. 18)

Putting this in other terms means that the infants' understanding of the first-person perspective—their own perspective—and the understanding that others can have another perspective—a third-person perspective—arise at the same time in the development of infants. Although it is clear that they not have yet the concept of 'perspective', infants understand that what they perceive, feel, see, etc., is only a perspective about an independent world. Moreover, infants understand that the perspective of the mother may be different from their own perspective, that there is a separate world that may not be how they 'think' it is, a world about which they (the infant and the mother) can have different perspectives. It can be argued that the beginning of a sense of an objective world and thus of objectivity arises in infants at this stage and that this may be the reason why infants begin to comment on the object with their mother. Therefore, the first- and the third-person perspectives are based on—or constituted by—the second-person perspective, the perspective of interaction.

In this section I described the empirical data on the development of joint attention and the philosophical definition of joint attention. I contrasted Tomasello's approach with the second-person approach that I advocate. Both approaches can be considered as rich accounts of joint attention, i.e., approaches that sustain that the kind of casual co-ordinations of attention we find in infants between one and two years old provides for mutual awareness. I argued that in Tomasello's account infants are endowed with rational capabilities that exceed those we can attribute to them at 12

months of age. Thus, Tomasello over-mentalizes the phenomenon of joint attention. On the contrary the second-person approach explains in a more natural and appropriate way this phenomenon by sustaining continuity between mutual and joint attention. Infants are aware of the third element through the directedness of the mother's expressions. They engage now with the external object not only in a direct but also in an indirect way, which implies that they *distance* themselves from both the mother and the object. Infants start, then, considering the external object as something to contemplate and comment with their mother. This process allows infants to understand both the world as something independent of their own perspective—the first-person perspective—and the existence of other perspectives about it—the third-person perspective.

In the next section, I will explain what is going on in expressive communication when the third element comes into play, the significance of the comments between the infant and the mother. I will argue that the mother becomes a sort of authority for the infant—which I will call 'second-person authority'—and that it is related to the emergence of a basic awareness of normativity in infants.

5. Expressive Communication and Joint Attention: Second-Person Authority and Normative Regulation

As we have seen, experiments carried out by Franco (2005) show that infants begin to comment on the world with the mother once they check that she is attending together to the same object after they have performed a proto-declarative. Infants are now engaged in an expressive communication that includes a third element in which some exchange of information concerning it takes place. As in dyadic expressive communication, where infants catch the difference between attuned or engaged affective interaction, on the one hand, and non-attuned, or non-engaged, interaction on the other (Reddy & Trevarthen, 2004), in triadic interactions infants also catch whether the expression of the mother *matches*, or is *attuned*, or does not match with how they feel about the third element. At this stage infants engage in expressive communication not only to regulate their emotions but also to regulate their perspective about something external to them, something new, to find a particular object interesting or scary, or pleasant, or funny, or

disgusting. Infants want to 'ask' about objects and situations not only to know the mother's perspective—Is it scary *to you*? Is it funny *to you*?—but also the appropriateness of the response to the world—*Should I* be scared or not? *Is this* funny or not? The affective regulation that takes place in dyadic expressive communication turns out in a world—or social—regulation in triadic expressive communication. Now the mother's expression regulates the infant's affective stance—and, therefore, her actions––towards the world.

A series of four studies of the communicative and regulatory functions of emotional expressions indicate that, by 12 months of age, infants seek out and use facial expressions to disambiguate situations (Sorce, Emde, Campos, & Klinnert, 1985). Sorce and his colleagues describe what they have called 'social referencing' as "a process whereby an infant seeks out emotional information in order to make sense of an event that is otherwise ambiguous or beyond that infant's own intrinsic appraisal capabilities" (1985, p. 199). In the experiment, called "Visual Cliff", the infant in placed in order to crawl on a Plexiglas top table at the end of which there is an object of interest (e.g., a toy) and around which is her mother, in a way that the infant has visual access to both. The beginning of the table has a checkerboard pattern just underneath the surface but halfway across there is nothing underneath the surface, so the infant can only see the great height that separates her from the ground without being aware that it is covered by Plexiglas. In the tests carried out some infants did not notice the drop and continued crawling towards the object. Likewise, many others did notice it and, consequently, stopped before it. Of the latter, a large number of them raised their heads in search of their mother's gaze. The results of these studies indicate that, in this situation, 74% of babies continue their way if the mother responds to the infant's eyes with a favourable expression, which communicates a positive value, such as a smile gesture and amiability. On the contrary, if the mother responds with a negative gesture, as an expression of fear, none of them continues their way. When the mother puts on an anger expression only 11% of infants continue, while about 80% go backwards.

These results show that infants around the age of twelve months are able to grasp the expression of their mother, relate it to a certain situation and react emotional and actively accordingly. Infants recognize that the mother's expression has a meaning related to an environment they both share. The psychological bond that is created between the mother and the infant, through expressive communication, enables the psychological appropriation of the value of the events by the infant: I regulate my

emotions with her (it works!), so she can also help me with this new situation. When the infant faces an experience unknown to her, to which she does not have a value assigned, it is the expression she grasps in her mother what makes her act in one way or another. Whereas gaze expresses interest and goal-intention, as gaze following shows, facial expression of emotion communicates attitude and evaluation. (Brinck, 2008; Reddy, 2008). Facial expressions of emotion, tone of voice, body movements, "are not merely responses indicative of internal states, they are also stimulus patterns that regulate the behaviour of others" (Sorce et al., 1985, p. 195). We can also see social reference in a daily life's example: imagine an infant learning to walk who suddenly trips and harmlessly falls. After falling, the infant immediately looks at her mother. If the mother stays calm (perhaps saying "oops!" while making a funny face) the infant will much more likely feel calm about what happened and react less dramatically than when the mother is overtly frightened and concerned (perhaps saying "ouch!" while making a scary face).

Now the question is: how does the infant understand the role of the mother at this stage? As in dyadic interactions, where infants regulate their feelings with the mother, in triadic interactions infants seek an answer to the question "How *should* I feel about that?" to regulate their feelings about the world. The emotional confidence that infants feel towards their mothers makes the latter be the ones who determine for infants how to deal with an unknown situation for which they have no answer or in which they are not sure what course of action to follow. Infants assimilate the value of the information that they obtain from their mothers not just in the descriptive but also in the *normative* sense. From then on, they will respond to similar situations in the same way as learned from their mothers. Thus, it can be argued that the role of the mother at this stage of the infant's development can be described as a sort of *authority*. Philippe Rochat in psychology and Johannes Roessler in philosophy have recently explored the idea of an authority in infants' development when they look at their mother for social reference. According to Rochat (2001), from birth infants are able to discriminate their own body from other entities in the world and from at least two months of age develop an interpersonal sense of themselves and begin to learn what to expect from others as social partners and communicative agents. From nine months of age, when joint attention comes into play, infants recognize others as more knowledgeable and advanced in their skills. At that stage infants begin to understand others as a source of help and instruction, they see them as having an authority:

> In all, it appears that from 9 months of age, infants become explicitly aware that others are potential teachers and informants to solve problems. […] In particular, infants construe others as having *authority* with the power to judge because they know better. (Rochat, 2001, p. 356, italics added)

Roessler (2005) states that when infants look for an appraisal of a perceived object they look not only for information of an object because they do not know what the object is like but also for new aspects of it—or new kinds of evaluative features of it. Infants recognize in some way the mother's authority because she is a potential sharer of knowledge:

> An alternative picture here would be that infants relate to caregivers as to something like an *authority*: a source, and a potential sharer, of knowledge of what objects are like. (Roessler, 2005, p. 248, italics added)

Since this kind of authority takes place when the mother is interacting with the infant through expressive communication and in accordance with the second-person approach I have been advocating, I will call this authority *second-person authority*[12]: the authority that infants confer to the person with whom they usually interact. It is through the mother's expressive answers as the infant gives value to unknown events and also as "some truths may become plain ones for the infant in virtue of her coming to share the caregiver's attitude" (Roessler 2005, p. 249). Interpersonal relations assume a normative significance at this stage of the development of infants. This entails that second-person authority is used by infants not only to regulate their behaviour in the occasional moments in which expressive communication takes place but also to act similarly in future occasions in which they find themselves in similar contexts. Consider the daily life's example showed above where the infant, while learning to walk, suddenly trips and harmlessly falls. Imagine that the answer the infant sees in the mother's expression when she looks at her is a positive one—e.g., the mother smiling says: "It's ok, get up!". After a while, when the infant falls again in the same circumstances she already expects the positive response of the mother and perhaps looks at her to confirm the answer. The reaction of the infant begins, then, to be less dependent of the mother's consultation. The infant no longer seeks to confirm it each time in the mother's

[12] It is worth noting that the sense in which the term 'authority' has to be understood in this context is not related to *first-person authority,* the technical sense used to explain the distinctive characteristic of the subject's knowledge of their own mental states. In the next chapter I will describe how the different accounts of self-knowledge give sense of the term 'authority' in this technical sense.

expression and at some point she does no longer require it. In this regard, Taipale (2016) sustains that what is happening is that the infant *internalizes* the mother:

> Now, along with the increasing capacity to track how she feels, the infant begins to anticipate dyadic mirroring of her affective experiences: parental mirroring is gradually internalized. […] The presence of the actual caregiver in flesh and blood is no longer continually required, because the infant has internalized her and carries her with her, as it were. (pp. 7–8)

The infant comes to know how she has to behave in these circumstances without consulting the mother for an answer because she has already internalized her. The mother has become *the norm* of how she should feel, and therefore behave, in these circumstances.

As showed above, through expressive communication infants perceive contingencies between their own and the mothers' emotions towards a third element—an external object or event. They perceive that the perspective of the mother may be different from their own perspective. Infants understand that here is a world about which they (the infant and the mother) can have different perspectives. This process allows infants to *distance* themselves from both the mother and the third element. At this stage, infants begin to see the mother as an authority and to internalize her as the norm of how they should feel, and therefore act, in such-and-such circumstances. Insofar as infants internalize the mother they are increasingly capable of affect regulation also without interacting with her. That means that the more internalization of the mother, the more distance of the infant from the necessity of interacting with her. Infants gradually become more independent of the dyadic relation with the mother, which enables further dyadic relationships in which infants constitute others as social mirrors. As different people respond differently, infants begin to differentiate between responses and consequently to favour some over others. In Taipale (2016) words:

> By entering the social world, the child no longer *uncritically* takes the caregiver as *the norm* of how one should feel in such-and-such circumstances, and in this sense she grows more independent: by internalizing alternative social mirrors, as it were, she also learns to disagree with the primary mirror. To know how she is expected to feel, the child no longer exclusively consults the (actual or internalized) *caregiver*; her question is rather more social and general: "How am I expected to feel *by these-and-these others*, and how do *I myself* feel?". (p. 8)

At this stage of the development infants understand that there are not only those of the mother but also alternative perspectives to the appropriateness and inappropriateness of

their affective expressions. Infants' social self-awareness begins to develop as well as their sense of independence. Mutual emotional regulation becomes a world regulation, a world that is ruled by *social norms*. Now the development of a basic awareness of normativity in infants goes beyond the mother's authority, becoming an awareness of social norms.

6. Summary

In this block I showed the importance of the second-person perspective in both the ontogenetic development of infants and in the constitution of the awareness of themselves and of the others' perspective. I embrace a second-person approach rather than one of the traditional approaches based on the first or third person.

In the second section, I have briefly sketched infants' ontogenetic development from birth until they begin to be considered as conceptual beings. Section 3 was focused on the period of dyadic interactions. Regarding neonatal imitation I have argued that newborns have some capacity to engage in 'intentional'—non-purposeless—interactions. According to a second-person approach the kind of awareness of the self we can attribute to newborns is of a *self in relation with the other*, a sort of interpersonal unit in which the infant's 'self' is detached from the other but within the same *interactive, interpersonal unit*. Thus, the second-person approach explains in a more parsimonious way the recognition of others by infants since it does not need the bridge between self and other. Moreover, the second-person approach accounts for the behaviour of the infant when imitating since it explains not only *how* but also *why* infants imitate: infants imitate *because* they are responding to the communicative attempt by the mother. The dialogical characteristic of neonatal imitation, the innate sense of intersubjectivity, is shown through the direct recognition by infants of the mother's expressions as directed at them and their ability to 'provoke' her responses as well as through their ability to know when to take turns in the interpersonal interactions with the mother.

With regards to protoconversations I have argued for the existence of an *expressive communication* between infants and the mother—which is *mutually manifest* for both of them—grounded on the infant's ability to detect and exchange *significant*

expressive information with the mother, rather than on the theoretical knowledge of other's minds. Expressive communication allows understanding protoconversations as entailing intersubjectivity. I have put forward a philosophical explanation of how expressions have to be understood in order for this to be possible. Instead of understanding mental states as invisible objects that cannot be perceived but *theoretically conceived* or *understood* by the infant through the mastery of the relation between them by inferences, we should understand expressions as the *communicative condition of mental states*. There is no gap between mental state and expression because they are not in different ontological realms. Facial gestures, body movements, tone of the voice, are the communicative condition of mental states. Infants see, hear, feel the mental states of the mother as well as perceive their intentionality when interacting with her since they 'know' the mental state in its communicative condition.

Section 4 focused on triadic interactions, on the so-called joint attention. The second-person approach gives a more natural and appropriate explanation of the phenomenon of joint attention by highlighting continuity between mutual and joint attention, while others accounts—as Tomasello's—over-mentalize the phenomenon. The phenomenon of joint attention comes from the introduction of a third element in mutual attention—in expressive communication. Infants follow the mother's gaze and grasp the directedness of her expressions toward the object. Engaging with the object in an indirect way enables infants to *distance* themselves from both the mother and the object. Infants begin to understand both the world as something independent of their own perspective—the first-person perspective—and the existence of other's perspectives about it—the third-person perspective. Second-person perspective is, therefore, prior and constitutive of both the first- and the third-person perspective. Since joint attention comes into play infants begin to perform the so-called proto-declaratives by which they start to share the contemplation of an object and to comment on it with the mother.

In section 5 I have argued that through expressive communication infants look for social reference in the mother by asking her when they face situations in which they doubt or do not know how to proceed and employ her answers to regulate their behaviour. The mother becomes a sort of authority for infants—that I have called 'second-person authority'—which shows the relation with the emergence of a basic awareness of normativity in infants. Infants begin to internalize the mother as *the norm* of how they *should* feel, and, therefore, behave, in such-and-such circumstances. Insofar

as the mother is being internalized, infants begin to require her presence less, becoming more independent. At this stage, infants begin to internalize alternative perspectives and no longer take the mother as the norm uncritically. Their entering in the social world is what makes infants develop a basic awareness of 'social' normativity, i.e., they change the question: "How should I feel?" for the more general question: "How am I expected to feel by these-and-these others, and how do I myself feel?".

In the next block, I will analyse some of the accounts about the so-called "problem of self-knowledge" and highlight their shortcomings about the issue. This is the problem of explaining the distinctive characteristics of the first-person self-adscriptions, namely, transparency and first-person authority. To carry out this analysis, I will divide the approaches into three large groups: the introspectionist, the agentialist, and the expressivist. I will sustain that all of these accounts are grounded on individualist framework, so they have significant deficiencies regarding the understanding of this phenomenon even when they contain relevant and significant elements.

BLOQUE II
Autoconocimiento y Auto-Atribución de Estados Mentales

1. Introducción

En la filosofía contemporánea el fenómeno del autoconocimiento, entendido como el conocimiento de las propias sensaciones, deseos, creencias y otros estados mentales, ha sido y sigue siendo un tema de controversia. Una de las principales discrepancias proviene de una afirmación compartida con nuestras intuiciones de sentido común: el autoconocimiento se distingue del conocimiento del mundo externo, lo cual incluye el conocimiento de otras mentes, por poseer unas características especiales distintivas, a saber, la *autoridad* y la *trasparencia*. Aun cuando esta misma afirmación también ha sido puesta en duda por algunos autores (Ryle, 1949, Gopnik 1993, Gopnik & Wellman, 1994, entre otros) el debate actual se centra en las diversas formas de entender dichas características.

En este bloque expondré cómo entienden el autoconocimiento los distintos enfoques en la actualidad basándome en su explicación de la autoridad y la transparencia, las cuales serán introducidas en la sección 2. Para ello dividiré los enfoques en tres grandes grupos: los introspeccionistas, que defienden que la autoridad proviene de acceso epistémico especial de primera persona a los propios contenidos mentales, serán tratados en la sección 3; los agencialistas, para quienes la autoridad no es una cuestión de acceso privilegiado a los contenidos mentales sino que tiene que ver con lo que hacemos como agentes responsables, serán tratados en la sección 4; finalmente, los expresivistas, para quienes la autoridad de primera persona proviene de nuestras capacidades expresivas, serán tratados en la sección 5. Sostendré que todos estos enfoques parten de la dicotomía entre primera y tercera persona y obvian las interacciones en su explicación del autoconocimiento, aun cuando algunos de ellos están relacionados en algunos aspectos con el enfoque de segunda persona que defiendo en este trabajo. Señalaré los problemas o carencias que conlleva para la comprensión

del autoconocimiento partir de estos presupuestos individualistas. Finalmente, expondré algunas conclusiones sobre lo expuesto.

2. Autoridad y transparencia

En nuestras interacciones diarias las auto-adscripciones (o auto-atribuciones)[13] de estados mentales en primera persona del presente de indicativo, los denominados 'avowals', tales como "Me duele la cabeza", "Creo que va a llover", "Quiero terminar pronto", parecen gozar de una presunción de verdad de la que carecen nuestros juicios acerca del mundo externo, nuestras auto-adscripciones de estados corporales o nuestras adscripciones de estados mentales a los demás. Si digo de mí mismo que peso tantos kilos, o que aquello que se ve en la distancia es un perro y no una oveja, mi afirmación es, normalmente, susceptible de duda y corrección puesto que estas afirmaciones no se consideran verdaderas por defecto, mi báscula puede estar estropeada y mi visión puede confundir objetos lejanos. Sin embargo, si uno es sincero y competente en el uso de los conceptos involucrados en las auto-adscripciones que usa para expresar sus estados mentales, no hay razón, en principio, para dudar de ellas. Si una persona tal contesta, en circunstancias normales, a la pregunta "¿Qué estás pensando?" afirmando "Estoy pensando que un ordenador sería un buen regalo para mi hija", parece que no tiene sentido preguntarle "¿Es eso seguro en lo que estás pensando?". Por supuesto, podemos interpelarla argumentando que no es un buen regalo, pero esto ya incluye nuestra creencia de que esa persona cree, erróneamente, que un ordenador sí sería un buen regalo. Normalmente asumimos que estamos en una posición privilegiada para hablar acerca de lo que nosotros mismos pensamos, sentimos, creemos, etc., pero no así acerca de lo que piensa, siente, desea, cree, etc., cualquier otra persona. Si quiero saber lo que alguien piensa debo realizar inferencias con base en la mejor explicación de su comportamiento o, por supuesto, preguntarle. Sin embargo, si quiero saber lo que yo mismo pienso no necesito normalmente inferirlo de mi comportamiento ni preguntarme a mí mismo, lo sé de una manera directa, '*transparente'*. Se nos supone *autoritativos* con respecto a nuestros propios estados mentales, nadie mejor que nosotros mismos sabe lo que queremos, deseamos, creemos, etc.

[13] En este bloque, y a lo largo de toda la tesis, usaré indistintamente los términos "auto-atribución" y "auto-adscripción" para referirme al mismo fenómeno.

Estas dos nociones brevemente esbozadas de *autoridad* y *transparencia* han tenido un largo desarrollo por parte de los filósofos contemporáneos. Con objeto de dilucidar estas características distintivas del conocimiento del sujeto de sus propios estados mentales, la investigación se ha centrado en la conexión entre los estados mentales de primer orden y los de segundo orden (o auto-atribuciones). La manera en la que esta conexión se conciba dará lugar a uno u otro enfoque, a una u otra definición de la autoridad de primera persona y de la transparencia, e incluso a la negación de alguna o de ambas nociones.

Una larga tradición que se remonta al menos a Descartes y se extiende a enfoques contemporáneos no cartesianos, considera esta conexión como una cuestión de un *acceso epistémico especial* que los sujetos poseen a ciertos acontecimientos dentro de ellos. Estos enfoques, denominados *introspeccionistas*, sostienen que el autoconocimiento es un hecho *empírico*, en el sentido de que existe una conexión *causal* y *contingente* entre los estados mentales de primer orden y los de segundo orden. (Armstrong, 1968, 1981, Lycan, 1987, 1996). De otra parte, hay autores que defienden que el autoconocimiento no es un hecho empírico sino *racional o conceptual,* en el sentido de que es algo que el sujeto hace, en tanto que *agente*, bien mediante razonamientos o mediante la constitución del estado mental, dándose, por tanto, una *conexión a priori* entre los estados mentales de primer orden y los de segundo orden ((Bilgrami, 1998, 2006, 2012; Wright, 1984, 1986 1987, 1989a, 1989b, 1991, 1996a, 1996b, 2001b, 2015; Wright, Smith, & Macdonald, 1998). Asimismo, otros autores sostienen que el autoconocimiento no es un hecho ni empírico ni conceptual sino que está relacionado con la *capacidad de expresar* nuestros estados mentales (Bar-On, 2004, 2009, 2011, 2013, 2015; Finkelstein, 1999, 2003, 2010, 2012).

A continuación presentaré dos de las propuestas más representativas del enfoque introspeccionista señalando las principales críticas al mismo.

3. Enfoques introspeccionistas

Los defensores de los enfoques introspeccionistas sostienen que para que el autoconocimiento pueda ser catalogado como una clase de conocimiento ha de ser caracterizado en términos epistémicos. Esta afirmación está relacionada con la

suposición de que los estados mentales son como los objetos del mundo, cognoscibles mediante los sentidos. Es por ello que estos autores defienden que el conocimiento de nuestros estados mentales se realiza mediante un método especial de acceso similar a la percepción, es decir, mediante la 'observación interna'. Sin embargo, dado que los objetos de conocimiento, los propios estados mentales, son considerados internos al sujeto, se asume que dicho acceso es exclusivo de este, es decir, privado, y que está exento de los errores perceptuales comunes tales como escasez de luz, lejanía, distorsión visual, etc. Nuestro acceso a ellos resulta, por tanto, más directo, inmediato, seguro, veraz que a los objetos externos. No obstante, el grado de seguridad que se asigna a la introspección varía dependiendo del enfoque en cuestión. En un extremo está el enfoque cartesiano, según el cual el acceso a nuestros estados mentales es certero, infalible, indubitable, incorregible. De otra parte, están los enfoques materialistas contemporáneos, los cuales han rechazado la infalibilidad[14] cartesiana en el acceso a nuestros estados mentales, aceptando que la introspección puede dar lugar a errores y postulando otro tipo de mecanismo de detección interna de estados mentales. En el caso de Armstrong (1968, 1981) y Lycan (1987, 1996), que veremos a continuación, el sistema de detección conlleva un mecanismo de monitorización o "auto-escaneo" que toma los estados mentales como *input* y produce representaciones de esos mismos estados como *output*. Los defensores de este tipo de introspección entienden este mecanismo como un sentido mediante el cual detectamos nuestros estados mentales, un *sentido interno* cuyo funcionamiento es similar al de los sentidos externos. Así, coinciden en entender la conexión del estado mental objeto de la introspección, el input, con el estado mental resultado de la introspección, el output, de una manera causal y contingente. Veámoslo con más detalle.

[14] La infalibilidad cartesiana y otros aspectos de esta concepción del autoconocimiento han sido ampliamente criticados y prácticamente rechazados en la tradición filosófica. Una de las principales objeciones que se han presentado en contra de la infalibilidad proviene de las criticas wittgensteinianas a la posibilidad de un lenguaje privado así como de su análisis del seguimiento de reglas: El lenguaje es una práctica gobernada por reglas que conlleva una serie de criterios para distinguir entre la aplicación correcta y la incorrecta de los términos. En el enfoque cartesiano el significado de los términos psicológicos está determinado por estados mentales privados de cada individuo, únicamente accesibles por él mismo. De esta manera, cualquier aplicación que le parezca correcta al individuo será (para él) la correcta, desapareciendo, con ello, la posibilidad de distinguir entre la aplicación correcta o incorrecta del término, o lo que es lo mismo, entre "es correcto" y "me parece correcto": "Y por lo tanto, también 'obedecer una regla' es una práctica. Y pensar que uno obedece una regla no es obedecer una regla. Por lo tanto, no es posible obedecer una regla 'privadamente': de lo contrario, pensar que uno está obedeciendo una regla sería lo mismo que obedecerla" (Wittgenstein, 1953, §202). Por tanto, para que pueda haber conocimiento debe existir la posibilidad de error, si uno no puede estar equivocado tampoco puede estar en lo correcto.

3.1 Enfoques del sentido interno: D. Armstrong y W. Lycan

David Armstrong (Armstrong, 1968, 1981), uno de los principales referentes del enfoque del sentido interno, sostiene que del mismo modo en el que conocemos el mundo externo mediante la percepción sensorial, conocemos lo que ocurre en nuestra mente mediante el sentido interno. Siguiendo a Kant, escribe:

> I believe that Kant suggested the correct way of thinking about introspection when he spoke of our awareness of our own mental states as the operation of "inner sense". He took sense perception as the model for introspection. By sense-perception we become aware of current physical happenings in our environment and our body. By inner sense we become aware or current happenings in our own mind (1968, p. 95).

Para Armstrong existe, por tanto, un paralelismo entre la percepción sensorial y el sentido interno. El autoconocimiento es el resultado de la introspección que se lleva a cabo mediante un mecanismo cognitivo *fidedigno*, una especie de mecanismo de monitorización o proceso de "auto-escaneo en el cerebro" (1968, p. 324) que toma un estado mental de primer orden como input dando otro estado mental de segundo orden como output. De esta manera, los estados mentales de segundo orden son producidos de manera confiable, mediante dicho mecanismo, por los correspondientes de primer orden, es decir, las creencias de segundo orden que se producen por medio de su correcto funcionamiento son, normalmente, correctas. Y puesto que son correctas y confiables, se convierten en (auto)conocimiento.

Armstrong sostiene un enfoque materialista, por lo que rechaza el dualismo cartesiano a favor de un monismo físico. Para Armstrong los estados mentales se corresponden con estados cerebrales. Así, tener un estado mental de primer orden, por ejemplo la creencia de que hay una pantalla de ordenador frente a mí, es estar en un estado cerebral determinado. Nuestro sistema de monitorización es también físico y funciona de manera tal que dado dicho estado cerebral, el estado mental de primer orden, produce otro estado cerebral, un estado mental de segundo orden, es decir, la creencia de que creo que hay una pantalla de ordenador frente a mí. De esta forma, el acceso a los estados mentales no es sólo epistémico sino también empírico, siendo, por tanto, causal y contingente.

La propuesta de Lycan (1987, 1996) es, en casi todos los sentidos, similar a la de Armstrong. Lycan coincide en postular un sistema interno de monitoreo mediante el

cual los estados de segundo orden son producidos causalmente por los de primer orden. Sin embargo, para Lycan el proceso de monitoreo requiere de mayores recursos cognitivos, dado que involucra mecanismos relacionados con la atención[15].

> I construe Active Introspection as a monitoring phenomenon. That is, I fall in with Armstrong's (1968b) notion of Active Introspection as self-scanning (Lycan, 1987, p.72).
>
> As I would put it, consciousness is the functioning of internal *attention mechanisms* directed at lower-order psychological states and events (Lycan, 1996, p. 14).

Ambos autores coinciden en sostener que el acceso a los estados mentales de primer orden no es inferencial en el sentido de que el individuo no tiene que acceder a las conexiones causales para tener autoconocimiento. El proceso de monitorización ocurre sin necesidad de que el sujeto sea consciente ni de él ni de su funcionamiento. Así, saber que se cree que *p* es el resultado del mecanismo causal fiable del sistema de monitoreo interno.

Como acabamos de ver, los introspeccionistas del sentido interno sostienen que tenemos un acceso epistémico privilegiado a nuestros estados mentales. Dado que dicho acceso a los estados mentales es empírico, que la conexión entre los estados de primer y segundo orden es causal y contingente, estos autores rechazan que este privilegio esté relacionado con la infalibilidad, objetando, por tanto, el introspeccionismo de la tradición cartesiana y admitiendo la posibilidad de error (Armstrong, 1968, p. 326; Lycan, 1996, p. 17). Para estos autores,, hablar autoritativamente acerca de nuestros propios estados mentales no es el resultado de un *perfecto* acceso epistémico a ellos, sino de un acceso epistémico *confiable*, es decir, los estados mentales detectados por nuestro sistema de monitorización resultan, *normalmente*, correctos. La autoridad de primera persona es entendida en este enfoque, por tanto, como el resultado de un acceso epistémico privilegiado, pero falible, a nuestros estados mentales. El funcionamiento confiable de nuestro sistema de monitoreo interno asegura que dado un estado mental de primer orden, se produzca el correspondiente de segundo orden. Y dado que no tenemos que ser conscientes de dicho funcionamiento, el acceso a nuestros estados mentales resulta transparente para nosotros. Bajo este enfoque, por tanto, la autoridad y la transparencia son *a posteriori*, ya que se supone que el estado mental de primer orden que el mecanismo de monitoreo escanea está en nuestro cerebro de manera previa al

[15] En el segundo apartado del Bloque III, en la parte dedicada a la Teoría de la Simulación, veremos en Alvin I. Goldman (2006) un enfoque similar al de Lycan en lo referente a la aprehensión de los estados mentales propios.

escaneo (de lo contrario no habría nada que escanear) y que lo conocemos, de manera contingente, sólo tras dicho escaneo.

A primera vista, el enfoque introspeccionista del sentido interno parece correcto, plausible, incluso científicamente, pues encaja muy bien con nuestras intuiciones de sentido común al dar una explicación tanto de la autoridad como de la transparencia que se supone será posible demostrar de manera científica: dado que tanto los estados mentales de primer orden como el sistema interno de monitoreo mediante el cual formamos los de segundo orden se suponen físicos, sólo es cuestión de tiempo el que la neurociencia avance y sea capaz de mostrarnos la localización y el funcionamiento del mismo. Sin embargo, esta plausibilidad es solo aparente. En primer lugar, dado que el sentido interno es, al igual que los externos, físico y falible podría darse el caso de que un individuo sufriera de 'auto-ceguera' interna, es decir, que el sentido interno, al igual que puede ocurrir con los sentidos externos, dejara de funcionar en un él. Un individuo así debería utilizar otros mecanismos para conocer sus estados mentales, al igual que, por ejemplo, un ciego utiliza el tacto para reconocer la forma de un objeto. Imaginemos que para conocer sus estados mentales un individuo auto-ciego necesita inferirlos de la observación de su propio comportamiento. Parece un poco extraño, por no decir absurdo, que este individuo supiera que tiene dolor de cabeza cuando se viera llevando a cabo el comportamiento típico relacionado con alguien que sufre un dolor de cabeza, por ejemplo yendo por una aspirina.[16]

De otra parte, del mismo modo que el introspeccionismo parece coincidir con nuestras intuiciones de sentido común, sus propios supuestos conllevan consecuencias que lo alejan del mismo. Aun cuando no creemos que podamos conocer en detalle la vida mental de otras personas no dudamos de que la tienen, de que experimentan estados mentales al igual que nosotros lo hacemos, de que sufren, aman, desean, tienen creencias, sentimientos, etc. Ahora bien, como he señalado anteriormente, los introspeccionistas sostienen que, puesto que los estados mentales son estados internos del individuo, solo este puede acceder a ellos. Si solo yo puedo, por tanto, acceder a mis estados mentales, si la experiencia que tengo de ellos es privada, ¿cómo puedo saber que los demás también los tienen? Y en el caso de que así sea, de que los demás también tengan estados mentales, ¿cómo puedo saber que *sus* estados mentales son

[16] Sidney Shoemaker (1994) argumenta la imposibilidad de que una persona pueda ser auto-ciega en este sentido. Para una crítica a dicho argumento véase Finkelstein (1999, 2003).

iguales a los míos?[17]. Armstrong (1968) sugiere la existencia de un "conocimiento telepático" en el sentido de que es posible tener acceso directo a los estados mentales de los demás mediante la introspección:

> I take the claim that telepathic knowledge exist to be the claim that we do in fact have some direct awareness of the mental states of others. […] We can say that telepathic knowledge is introspective awareness of other people's mental states, or, if preferred, that introspection is telepathic knowledge of our own mental states (p. 124).

Sin embargo, aún cuando aceptáramos que tal conocimiento existiera, todavía no podríamos asegurar que los otros tienen estados mentales puesto que aún faltaría el conocimiento directo de que aquello que introspectamos es, de hecho, el estado mental de otro. Sabríamos que hay un dolor que estamos experimentando pero no tendríamos la garantía de que, de hecho, alguien más está experimentando ese dolor.

Finalmente, una de las principales críticas de la tradición filosófica al introspeccionismo, proveniente de las discusiones que tuvieron lugar a mediados-finales del siglo pasado entre internistas y externistas en relación al contenido de nuestro conocimiento, cuestiona la posibilidad de autoconocimiento introspeccionista. Los externistas mostraron que los significados de nuestras palabras y el contenido de nuestros pensamientos están determinados en parte por factores externos a nosotros, factores tales como esencias subyacentes de productos con los que interactuamos (Putnam, 1975) o "convenciones sociales del entorno" (Burge, 1979, p. 109). El problema con el que se enfrentan los introspeccionistas es que, siendo esto así, si existen estados mentales que están determinados, en parte, por factores externos, podríamos no conocerlos, lo cual implica que si uno es introspeccionista y sostiene que el conocimiento de los nuestros estados mentales se realiza únicamente mediante un acceso interno, ha de rechazar que haya autoconocimiento (cuanto menos, de los estados mentales externamente diferenciados).

[17] Estas dos preguntas se corresponden con las cuestiones epistemológica y conceptual del denominado en la tradición filosófica "problema de las otras mentes". La cuestión epistemológica está relacionada con la naturaleza de la experiencia, con la suposición del introspeccionismo de que la experiencia mental es privada, y conduce al problema acerca de la imposibilidad de saber si los demás también tienen estados mentales como nosotros. La cuestión conceptual, por su parte, está relacionada con la naturaleza del lenguaje, con la suposición de que las palabras pueden adquirir su significado por simple definición ostensiva, es decir, con la suposición que los conceptos de estados mentales, entendidos estos por el introspeccionismo como objetos internos, son formados mediante la observación interna de los mismos, mediante la introspección. La principal crítica a estos dos supuestos la encontramos en el argumento del lenguaje privado de Wittgenstein (1953, §§243–315), según el cual un lenguaje ininteligible para cualquier persona salvo su originador es imposible puesto que un lenguaje tal sería, necesariamente, ininteligible también para el mismo originador.

En este apartado he presentado los enfoques introspeccionistas del sentido interno de Armstrong y Lycan. Estos autores consideran el autoconocimiento como un hecho empírico en el que se da una conexión *causal* y *contingente* entre los estados mentales de primer orden y los de segundo orden. La autoridad de primera persona es entendida como el resultado de un *acceso epistémico especial* que los sujetos poseen a ciertos acontecimientos que ocurren en su interior. Este acceso privilegiado es transparente para el sujeto puesto que no ha de ser consciente del proceso llevado a cabo por el sentido interno, entendido este como un escáner que monitorea de manera confiable los estados mentales de primer orden dando como resultado estados mentales de segundo orden. De esta manera, saber que se cree que *p* es el resultado del mecanismo causal fiable del sistema de monitoreo interno.

Asimismo, he señalado las críticas más relevantes a este enfoque provenientes tanto del sentido común como de la reflexión filosófica. El introspeccionismo conlleva la posibilidad de que el sentido interno no funcione, al igual que ocurre con los sentidos externos, y que existan individuos 'auto-ciegos', es decir, individuos incapaces de 'observar' internamente sus estados mentales. Igualmente, el hecho de que solo el individuo sea capaz de acceder mediante la introspección a sus estados mentales implica la incertidumbre acerca de si las demás personas también los tienen. Finalmente, el externismo mostró que los significados de nuestras palabras y el contenido de nuestros pensamientos están determinados, en parte, por factores externos a nosotros, lo cual conlleva la incompatibilidad entre introspeccionismo y autoconocimiento.

El enfoque introspeccionista es, como acabamos de ver, completamente individualista en su interpretación de las características distintivas del autoconocimiento, es decir, solo tiene en cuenta el acceso del individuo a sus estados mentales privados y no a sus interacciones con los demás. Este enfoque es, por así decirlo, un enfoque opuesto al de segunda persona que sostengo en este trabajo, ya que el análisis de este último está centrado en su relación con los demás y no única y exclusivamente en el individuo. Centrarse en el individuo introspectivo conlleva una serie de problemas como acabamos de ver, siendo uno de los principales la incertidumbre acerca de si las demás personas también tienen estados mentales al igual que uno mismo.

A continuación, presentaré algunos de los autores más relevantes del enfoque agencialista, un enfoque que, como veremos, sí está relacionado con el enfoque de

segunda persona, aunque continua siendo un enfoque que parte de presupuestos individualistas en su explicación del autoconocimiento.

4. Enfoques agencialistas

Los *agencialistas*, al contrario que los introspeccionistas, sostienen que el autoconocimiento no consiste en examinar o detectar las propiedades de objetos internos puesto que no existen tales objetos internos. El problema de explicar cómo tenemos autoconocimiento autoritativo de nuestros estados mentales no es, en esencia, un problema de *acceso* a contenido. Cuando realizamos una auto-atribución de un estado mental no estamos *describiendo* el contenido de un objeto interno al cual tenemos un acceso privilegiado. La autoridad de primera persona no tiene que ver con un mecanismo de acceso privilegiado a un contenido que ya está ahí presente respecto al cual el sujeto posee un rol pasivo sino que está relacionada con la acción de un *agente*, es decir, los estados mentales no son algo que, meramente, observamos sino algo que *hacemos* en tanto que seres racionales y conceptualmente competentes. La autoridad de primera persona y el autoconocimiento están relacionados, según estos autores, con nuestra capacidad para formar, constituir nuestros estados mentales, es decir, con nuestra *agencia*. Dentro de este enfoque encontramos diversas propuestas de entre las cuales presentaré las que considero más representativas y relacionadas con el enfoque de segunda persona. El primer sub-apartado estará centrado en el enfoque deliberativo de Richard Moran, también denominado 'de la transparencia'. En el segundo sub-apartado presentaré los enfoques constitutivistas de Crispin G. J. Wright y Akeel Bilgrami.

4.1 El enfoque deliberativo de Richard Moran

El enfoque deliberativo de Moran relaciona la agencia racional con el denominado "método de la transparencia" de Gareth Evans (1982)[18]. Moran argumenta que hechos

[18] Aunque el propósito de Evans no es la creación de ningún método de autoconocimiento sino la continuación de la crítica wittgensteiniana a la concepción cartesiana de la mente, el "método de la transparencia" se le atribuye por ser el primero en desplazar la "mirada a lo interno" hacia la "mirada a lo externo" con respecto a las creencias, aún cuando Evans en su obra no hace uso del término "transparencia". Como veremos, este "giro" de lo interno a lo

acerca de la agencia racional justifican nuestro uso de este método ya que "cuando me concibo a mí mismo como un agente racional [...] puedo informar sobre mi creencia acerca de X al considerar (nada más que) a X mismo" (Moran 2001, p. 84). La idea de la transparencia de Evans proviene de su crítica a la visión introspeccionista del autoconocimiento. Evans se propone abandonar la idea de que captar nuestros estados mentales "implica siempre una mirada interna a los estados y hechos de algo a lo que sólo la persona tiene acceso" (1982, p. 225). Evans afirma:

> [I]n making a self-ascription of belief, one's eyes are, so to speak, or occasionally literally, directed outward—upon the world. If someone asks me 'Do you think there is going to be a third world war?' I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question 'Will there be a third world war?' (1982, p. 225).
>
> It is of the utmost importance to appreciate that in order to understand the self-ascription of experience we need to postulate no special faculty of inner sense or internal self-scanning (1982, p. 230).

El propósito de Evans es invertir la idea introspeccionista del 'interior' al 'exterior', de la 'transparencia hacia lo interno' a la 'transparencia hacia lo externo'. Los estados mentales no son transparentes hacia sí mismos, como afirman los introspeccionistas, sino que son transparentes 'hacia el mundo'. Para conocer nuestras creencias no debemos mirar dentro de nosotros mismos, no debemos atender interiormente a los contenidos de nuestra mente. Más bien necesitamos mirar al mundo exterior, a los mismos *hechos externos, objetos o propiedades* a los que atenderíamos si estuviéramos considerando, por ejemplo, si está lloviendo, o si hay algún objeto delante de nosotros. Del mismo modo que respondo a una pregunta sobre el mundo, puedo responder a la pregunta sobre mis propias creencias. Para auto-atribuirme una creencia, de acuerdo con Evans, no me involucro en una percepción interior. Por el contrario, "me pongo en condiciones de responder a la pregunta de si creo que *p*, poniendo en práctica cualquier procedimiento que tenga para responder a la pregunta de si *p*" (1982, p. 225). La idea central es que la cuestión de si creo que *p* es, para mí, transparente a la pregunta de si *p*. Puedo responder a la primera pregunta respondiendo a esta última.

---

externo modifica el sentido en el que la transparencia es entendida, ya que no se trata de la transparencia interna hacia los estados mentales, es decir, que podamos acceder a ellos de manera interna sin necesidad de inferencias y observación, sino de que la creencia acerca de un contenido es transparente a la afirmación de dicho contenido, es decir, que el paso de la afirmación de un contenido a su creencia se realiza de manera directa y no inferencial.

En su ampliamente discutido libro *Authority and Estrangement* (Moran, 2001), Moran recoge la idea de Evans[19] formulando su 'reivindicación de la transparencia':

> With respect to the attitude of belief, the claim of transparency tells us that the first-person question "Do I believe P?" is "transparent" to, answered in the same way as, the outward-directed question as to the truth of P itself (2001, p. 66).

Dado que la pregunta acerca de *p* y la pregunta acerca de si *creo que p* parecen corresponder a temas diferentes, Moran sostiene que tenemos derecho a asumir que la reflexión sobre las razones a favor de *p* proporciona una respuesta a la pregunta sobre nuestra creencia de que *p*, si asumimos que nuestra creencia es algo determinado por la conclusión de nuestra reflexión sobre esas razones ya que "[u]na suposición de este tipo proporcionaría el tipo correcto de vínculo entre las dos preguntas" (2003, p. 405).

Moran presenta su enfoque de la transparencia como una alternativa a los enfoques introspeccionistas a los que denomina 'modelos perceptuales'. Para Moran, cuando una afirmación de una creencia propia sobre *x* se hace considerando los hechos acerca de *x* en sí, y no por una 'mirada interna' dicha afirmación obedece a la Condición de la Transparencia:

> A statement of one's belief about X is said to obey the Transparency Condition when the statement is made by consideration of the facts about X itself, and not by either an "inward glance" or by observation of one's own behaviour. An avowal is a statement of one's belief which obeys the Transparency Condition (2001, p. 101).

Según Moran, concluir que *p* es verdadero es suficiente para *justificar* que se crea que *p*, no se requiere evidencia adicional. El procedimiento de la transparencia para descubrir lo que uno cree es "inmediato", no "implica inferencia a otra cosa" (2001, pp. 10–12) [20].

---

[19] Robert. M. Gordon (2007), otro de los defensores de la Teoría de la Simulación que veremos en el apartado 2 del Bloque III, toma también esta idea de Evans describiéndola como una "rutina de ascenso semántico".

[20] Alex Byrne (2005, 2011) defiende, por el contrario, un enfoque inferencialista de la transparencia. Para Byrne, aun cuando Evans formula el método de la transparencia, no aclara que este involucra una 'inferencia del mundo a la mente':

> Evans does not explicitly address this question, but the natural answer is that the next step involves an *inference from world to mind*: I infer that I believe that there will be a third world war from the single premise that there will be one (2011, p. 203).

Siguiendo a Gallois (1996), Byrne sostiene que el argumento que corresponde a la que él llama "inferencia transparente" es el "esquema doxástico" (2011, p. 204):

> *p*
> ---
> Creo que *p*

Ahora bien, dado que una inferencia como esta no parece ser válida, como el mismo Byrne admite al reconocer que no es "ni deductivamente válida ni inductivamente fuerte" (2011, p. 204), ¿cómo es posible que produzca auto-conocimiento? Byrne argumenta que el esquema doxástico es *fuertemente auto-verificable* en el sentido de que "si uno razona de acuerdo al esquema doxástico, e infiere que cree que *p* de la premisa *p*, entonces la creencia de segundo orden es *verdadera*, puesto que la inferencia de una premisa implica la creencia en dicha premisa" (2011, p. 206). Asimismo, este tipo de razonamiento conduce a creencias *seguras*, es decir, a creencias que no es fácil que sean

La inmediatez del acceso en primera persona, "como afirmación sobre el modo de conciencia, sólo significa que tales juicios no se deducen de nada epistémicamente más básico" (2001, p. 10). Más que desafiar los enfoques introspeccionistas "los filósofos pueden estar más preparados para negar la sustancialidad de la introspección en sí" (2001, p. 13).

La idea de Moran es que la reivindicación de la transparencia tiene su "fuente en la primacía de una postura *deliberativa* más que teórica sobre nuestros estados mentales" (2001, p. 64). El autoconocimiento *autoritativo* es, para Moran, una cuestión de "formar la mente" (*making up one's mind*), es decir, es algo que hacemos, no algo que simplemente ocurra. Tenemos autoconocimiento autoritativo en primera persona cuando formamos creencias (y otras actitudes proposicionales) como agentes responsables, es decir, creemos que *p* cuando lo hacemos sobre la base de las razones a favor y en contra de *p* y deliberamos, en base de tal evaluación, que *p* es el caso. Según Moran, cuando reflexiono sobre mis estados mentales puedo suspender tanto su fuerza normativa como su fuerza psicológica. La autorreflexión introduce una brecha entre mis estados mentales ocurrentes y mi yo. Esta brecha muestra que no estoy simplemente determinado por los deseos y creencias que tengo. Puedo elegir si los tengo. Formamos nuestras creencias, por tanto, decidiendo, racionalmente, qué creer. Un agente responsable considera sus creencias una expresión de su criterio de evaluación y ponderación de razones prácticas y teóricas (Moran, 2001).

Para Moran, el juicio psicológico resultante de la deliberación, 'creo que *p*', no es una descripción, como sí lo es en el caso de las atribuciones de estados mentales a otros, sino un avowal, es decir, una expresión directa de la creencia de que *p* que uno ha formado a través de la deliberación racional y con la que uno está comprometido. Según

---

falsas (2005. pp. 96–98). Byrne explica este razonamiento en términos de seguimiento de reglas epistémicas. Así, el esquema doxástico puede interpretarse a través de la siguiente regla BEL: Si *p*, cree que crees que *p*. Para Byrne, seguimos la regla BEL al igual que seguimos, por ejemplo, la regla DOORBELL: Si el timbre de la puerta suena, cree que hay alguien en la puerta. Si creemos que hay alguien detrás de la puerta es *porque* el timbre ha sonado. Este 'porque' marca la *relación básica* epistémica: el reconocimiento de que *p* (el timbre sonando en este caso) sirve tanto de relación causal como de fundamento racional para nuestra creencia de que creemos que *p*. Esta relación básica epistémica, es, por tanto, la que sustenta la transición cognitiva en forma de inferencia de *p* a creo que *p*. La diferencia entre DOORBELL y BEL radica en que BEL es auto-verificable, mientras que DOORBELL no lo es. En el caso de BEL la verdad de la creencia de segundo orden está garantizada, sin embargo en el caso de DOORBELL no lo está. Esto es así puesto que a un cuando por una extraña casualidad no hubiera nadie detrás de la puerta, y nuestra creencia de que hay alguien detrás de la puerta fuera falsa, la creencia de que creemos que el timbre ha sonado sería verdadera (2005, pp. 96–97).

Este contraste puede verse, asimismo, en el caso de la atribución de creencias a otros, donde la regla BEL_3 puede no conducir a conocimiento. Así BEL_3 Si *p,* cree que él/ella cree que *p* no es auto-verificable ya que el resultado de seguir esta regla puede producir una falsa creencia.

Byrne sostiene que la conclusión de este razonamiento es seguir la regla BEL, razonar a través del esquema doxástico produce conocimiento (*is knowledge-conductive*) y explica el carácter peculiar y privilegiado del autoconocimiento autoritativo y transparente.

Moran, la transición de la atribución al avowal es una expresión de la libertad racional de la persona, una afirmación de autoridad que conlleva un compromiso como agente *responsable* de esa misma auto-atribución. Es esta afirmación de autoridad, "este compromiso, lo que hace posible que su avowal se ajuste a la Condición de Transparencia, el anuncio de su creencia sin depender de la evidencia psicológica sobre sí misma" (2001, p. 151). Al realizar un avowal, por tanto, no sólo se registra la presencia de la creencia sino que se *expresa* un compromiso con ella. Según Moran, la decisión del agente "no es solo evidencia (empírica) sobre lo que hará, sino una resolución de la cual él es el *autor* y el responsable de llevar a cabo" (2001, p. 79). Es por ello que Moran rechaza el rol pasivo de los enfoques introspectivos. Si, como estos enfoques sostienen, el conocimiento de nuestros estados mentales proviene de su observación interna, esto nos deja en una posición de meros espectadores respecto a ellos obviando, por tanto, la responsabilidad que conlleva su auto-atribución. No importa lo privilegiado que sea el punto de vista del observador pues, en esencia, "lo que tenemos aquí es una imagen del autoconocimiento como una especie de lectura de la mente (*mind-reading*) aplicada a uno mismo" (2001, p. 91). Según Moran, solo con respecto a mi propio estado mental estoy en posición de *determinar* qué hacer con él mediante reflexión deliberativa sobre el mismo. La inmediatez epistémica del espectador no le otorga control alguno sobre la *autoría* de la mente observada, su relación con esta está *alienada*. Morán admite que todas las personas estamos alienadas con respecto a determinados estados mentales como, por ejemplo, cuando hacemos afirmaciones sobre nosotros mismos basadas en observaciones de nuestro propio comportamiento tales como 'creo que estoy un poco enojado con ella, dado lo que acabo de hacer'. Sin embargo, aún cuando esto pueda, en ocasiones, ocurrir, no significa que sea posible estar alienado con respecto a todos nuestros estados mentales, como en el caso del espectador pasivo el cual carecería de agencia y, por tanto, de responsabilidad.

Como acabamos de ver, Moran desarrolla un enfoque del autoconocimiento en el que el sujeto es visto como un agente responsable cuyas auto-atribuciones son resultado de la deliberación. La perspectiva del agente es tratada como autor en oposición a la del otro, el cual es visto como un mero observador. Es por ello que Moran rechaza el introspeccionismo, ya que según él este enfoque está basado en la mera observación de los estados mentales, lo que convierte al individuo en un sujeto pasivo carente de agencia y responsabilidad. Sin embargo, aun cuando Moran rechaza el

enfoque introspeccionista, comparte con él el mismo presupuesto individualista. Al igual que los introspeccionistas, para quienes la autoridad proviene de la introspección del individuo de sus estados mentales, para Moran la autoridad proviene del *individuo* racional y responsable que justifica sus acciones mediante razones provenientes de la deliberación. Uno de los principales motivos por los cuales Moran comparte este individualismo es la idea de que para dilucidar las características del autoconocimiento la estrategia a seguir es la de mostrar lo distintivo de las auto-atribuciones de estados mentales *en oposición* a las atribuciones de estados mentales a los demás (idea que, como mostraré más adelante, encontramos también tanto en los constitutivistas como en los expresivistas). Su enfoque está centrado, al igual que el introspeccionista, en la dicotomía primera-tercera persona, lo cual descarta las interacciones en segunda persona como relevantes para el problema del autoconocimiento y sitúa al individuo en el centro de la explicación. Tanto una cosa como la otra, situar al individuo en el centro de la explicación como descartar las interacciones en segunda persona, resultan problemáticas.

Situar al individuo en el centro de la explicación conlleva una serie de problemas tanto en la explicación del individuo mismo como en la explicación del problema del auto-conocimiento, como hemos visto que ocurre en el caso del introspeccionismo. En el caso de Moran, conduce a una concepción del individuo excesivamente teórica e irreal, como señala Quassim Cassam (2014), quien denomina "racionalismo" al agencialismo de Moran por ser, según él, demasiado exigente para con lo que realmente somos los seres humanos. Para Cassam, el enfoque deliberativo de Moran es, en el mejor de los casos, una descripción del autoconocimiento para un *homo philosophicus* mítico, cuyas creencias y otras actitudes son siempre lo que deberían ser racionalmente, más que para los seres humanos (*homo sapiens*) quienes "no son ciudadanos epistémicos modelo y está lejos de ser obvio que sus actitudes sean como deben ser racionalmente" (2014, p. 7). La critica de Cassam está centrada en la existencia de las actitudes recalcitrantes, actitudes que seguimos teniendo aun cuando racionalmente creamos que no deberíamos hacerlo, por ejemplo, creencias que están profundamente arraigadas (como supersticiones) o que nos reconfortan (como las ilusiones) y que en ocasiones perseveran ante la evidencia contraria, deseos que a veces chocan con nuestros valores más profundos o, sencillamente, fobias de las cuales no podemos deshacernos (como, por ejemplo, el miedo irracional a las arañas).

En una línea similar respecto al autoconocimiento de nuestros estados mentales, Finkelstein sostiene que "hablamos con autoridad de primera persona sobre un amplio abanico de estados y sucesos mentales que no pueden ser declarados (*avowable*) en el sentido de Moran" (2003, p. 162). Según Finkelstein, entre estos estados figuran las sensaciones y otros procesos mentales ocurrentes (que Moran no pretende incluir en su enfoque), estados mentales pasados, actitudes que no se pueden declarar porque las razones abogan contra ellas y todas las actitudes que la reflexión deliberativa ni requeriría ni prohibiría, como, por ejemplo, el desdén, la veneración, los celos, el pesar, el asco y el odio (2003, pp. 162 y ss.).

Por otro lado, descartar las interacciones en segunda persona como relevantes para el problema del autoconocimiento también conlleva una serie de problemas. En primer lugar, se está obviando que el otro no es simple y únicamente un observador sino también un *interlocutor*. Tratar al otro como un interlocutor resulta fundamental para la comprensión del agente deliberativo pues revela una concepción diferente y más amplia de las razones y la justificación del mismo. Para Moran, el agente justifica sus acciones o auto-atribuciones mediante las razones que, tras deliberar, aduce (o puede aducir). Sin embargo, Moran no parece tener en cuenta el reconocimiento por parte del otro, en cuanto interlocutor, de dichas razones. Dicha justificación, dichas razones, no pueden estar única y exclusivamente dirigidas a uno mismo puesto que razonamos también para justificar nuestras acciones frente a los demás. Moran parece obviar que la responsabilidad, la justificación de nuestras acciones y auto-atribuciones, no son esgrimidas solo para uno mismo, sino también para los demás, que el deber hacer algo implica el reconocimiento del otro, de dar cuenta a otros de mis acciones y no solo a mí mismo, al igual que los otros deben dar cuenta de sus acciones para con uno. Si las razones que el agente aduce para su acción o auto-atribución no incluyen el reconocimiento del otro de esas mismas razones, ¿qué sentido tienen? Si lo que uno ve como una razón no es visto por nadie más como tal, dicha razón sería una razón 'privada' y al igual que no puede haber un lenguaje privado, como ya mostró Wittgenstein (1953, §§243–315), tampoco puede haber una razón 'privada'. El lenguaje es constitutivamente público, lo que cuenta como una razón para mí ha de contar, asimismo, como una razón para el otro y lo que cuenta como una razón para el otro ha de contar como una razón para mí, esté o no de acuerdo con ella. Esto no quiere decir que todos compartamos las mismas razones para los mismos casos, sino que las razones que aducimos han de ser vistas como razones en sí mismas, con las que, de otra parte, se

puede o no estar de acuerdo. Moran no tiene en cuenta la necesidad de este reco*nocimiento mutuo* que, como argumentaré más adelante, resulta crucial para la comprensión del problema del autoconocimiento.

En este apartado he descrito la propuesta agencialista de Moran, el cual rechaza el rol pasivo que los enfoques introspectivos asignan al sujeto. Para Moran, como hemos visto, el agente utiliza el 'método de la transparencia' para, a través de la deliberación, formar sus estados mentales (*making up one's mind*). La autoridad de primera persona es entendida por Moran como la autoridad del agente como responsable de sus auto-atribuciones. He sostenido que, al igual que el enfoque introspeccionista, Moran cae en un individualismo que resulta problemático tanto para la comprensión del individuo como para la comprensión del autoconocimiento. En el siguiente apartado describiré otro enfoque agencialista, el constitutivismo, que sostiene una forma distinta de entender el rol activo del sujeto. Los autores que veremos en el siguiente apartado, Wright y Bilgrami, subrayan la parte constitutiva, más que la deliberativa, del agente respecto a sus propios estados mentales y entienden el autoconocimiento como una condición *a priori* de las reglas gramaticales que gobiernan el uso de las (auto-)atribuciones de estados mentales. Estos autores, a diferencia de Moran, sí incluyen las interacciones en su enfoque aunque, como veremos, no dejan de ser enfoques individualistas del autoconocimiento.

4.2 Enfoques constitutivistas

El autoconocimiento desde el punto de vista constitutivo, al igual que en el enfoque deliberativo de Moran, no es el resultado de un acceso privilegiado y privado a nuestros estados mentales. Para los constitutivistas, el autoconocimiento es una característica necesaria de que tengamos estados mentales sobre nosotros mismos. Es una necesidad conceptual, *a priori*, que concierne, según Wright, a la "gramática", es decir, a las reglas, de nuestro 'juego de lenguaje' de nuestras (auto-)atribuciones psicológicas. Según estas reglas, que rigen las prácticas lingüísticas de la comunidad, no hay por qué dudar de las auto-atribuciones psicológicas de un agente que cumple con las condiciones normales de racionalidad: un agente que es sincero, cognitivamente lúcido y alerta, con el repertorio conceptual relevante y sin razón para pensar que podría auto-

engañarse. Las auto-atribuciones de estados mentales son "autoritativas por defecto" (*default-authoritative*) (Wright, 1996b, p. 369). Esto es así, según los constitutivistas, porque las auto-atribuciones psicológicas no son informes de observación, en los que el observador puede errar en la observación, sino que son constitutivas de los hechos a los que se refieren. En ese sentido, lo que uno dice o piensa sobre sus propios estados mentales desempeña un papel constitutivo respecto a los mismos puesto que establece lo que son. La autoridad de primera persona, por tanto, consiste en la constitución por parte del agente de los estados mentales auto-atribuidos. En otras palabras, uno tiene autoridad sobre su vida mental pues es uno mismo quien la constituye.

A continuación presentaré dos de las propuestas más representativas del enfoque constitutivo, la de Crispin Wright, el cual desarrolla su propuesta a partir del tratamiento wittgensteiniano del seguimiento de reglas y la de Akeel Bilgrami, inspirada en la concepción strawsoniana de la libertad y la agencia y centrada en el compromiso y la responsabilidad.

4.2.1 Crispin Wright

El enfoque constitutivista de Wright está relacionado con, o motivado por, el análisis acerca del seguimiento de reglas que Wittgenstein realiza en las *Investigaciones Filosóficas* (Wittgenstein, 1953). Wright presenta su lectura del mismo como una alternativa a la lectura de Saul A. Kripke (1982), con el que, sin embargo, coincide en algunos aspectos. Comenzaré, por tanto, presentando brevemente los análisis wittgensteiniano y kripkeano acerca del seguimiento de reglas para, tras ello, presentar la postura de Wright al respecto y su relación con el autoconocimiento.

Con objeto de dilucidar el problema de la aplicación correcta o incorrecta de una regla Wittgenstein utiliza el ejemplo del seguimiento de una instrucción del tipo "+n" para mostrar, en primer lugar, que su significado no puede residir en la interpretación que de ella haga un sujeto, ya que la interpretación requeriría de otra interpretación y esta de otra a su vez y así sucesivamente, dando lugar a un regreso al infinito. Es por ello que Wittgenstein (1953) se pregunta:

> "But how can a rule shew me what I have to do at *this* point? Whatever I do is, on some interpretation, in accord with the rule."— That is not what we ought to say, but rather:

> any interpretation still hangs in the air along with what it interprets, and cannot give it any support. Interpretations by themselves do not determine meaning (§198).

Asimismo, nada impide la posibilidad de que un sujeto interprete la instrucción de cierta manera hasta cierto punto y luego lo haga de manera diferente después. De ahí que nos encontremos con la conocida paradoja de las *Investigaciones*:

> This was our paradox: no course of action could be determined by a rule, because every course of action can be made out to accord with the rule. The answer was: if everything can be made out to accord with the rule, then it can also be made out to conflict with it. And so there would be neither accord nor conflict here (§201).

Si el significado de una regla no depende de la interpretación que de ella haga el sujeto ni de la disposición del mismo a usarla de determinada manera, debemos concluir, como sugiere Kripke que Wittgenstein también hace, que no hay ningún hecho predeterminado que sea intrínsecamente significativo y que comprendamos cuando entendemos una instrucción dada como "+n", es decir, no hay ningún hecho que fundamente o justifique nuestras atribuciones de significado (Kripke, 1982). Un hecho tal, de existir, pertenecería a la mitología platónica, haciendo que fuera misterioso cómo mentes como la nuestra pudieran captarlo[21]. Kripke trata de ilustrar esta conclusión a través de la utilización del símbolo matemático "+", proponiéndonos que imaginemos un escéptico que reta a su interlocutor a probar que la respuesta correcta a la pregunta "¿Cuánto son 68 más 57?" es "125" y no "5", teniendo en cuenta lo que, en el pasado, siempre ha querido decir con 'más'. La idea es que, aduciendo hechos de su vida pasada en los que con 'más' quiso decir 'más', nuestro interlocutor ha de probar que en el pasado con 'más' siempre ha querido decir 'más' y no otra función matemática como, por ejemplo, 'quas', cuyo valor resultante de los argumentos 68 y 57 es 5. Según Kripke, dado el número infinito de posible aplicaciones o, en sus propias palabras, la "cantidad infinita de candidatos que el escéptico puede proponer para desempeñar el papel de la quadición" (1982, p. 27), ningún hecho acerca de cómo el interlocutor ha interpretado el término 'más', ni acerca de las circunstancias en las que lo ha usado, ni

[21] Al respecto, Finkelstein (2003) señala dos problemas del platonismo. Uno de ellos es que sostener que nuestras palabras y gestos obtienen su significación semántica de ítems que permanecen ocultos tras ellas hace que la comunicación parezca un milagro puesto que "¿cómo puede ser que cuando te digo algo, tú no sólo oigas mis palabras, sino que generalmente captes mi significado?" (p.86). El otro problema es que el platonista no sabe realmente cómo algo podría bloquear el regreso de interpretaciones para convertirse en la 'interpretación definitiva', no sabe cómo algo podría ser una fuente capaz de dotar de significado a nuestras palabras, en vez de ser otro elemento que carece de contenido intrínseco. El platónico se ve obligado a admitir que "*debe* haber, tales ítems (ya que de otra forma nuestras palabras serían meros ruidos y marcas vacías), pero esos ítems le resultan misteriosos incluso a él" (p.88).

acerca de su disposición a usarlo de una u otra manera, ni acerca los episodios mentales ocurrentes que experimentó en torno a él puede determinar qué aplicaciones concordarían con lo que él quería decir con tal término. Ante esto, según Kripke, Wittgenstein ofrece una "solución escéptica"[22] al problema: solo el acuerdo de la comunidad lingüística de usar un símbolo como "+" de cierta manera determina lo que significa al establecer qué regla expresa y, por lo tanto, cuáles deben considerarse aplicaciones correctas o incorrectas de la misma.

Tras esta breve explicación de la solución escéptica que Kripke atribuye a Wittgenstein, a mi juicio correctamente, veamos la opinión Wright al respecto.

Según Wright, Kripke interpreta erróneamente a Wittgenstein pues las objeciones que presenta a todos los candidatos a hechos determinantes del significado están basadas en una epistemología inferencial de los mismos. La clave está en abandonar esta idea y argumentar a favor de la existencia de una capacidad de conocer significados *no-inferencialmente*. Según Wright, en lo que respecta a la relación entre la intención y el pensamiento, "puede llegar a ser cierto para mí que tengo una cierta intención sin que me involucre en ningún proceso de deliberación consciente ni piense en ningún pensamiento que especifique el contenido de esa intención" (1987, p. 126)[23]. Por tanto, en contra de Kripke, Wright sostiene que, en general, un sujeto tiene acceso autoritativo y no inferencial al contenido de sus propias intenciones y, dado que este contenido puede ser abierto, general y 'potencialmente infinito', puede extenderse a todos los posibles usos de ese símbolo particular. Así pues, para refutar el argumento escéptico, en el punto en que el este nos reta a aducir algún hecho mental para descartar las interpretaciones de 'quas', basta recordar precisamente nuestra intención anterior con respecto al uso de 'más'. En otras palabras, el hecho sobre mi anterior uso de 'más' que determina que lo que ahora estoy haciendo concuerda con lo que en aquel momento quería decir con 'más', es mi afirmación de que con 'más' quería decir 'más' (Wright, 1989b). Para Wright, "una filosofía satisfactoria sobre la intención tiene que validar

[22] Una solución 'escéptica', entendida en el sentido humeano, es aquella que acepta los resultados negativos de las dudas escépticas—es decir, que aquello que intentábamos encontrar es algo que simplemente no podremos encontrar—e intenta desarrollar una teoría compatible con la inexistencia de hechos intrínsecamente significativos. Por el contrario una solución 'directa', como la que sostiene Wright, es aquella que muestra que el argumento del escéptico es erróneo y señala un hecho intrínsecamente significativo, una "condición en el mundo" que refuta al escéptico (Kripke, 1982, pp. 68–69).

[23] Y continúa: "Más bien, simplemente puedo encontrarme con mi mente hecha (*made up*), por así decirlo, capaz de dar cuenta de mis intenciones si me lo piden, pero sin ninguna historia que contar sobre el cuándo o el por qué de su aparición" (1987, p. 126). Vemos en este pasaje una diferencia fundamental con en agencialismo de Moran, para quien, como hemos visto, el autoconocimiento autoritativo es una cuestión de "formar la mente" (*making up one's mind*), es decir, es algo que hacemos mediante la deliberación, no algo que simplemente ocurra.

nuestra reivindicación de la autoridad no inferencial sobre nuestras intenciones presentes (y previas) sin sucumbir a la mitología del contenido infinito, explícito e introspectable" (1984, p. 115). Debemos, pues, rechazar la solución escéptica sin caer en un platonismo según el cual intuimos o percibimos un significado oculto tras el enunciado de una regla que, de manera autónoma, determina el curso de acción a seguir. La solución de Wright pasa por sostener que las respuestas al seguimiento de reglas provienen de nuestra *decisión* acerca de cómo seguir la regla (1989b). Esto no quiere decir que quien sigue una regla pueda decidir que cualquier cosa que pueda acabar haciendo es lo que la regla exige. Solo nuestros *mejores* juicios, nuestras mejores decisiones, sobre el enunciado de una regla son los que *determinan* su significado, el curso a seguir[24]. Para Wright, nuestros mejores juicios son "juicios hechos en lo que son, con respecto a su tema particular, condiciones *cognitivamente ideales* tanto del juez como de la circunstancia" (1989b, p. 192), condiciones que Wright denomina 'condiciones–C': un sujeto sincero, cognitivamente lúcido y alerta, que tiene el repertorio conceptual relevante y no hay ninguna razón para pensar que pueda engañarse a sí mismo. La verdad para tales juicios es, por tanto, lo que juzgamos que es verdad cuando operamos bajo condiciones cognitivamente ideales ya que "las mejores opiniones de los sujetos determinan, en lugar de reflejar, lo que es verdadero acerca de sus estados intencionales" (1989b, p. 200)[25]. Wright, sin embargo, aclara que este tipo de juicios pueden ser rechazados, o cuestionados, bajo circunstancias especiales en las que una o más de las condiciones-C no se cumplen y apela, al igual que Kripke, a nuestra pertenencia a una comunidad lingüística como garante de la validez de los mismos[26].

---

[24] Wright distingue entre los juicios en los que nuestras *mejores opiniones* determinan la extensión del predicado de verdad (*extension-determining*) y aquellos que, como máximo, reflejan una extensión determinada independientemente (*extension-reflecting*) (1989b, p. 192).

[25] En este sentido, Wright compara los significados de un sujeto con las propiedades secundarias. Wright utiliza los colores como ejemplo para su comparación: "la relación entre qué color tiene algo y cómo parece visualmente a los humanos con visión normal es una relación constitutiva: el modo apropiado de impresionabilidad de los humanos con visión normal en las circunstancias correctas no es mero indicador de un color auto-estable, sino que pertenece a la esencia de lo que consiste tener ese color. El pensamiento correspondiente sobre el significado –y uno que proporcionaría la réplica perfecta al escéptico de Kripke– es que, en circunstancias normales, pertenece a la esencia del significar esto o aquello con una expresión particular, ahora o en el pasado, que nos parezca a nosotros mismos que así es" (2001, pp. 86-87)

[26] Sin embargo, al contrario que Kripke, su apelación a la comunidad lingüística conlleva la aceptación de la existencia de hechos acerca del significado: "De hecho, hay hechos sobre lo que quiero decir, en contra del escéptico de Kripke, y están constitutivamente restringidos por lo que considero que son; pero la validez de estas auto-impresiones está, a su vez, constitutivamente limitada por su contribución a mi capacidad de darme sentido a mí mismo y a los demás en mi comunidad (de hablantes)" (Wright, 2001, pp. 87-88). John McDowell ha sostenido un amplio debate con Wright respecto a los de hechos acerca del significado. McDowell también sostiene su existencia pero, a diferencia de Wright, para McDowell estos hechos son independientes del sujeto. Para una discusión al respecto véase McDowell (1982, 1984, 1991, 1994, 1998) y Wright (1986, 1996a, 1996b, 2001a).

Como señalé al principio del apartado, Wright realiza su análisis del seguimiento de reglas para desarrollar su enfoque constitutivista del autoconocimiento. En coherencia con este análisis y evitando caer en una concepción cartesiana de la mente, Wright rechaza que la autoridad de primera persona sea consecuencia de la naturaleza de los estados intencionales, de una relación epistémica privilegiada del sujeto para con los mismos, y nos exhorta a considerar el siguiente bicondicional:

> X intends that P if and only if X is disposed to avow the intention that P, and would be sincere in so doing, and fully grasps the content of that intention, and is prey to no material self-deception, and ... and so on (1987, p. 139).

Este bicondicional, como cualquier otro, puede descomponerse en dos condicionales. Si otorgamos prioridad al lado izquierdo del bicondicional obtenemos lo que Wright denomina una 'lectura detectiva': Dadas las condiciones-C[27], si X tiene intención de P entonces X está dispuesto a declarar su intención de que P. Según esta lectura, el lado izquierdo del condicional describe un estado de cosas determinado que el sujeto es está en condiciones de *captar* si se cumplen todas las cláusulas del lado derecho (1987, pp. 139 y ss.). Si, por el contrario, le otorgamos prioridad al lado derecho del condicional, obtenemos lo que Wright llama una 'lectura constitutiva': Dadas las condiciones-C, si X está dispuesto a declarar (*to avow*) su intención de que P entonces X tiene intención de P. Según esta lectura, la *disposición* a declarar *constituye* el estado de cosas del que el lado izquierdo informa cuando las cláusulas se cumplen. Wright rechaza la lectura 'detectiva' sosteniendo que la autoridad de primera persona, "la autoridad *otorgada* de manera estándar a las propias creencias del sujeto, o a los avowals expresados, sobre sus estados intencionales es un *principio constitutivo"* (1991, p. 312, primeras cursivas añadidas).

Llegados a este punto, es importante detenernos en varias de las afirmaciones de Wright.

En primer lugar, es necesario aclarar que la *disposición* a declarar (*to avow*) que señala Wright no hay que entenderla como una disposición natural en sentido tradicional, como pueda serlo, por ejemplo, la disposición a romperse de un cristal, sino que este tipo de disposiciones son, según él, comunitarias y *normativas*, es decir, son disposiciones al acuerdo en el juicio y la acción de toda la comunidad lingüística. Estas disposiciones provienen de nuestra introducción al lenguaje, del ser entrenados cuando

[27] Es decir, "es sincero al hacerlo, y entiende completamente el contenido de esa intención, y no es víctima de ningún autoengaño material, y … y así sucesivamente" (1989a, p. 139).

aprendemos a compartir un lenguaje y tienen sus raíces en el éxito de las prácticas derivadas del esquema interpretativo cooperativo de los miembros de la comunidad lingüística. De ahí que, según Wright, las (auto-)adscripciones tiendan a estar de acuerdo con las evaluaciones de los otros miembros de la comunidad (Wright, 1987).

Asimismo, Wright sostiene que la autoridad es "una *concesión otorgada* extraoficialmente a cualquiera a quien uno tome en serio como sujeto racional" (1987, p. 138, cursivas añadidas), lo cual supone la aceptación que la autoridad no es algo meramente individual sino que conlleva la inclusión de otro miembro de la comunidad que es quien, extraoficialmente, concede (o reconoce, atribuye) dicha autoridad. Sin embargo, Wright entiende la figura de ese otro miembro de la comunidad de forma impersonal, es decir, como una característica de la "gramática" (*grammar*) en el sentido wittgensteiniano del término, es decir, del conjunto de reglas y criterios que constituyen nuestros conceptos psicológicos y determinan el significado de las palabras usadas para expresarlos. De esta manera, Wright sitúa la autoridad en un plano teórico, como algo que se otorga por defecto (*by default*) a cualquier miembro de la comunidad que cumpla con las condiciones-C y que no puede ser cuestionada a menos que hayan razones positivas para ello, es decir, circunstancias especiales en las que una o más de las condiciones-C no se cumplen (Wright, 1996b).

Wright, al igual que otros muchos autores, como señala Borgoni (2018), no parece tener en cuenta que esta última afirmación descansa en una concepción idealizada del autoconocimiento en la que los avowals son realizados en un "entorno esterilizado" que obvia los factores sociales (2018, p. 6). En este sentido la crítica de Borgoni se asemeja a la que Cassam realiza a Moran. Tanto el enfoque de Moran como el de Wright están desarrollados desde un punto de vista teórico, idealizado, que deja fuera aspectos de la conducta humana que están lejos de ser 'situaciones cognitivamente ideales de racionalidad'. Del mismo modo que los seres humanos estamos lejos de ser ciudadanos epistémicos modelo cuyas actitudes son como racionalmente deben de ser, nuestras interacciones con los demás están, asimismo, lejos de ser como racionalmente deberían de ser. Borgoni apela a los casos de injusticia epistémica (en concreto de injusticia testimonial) descritos por Miranda Fricker (2007), casos en los que el interlocutor da un bajo nivel de credibilidad al testimonio de una persona debido a determinados prejuicios (racistas, sexistas, etc.). Así, por ejemplo, un policía puede no creer el testimonio de una persona frente al de otra por ser la primera de raza negra y la segunda de raza blanca. O una persona puede no tomar en serio las palabras de una

mujer por el simple hecho de ser mujer debido a sus prejuicios machistas. De esta manera, según Borgoni, la suposición de que las dudas acerca de los avowals están fuera de lugar, es sencillamente falsa (2018, p. 6). La autoridad de primera persona no está garantizada (*granted*) a menos que nuestros interlocutores, los demás miembros de la comunidad, la *reconozcan*. Wright, por tanto, presupone el reconocimiento por parte del otro sin tener en cuenta lo que *de facto* ocurre en nuestras interacciones en segunda persona, en las que el reconocimiento de la autoridad puede ser unilateral y no mutuo, es decir, en las que la autoridad no es reconocida por uno de los participantes en la interacción (volveré a este argumento en el siguiente apartado dedicado a Bilgrami).

Finalmente, en lo que se refiere a la estipulación, la decisión mediante la cual estipulamos tanto el significado de una regla como nuestros estados mentales, la idea de que lo que uno dice o piensa sobre sus propios estados mentales desempeña un papel constitutivo respecto a los mismos puesto que establece lo que son, Finkelstein sostiene acertadamente, que Wright cae en la crítica que él mismo señala. Como hemos visto, según Wright, cada regla y cada estado intencional adquiere su contenido mediante cierto tipo de estipulación. Asimismo, Wright afirma que esta apelación a la estipulación evita el regreso *ad infinitum* de la interpretación, ya que esta última siempre está sujeta a una nueva interpretación y esta, a su vez a otra y así sucesivamente. Sin embargo, según Finkelstein la estipulación no evita este regreso al infinito puesto que "esta concepción conlleva una forma de platonismo poco común, según la cual, si bien las reglas, intenciones, recetas y deseos son vacíos en sí mismos, las estipulaciones, de alguna manera, tienen contenido intrínsecamente" (2003, p. 44)[28].

En este apartado hemos visto cómo Wright desarrolla su enfoque constitutivo a partir del análisis del seguimiento de reglas wittgensteiniano. Según Wright, los avowals son autoritativos, es decir, si un sujeto, en condiciones cognitivamente ideales realiza una auto-atribución de un estado mental, ya sea de intención, de creencia, deseo, etc., dicha auto-atribución debe ser tomada como verdadera por defecto, es decir, no deben ser cuestionada a menos que existan razones positivas para hacerlo. Dicha autoridad no está basada en una ventaja cognitiva por parte del sujeto sino, más bien, es

[28] De otra parte, Finkelstein afirma que el constitutivismo no puede dar cuenta de los estados mentales fenoménicos, como, por ejemplo, de una jaqueca. Finkelstein plantea una objeción que denomina "la objeción de la responsabilidad" según la cual "(s)i mi jaqueca estuviese constituida en alguna medida significativa por mi avowal, tendría sentido que se me culpase por ella" (p. 46). Según Finkelstein, aun cuando en ocasiones se nos pueda culpar por nuestras jaquecas (por ejemplo, si hemos trasnochado y bebido en exceso) no se nos puede culpar por el hecho de auto-atribuírnoslas o por creer que las tenemos cuando las tenemos.

"una *concesión* otorgada extraoficialmente a cualquiera a quien uno tome en serio como sujeto racional" (1987, p. 138). Para Wright, no hay nada oculto respecto a lo mental y la apariencia de que las cosas son de otra manera se debe a una comprensión errónea de los aspectos constitutivos de nuestros conceptos psicológicos. Los avowals son un ejemplo de cómo se "juega nuestro juego del lenguaje", de nuestra "gramática", y su significado depende del *juicio de los hablantes*, está constituido por los hablantes mismos, es decir, por los miembros de una comunidad lingüística.

Sin embargo, como he señalado, aun cuando Wright incluye en su explicación a los otros miembros de la comunidad, esta inclusión es realizada desde un plano teórico en el que no se tienen en cuenta aspectos factuales de nuestra conducta, situaciones en las que nuestra interacción con los demás no encaja en lo 'cognitivamente ideal' debido a determinados prejuicios que escapan a los razones positivas que Wright señala. Wright, por tanto, parte de una concepción del problema del autoconocimiento en el que la estrategia a seguir es la de mostrar lo distintivo de las auto-atribuciones de estados mentales *en oposición* a las atribuciones de estados mentales a los demás. Su enfoque está centrado, al igual que el introspeccionista y el enfoque deliberativo de Moran, en una dicotomía primera-tercera persona que sitúa al individuo en el centro de la explicación. En ese sentido, el enfoque de Wright es también un enfoque individualista pues aun cuando incluye a los demás miembros de la comunidad en su explicación, estos son vistos desde un punto de vista impersonal, teórico, desde el cual la autoridad de primera persona sigue siendo una característica que depende exclusivamente del individuo y que está relacionada con la constitución por el mismo de sus estados mentales Al respecto, he señalado que Finkelstein sostiene acertadamente que la constitución de los estados mentales mediante la estipulación cae en la misma crítica que Wright hace de la interpretación, ya que la estipulación cae igualmente en un regreso al infinito, so pena de entenderla como poseyendo un contenido intrínseco, lo cual se acerca peligrosamente al platonismo que el mismo Wright rechaza.

En el siguiente apartado veremos el enfoque de Bilgrami, quien también desarrolla un enfoque agencialista similar al de Wright y Moran, pero con algunos matices sustanciales.

4.2.2 Akeel Bilgrami

El enfoque agencialista respecto al autoconocimiento de Bilgrami combina elementos de los enfoques de Moran y de Wright. Al igual que Moran, Bilgrami centra su enfoque en la agencia responsable, sin embargo, a diferencia de este y al igual que Wright, sostiene que el autoconocimiento es un hecho *a priori* de la agencia racional y responsable. Para Bilgrami, no podríamos ejercer la agencia involucrada en nuestras actitudes intencionales a no ser que ya conozcamos dichas actitudes. Su enfoque está motivado, o inspirado, en la interpretación del libre albedrío que Peter F. Strawson desarrolló en su artículo "Freedom and Resentment" (1974). De hecho, el título del libro en el que desarrolla su enfoque, *Self-Knowledge and Resentment* (Bilgrami, 2006), es un reflejo, como vemos, del título del artículo de Strawson. Bilgrami desarrolla y extiende la idea de Strawson, de que la libertad de la acción humana es una *presuposición* en nuestras prácticas en torno a la responsabilidad y las "actitudes reactivas" que las sustentan, al autoconocimiento. Asimismo, suscribe el bicondicional de Wright pero aportando una nueva lectura del mismo. Veámoslo con más detalle.

La tesis de Strawson se centra en las relaciones interpersonales, en concreto en las denominadas "actitudes reactivas", ya sean hacia los demás (como el resentimiento y la gratitud, la indignación o la crítica y la aprobación moral) o hacia nosotros mismos (como la autocrítica, el remordimiento o el orgullo). Strawson considera estas actitudes centrales para la vida humana ya que, según él, "todos estos tipos de actitudes tienen raíces comunes en nuestra naturaleza humana y en nuestra pertenencia a las comunidades humanas" (1974, p.17). Es el hecho de que en *nuestra naturaleza* está el reaccionar normativamente (es decir, el hecho de que tengamos actitudes reactivas que hacen que sintamos que hemos de castigar o premiar la acción del otro en nuestras relaciones interpersonales) lo que subyace y justifica la atribución general de responsabilidad y libertad a nuestras acciones y las prácticas de premio y castigo relacionadas. Estas actitudes son inevitables y centrales en nuestras relaciones interpersonales, son parte de lo que somos como seres humanos que se relacionan entre sí. Involucrarse en "relaciones interpersonales, como normalmente las entendemos, es estar expuesto al rango de actitudes y sentimientos reactivos en cuestión" (1974, p.12). Para Strawson, por tanto, nuestras actitudes reactivas son lo que subyace, lo que

justifica nuestras prácticas normativas relacionadas con la responsabilidad que conlleva nuestro estatus de agentes libres.

Bilgrami, siguiendo a Strawson, sostiene que si tratamos de fundamentar las actitudes reactivas de manera metafísica, es decir, de manera no normativa, corremos "el peligro de caer en la falacia naturalista" (1998, p. 229, 2006, p. 75). Sin embargo, no cree que el análisis deba detenerse en la conclusión de Strawson de que las actitudes reactivas son parte de lo que somos como seres humanos que se relacionan entre sí, es decir, que tenemos actitudes reactivas porque así es como somos los seres humanos. Para Bilgrami, debemos continuar el análisis para darnos cuenta de que nuestras evaluaciones, la normatividad, van mucho más allá de las actitudes reactivas y de que si los seres humanos somos así, si de hecho tenemos dichas actitudes, es debido a ciertos valores y objetivos que adoptamos y perseguimos (1998, pp. 216–217). En lo referente a la justificación de las actitudes reactivas, de los valores y de la misma agencia no podemos salir del ámbito de los valores ni tampoco existe un fundamento primitivo dentro de los valores mismos.[29] En palabras de Bilgrami: "Podemos justificar que somos agentes con actitudes reactivas, es decir, agentes que son sujetos normativos, citando las normas particulares o los valores que promueven" (2012, p. 275). De esta manera, Bilgrami extiende la afirmación de Strawson sobre la normatividad de las actitudes reactivas a la idea de la agencia humana como, asimismo, normativa (volveré más adelante a este argumento).[30]

Bilgrami continúa la extensión de la idea strawsoniana relacionándola, finalmente, con el autoconocimiento. Como hemos visto, Strawson sostiene que la libertad de la acción humana hay que considerarla como una condición necesaria en nuestras prácticas en torno a la responsabilidad y las "actitudes reactivas" que las sustentan. Culpamos o sentimos resentimiento *solo* en la medida en la que consideramos que la otra persona ha actuado libremente. Esta presunción de libertad conlleva que las actitudes reactivas que sustentan y justifican nuestras prácticas relacionadas con la responsabilidad sean, asimismo, justificables (mediante otros

[29] Bilgrami compara su argumentación con la del coherentismo anti-fundacionalista de otros dominios de estudio (1998, pp. 216–217).

[30] El análisis de Bilgrami sobre la normatividad de la agencia humana, así como el de Strawson, están centrados en seres adultos ya introducidos en el lenguaje. Sin embargo, y sin prejuicio de los análisis de Strawson y Bilgrami, nótese la relación de la idea strawsoniana acerca de que está *en nuestra naturaleza* el tener actitudes reactivas que hacen que sintamos que hemos de castigar o premiar la acción del otro en nuestras relaciones interpersonales, con los experimentos como el de "Still-face" descritos en el primer bloque. La reacción de los bebés ante la falta de respuesta de la madre es una muestra de que está en nuestra naturaleza el reaccionar ante determinadas actitudes o comportamientos ya desde los 2-3 meses de edad, mucho antes, por tanto, de nuestra introducción en el lenguaje.

valores). Por consiguiente, dado que estas prácticas han de ser justificables, tanto las mismas actitudes como los valores que las sustentan y mediante los cuales las justificamos tienen, necesariamente, que ser conocidos, es decir, no podemos justificar tener actitudes reactivas ante acciones o estados intencionales que no son conocidos (2012, p. 267). Según Bilgrami, por tanto, el autoconocimiento es una condición necesaria para la agencia responsable (1998, p. 222). Para que consideremos al otro como un agente libre, hemos de considerar como una condición previa y necesaria que *conoce* tanto sus practicas relacionadas con la responsabilidad como las actitudes reactivas que las sustentan y que es, asimismo, capaz de justificarlas.

Con objeto de desarrollar esta idea, Bilgrami suscribe el enfoque de Wright, aunque solo parcialmente pues, en general, lo considera un enfoque controvertido. De ahí que aclare que su argumentación no seguirá ninguna de las líneas polémicas que se encuentran en Wright[31] evitando adoptar "ningún compromiso explícito con nada más que con el núcleo de su punto de vista" (1998, p. 211). En este núcleo, Bilgrami sitúa el rechazo a los enfoques introspeccionistas, el término 'constitutivo' como apropiado para describir la relación entre los estados de primer y segundo orden y el bicondicional del que Wright extrae la lectura constitutiva de la autoridad de primera persona. Sin embargo, Bilgrami hace una nueva lectura de este bicondicional en la que identifica las condiciones-C con las "condiciones para la agencia responsable" y cada uno de los dos condicionales (en los que podemos descomponerlo) con una característica distintiva del autoconocimiento, a saber, la transparencia y la autoridad, aunque, aclara, referidas exclusivamente a "los estados más canónicos, las creencias y los deseos, principalmente, y no a toda la gama de estados intencionales" (2006, p. 1). De esta manera, la lectura de izquierda a derecha, "dadas las condiciones-C, si uno desea o cree que *p*, uno cree que uno desea o cree que *p*", lejos de ser una lectura 'detectiva', como afirma Wright, es la que caracteriza la *transparencia* de dichos estados mentales. De otra parte, la lectura de derecha a izquierda, "dadas las condiciones-C, si uno cree que uno desea o cree que *p*, entonces uno desea o cree que *p*", es, coincidiendo con Wright, la que corresponde a la *autoridad*.

Según Bilgrami, para conocer el alcance de estos condicionales, es decir, para saber a qué estados mentales se aplican, es preciso hacer una distinción entre tipos de

[31] Entre las líneas generales polémicas que Bilgrami rechaza se encuentran la equiparación entre estados intencionales y propiedades secundarias y su conclusión anti-realista sobre ellos y, "finalmente, el contraste con el autoconocimiento del dolor y de las sensaciones en general" (1998, p. 211).

estados mentales. Así, para poder diferenciar entre los estados transparentes y/o autoritativos y los que no lo son, hay que distinguir tanto entre estados mentales de primer orden y estados mentales de segundo orden como entre estados mentales como *disposiciones* y estados mentales como *compromisos*. El condicional de la transparencia, se aplica a los estados mentales de primer orden (restringidos, como dijimos, a las creencias y deseos), mientras que la autoridad se aplica a las creencias y deseos de segundo orden, (es decir, a las auto-atribuciones) sobre la existencia de esos estados mentales de primer orden. De otra parte, los términos "creencia" y "deseo" pueden usarse para describir estados mentales normativos o estados mentales como disposiciones. Puedo usar el término "deseo", por ejemplo, para *describir* una tendencia o impulso, es decir, una *disposición* a realizar una acción, comportamiento, etc. Asimismo, "creencia" puede usarse para describir una disposición, ya que cuando las creencias 'anidan' (*nest*) con "deseos (también concebidos como disposiciones) *tienden*, en circunstancias adecuadas, a provocar un comportamiento que puede describirse como apropiado para los contenidos proposicionales mediante los cuales se especifican esas creencias y deseos" (2012, p. 264). Sin embargo, ambos términos pueden también ser usados de manera normativa, como indicando algo que debemos (o creemos que debemos) hacer, es decir, como un *compromiso*. Las creencias o deseos como compromisos nos comprometen a creer en otras cosas o a hacer ciertas cosas, respectivamente. Si creo que hay una mesa enfrente de mí, he de creer otras ciertas cosas, como, por ejemplo, que hay algo delante de mí, o que si deseo pasar por ahí voy a tropezar con la mesa a no ser que la aparte.

La diferencia fundamental que Bilgrami señala entre compromisos y disposiciones es que los estados mentales como compromisos, por su propia naturaleza normativa, son un tipo de estados que tal vez no seamos capaces de cumplir (*to live up to*) sin que dejen de ser compromisos, puesto que, como afirma Bilgrami, "está en la naturaleza de las normas el que podamos no llegar a cumplirlas" (2012, p. 265). En ese caso, el de no poder cumplir con nuestros compromisos, hemos de estar dispuestos a ser críticos con nosotros mismos (y/o aceptar las críticas de los demás). Sin embargo, al contrario que con los compromisos, la misma existencia de los estados mentales como disposiciones sería puesta en duda si, dadas las condiciones ideales para su producción,

no se produce lo que se disponía a producir (por ejemplo si afirmo de mí que tengo sed y no bebo teniendo la posibilidad de hacerlo, mi sed misma sería puesta en duda)[32].

Aclaradas estas diferencias, veamos qué tipos de estados sitúa Bilgrami como transparentes y cuáles como autoritativos.

Bilgrami desarrolla la explicación del condicional de la transparencia siguiendo la misma línea argumentativa strawsoniana: si, como hemos visto, para que una acción sea libre y responsable, para que podamos tener actitudes reactivas justificables ante ella, tanto la misma acción como las actitudes reactivas que la sustentan y los valores que las justifican ha de ser conocidos necesariamente por el agente, este conocimiento ha de ser, necesariamente, transparente para él. Según Bilgrami, tanto los compromisos como las disposiciones son transparentes. De una parte, por su propia naturaleza normativa, no puedo desconocer mis propios compromisos ya que no puedo tratar de cumplir con algo que no sé que tengo. Los compromisos son, por tanto, necesariamente transparentes, si los tengo, los conozco. De otra parte, las disposiciones, según Bilgrami, son también transparentes cuando están (potencialmente)[33] vinculadas a la producción, por parte de un *agente*, de acciones libres y responsables, que son los objetivos (potenciales) de las actitudes reactivas justificadas, ya que solamente estoy justificado por sentir resentimiento "por las acciones intencionales (por ejemplo, aquellas que causan daño) que fluyen de las disposiciones de alguien" (2012, p. 267), si tiene conocimiento propio de las mismas. Y dado que, de hecho, sentimos resentimiento

[32] Asimismo, los estados mentales como compromisos no pueden ser reducidos a disposiciones. Para mostrar esto, Bilgrami combina dos argumentos, uno de Moore y Kripke y otro de Frege. Según Bilgrami, "un argumento mooreano complementado con un argumento fregeano, construyen un efecto pinza contra la identificación naturalista de los estados intencionales y las disposiciones" (2012, p. 273). El argumento mooreano-kripkeano se apoya en la pregunta 'abierta' de Moore (1903), con la que argumenta en contra de la posibilidad de las reducciones de los términos normativos como "bueno" a propiedades naturales externas [a conceptos de propiedades naturales: creo que Moore acepta la posibilidad de que los términos normativos refieran a propiedades naturales, lo que rechaza es que sea reducibles a términos no normativos], y en la consideración kripkeana sobre la irreductibilidad de las normas a disposiciones de primer orden (ni, como vimos en el otro apartado, a ningún tipo de hecho). Según Bilgrami, del mismo modo que al tratar de reducir los términos normativos a propiedades naturales la pregunta siempre queda abierta (siempre podemos preguntar, por ejemplo, "esta acción maximiza la utilidad, pero ¿es buena?"), si tratamos de reducir los estados mentales como compromisos a disposiciones ocurre lo mismo: el agente siempre podría, no trivialmente, preguntarse "tengo tal y cual disposición de primer orden, pero ¿debo φ?". La idea de Bilgrami es que en esta pregunta, el estado mental involucrado en "¿debo φ?" no puede ser la disposición en sí misma pues, de ser así, la objeción mooreana seguiría aplicándose. De otra parte, el argumento fregeano trata de responder a la restricción de la aplicación del argumento mooreano al ámbito "definicional" (al ámbito de las definiciones de los conceptos involucrados) y a la objeción naturalista que sostiene la posibilidad de la identidad a posteriori entre compromisos y disposiciones, del mismo modo que ha ocurrido con agua = H_2O o Hesperus = Phosphorus. Bilgrami utiliza la diferencia fregeana entre sentido y referencia para afirmar que podemos negar que los compromisos sean disposiciones sin ser irracionales o inconsistentes aun cuando no conozcamos la posible identidad a posteriori entre ambas. Según Bilgrami, no importa que no conozcamos todos los sentidos de un mismo referente (en este caso todos los sentidos del referente de "estado mental intencional, o compromiso") porque siempre y cuando el sentido refiera a una propiedad natural (en este caso a una disposición entendida en términos naturalistas) el argumento mooreano-kripkeano de la pregunta abierta puede aplicarse de nuevo.

[33] La expresión "*potencialmente* vinculadas" significa que podrían formar parte de la justificación racional de una acción que pudiera dar lugar a una actitud reactiva, aun cuando el agente todavía no la haya realizado.

ante las acciones intencionales que tienen estados disposicionales vinculados a ellas, dichos estados han de ser, necesariamente, transparentes para el sujeto de la acción.

Bilgrami define la transparencia como una propiedad de los estados mentales de primer orden y afirma que, "por su propia naturaleza y no por mera contingencia" (2012, p. 263), un estado mental de primer orden, una creencia o un deseo, es transparente si se puede decir que es conocido por su poseedor y, como hemos visto, este ha de serlo si forma parte de las acciones con respecto de las cuales tenemos actitudes reactivas.

En lo referente a la autoridad Bilgrami resalta su relación con la verdad. Según Bilgrami, "una creencia de segundo orden sobre la presencia de una creencia o deseo de primer orden tiene autoridad si, por su propia naturaleza, puede decirse que es una creencia verdadera" (2012, p. 264). Con "por su propia naturaleza" Bilgrami está identificando la autoridad con una propiedad de los estados mentales de segundo orden, es decir, de las auto-atribuciones, y, por tanto, identificándolos como siempre verdaderos. Si un avowal, una auto-atribución sincera de un agente responsable, es siempre verdadero, también el condicional "dadas las condiciones-C, si uno cree que uno desea o cree que *p*, entonces uno desea o cree que *p*", es siempre verdadero. Ahora bien, si nuestros estados mentales como compromisos son siempre autoritativos y, por tanto, verdaderos, ¿dónde queda la posibilidad de engaño?, o más concretamente, ¿cómo puede ser que como agentes responsables, en ocasiones, estemos (o podamos estar) equivocados respecto a ellos por el autoengaño? Aún cuando su enfoque parece indicar una infalibilidad respecto al autoconocimiento que no da lugar a la posibilidad de autoengaño, Bilgrami sostiene la veracidad del condicional incluso en los casos en los que somos víctimas del mismo. Según Bilgrami, los casos de autoengaño no son debidos a que el agente se equivoque respecto al estado mental de primer orden, el compromiso que es objeto de la auto-atribución, y realice una auto-atribución falsa. Más bien, lo que el autoengaño nos muestra es que el agente no ha sido capaz de cumplir en su comportamiento con el compromiso en cuestión, no que no sea verdad que tiene dicho compromiso. Esto es así, según Bilgrami, debido a que el estado mental de primer orden que es objeto de la auto-atribución, el compromiso, es inconsistente con otro estado mental de primer orden, en este caso de una disposición, que "no es *transparente* para su poseedor" (2012, p. 269), lo cual hace que ambos entren en conflicto y que el agente se comporte de formas que van en contra de su compromiso. Para Bilgrami, esto no es contradictorio con la verdad del condicional porque, en el caso del autoengaño, el

estado mental en cuestión (la disposición no transparente que entra en conflicto y crea un comportamiento contradictorio) no cumple las condiciones para la agencia responsable, a saber, no es conocida por su poseedor y por lo tanto, el condicional no se aplica en este caso. De esta manera, Bilgrami escapa a la crítica que Cassam realiza a Moran, ya que según Bilgrami tener actitudes recalcitrantes no es una cuestión de irracionalidad sino algo normal en los seres humanos, resultado de la existencia de dos actitudes contradictorias. El hecho de que existan este tipo de actitudes no es una amenaza para la agencia responsable según la entiende Bilgrami sino, más bien, un indicador de nuestra obligación a estar dispuestos a ser críticos con nosotros mismos (o aceptar las críticas de los demás) y tratar de hacerlo mejor para poder cumplir con nuestros compromisos (en este caso, una obligación a estar dispuesto a desprendernos de las actitudes recalcitrantes que contradicen lo que afirmamos sobre nosotros mismos).

Sin embargo, aunque Bilgrami escape a la crítica de Cassam, también sitúa la autoridad en un plano teórico, como una característica *a priori* que los demás han de reconocer en toda persona que cumpla las características para la agencia responsable. Su análisis, al igual que el de Wright y el de Moran, descansa en una concepción idealizada del autoconocimiento en la que los avowals son realizados en un "entorno esterilizado" que, como he señalado anteriormente, obvia los factores sociales. Asimismo, aun cuando Bilgrami, siguiendo a Strawson, incluye las interacciones en su explicación del autoconocimiento, su enfoque sigue siendo, principalmente, un enfoque individualista basado en la oposición entre la primera y la tercera persona. Su objetivo es explicar las características distintivas del autoconocimiento desde y en relación al individuo. La transparencia es vista únicamente como algo relativo a la relación del individuo con sus estados mentales, sin tener en cuenta si existe otro tipo de transparencia relacionada con las interacciones y el acceso directo del otro a nuestros estados mentales, como he defendido en el primer bloque que ocurre en la comunicación expresiva. Del mismo modo, la autoridad de primera persona es vista como algo inherente al individuo, como una característica que, una vez dadas las condiciones para la agencia responsable no puede ser cuestionada. Al respecto conviene recordar, como nos exhorta Strawson, el olvido filosófico contemporáneo de las relaciones interpersonales:

> The object of these commonplaces is to try to keep before our minds something it is easy to forget when we are engaged in philosophy, especially in our cool, contemporary style, viz. what it is actually like to be involved in ordinary interpersonal relationships, ranging from the most intimate to the most casual (1974, p. 7).

La llamada de atención de Strawson está enfocada a las actitudes reactivas, sin embargo, la importancia de nuestras relaciones interpersonales no se limita a estas. Como hemos visto que ocurre en el casos de la injusticia epistémica, en nuestras interacciones con los demás no solo reaccionamos con resentimiento, aprobación, gratitud, etc., sino que también se dan situaciones en las que la autoridad de primera persona no le es reconocida a una persona aun cuando esta no deja, de hecho, de tener autoconocimiento. Al respecto, Borgoni propone que imaginemos situación en la que una mujer y un hombre están revisando solicitudes de empleo y, tras haber revisado las solicitudes de todos candidatos, la mujer le dice a su compañero: "No estoy de acuerdo con tu propuesta. Creo que es esta persona la que deberíamos contratar". Ante ello, menospreciando lo que su compañera acaba de decir y debido a prejuicios sexistas, el hombre le responde: "No, no lo crees, a las mujeres simplemente os gusta meteros con las opiniones de los hombres" (2018, p. 4). En este caso, el hombre está planteando dudas sobre la creencia que la mujer acaba de formarse basándose en la consideración de las razones oportunas para ello, es decir, le está sugiriendo que se ha equivocado respecto a su propia creencia, que no tiene la creencia de que su propuesta sea mala sino un deseo de meterse con lo que él, como hombre, opina —o, en caso de tenerla, que su creencia es falsa. Sin embargo, la mujer conserva, de hecho, su autoconocimiento pues no deja de saber lo que cree realmente, independientemente de que su compañero cuestione su auto-adscripción y no le reconozca autoridad sobre sus propios estados mentales[34]. Estos casos muestran que la autoridad de primera persona conlleva necesariamente un elemento atribucional, de reconocimiento del otro de tal estatus y que la ausencia de reconocimiento de la autoridad no conlleva la ausencia de autoconocimiento. Por tanto, aunque la autoridad sea, en el plano teórico, una característica *a priori* del autoconocimiento, nuestras interacciones personales con los demás no se ajustan al ideal racional propuesto por los agencialistas, por lo que su análisis resulta, cuanto menos, incompleto.

Además de con las interacciones personales, en el sentido limitado que he señalado, el agencialismo de Bilgrami está relacionado con otro aspecto importante del

[34] Como vemos, esta situación no es como la descrita por Bilgrami en la que se dan dos estados mentales en contraposición que entran en conflicto y uno de los cuales no es transparente para su poseedor. En esta situación, la mujer del ejemplo tiene un solo estado mental y lo conoce perfectamente (la creencia de que el candidato elegido por su compañero no es el más adecuado sino que lo es otro con mejores cualidades), al igual que su compañero también conoce perfectamente su creencia de que las mujeres, en ocasiones, carecen de autoridad de primera persona y está dispuesto a defender dicha creencia con los correspondientes argumentos sexistas.

enfoque de segunda persona: la normatividad. Como vimos al final del primer bloque, dedicado al desarrollo ontogenético, el bebé comienza a normativizar su conducta a partir, aproximadamente, de los 12 meses de edad mediante las evaluaciones de su madre ante las situaciones que desconoce o para las cuales no tiene una respuesta concreta (la cual es vista por este como una autoridad que he denominado 'autoridad de segunda persona'). La normatividad, por tanto, es un rasgo constitutivo en los seres humanos ya desde muy temprana edad y está directamente relacionada con las interacciones personales. Como señalé anteriormente, según Bilgrami, las personas somos seres normativos, la normatividad es un rasgo constitutivo de la agencia humana. Siguiendo la idea de Strawson de que la libertad y la acción humanas están constituidas por las prácticas normativas relacionadas con la noción de responsabilidad y estas prácticas, a su vez, están basadas en nuestras reacciones normativas para con el comportamiento del otro y de nosotros mismos, Bilgrami considera la agencia humana como normativa en sí misma. Según Bilgrami, la agencia no puede explicarse desde ningún punto de vista teórico o marco metafísico, sino solo desde un punto de vista práctico y normativo, ya que "somos criaturas con responsabilidad y con estados mentales que se describen en términos radicalmente normativos y no meramente como motivos y disposiciones" (2006, p. xiii).

Sin embargo, hay un aspecto de la normatividad que, aun cuando está implícito en sus afirmaciones, no forma parte central ni de su enfoque ni de los otros enfoques agencialistas: la normatividad social. Según el enfoque de segunda persona, este elemento ausente en los análisis agencialistas, la normatividad social, resulta crucial para la comprensión del fenómeno del autoconocimiento y está íntimamente relacionado con los compromisos que adoptamos como agentes responsables al auto-adscribirnos estados mentales. Como he señalado en el apartado dedicado a Moran, este sostiene que al realizar un avowal de creencia no sólo se registra la presencia de la creencia sino que se *expresa un compromiso* con su verdad. Según Moran "al concebirme a mí mismo como un agente racional, mi conciencia de mi creencia es la conciencia de mi compromiso con su verdad, un compromiso con algo que trasciende cualquier descripción de mi estado psicológico" (2001, p. 74). Al realizar un avowal no describo ni expreso mi estado mental de creencia, como sí puedo expresar, por ejemplo, mi estado mental de dolor. Al realizar un avowal de creencia, *afirmo* dicha creencia y *expreso* un compromiso con su verdad. Moran entiende que este compromiso es el resultado de la deliberación, es algo que adquiero tras haber deliberado acerca del hecho

en cuestión. Según Moran, delibero sobre un hecho, llego a la conclusión *p* acerca de ese hecho y, dada la Condición de la Transparencia, obtengo la creencia acerca de que *p* de manera directa sin depender de ninguna evidencia psicológica sino sólo de mi deliberación acerca de ese hecho. El avowal de dicha creencia es una expresión de mi libertad racional, una afirmación de mi autoridad que conlleva un compromiso como agente responsable de esa misma auto-adscripción. De esta manera, la autoridad implica deliberación, es decir, mis auto-adscripciones tienen autoridad puesto que provienen de mi deliberación, son una decisión que he tomado tras deliberar, y esta decisión no es solo evidencia (empírica) de lo que haré, sino una resolución de la que soy el autor y responsable de llevar a cabo (2001, p. 79).

De otra parte, Bilgrami sostiene que una creencia o deseo como compromisos son compromisos de creer en otras cosas o de hacer ciertas cosas (2012, p. 265). Los compromisos, para Bilgrami, son tanto teóricos (o inferenciales) como prácticos y han de ser necesariamente transparentes, si los tengo, los conozco, ya que forman parte de las acciones con respecto a las cuales tenemos actitudes reactivas. Según Bilgrami, la diferencia entre los estados mentales como compromisos y los estados mentales como disposiciones radica en que los primeros son algo que debemos (o creemos que debemos) hacer, es decir, algo con lo que estamos comprometidos, mientras que los segundos son tendencias o impulsos a realizar una acción, comportamiento, etc. De esta manera, los compromisos, por su naturaleza normativa, son algo que pueden no llegar a cumplirse, sin que dejen de ser compromisos, es decir, no por el hecho de no llevarlos a cabo han de ser necesariamente puestos en duda ya que, como hemos visto, "está en la naturaleza de las normas el que podamos no llegar a cumplirlas" (2012, p. 265). Sin embargo, según Bilgrami, la misma existencia de las disposiciones sería puesta en duda si, dadas las condiciones ideales para su producción, no se produce lo que se disponía a producir.

Ambos autores, por tanto, basan su análisis del compromiso en términos de algo que es resultado de nuestra decisión. En este sentido, para Moran estoy comprometido con la verdad de mis creencias cuando me las auto-adscribo ya que, dada la Condición de la Transparencia, no puedo no creer aquello que afirmo como producto de mi deliberación. Bilgrami, por su parte, sostiene que los estados mentales como compromisos son intrínsecamente normativos, es decir, al auto-adscribirme un estado mental como compromiso me comprometo a creer ciertas cosas y/o a hacer ciertas cosas, es decir, es uno mismo como agente quien decide asumir dichos compromisos al

hacer un uso normativo de los términos de creencia o deseo. Sin embargo, estos autores no tienen en cuenta la existencia de unos compromisos que no son resultado directo de nuestra decisión personal, sino que son compromisos que todos los miembros de la comunidad asumimos al participar en el "juego del lenguaje" de las (auto-)atribuciones de estados mentales. Este tipo de compromisos son el resultado de las normas sociales que regulan ese tipo de auto-adscripciones y están relacionados, no con algo que decidimos hacer, sino con cómo se supone que debemos comportarnos de acuerdo a dichas normas. De esta manera, el compromiso no es solo adquirido por una decisión del agente sino que es algo inherente a su interacción con los demás miembros de la comunidad como copartícipe de las normas que regulan este tipo de auto-adscripciones y los comportamientos que de ellas se derivan.

Siendo esto así, ¿por qué limitar el análisis del compromiso a las creencias y los deseos?, ¿las auto-atribuciones de emociones o sensaciones no comprometen, asimismo, al agente responsable a actuar de acuerdo a lo que se espera de alguien que está en el estado mental auto-atribuido? Bilgrami sostiene que las disposiciones son estados mentales que *tienden* a provocar un comportamiento que no es una decisión del agente, mientras que los estados mentales como compromisos son algo que decidimos creer o hacer. De esta manera, al no llevar a cabo una disposición se dudaría de la existencia de la disposición misma, mientras que al no llevar a cabo un compromiso no se dudaría del mismo sino de nuestra capacidad para cumplir con él. Ahora bien, que una persona *tienda* a tener un comportamiento debido a una disposición no quiere decir que tenga, *necesariamente*, que actuar de acuerdo a dicha tendencia, so pena de dudar de su existencia. Del mismo modo que podemos no cumplir con los compromisos que decidimos adoptar, podemos no realizar una acción que tendemos a realizar. Pongamos el ejemplo de una auto-atribución de dolor. Supongamos que expreso el avowal "Me duele la cabeza" y acto seguido voy al aparato de radio y subo el volumen al máximo. En este caso mi auto-adscripción será, con casi toda seguridad, puesta en cuestión, es decir, la veracidad de la misma será puesta en duda, pero no porque no me duela la cabeza, sino por el hecho de que no me comporto como se espera que se comporte alguien a quien le duele la cabeza. Lo mismo ocurre si, por ejemplo, digo de mí que no tengo frío y, tras ello, me pongo un abrigo. El motivo por el cual me pongo el abrigo puede que no sea porque tenga frío sino, por ejemplo, que lo he encargado por internet y me he acordado que he de probármelo porque es el último día en que puedo devolverlo si no se ajusta a mis medidas. El ejemplo de la sed y el del tabaco pueden resultar muy

esclarecedores al respecto. Supongamos que digo de mí que tengo sed y frente a un vaso de agua decido no beber. Según Bilgrami lo que sería puesto en duda sería mi sed misma, puesto que la sed es una disposición que *tiende* a hacerme actuar de determinada manera (en este caso bebiendo agua). Ante esto, según Bilgrami, mi interlocutor podría dudar de la sed misma, es decir, podría dudar de que tengo sed y cuestionarla diciendo "¿Tú no tenías sed?". Sin embargo ante una auto-adscripción de un estado mental como compromiso mi interlocutor no tendría por qué dudar del mismo. Si, por ejemplo, digo acerca de mí que quiero dejar de fumar, y me fumo un cigarro, mi compromiso de dejar de fumar no sería puesto en duda, sino mi capacidad para cumplir con dicho compromiso. Ante esta situación mi interlocutor lo que cuestionaría no es mi compromiso de dejar de fumar sino mi capacidad para hacerlo.

Ahora bien, ¿qué diferencia hay entre no cumplir con lo que digo acerca de mí mismo respecto a mis compromisos (como el de dejar de fumar) y no cumplir con lo que se espera que haga al auto-adscribirme cualquier otro estado mental? De una parte, mi compromiso de dejar de fumar puede ser cuestionado no porque no pueda cumplir con él sino porque, de hecho, no tenga dicho compromiso. En este caso mi interlocutor podría preguntar "¿Tú no querías dejar de fumar?", cuestionando mi compromiso de dejar de fumar, y no afirmar, "No puedes dejar de fumar", asumiendo que es mi capacidad para dejar de fumar la que está en cuestión. Asimismo, en el caso de la sed, mi interlocutor no tiene por qué cuestionar mi sed, sino el hecho de que no beba teniendo sed. Mi comportamiento puede ser puesto en duda porque no actúo de acuerdo de lo que se espera de alguien que tiene sed. Puede que decida no beber porque no me fío de las condiciones de salubridad del agua que tengo delante, no necesariamente porque no tenga sed. En este sentido, ser un agente responsable también implica ser un agente que es capaz de ajustar los estados mentales que se auto-adscribe con sus acciones y sus acciones con los estados mentales auto-adscritos. El compromiso, por tanto, puede ser entendido como el compromiso a actuar de acuerdo a lo que decimos de nosotros mismos, es decir, como el compromiso a actuar según se espera que actúe alguien que se auto-adscribe determinado estado mental, sea un deseo, una creencia, una sensación o una emoción. En este sentido, podemos afirmar que expresamos este tipo de compromiso con todas y cada una de nuestras auto-atribuciones de estados mentales.

En este apartado he analizado tres de los autores más representativos del enfoque agencialista: Moran, que tomando la idea de Evans sostiene la Condición de la

Transparencia como método de autoconocimiento de las creencias y defiende la agencia deliberativa como garante de la autoridad de primera persona; Wright, que desarrolla su enfoque siguiendo el análisis wittgensteiniano del seguimiento de reglas y para el que la autoridad de primera persona es una característica de las reglas gramaticales, del "juego del lenguaje" de la comunidad lingüística, las cuales tienen sus raíces en las prácticas de interpretación y cooperación de los miembros de dicha comunidad; y Bilgrami, quien, siguiendo y extendiendo el análisis de Strawson sobre la libertad, defiende la naturaleza normativa tanto de la agencia como de los estados mentales intencionales y sostiene la autoridad y la transparencia como elementos *a priori* de la agencia responsable.

Asimismo, he sostenido que todos estos enfoques son enfoques individualistas que parten de la idea de que para dilucidar las características del autoconocimiento la estrategia a seguir es la de mostrar lo distintivo de las auto-atribuciones de estados mentales en oposición a las atribuciones de estados mentales a los demás, es decir, parten de la dicotomía primera-tercera persona, lo cual descarta las interacciones en segunda persona como relevantes para el problema del autoconocimiento y sitúa al individuo en el centro de la explicación. He argumentado que tanto una cosa como la otra, situar al individuo en el centro de la explicación como descartar las interacciones en segunda persona, resultan problemáticas. En el caso de Moran conduce a una idea del ser humano excesivamente racional, algo muy alejado de lo que, en realidad, somos los seres humanos y cuya actitud deliberativa obvia el reconocimiento por parte del interlocutor. En el caso de Wright y de Bilgrami resulta en una concepción idealizada del autoconocimiento en el que los avowals son realizados en un "entorno esterilizado" que obvia factores sociales tales como la injusticia epistémica.

Finalmente, he señalado la importancia de la normatividad social para la comprensión de los compromisos del agente. Tanto Moran como Bilgrami entienden los compromisos como algo que es resultado de nuestra decisión, sin tener en cuenta la existencia de unos compromisos que no son resultado directo de nuestra decisión personal, sino que son compromisos inherentes a la interacción con los demás miembros de la comunidad como copartícipes de las normas que regulan las (auto-)adscripciones y los comportamientos que de ellas se derivan. En este sentido, he sostenido, expresamos compromisos con todas las auto-atribuciones de estados mentales, ya que son compromisos a comportarnos según se espera de alguien que se auto-atribuye el estado mental en cuestión.

En el siguiente apartado, dedicado al expresivismo, veremos también

importantes relaciones, aunque no sin diferencias considerables, entre el enfoque de segunda persona y los enfoques expresivistas. Tanto su interpretación de la adquisición y uso del lenguaje por parte del niño, como de la transparencia y de algunas características de la expresión, se ajustan a lo sostenido en el primer bloque respecto a la interpretación del desarrollo ontogenético bajo el enfoque de segunda persona.

5. Enfoques expresivistas

El término "expresivismo" designa una serie de enfoques que comparten una idea básica: algunas expresiones significativas no tienen como función representar o describir aspectos de la realidad (es decir, un rango de hechos), sino *expresar* sentimientos, sensaciones, compromisos u otras actitudes o estados mentales no cognitivos (o no representativos). El expresivismo es, por tanto, una posición pragmatista, en el sentido de que asume que hay diferentes propósitos para los que usamos el lenguaje, así como anti-representacionalista (anti-cognitivista, anti-descriptivista), puesto que rechaza la idea de que todos los términos (o expresiones) representan (o describen) estados de cosas factuales.

El expresivismo comenzó su desarrollo, a mediados del s. XX, siendo una postura semántica relativa al tratamiento de los términos morales. Sin embargo, sus orígenes pueden remontase hasta Hume (1739), quien señaló que de premisas descriptivas no se siguen premisas normativas, es decir, que del "ser" no se puede inferir el "deber ser". Según Hume, "la moralidad no es un objeto de la razón, […] la distinción entre vicio y virtud no se encuentra simplemente en las relaciones entre objetos, ni es percibida por la razón" (1739, Libro III, Parte I, Sección I), las distinciones morales provienen de las pasiones, no de las relaciones conceptuales de la razón.

A principios del s. XX y en relación con esta misma idea, Moore (1903) sostuvo que los predicados normativos son primitivos y, por tanto, no pueden definirse apelando a predicados más básicos. Su renombrado "argumento de la pregunta abierta" muestra que la pregunta acerca de los predicados normativos siempre queda abierta, no importa la cantidad de adjetivaciones o descripciones que hagamos de una acción, una situación, una persona, siempre tendrá sentido preguntar "pero ¿es buena?".

Casi a mediados de siglo, en pleno auge del positivismo lógico, Ayer (1952) señaló la diferencia en la manera de significar que tienen los enunciados éticos y la de los de las ciencias naturales. Según Ayer, al contrario que los enunciados de las ciencias naturales, los enunciados como "robar es malo" o "es bueno decir la verdad" no representan estados de cosas, no describen el mundo. Lo que hacen este tipo de enunciados es *expresar una actitud* de aprobación o desaprobación:

> Thus if I say to someone, 'You acted wrongly in stealing that money', I am not stating anything more than if I had simply said, 'You stole that money'. In adding that this action is wrong, I am not making any further statement about it. I am simply evincing my moral disapproval of it. It is as if I had said, 'You stole that money', in a peculiar tone of horror, or written it with the addition of some special exclamation marks (1952, p. 107).

De otra parte, Stevenson (1937, 1944) desarrolló esta misma idea y propuso una distinción entre dos usos del lenguaje: el uso descriptivo, cuyo objetivo es registrar, describir, cómo son las cosas, y el uso dinámico, cuyo objetivo es expresar sentimientos o actitudes y mover a la acción. Asimismo, Stevenson, distinguió entre el significado descriptivo y el significado emotivo de los términos. El significado descriptivo se corresponde con el uso descriptivo y el significado emotivo se corresponde con el uso dinámico. Dependiendo del propósito que tengamos, así usaremos las oraciones: si nuestro propósito es decir algo verdadero o falso acerca de cómo son las cosas estaremos haciendo un uso descriptivo, si, por el contrario, nuestro objetivo es expresar un sentimiento o una actitud haremos un uso emotivo. Para Stevenson, los términos evaluativos como "bueno", "malo" no están relacionados con estados de cosas, no refieren a objetos ni a propiedades del mundo. Por tanto, las expresiones morales, es decir, las oraciones de la ética, no tienen un uso descriptivo, no describen cosas sino que expresan un sentimiento o una actitud, no podemos usarlas para hacer afirmaciones verdaderas o falsas.

En la actualidad, el expresivismo ético continua siendo desarrollado por autores como Gibbard (2003, 2012) y Blackburn (1993, 1998), entre otros, y su análisis se ha extendido a otros ámbitos del discurso más allá del de la ética como, por ejemplo, al de la estética (Hopkins, 2009; Prinz, 2004), al epistémico (Chrisman, 2007, 2012), al de la lógica (Brandom, 1994) e incluso de manera global (Price, 2011). Sin embargo, aun cuando el expresivismo comenzó siendo una postura semántica relativa al tratamiento de los términos morales, podemos encontrar posiciones expresivistas fuera del ámbito

de la ética ya desde mediados del s. XX. Strawson (1949), por ejemplo, realizó un tratamiento expresivista de los predicados de verdad:

> The sentence "What the policeman said is true" has no use *except* to confirm the policeman's story; but… [it] …does not say anything further *about* the policeman's story… It is a device for confirming the story without telling it again. So, in general, in using such expressions, we are confirming, underwriting, admitting, agreeing with, what somebody has said; but … we are not making any assertion additional to theirs; and are *never* using "is true" to talk *about* something which is *what they said,* or the sentences they used in saying it (p. 93).

De otra parte, Austin (1970) [35] ofreció una interpretación expresivista de las afirmaciones de conocimiento:

> [S]aying 'I know' is taking a new plunge. But it is *not* saying 'I have performed a specially striking feat of cognition, superior, in the same scale as believing and being sure, even to being merely quite sure': for there *is* nothing in that scale superior to being quite sure. Just as promising is not something superior, in the same scale as hoping and intending, even to merely fully intending: for there is nothing in that scale superior to fully intending. When I say 'I know', I *give others my word: I give others my authority for saying* that 'S is P' (1970, p. 99).

Asimismo, Wittgenstein sostuvo una posición expresivista acerca del significado de las adscripciones de estados mentales, tales como creencias, deseos, sensaciones, expectativas, etc. El análisis de Wittgenstein está motivado en su crítica al introspeccionismo y es, quizás, el análisis expresivista más destacable fuera del ámbito de la ética. Según Wittgenstein la función de las auto-atribuciones de estados mentales no es la de describir dichos estados sino, más bien, la de expresarlos:

> How do words *refer* to sensations? —There doesn't seem to be any problem here; don't we talk about sensations every day, and give them names? But how is the connection between the name and the thing named set up? This question is the same as: how does a human being learn the meaning of the names of sensations? —of the word 'pain' for example. Here is one possibility: words are connected with the primitive, the natural, expressions of the sensation and used in their place. A child has hurt himself and he cries; and then adults talk to him and teach him exclamations and, later, sentences. They teach the child new pain-behaviour. So you are saying that the word "pain" really means crying?' —On the contrary: the verbal expression of pain replaces crying and does not

[35] En Austin podemos comprobar la relación del expresivismo con la corriente pragmática de la filosofía del lenguaje. Austin, uno de los iniciadores de dicha corriente, rechazó la idea de que los enunciados sirven solo para describir estados de cosas y realizó un análisis de los diferentes usos del lenguaje en su obra *How to do things with words* (1962), que dio lugar a su Teoría de los actos de habla. Recordemos su célebre pasaje al comienzo del libro en el que llama la atención sobre el error en el que los filósofos incurrían:

> It was for too long the assumption of philosophers that the business of a 'statement' can only be to 'describe' some states of affairs, or to 'state some fact', which it must do either truly or falsely. Grammarians, indeed, have regularly pointed out that not all 'sentences' are (used in making) statements; there are, traditionally, besides (grammarians') statements, also questions and exclamations, and sentences expressing commands or wishes or concessions (1962, p. 1).

describe it (1953, §244).

De esta manera, cuando decimos "me duele la cabeza" no estamos describiendo nuestro estado mental, sino expresando nuestro dolor. Las auto-atribuciones no describen un objeto interno al cual solo nosotros tenemos acceso sino que sustituyen a las expresiones naturales[36] de dolor, alegría, tristeza, etc.

Ahora bien si, como acabamos de ver, las expresiones no representan estados de cosas, no describen el mundo ni tienen valor de verdad, el expresivismo parece conllevar una visión deflacionaria del autoconocimiento: si las auto-atribuciones no describen sino que expresan el estado mental y no son susceptibles de ser calificadas como verdaderas o falsas y, asimismo, un estado mental no es un objeto interno que podamos describir, ¿cómo podemos decir que tenemos conocimiento de nuestro estado mental? Es por ello que, con objeto de ofrecer una explicación expresivista no deflacionaria del autoconocimiento, los autores contemporáneos que trabajan en el tema sí consideran las auto-atribuciones de estados mentales como manifestaciones del conocimiento (con algunos matices, como veremos más adelante). Estos autores parten del análisis expresivista wittgensteiniano pero sostienen que las auto-adscripciones de estados mentales en primera persona del presente de indicativo, avowals como "me duele la cabeza" o "quiero café", son expresiones del mismo estado mental expresado, en lugar de una creencia auto-adscriptiva de segundo orden y, *asimismo*, poseen condiciones de verdad.

En este apartado discutiré los enfoques de Dorit Bar-On (2004, 2009, 2011, 2013, 2015) y de David Finkelstein (1999, 2003, 2010, 2012), los cuales, tomando prestado el nombre con el que Bar-On denomina a su enfoque, agruparé bajo el término "neo-expresivismo", ya que, a pesar de ser enfoques diferentes, ambos comparten ciertas similitudes: en primer lugar, como acabamos de ver y en contraste con el expresivismo clásico, ambos neo-expresivistas sostienen que las expresiones de estados mentales tienen condiciones de verdad. Según los neo-expresivistas, al igual que otras expresiones lingüísticas con las que atribuimos estados mentales a los demás, los avowals también son expresiones lingüísticas con una estructura semántica. Por lo tanto, al igual que ellas, los avowals pueden ser verdaderos o falsos[37]. Asimismo, ambos neo-

[36] En adelante, cuando refiera a la expresión de un estado mental, como en este caso, entiéndase el término "natural" en el sentido de primitivo, primigenio, originario, no aprendido, es decir, referido a las primeras expresiones con las que el bebé se comunica con su madre de manera no conceptual.

[37] Aun cuando el neo-expresivismo es considerado un enfoque actual, esta idea ya fue previamente sostenida por Rudolf Carnap a principios del s. XX:

expresivistas concuerdan en rechazar tanto los enfoques introspectivos como los enfoques agencialistas que acabamos de ver en el apartado anterior. De esta manera, los neo-expresivistas rechazan tanto que el autoconocimiento sea el resultado de un sistema epistémico interno mediante el cual captamos y conocemos nuestros estados mentales, como que provenga de la deliberación racional o del hecho de ser constituidos por el agente, o que haya que considerarlo como un elemento *a priori* de la agencia responsable. Según los neo-expresivistas, nuestro autoconocimiento proviene del ejercicio de *nuestra capacidad para expresar* nuestros propios estados mentales: conocemos nuestros estados mentales mediante su expresión. Eso no significa que para que un individuo conozca su estado mental necesite expresarlo en voz alta pues también podemos expresarlo "en silencio", para nosotros mismos (Bar-On, 2004, p. 9; Finkelstein, 2003, p. 103).

Por último, el objetivo principal de los neo-expresivistas a la hora de desarrollar su enfoque sobre el autoconocimiento es explicar la autoridad de primera persona: el hecho de que, generalmente, los avowals no se pongan en tela de juicio ni se espere que el sujeto dé razones de lo que dice sobre sí mismo, el hecho de que les confiramos el estado de ser *verdaderos* por defecto. Veamos estos enfoques con más detalle.

5.1 Dorit Bar-On

El enfoque neo-expresivista de Bar-On hereda la idea clave del expresivismo de Wittgenstein de que las auto-atribuciones psicológicas expresan los estados mentales presentes mediante elementos lingüísticos que reemplazan las expresiones naturales de

---

> In special cases, this asserted state may be the same as that which is inferred from a certain expressive utterance; but even in such cases we must sharply distinguish between the assertion and the expression. If, for instance, somebody is laughing, we may take this as a symptom of his merry mood; if on the other hand he tells us without laughing: "Now I am merry", we can learn from his words the same thing which we inferred in the first case from his laughing. Nevertheless, there is a fundamental difference between the laughter and the words: "I am merry now." This linguistic utterance *asserts* the merry mood, and therefore it is either true or false. The laughter does not assert the merry mood but *expresses* it. It is neither true nor false, because it does not assert anything, although it may be either genuine or deceptive (1935, p. 28).

Asimismo, en Carnap también encontramos la diferencia entre la función expresiva y la representativa (o cognitiva):

> We have here to distinguish two functions of language, which we may call the expressive function and the representative ⟨or cognitive⟩^{var} function. Almost all the conscious and unconscious movements of a person, including his linguistic utterances, express something of his feelings, his present mood, his temporary or permanent dispositions to reaction, and the like. Therefore we may take almost all his movements and words as symptoms from which we can infer something about his feelings or his character. That is the expressive function of movements and words. But besides that, a certain portion of linguistic utterances (e. g., "this book is black"), as distinguished from other linguistic utterances and movements, has a second function: these utterances represent a certain state of affairs; they tell us that something is so and so; they assert something, they predicate something, they judge something (1935, pp. 27–28).

los mismos. Sin embargo, para Bar-On esto no es incompatible con un enfoque expresivista según el cual las auto-atribuciones son susceptibles de ser manifestaciones verdaderas (o falsas) de autoconocimiento.

Según Bar-On, hemos de rechazar la presuposición, rara vez seriamente cuestionada en las discusiones sobre autoconocimiento, de que la única manera de reivindicar la autoridad en primera persona es identificar las características epistémicas específicas que hacen que los avowals sean conocidos de una manera especial, ya sea a través de un acceso epistémico privilegiado o mediante un método epistémico distintivo[38]. La mejor manera de abordar la cuestión acerca del autoconocimiento es, para Bar-On, separar la cuestión acerca de la seguridad distintiva de los avowals (el hecho de que les confiramos el estado de ser *verdaderos* por defecto) de la cuestión acerca autoconocimiento privilegiado, es decir, separar la cuestión semántica de la cuestión epistémica puesto que "el logro semántico no necesita ser suscrito epistémicamente" (Bar-On and Long, 2003, p. 16).

De esta manera, Bar-On separa la cuestión sobre el autoconocimiento en dos preguntas: (i) ¿Cómo se puede explicar la seguridad distintiva de los avowals? (ii) ¿Qué hace que los avowals sean instancias de un tipo de conocimiento privilegiado?

De acuerdo con Bar-On, ambas preguntas pueden responderse apelando al carácter expresivo de los avowals. Para responder a la pregunta (i), es decir, para explicar el por qué nuestros avowals son autoritativos (por qué disfrutan de una seguridad distintiva), no tenemos necesidad ninguna de apelar a justificaciones (inferenciales o de otro tipo) o a razones positivas. Tampoco necesitamos apelar a un sistema interno de reconocimiento de nuestros estados mentales. La explicación sobre nuestra autoridad respecto a nuestros estados mentales no es una cuestión epistémica sino expresiva. Siguiendo el expresivismo wittgensteiniano, al que ella llama "Enfoque Expresivista Simple" (*Simple Expressivist Account*) (2004, pp. 228 y ss.), Bar-On sostiene que nuestros avowals expresan (dan voz, muestran, exhiben), los estados mentales en los que el sujeto se encuentra. Sin embargo, Bar-On señala que el expresivismo "simple" no ha sabido diferenciar entre la capacidad natural de expresar nuestros estados mentales, que compartimos con los animales no humanos y los bebés, y la capacidad lingüística propia de los seres humanos entrenados en una lengua. Según Bar-On, la autoridad de

[38] Es por ello que Bar-On introduce la noción de "privilegio de primera persona" frente a la de "autoridad de primera persona" como una noción más neutral que no prejuzga la cuestión de si la seguridad de los avowals es una cuestión de privilegio epistémico (Bar-On & Long, 2001).

primera persona proviene de la suma de dichas capacidades, es decir, de nuestra capacidad para expresar nuestros estados mentales junto con nuestra capacidad para hacerlo de manera lingüística (2015, p. 147).

Esta distinción entre dos capacidades expresivas, la natural y la lingüística, resulta fundamental para el neo-expresivismo de Bar-On y es, según ella lo que permite salvar la *continuidad semántica* de las auto-adscripciones con otras atribuciones de estados a otros y a uno mismo. Con este propósito, Bar-On distingue entre el *acto* de expresar y su *producto* (2004, pp. 216 y ss., 2009, pp. 67 y ss., 2011, p. 194, 2013, p. 706, 2015, p. 140). Los avowals son actos expresivos cuyos productos son las auto-adscripciones. Como *actos*, los avowals son expresiones en el sentido de la acción (*a-expresiones*). Cuando un sujeto a-expresa su estado está haciendo algo intencionadamente, es decir, está utilizando un medio expresivo para realizar una acción, por ejemplo, expresar su afecto con un abrazo[39]. De esta manera, la *a-expresión* consiste en una relación triádica entre un agente *X*, un estado mental *M* y un medio expresivo o vehículo *E* (*E* puede ser una postura corporal, una expresión facial, un gesto o incluso un comportamiento verbal mínimo). De otra parte, como *productos* los avowals son expresiones en sentido semántico (*s-expresiones*). Las s-expresiones consisten en una relación diádica entre vehículos *lingüísticos* de expresión, como oraciones, y sus contenidos semánticos. El avowal como producto, "a diferencia de una sonrisa o una mueca de dolor, o incluso un grito verbal como "¡Ay!", es una auto-adscripción semánticamente articulada con estructura semántica y condiciones de verdad" (2009, p. 68). De esta manera, para Bar-On, son los avowals como productos y no como actos los que poseen propiedades semánticas. Esta distinción, por tanto, permite a Bar-On salvar la continuidad semántica de nuestros avowals con otras atribuciones de estados a otros y a uno mismo, dado que el hecho de que podamos expresar nuestros estados mentales mediante un vehículo lingüístico, es decir, el hecho de que podamos expresar nuestros estados mentales mediante proposiciones veritativo-evaluables, permite que nuestros avowals tengan condiciones de verdad, es decir, permite que puedan ser considerados verdaderos o falsos.

Ahora bien, si esto es así ¿qué es lo que hace que un avowal sea verdadero? Según Bar-On, las auto-adscripciones que son producto de los avowals, las s-

[39] Bar-On quiere subrayar la distinción entre los actos expresivos intencionados de los sujetos de los actos no intencionados o reflejos del cuerpo como, por ejemplo, expresiones faciales involuntarias o gestos incontrolados que reflejan nuestro estado mental. Para Bar-On sólo los primeros pueden ser considerados *avowals* (v. 2004, 216–17, 249 y ss., 289, 315).

expresiones, no son meramente verdaderas o falsas, sino que son especialmente aptas para ser verdaderas, ya que si los avowals, al igual que las expresiones naturales, provienen y, por tanto, muestran los estados mentales en los que el sujeto se encuentra, comparten la misma inmediatez y fiabilidad que estas (2003, p. 23). Asumir que estoy realizando un avowal es, entonces, asumir que estoy en el estado designado, es decir, asumir el avowal como verdadero. Al expresar un avowal a-expreso el *mismo estado* cuya presencia *hace verdadera* la proposición s-expresada por el avowal (entendido como producto). Sin embargo, si los avowals son, al igual que las expresiones naturales, expresiones sinceras, directas, no inferenciales, de los estados mentales en los que el sujeto se encuentra ¿cómo es posible expresar un avowal falso? Según Bar-On, el hecho de que dotemos a los avowals de la misma fiabilidad que a las expresiones naturales, el hecho de que les concedamos el estatuto de verdaderos por defecto, no implica la imposibilidad de proferir un avowal falso. Aquí, de nuevo, la distinción entre avowal como acto y avowal como producto resulta crucial en la explicación de Bar-On, ya que, según ella, existe la posibilidad de que un sujeto exprese, como acto, un avowal sincero pero falso como producto. Bar-On argumenta que los casos en los que esto ocurre *no* son casos de *errores epistémicos* en los que el sujeto se equivoque a la hora de auto-adscribirse el estado mental en que se encuentra, no es una cuestión de que el sujeto falle al reconocer el estado mental en que se encuentra, o que lo confunda o se equivoque con otro, sino que son casos de *fallos expresivos*, "donde el fallo tiene causas psicológicas identificables" (2015, p. 146) y el sujeto expresa un estado mental pero no su estado mental (2004, p. 320–39). Por ejemplo, un estudiante de primer curso que está siendo iniciado en una fraternidad y teme que le hagan algún tipo de daño puede gritar "¡Basta! ¡Me duele el cuello!" cuando, teniendo los ojos vendados, le tocan con un trozo de hielo en el cuello. Si asumimos, como es lógico y razonable hacerlo, que el estudiante no siente dolor, su auto-atribución resulta ser falsa como producto, pero puede contar como una expresión sincera como acto (en lugar de una simulación o engaño), ya que ha expresado sinceramente *el* dolor, aunque no ha expresado *su* dolor.

Sin embargo, Coliva (2016) sostiene que la posición de Bar-On en este punto es extremadamente problemática. Coliva toma otro ejemplo de Bar-On en el que un sujeto se encuentra en la consulta del dentista y expresa dolor antes que el dentista toque su boca. Según Coliva, dada la explicación de Bar-On (según la cual en tal caso la auto-adscripción de dolor sería tanto injustificada como falsa), surge la duda sobre cómo "podríamos reconocer un cierto grado de racionalidad en el sujeto, al menos desde su

propio punto de vista" (Coliva, 2006, p. 157). Según Coliva, afirmar, como hace Bar-On, que los avowals verdaderos y falsos comparten el ser actos de "comunicar mediante el habla la propia mente (*speaking one's mind*) cuya caracterización como actos razonables puede requerir mencionar el (quizás ausente) estado mental asignado" (Bar-On, 2994, p. 395), no aporta nada a la solución del problema por que no hay nada que garantice el avowal falso, es decir, no hay un estado mental que sea la causa *racional* de la auto-adscripción falsa. Asimismo, tampoco aporta nada decir que "incluso si tomamos un avowal como falso, y por lo tanto consideramos que el sujeto está equivocado, aún no podemos considerar que *ha cometido* un error, *epistémicamente* hablando" (Bar-On, 2004, p. 396), puesto que "si la causa racional es idéntica a lo que lo hace verdadero [*truth- maker*] y el segundo está ausente, también lo está la primera" (Coliva, 2016, p. 158). Según Coliva, por tanto, los avowals falsos pero sinceros en el sentido de Bar-On, no pueden ser catalogados como avowals sino, a lo sumo, como un *intento* de expresar un avowal, ya que si, según Bar-On, el avowal incorpora, de alguna manera, el estado mental auto-adscrito y este no está presente en el sujeto, la auto-adscripción puede *parecer* un avowal pero no serlo y afirmar que el sujeto logra expresar un estado mental, pero no su estado mental, es, según Coliva, muy confuso: "Sin duda, no ha expresado el estado mental de otra persona. Además, es cierto que podemos entender el avowal fallido de dolor como un síntoma del miedo del sujeto al dentista" (Coliva, p. 159).

Como he señalado anteriormente, Bar-On sitúa la seguridad distintiva de nuestros avowals en nuestra capacidad expresiva junto con nuestra capacidad de hacerlo de manera lingüística. Sin embargo, si bien nuestra capacidad expresiva utiliza vehículos expresivos que no son aprendidos (es decir, que son innatos), no es el caso de nuestra capacidad lingüística, la cual utiliza vehículos adquiridos (o aprendidos). Por tanto, ¿cómo desarrollamos esa capacidad lingüística que nos permite reemplazar nuestras expresiones naturales por expresiones con contenido semántico? Para explicar cómo es posible que lleguemos a desarrollar la capacidad para expresar nuestros estados mentales de manera lingüística de la misma forma que los expresamos de manera natural, Bar-On apela a la *transparencia* de los estados mentales. De acuerdo con Bar-On "en la medida que se puede decir que todos los avowals expresan —y, por tanto, muestran— los estados adscritos (en virtud de los vehículos expresivos auto-adscriptivos que usan), se puede decir que los avowals gozan de una cierta transparencia" (2015, p. 142). Bar-On (2004, 2015) distingue entre la transparencia-al-

mundo (*transparency-to-the-world*) y la transparencia-al-estado-mental-del-sujeto (*transparency-to-the-subject's-mental-state*[40]). La transparencia-al-mundo es, para Bar-On, la transparencia en el sentido de Evans que vimos en el apartado dedicado a Moran. Según este tipo de transparencia para responder a una pregunta como "¿Crees que va a llover?", miramos hacia fuera, a los mismos hechos reales en el mundo a los que miraríamos si estuviéramos respondiendo la pregunta "¿Va a llover?". Sin embargo, según Bar-On, un avowal como, por ejemplo, "Estoy pensando en mi amigo Alex", no va precedido de una consideración directa del mundo. Aunque el estado de primer orden esté dirigido al mundo, parece extraño pensar que mi auto-adscripción procede de mi consideración directa de mi amigo Alex. Por el contrario, parece más plausible pensar que procede de mi consideración de lo que creo que es Alex, de la idea que tengo sobre él, no del propio Alex en sí mismo. Para Bar-On, la transparencia-al-mundo puede verse como una forma de dar voz a *ciertos* estados mentales, como las creencias de primer orden. Pero expresar un avowal significa estar en situación de dar una voz directa a *todos* nuestros estados mentales actuales. Por tanto, "en cierto sentido, la transparencia en el mundo de las auto-adscripciones de creencias deviene un caso o síntoma especial de un fenómeno más amplio: la transparencia-al-estado-mental-del-sujeto de los avowals entendidos como actos expresivos" (2015, p. 140).

Para explicar la transparencia-al-estado-mental-del-sujeto Bar-On (2004; Bar-On & Long, 2001) describe, en primer lugar, las expresiones naturales en oposición a los actos reflejos y las acciones básicas no intencionadas (tales como una mueca de dolor o sentarse cuando se está cansado), que solo proporcionan una base para inferir el estado mental del sujeto. Las expresiones naturales no son simplemente indicadores de un estado mental, sino que exhiben la naturaleza misma del estado mental expresado. Son algo que hacemos intencionadamente manifestando la cualidad, el grado y el objeto del estado mental expresado de una manera directa y no reflexiva. Una vez aclarada la diferencia entre las expresiones naturales y los actos reflejos, Bar-On mantiene la continuidad entre las primeras y los avowals. Al igual que las expresiones naturales, los avowals no son la culminación de actos reflexivos del sujeto, sino "expresiones

[40] Para referirse a este tipo de transparencia, Bar-On utiliza también los términos "transparencia-a-la-condición-del-sujeto" (*transparency-to-the-subject's-condition*), "transparencia-a-la-condición-mental-del-sujeto" (*transparency-to-the-subject's-mental-condition*) y "transparencia-al-estado-mental-del-sujeto" (*transparency-to-the-subject's-state-of-mind*) (v. Bar-On, 2004, pp. 264–284; Bar-On & Long, 2001). El hecho de que Bar-On use el término "condición" en vez de "estado mental" es debido a que, según ella, el concepto de "estado mental […] es el concepto de la condición de un sujeto, expresable de varias maneras características, más que el concepto de un estado dentro del sujeto que tiene tales y cuales (causas y) efectos típicos" (2004, p. 424).

sinceras, espontáneamente voluntarias e irreflexivas (sonoras o silenciosas) como "Tengo un terrible dolor de cabeza" o "Estoy muy asustado" [...] similares a soltar un grito o aferrarse con miedo" (Bar-On & Long, 2001, p. 326). En la medida en que los avowals son continuos con las expresiones naturales, permiten, asimismo, que otros vean a través de ellos. De esta manera, los avowals, al igual que las expresiones naturales, sirven para *mostrar* el estado mental del sujeto *permitiendo a otros "verlo"* de manera directa, transparente, en lugar de simplemente permitirles hacer inferencias al respecto (Bar-On 2004, pp. 278 y ss.).

Esta ultima idea está directamente relacionada con la explicación de la comunicación expresiva defendida en el primer bloque. En el apartado 3.1, dedicado a la imitación neonatal, he sostenido que los bebés captan el estado mental expresado por la madre de manera directa, lo cual, entre otras cosas, explica su capacidad para las interacciones intersubjetivas sin necesidad de realizar inferencias. He rechazado, por tanto, la afirmación que sostienen los que he denominado "enfoques individualistas" según los cuales los bebés solo son capaces de percibir características físicas a partir de las cuales realizar inferencias a un supuesto objeto invisible (el estado mental) dentro de su madre. Asumir esta idea de los enfoques individualistas es asumir un dualismo mente-cuerpo que es necesario salvar mediante inferencias y que, como he sostenido, resulta innecesario si adoptamos un enfoque de segunda persona acerca del desarrollo ontogenético. Bar-On sostiene una idea similar al enfoque de segunda persona respecto a los estados mentales. Según Bar-On, lo que entendemos como expresado "son *condiciones en las que el sujeto se encuentra, no estados que están en los sujetos* [...] el concepto común de estado mental es, entonces, el concepto de la condición de un sujeto, expresable de varias maneras características, en vez del concepto de un estado dentro del sujeto que tiene tales y cuales (causas y) efectos típicos" (2004, p. 424). Asimismo, he defendido que el bebé capta tanto el estado mental de la madre como el intento comunicativo de esta, es decir, a la madre comunicándole dicho estado mental. Bar-On sostiene, de nuevo, una idea similar pero con una diferencia fundamental como veremos a continuación.

Bar-On utiliza un ejemplo del desarrollo lingüístico de los niños para explicar su idea. Bar-On (2004, pp. 248 y ss.; 2015, p. 140) nos pide que imaginemos a una niña pequeña que intenta con ahínco alcanzar su osito de peluche. Ante esto, la madre no tiene necesidad ninguna de hacer una inferencia para saber en qué tipo de estado mental se encuentra la niña. La madre "ve" de manera directa, *transparente*, el deseo de la niña

de alcanzar el juguete. Asimismo, la madre no considera que el comportamiento de la niña sea el efecto posterior de haber juzgado que quiere el juguete. Su expresión proviene directamente de su estado mental, de su deseo en este caso. Tanto para la madre como para la niña no existe una distancia epistémica entre ella (la niña) y su estado mental. La niña simplemente quiere el juguete y expresa su deseo de alcanzarlo, lo que permite a la madre conocer directamente el estado mental expresado, es decir, de manera *transparente*. En esta situación, la madre, normalmente, se dirigirá a la niña diciéndole "Es tu osito", "¿Quieres tu osito?", transfiriendo a la niña, de esta manera, un nuevo vehículo expresivo para articular aspectos del estado que ha expresado a través del comportamiento no verbal. Posteriormente, llegará el momento en el que la niña podrá "dar voz" a su deseo exclamando "¡Osito!" acompañándolo con un gesto (con un proto-imperativo), y más tarde, quizás "¡Quiero osito!" con un tono de voz ansioso, y finalmente simplemente expresando "¡Quiero mi osito!". Según Bar-On, el intento de la niña de alcanzar el osito y sus expresiones verbales pueden considerarse el mismo tipo de acto (en el sentido de que son comportamientos espontáneos, no reflexivos o no "estudiados") que, sin embargo, emplean diferentes vehículos de expresión (natural y lingüístico respectivamente).

Según Bar-On, por tanto, durante nuestro desarrollo ontogenético (lingüístico) adquirimos nuevos vehículos expresivos que reemplazan a las expresiones naturales. Sin embargo, este proceso no se limita a la infancia. Nuestra cultura, la exposición al lenguaje, las reglas sociales y las experiencias nos proporcionan nuevos medios para expresar nuestros estados mentales (los tacos, palabrotas, son un ejemplo, pero no el único). El uso de un vehículo lingüístico no implica que estemos haciendo una descripción, sigue siendo una expresión. Recordando las palabras de Wittgenstein: "¿Entonces estás diciendo que la palabra 'dolor' realmente significa llorar? —Al contrario: la expresión verbal del dolor reemplaza al llanto y no lo describe" (1953, §244). Cuando alguien se auto-adscribe un estado mental mediante un avowal, consideramos que el avowal nos muestra directamente el estado mental expresado. Todo lo que se requiere es la asimilación lingüística, que todos los adultos tenemos. De la misma forma que no necesitamos inferir el contenido de una expresión natural no necesitamos hacer inferencias para saber qué estado mental está expresando el sujeto mediante un vehículo lingüístico. También conocemos el estado mental expresado de una manera directa, no inferencial y no reflexiva. Según Bar-On, la carga de mostrar y

reconocer inmediatamente los estados expresados y sus contenidos intencionales está respaldada por las características *semánticas* del vehículo expresivo auto-adscriptivo:

> [A]vowals wear the conditions they are supposed to express on their linguistic sleeve, as it were. An avowal such as "I wish we'd get some rain today" explicitly names a kind of condition (a hope) and articulates its content (that it rain today), as well as ascribing it to a certain individual; it reveals the kind of state the avower expresses (as well as its intentional content, when it has one) through what the sentence expresses in the semantic sense (2015, p. 142).

De esta manera, según Bar-On, el hecho de que no cuestionemos los avowals, el hecho de que los consideremos verdaderos por defecto, su seguridad distintiva, se debe a que consideramos que expresan genuinamente el estado mental de manera transparente, tal y como lo hacen las expresiones naturales, ya que, como hemos visto, nada en la transición a la expresión verbalmente articulada modifica dicho estatus. La diferencia con el enfoque de segunda persona es que, a la hora de explicar la transparencia de las expresiones naturales, Bar-On omite, o no percibe, la condición comunicativa[41] del estado mental. De ahí que sostenga que la madre ve a la niña simplemente estando en dicho estado mental y no comunicándolo, como sostiene el enfoque de segunda persona que defiendo en este trabajo. Sin embargo, aun cuando Bar-On no lo perciba, esta idea está implícita en su explicación de su noción de la transparencia-al-estado-mental-del-sujeto: dado que la madre ve de manera transparente el deseo de la niña por el osito al ser expresado por esta, ¿qué otra función puede tener la expresión que no sea la comunicativa? En otras palabras, si la madre es capaz de ver a la niña estando en el estado mental sin necesidad de realizar inferencia alguna es porque la niña está, de alguna forma, comunicando —de manera directa— el estado mental en que se encuentra, aun cuando no lo haga con la intención de comunicárselo a la madre. En el caso de que la niña hubiera establecido primero un triángulo de atención conjunta con la madre, no solo estaría comunicando el estado mental en que se encuentra sino que estaría comunicándoselo a la madre. Podemos afirmar, por tanto, que la transparencia-al-estado-mental-del-sujeto subraya o suscribe la captación directa, es decir, no inferencial, de la condición comunicativa los estados mentales.

[41] En el primer bloque elegí usar el término "condición" en lugar de "aspecto" para evitar el sentido que este último término tiene en Wittgenstein (1953, parte II, sección xi, pág. 193), ya que, según él, "notar un aspecto" está relacionado con ver algo a través de ver otra cosa (ver nota 5). Entiéndase en este contexto, por tanto, "condición" en el sentido común de "aspecto", referido a la forma en la que algo se aparece o se presenta directamente, no como algo a través de lo cual vemos o entendemos otra cosa.

De otra parte y en relación al mismo tema, la noción de avowal como expresión lingüística de una expresión primitiva, natural, y el modo en el que aprendemos a reemplazar un tipo de expresión por otra están también íntimamente relacionados con el enfoque de segunda persona. Como señalé en el apartado 5 del primer bloque, dedicado a la autoridad de segunda persona y la regulación normativa, mediante los triángulos de atención conjunta el niño comienza a normativizar su conducta en base a las expresiones de su madre frente a situaciones que desconoce o para las cuales no tiene una respuesta definida. La madre, he sostenido, se torna en una suerte de autoridad para el bebé que he denominado 'autoridad de segunda persona'. Esta descripción encaja perfectamente con la explicación neo-expresivista (que, recordemos, proviene de las ideas de Wittgenstein sobre la expresión) acerca de la manera en la que aprendemos a sustituir nuestras expresiones naturales por expresiones lingüísticas: dada la autoridad de segunda persona de la madre, la niña incorpora las expresiones lingüísticas de la madre del mismo modo que incorpora sus respuestas expresivas para normativizar su conducta. La madre es una autoridad para la niña tanto en un caso como en el otro. De esta manera, la niña aprende a sustituir las expresiones naturales que capta de manera innata por las lingüísticas provenientes de su madre, incorporándolas, con el tiempo y reiteradas interacciones, a su repertorio expresivo y reconociéndolas, asimismo, en el de la madre. Sin embargo, es preciso aclarar que aprender a sustituir una expresión natural por otra lingüística no quiere decir que con la segunda eliminemos a la primera. Aunque es cierto que, en ocasiones, la expresión lingüística reemplaza a la natural, en la mayoría de ocasiones en las que la auto-adscripción sustituye a la expresión natural, expresamos, asimismo, el estado mental de manera natural. Por tanto, parece más adecuado decir que aprendemos a expresar nuestras emociones y sensaciones de otra manera, en este caso lingüística, que puede tanto sustituir como acompañar a la expresión natural.[42]

Asimismo, es necesario resaltar que Bar-On obvia o no da importancia al hecho de que el proceso de aprendizaje de las auto-adscripciones de estados mentales no es unilateral, no está separado del aprendizaje de las adscripciones de estados mentales a los demás. En otras palabras, la adquisición de un nuevo vehículo expresivo (lingüístico) no puede ser entendida como algo que la niña adquiere solo para expresar sus estados mentales de manera lingüística, sino *también* para reconocerlos en el

[42] Lo mismo ocurre en la edad adulta pero en sentido contrario. Las personas adultas aprendemos e incorporamos expresiones faciales o corporales que no son innatas sino producto de la cultura, son expresiones que hemos aprendido e incorporado para expresar determinados estados mentales, como, por ejemplo, la duda, o para sustituir expresiones lingüísticas, como, por ejemplo, determinados insultos, gestos de aprobación y desaprobación, etc.

repertorio expresivo de la madre. Esta ausencia de análisis muestra que el neo-expresivismo de Bar-On parte, asimismo, de presupuestos individualistas. Como señalé en el apartado 4, dedicado a los agencialistas, centrarse en el individuo conlleva una serie de problemas. En el caso de Bar-On le hace obviar aspectos fundamentales del proceso de aprendizaje de las auto-atribuciones de estados mentales. Aun cuando Bar-On reconoce que su explicación de la transparencia conduce a un replanteamiento de la asimetría entre primera y tercera persona respecto a las posturas tradicionales (Bar-On, 2004, p. 278), no desarrolla esta idea. Su acertada explicación del inicio del proceso de aprendizaje de las auto-atribuciones de estados mentales se limita al individuo porque su objetivo es explicar lo distintivo de los avowals *frente* a las atribuciones de estados mentales a los demás, es decir, explicar el privilegio de primera persona de las auto-atribuciones de estados mentales frente a las atribuciones de tercera persona. Bar-On obvia, por tanto, que nuestra capacidad para auto-atribuirnos estados mentales de manera lingüística es inseparable de la capacidad para atribuirlos a los demás. No solo su aprendizaje es simultáneo sino que, como he sostenido en el primer bloque, ambas capacidades (la de expresar nuestro estado mental de manera natural y la de captar los estados mentales de nuestro interlocutor) forman parte de un mismo fenómeno: nuestra capacidad para la comunicación expresiva (volveré a este argumento en el siguiente apartado dedicado a Finkelstein).

Al principio de este apartado señalé que Bar-On divide la cuestión acerca del autoconocimiento en dos preguntas. Hasta este momento hemos visto cómo responde Bar-On a la pregunta (1) acerca de la seguridad distintiva de los avowals. Pasemos ahora a su respuesta a la pregunta (2) ¿Qué hace que los avowals sean instancias de un tipo de conocimiento privilegiado?

Según Bar-On hay que rechazar la idea de que estar en un estado mental implica *ipso facto* tener conocimiento del mismo. Sin embargo, esto no quiere decir que el sujeto necesite una justificación (inferencial o de otro tipo) o razones positivas para verse a sí mismo estando en el estado que se auto-atribuye. En coherencia con su explicación de la seguridad distintiva de los avowals, Bar-On rechaza adoptar una explicación epistémica del autoconocimiento en la que el sujeto tenga que llevar a cabo un esfuerzo epistémico recognoscitivo, ya que, según ella, el autoconocimiento no tiene que ser considerado un logro epistémico (o cognitivo) en ese sentido. De ahí que rechace dar una explicación deflacionista —aunque, según ella, su neo-expresivismo puede endorsarla (cf. 2004, p. 345)— puesto que presupone un método epistémico de

reconocimiento de nuestros estados mentales presentes. Según Bar-On, los avowals carecen de fundamento (*groundless*), no precisan de justificación epistémica puesto que son inmunes a dos tipos de error, a saber, al error por identificación incorrecta y al error por adscripción incorrecta. Siguiendo los trabajos de Evans (1982) y Shoemaker (1968)[43], Bar-On propone que los avowals disfrutan, además de la inmunidad al error por identificación incorrecta (*immunity to error through misidentification*), de la inmunidad al error por adscripción incorrecta (*immunity to error through misascription*) (Bar-On, 2004, Cap. 3 y 6; 2009, 64 y ss.; 2011, pp 191 y ss.; 2015, p. 143). De manera resumida, la idea es que mi avowal es inmune al error por *identificación* incorrecta porque no tengo una razón independiente para pensar que soy yo quien se encuentra en el estado relevante. Este tipo de inmunidad refleja la ausencia de identificación por reconocimiento. Cuando digo o pienso: "tengo un dolor de muelas", la razón por la que no puedo equivocarme al identificarme es que mi adscripción no depende de un reconocimiento por separado de que alguien tiene un dolor de muelas junto con un juicio de que ese alguien soy yo, basado en las características que observo. Sin embargo, en el caso en el que, por ejemplo, digo cuánto dinero tengo en mi cuenta bancaria consultando la pantalla del cajero automático, la información que recibo de que alguien tiene, digamos, 500€ en su cuenta, junto con que tome esa información como siendo sobre mi cuenta, me da motivos para pensar que soy yo quien tiene los 500€ (Bar-On, 2004, p. 193). Asimismo, mi avowal es inmune al error por *adscripción* errónea porque no tengo razones para afirmar que *alguien* se encuentra en determinado estado mental *M* independientemente de las que tengo para afirmar o juzgar que *soy yo mismo* quien me encuentro en el estado mental *M* (2015, p. 143), es decir, al igual que no tengo razones para *reconocerme* como expresando un estado mental, tampoco las tengo para *reconocer* el tipo de estado mental expresado, lo conozco de manera directa, no inferencial, no recognoscitiva. De esta manera, ni puedo equivocarme respecto a quién está expresando el estado mental cuando soy yo quien lo expresa ni respecto a qué tipo de estado mental estoy expresando.

[43] Evans y Shoemaker toman como punto de partida la diferencia entre los usos del pronombre "yo" que Wittgenstein lleva a cabo en *The Blue and Brown Books* (1958, pp. 66 y ss.). Wittgenstein distingue entre el uso del pronombre de primera persona como sujeto y su uso como objeto. Lo que distingue al primero de estos usos, el subjetivo, es lo que Shoemaker ha bautizado como "inmunidad al error por identificación incorrecta" (*immunity to error through misidentification*) (1968, pp. 556–7), es decir, la imposibilidad de equivocarse con respecto a aquello a lo que nos referimos mediante ese pronombre. El uso objetivo, en cambio, opera mediante criterios de identidad personal y en su caso podemos equivocarnos al identificarnos a nosotros mismos entre otras personas. Para una discusión sobre el fenómeno de la inmunidad al error por identificación incorrecta véase Wittgenstein (1958, pp. 66–67), Evans (1982, Cap. 7, Sec. 2), Shoemaker (1968), y Wright (1998, p. 18–20).

El hecho de que los avowals sean inmunes a estos dos tipos de error implica, por tanto, la ausencia de necesidad de un medio para reconocerme a mí mismo como el que realiza la atribución o para reconocer qué tipo de estado es el (auto-)atribuido. Sin embargo, según Bar-On esto no es lo que explica el conocimiento privilegiado de nuestros estados mentales. Bar-On apela a la explicación de Evans sobre el tipo de autoconocimiento corporal que conlleva la inmunidad al error por identificación incorrecta para hacer un paralelismo con el de la inmunidad al error por adscripción incorrecta. Según Evans (1982) adquirimos conocimiento acerca de nuestro cuerpo, en cuanto algo físico y espacial, gracias a "la capacidad general para percibir nuestros propios cuerpos [compuesta por] nuestro sentido propioceptivo, nuestro sentido del equilibrio, del calor y frío, y la presión" (p. 220) y a la capacidad para determinar "nuestra posición, orientación y relación con otros objetos en el mundo sobre la base de nuestras percepciones del mundo" (p. 222). Bar-On extiende esta idea a las capacidades expresiva y lingüística descritas anteriormente. Según Bar-On, al igual que el ejercicio de la capacidad para obtener información sobre algunos de nuestros estados físicos o espaciales permite que los juicios pertinentes representen cierto tipo de autoconocimiento corporal y espacial (a pesar de que no implican auto-reconocimiento y, por lo tanto, son inmunes al error a través de una identificación incorrecta) es natural sostener "que es *el ejercicio de nuestra capacidad de dar voz a nuestros estados mentales presentes* —de comunicar mediante el habla nuestras mentes [to speak our minds]— la que permite que los avowals representen autoconocimiento mental" (2015, pp. 144).

En este apartado hemos visto como el neo-expresivismo de Bar-On queda, según ella, legitimado como una explicación del *autoconocimiento privilegiado*, ya que se trata de una clase de *conocimiento* que *sólo el sujeto* puede tener. Según Bar-On, tengo *conocimiento* de mis estados mentales gracias a la capacidad de expresarlos de manera lingüística, ya que no puedo equivocarme respecto a quién está en el estado mental (pues mi avowal es inmune al error por identificación incorrecta) ni respecto a qué tipo de estado mental me encuentro, ni cual es su contenido (ya que mi avowal es inmune al error por adscripción incorrecta) y es *privilegiado* porque, "como sujeto que está en un estado mental y que está dotado de las capacidades lingüísticas y conceptuales necesarias, *solo yo* estoy situado en una posición única para dar expresiones auto-adscriptiva de mis estados mentales presentes" (2004, p. 428).

He afirmado que la explicación de los avowals falsos de Bar-On resulta problemática. Como sugiere Coliva, si, como afirma Bar-On, el avowal incorpora, de alguna manera, el estado mental auto-adscrito y este no está presente en el sujeto, la auto-adscripción no puede ser un avowal, aunque pueda parecerlo. Asimismo, he sostenido que la transparencia-al-estado-mental-del-sujeto está directamente relacionada con la explicación de la comunicación expresiva defendida en el primer bloque dedicado al desarrollo ontogenético. Al igual que para el enfoque de segunda persona, para Bar-On las expresiones naturales son transparentes, es decir, expresan el estado mental de manera directa, no inferencial. Esto es así porque los estados mentales son condiciones en las que el sujeto se encuentra, expresables de varias maneras características, no estados que están dentro los sujetos y que inferimos mediante la expresión o "vemos" *a través* de ella. De esta manera, la transparencia-al-estado-mental-del-sujeto de Bar-On señala o subraya la captación directa, no inferencial, de la condición comunicativa de los estados mentales. La diferencia entre la explicación de Bar-On y la del enfoque de segunda persona radica en que Bar-On obvia la condición comunicativa de los estados mentales, aun cuando esta está implícita en su explicación ya que, como he argumentado, si la madre es capaz de ver a la niña estando en el estado mental sin necesidad de realizar inferencia alguna es porque la niña está, de alguna forma, comunicando —de manera directa— el estado mental en que se encuentra.

He señalado que Bar-On sostiene una continuidad entre las expresiones naturales y los avowals. Según Bar-On, los avowals expresan genuinamente el estado mental de manera transparente, es decir, directa, no inferencial, tal y como hacen las expresiones naturales. Siguiendo a Wittgenstein, Bar-On sostiene —acertadamente y en relación con la autoridad de segunda persona defendida en el primer bloque — que la niña aprende a sustituir (o acompañar) las expresiones naturales que capta de manera innata por expresiones lingüísticas provenientes de su madre, incorporándolas, con el tiempo y reiteradas interacciones, a su repertorio expresivo y reconociéndolas, asimismo, en el de la madre. Este último elemento, ausente en la explicación de Bar-On que, como he señalado, refleja el individualismo de su enfoque, resulta fundamental para la comprensión del proceso de aprendizaje de las (auto-)adscripciones de estados mentales y muestra que nuestra capacidad para expresar nuestros estados mentales no está separada de la capacidad para captar los estados mentales de nuestro interlocutor, ambas forman parte del mismo fenómeno: nuestra capacidad para la comunicación expresiva.

En el siguiente apartado veremos el enfoque neo-expresivista de Finkelstein. Al igual que Bar-On, Finkelstein sostiene que la cuestión central del autoconocimiento es explicar la autoridad de primera persona. Sin embargo, su enfoque, aun siendo similar en algunos aspectos, dista notablemente en otros del de Bar-On.

5.2 David Finkelstein

Finkelstein desarrolla su enfoque neo-expresivista en su libro *Expression and the Inner* (2003). En él realiza una crítica tanto del introspeccionismo, al que denomina "detectivismo" (tomando el término que Wright utiliza para describir la lectura de izquierda a derecha del bicondicional) como del constitutivismo y de las posturas intermedias, entre las que sitúa a McDowell (1994, 1996, 1998). Para Finkelstein, ninguno de estos enfoques da cuenta de manera fehaciente de la característica distintiva de las autoadscripciones de estados mentales: la autoridad de primera persona.

Según Finkelstein la autoridad de primera persona ha de caracterizarse "en términos de dos hechos: (1) Si deseas conocer mi condición psicológica, normalmente soy la mejor persona a quién preguntar y (2) no hay necesidad de que considere la evidencia de mi comportamiento para poder decir lo que estoy pensando o sintiendo" (2003, p. 124). La cuestión, por tanto, es explicar por qué uno es, normalmente, la mejor persona a la que preguntar acerca de sus propios pensamientos, emociones, sensaciones, etc., y por qué no es necesaria evidencia para ello. Para responder a estas preguntas Finkelstein apela a la similitud entre los avowals y las expresiones naturales, tales como expresiones faciales de alegría, dolor, tristeza, etc. Según Finkelstein, normalmente soy la mejor persona a quién preguntar (sobre en qué estado mental estoy) por, exactamente, la misma razón por la que mi cara es el mejor lugar al que puedes mirar (2003, p. 101). Su explicación está basada en las ideas de Wittgenstein acerca del seguimiento de reglas y de la expresión de estados mentales. En una línea muy similar a la de McDowell[44],

[44] En su discusión con las interpretaciones de Wright y de Kripke sobre el seguimiento de reglas, McDowell (1984, 1992) señala que la preocupación de Wittgenstein acabar con el "insidioso supuesto" de que una orden o la expresión de una regla requieren de una interpretación para seguir el curso de acción correcto (1984, p. 340). McDowell sugiere que la idea de que una regla siempre requiere una interpretación proviene de una concepción de las reglas que conlleva la existencia de una "brecha entre la instrucción y el uso competente" (1984, p. 333). Según McDowell, podemos evitar la idea de la existencia de esta brecha si dejamos de concebir las reglas como "elementos que en sí mismos son normativamente inertes", es decir, como meros sonidos o inscripciones carentes de significado por sí mismos y que, por tanto, no pueden contribuir a determinar el curso de acción correcto si no es mediante una interpretación (1992, p. 42).

Finkelstein señala que Wright malinterpreta el sentido en el que Wittgenstein desarrolla su propuesta acerca del seguimiento de reglas. Wright sostiene que el abismo entre una regla y su significado puede salvarse mediante la estipulación de este último por parte del agente (mediante "sus mejores juicios"). Finkelstein rechaza este argumento pues, como señalé anteriormente, o bien cae en el regreso al infinito o bien desemboca en un tipo platonismo. Según Finkelstein, el error de Wright es pensar que Wittgenstein está afirmando la existencia de este abismo y que, por tanto, hay que buscar una forma de salvarlo. Sin embargo, sostiene Finkelstein, lo que Wittgenstein nos está mostrando es que no existe tal abismo puesto que los significados de las palabras y los signos "viven en el uso", el abismo existe solo en la medida en que supongamos que los signos, las palabras, están separados de la actividad humana:

> At PSI432, Wittgenstein writes:
>
> > Every sign by itself seems dead. What gives it life? –In use it is alive.
>
> Wittgenstein does not agree that signs are dead until we interpret them or stipulate what they mean. A sign only seems dead if we consider it by itself—i.e., apart from the use that we make of it. In its use, a sign lives (2003, p. 81).

Según Finkelstein, de acuerdo con Wittgenstein, al igual que podemos ver el significado en las palabras de una página, podemos, asimismo, ver el estado mental en la expresión del mismo (2010, pp.188–189). Lo que Wittgenstein nos muestra es, según Finkelstein, que las expresiones solo parecen desprovistas de contenido psicológico si las vemos aisladas de la "trama de la vida". Los estados mentales no son algo oculto que debamos inferir de determinados movimientos físicos o contorsiones faciales, sino que son algo que vemos directamente: "Vemos la emoción — ¿En oposición a qué? — No vemos las contorsiones del rostro y hacemos inferencias de ellas (como el doctor que da un diagnóstico) a la alegría, aflicción o aburrimiento" (Wittgenstein, 1981, §225)[45]. Así pues, no existe un abismo que haya que salvar entre el estado mental y la expresión, "una expresión facial puede hacer el estado mental de alguien inmediatamente manifiesto a otros" (Finkelstein 2010, p. 188) .

[45] Esta es la misma idea que la de *la transparencia-a-la-condición-del-sujeto* que defiende Bar-On. Sin embargo, Finkelstein no usa el término "transparencia" puesto que él no diferencia entre este tipo de transparencia y la transparencia en el sentido de Evans o Moran (la *transparencia-hacia-el-mundo* de Bar-On). Según Finkelstein, hemos de rechazar la Condición de la Transparencia defendida por Moran puesto que no se cumple en un amplio rango de casos en los que hablamos con autoridad de primera persona acerca de nuestros estados mentales (como, por ejemplo, miedos irracionales o actitudes que no son ni dictadas por la razón ni irracionales) (v. Finkelstein, 2003, PostScript). Su crítica es, en este sentido, similar a la de Cassam señalada en el apartado anterior dedicado al agencialismo .

Como afirmé en el apartado anterior, esta idea es similar a lo sostenido en el primer bloque y a lo defendido por Bar-On respecto a los estados mentales[46]. Asumir el abismo es asumir un dualismo mente-cuerpo que es necesario salvar mediante inferencias y que, aunque es un presupuesto implícito en los enfoques individualistas, resulta innecesario si adoptamos un enfoque de segunda persona acerca del desarrollo ontogenético y entendemos las expresiones como la condición comunicativa de los estados mentales. Sin embargo, al igual que Bar-On, Finkelstein obvia esta última afirmación del enfoque de segunda persona pues entiende la expresión como "mostrando", "poniendo de manifiesto" el estado mental, no *comunicándolo*. Su enfoque adolece, por tanto, del mismo presupuesto individualista que el de Bar-On puesto que su objetivo es explicar lo distintivo de los avowals *frente* a las atribuciones de estados mentales a los demás, es decir, explicar la autoridad de primera persona en base a la asimetría entre primera y tercera persona obviando, por tanto, las relaciones interpersonales. Sin embargo, Finkelstein da una explicación de la autoridad y del conocimiento de los estados mentales que, como veremos más adelante, dista bastante de la explicación de Bar-On.

Como he señalado al principio del apartado, para Finkelstein, al igual que las expresiones faciales, nuestras auto-adscripciones normalmente expresan, y con ello muestran, aquello que adscriben. Según Finkelstein, "[p]restarle atención a mis auto-adscripciones y a mis expresiones faciales te sitúa en posición de percibir mi condición psicológica" (2003, p. 101) puesto que los avowals, al igual que las expresiones naturales, expresan y ponen de manifiesto el estado mental auto-adscrito. Sin embargo, sostiene que, al contrario que las expresiones naturales, los avowals, además de expresiones, son *aserciones con valor de verdad*. El hecho de que sean expresiones no impide que tengan valor de verdad puesto que también son aserciones y, como tales, en nuestro "juego del lenguaje" les aplicamos el estatus de evaluables veritativamente. Finkelstein se basa de nuevo en Wittgenstein quien, según él, no sostiene un expresivismo "simple" como afirma Bar-On, sino que en sus escritos podemos ver cómo apoya la idea de la dimensión asertiva de los avowals y, por tanto, aunque no directamente, su valor de verdad. Siguiendo a Jacobsen (1996), Finkelstein afirma que

[46] Sin embargo, al contrario que Bar-On, Finkelstein (2003. 2012) rechaza usar el término "transparencia" en este contexto pues no la entiende en el sentido de la transparencia-al-estado-mental-del-sujeto de Bar-On sino como la transparencia-hacia-el-mundo, en el sentido de Evans o Moran, la cual rechaza pues, como señalé anteriormente, según Finkelstein "hay una amplia gama de casos en los que una persona habla con autoridad de primera persona acerca de su estado mental que no pueda expresarlos (*avow*) en el sentido de Moran" (2012, p. 105).

del mismo modo que "aplicamos el cálculo de las funciones de verdad" a una proposición dada, p, lo aplicamos a las auto-atribuciones de estados mentales[47], por lo que no hay lugar para negar que las auto-atribuciones tengan condiciones de verdad (Finkelstein, 2010, p. 194). Finkelstein aclara que aun cuando Wittgenstein no afirma que los avowals tengan condiciones de verdad explícitamente, esto no supone que esté afirmando lo contrario: "una cosa es que un filósofo no aborde explícitamente una cuestión y otra muy distinta es entenderle como si hubiera dado una respuesta negativa a la misma" (2010, pp. 193).

De esta manera, para Finkelstein tenemos autoridad de primera persona respecto nuestros propios estados mentales puesto que podemos expresarlos mediante auto-adscripciones veritativo evaluables. Pero, ¿es, únicamente, el hecho de que sean expresiones y aserciones veritativo evaluables lo que las convierte en autoritativas? Finkelstein señala dos características más a tener en cuenta sobre los avowals para poder decir de ellos que son autoritativos. En primer lugar, sostiene que los avowals no son *meramente* aserciones sino que *contextualizan* la expresión del estado mental. Finkelstein compara esta afirmación con el principio del contexto de Frege (1953), el cual sostiene, de manera resumida, que la oración es la unidad primaria de inteligibilidad, es decir, que entender una oración no requiere captar los significados de partes independientes inteligibles de la oración puesto que una palabra solo tiene el significado que tiene en el contexto de una oración. Según Finkelstein, Wittgenstein extiende la idea de Frege al "juego del lenguaje" de manera que "donde Frege habla de oraciones, él habla de juegos del lenguaje" (2003, p. 107), el significado de nuestros avowals depende del juego del lenguaje en el que están insertos, de cómo están situados en nuestras vidas. Siguiendo a Wittgenstein, Finkelstein sostiene que cuando nos auto-adscribimos, por ejemplo, una expectativa, la adscripción "está 'inserta en una situación' de una manera muy semejante a como el significado de una palabra está inserta en el contexto de una oración o un parágrafo" (2003, p. 113).

Para Finkelstein, por tanto, la autoridad de primera persona no sólo está relacionada con su doble estatus de expresiones y auto-adscripciones veritativo evaluables, sino también con el hecho de que nuestros avowals contextualizan aquello que adscriben. Sin embargo, como señalé más arriba, hay otra característica necesaria

[47] Finkelstein pone el ejemplo de alguien que sostiene como falsa una afirmación acerca de un estado mental: "En respuesta a alguien que acaba de afirmar que nadie en la sala experimenta ningún dolor, podría decir: 'Estoy en la sala y siento dolor, por lo que lo que acaba de decir es falso'" (Finkelstein, 2010, p. 194).

de los avowals para ser considerados autoritativos. Dado que la autoridad se debe, principalmente, al hecho de que nuestros avowals expresan, y por tanto muestran, el estado mental auto-adscrito, Finkelstein sostiene que dicho estado mental ha de ser *consciente* para ser autoritativo. Según Finkelstein, existe una estrecha conexión entre la autoridad y la conciencia: solo tenemos autoridad de primera persona sobre los estados mentales conscientes. Para explicar qué entiende por "consciente", Finkelstein critica lo que denomina "la explicación simple de la conciencia" (2003, p. 114), según la cual un estado mental es consciente si uno sabe que está en él y es inconsciente si uno no sabe que está en él. Para Finkelstein este tipo de explicación es insatisfactoria puesto que no logra capturar los casos en los que somos conscientes de un estado mental inconsciente, por ejemplo, los casos en los que nuestro terapeuta nos hace saber que tenemos un miedo inconsciente a ser engañados por nuestra pareja. En este caso podríamos decir que *sabemos* que tenemos un miedo *inconsciente* a ser engañados por nuestra pareja. Al respecto Finkelstein distingue entre dos sentidos de "consciente": (i) ser conscientes *de que* tenemos tal o cual miedo, creencia, etc., el cual puede ser inconsciente y (ii) creer o temer, etc. algo conscient*emente*, esto es, que esa creencia o ese miedo sean conscientes. Para Finkelstein, solo el segundo de los casos es el apropiado para definir un estado mental consciente. Ahora bien, si la cuestión de la conciencia no tiene que ver con lo que sabemos o no sabemos, es decir, no es una cuestión epistémica, ¿cuándo podemos afirmar que un estado mental es consciente, es decir, que conscientemente creemos, tememos, etc. que *p*? Según Finkelstein, la conciencia de un estado mental tiene que ver con nuestra capacidad para expresarlo mediante auto-adscripciones. Cuando, y solo cuando, somos capaces de expresar un estado mental *meramente* auto-adscribiéndonoslo dicho estado mental es consciente, es decir, si precisamos de algún tipo de evidencia para auto-adscribirnos un estado mental, dicho estado mental no es consciente. En otras palabras, tengo autoridad sobre, por ejemplo, mi creencia de que *p*, sí, y sólo sí, creo conscientemente que *p*, es decir, si soy capaz de expresarlo con *tan solo* auto-adscribírmelo[48].

[48] Es importante señalar que, para Finkelstein, tanto la conciencia como la autoridad de primera persona deben entenderse como una cuestión de grados. A veces, una persona se auto-adscribe un estado mental que no es ni del todo consciente ni del todo inconsciente. En tales casos, puede decirse que la auto-adscripción se apoya, pero no descansa directamente en consideraciones evidenciales. Nuestras afirmaciones sobre estados mentales inconscientes son solo tan autoritativas como la evidencia que tenemos para apoyarlas. No son, por tanto, *plenamente* autoritativas como lo son los conscientes. Para Finkelstein, solo podemos hablar con plena autoridad sobre nuestros estados mentales conscientes (v. 2003, §5.5). Para una explicación más extensa sobre la relación entre estados mentales conscientes e inconscientes véase Finkelstein (1999; 2003, §5).

Llegados a este punto, y tras esta descripción de la conciencia de nuestros estados mentales, podemos pasar a la explicación de Finkelstein sobre la ausencia de fundamento epistémico de los avowals y sobre el autoconocimiento.

Como acabamos de ver, para Finkelstein la conciencia de los estados mentales no es una cuestión epistémica, por lo que no tenemos necesidad de un método epistémico interno de detección de nuestros estados mentales para que estos sean conscientes. Según Finkelstein, al igual que la conciencia de los estados mentales no requiere de un método epistémico, el hecho de que podamos expresar nuestros estados mentales mediante auto-adscripciones no implica que necesitemos un fundamento epistémico. Su afirmación está basada en la adquisición de expresiones lingüísticas descrita por Wittgenstein que vimos en el apartado anterior. Según Finkelstein, aunque no podemos negar que cuando un niño aprende a sustituir expresiones naturales por expresiones lingüísticas muchas cosas cambian, esto no quiere decir que, por ello, el niño necesite un fundamento epistémico para expresar su dolor (2003, p. 112). Por tanto, según Finkelstein, si, al igual que las expresiones naturales, nuestros avowals no precisan de un fundamento epistémico, la justificación acerca de los mismos no tiene tampoco por qué ser una cuestión epistémica. Finkelstein usa, de nuevo, los escritos de Wittgenstein para desarrollar su argumento. Según Wittgenstein, "[u]sar una palabra sin justificación no quiere decir usarla injustificadamente" (1953, §289). Finkelstein pone un ejemplo de un caso hipotético en el que una persona que tiene un dolor agudo en la parte posterior de la rodilla y que, basándose en una información fiable, infiere que es más propenso a padecer una determinada enfermedad. Esta persona no hubiera llegado a esa conclusión de no haber tenido el dolor, sin embargo, según Finkelstein, aun cuando el dolor no esté epistémicamente justificado está persona sí lo está en creer que es más propenso a contraer dicha enfermedad. Según Finkelstein, el error es pensar que "siempre que alguien tiene justificación para creer en algo en virtud de una inferencia efectuada a partir de cierto número de premisas también debe tener justificación epistémica para creer (o aseverar) todas y cada una de las premisas" (2003, p. 149). Una persona que expresa su dolor mediante una auto-adscripción como "Me duele la rodilla" lo hace "sin justificación epistémica, pero no sin derecho a hacerlo" (2003, 149). No tener una justificación epistémica no implica expresar un avowal injustificadamente. En este sentido, Finkelstein sostiene que estamos "por así decirlo, autorizados [*entitled*]" (2003, p. 150) a auto-adscribirnos un estado mental—por los motivos que hemos ido viendo a lo largo de su explicación de la autoridad de primera persona.

Ahora bien, siendo esto así, ¿cómo se puede decir de alguien, aun cuando según Finkelstein esté autorizado, que *sabe* en que estado mental se encuentra si no está epistémicamente justificado? Para responder a esta pregunta Finkelstein desarrolla una explicación de la conexión entre el sentir y el saber basada en los escritos de Wittgenstein que comienza con el siguiente pasaje:

> In what sense are my sensations *private*? — Well, only I can know whether I am really in pain; another person can only surmise it. — In one way this is wrong, and in another nonsense. If we are using the word "to know" as it is normally used (and how else are we to use it?), then other people very often know when I am in pain. — Yes, but all the same not with the certainty with which I know myself! — It can't be said of me at all (except perhaps as a joke) that I *know* I am in pain. What is it supposed to mean— except perhaps that I *am* in pain? (Wittgenstein, 1963, §246).

Finkelstein sostiene que lo importante de este pasaje es que no es que uno no pueda usar adecuadamente las palabras "Sé que tengo dolor", sino que uno *podría* usarlas o bien en broma o bien para *decir lo mismo* que "*Siento* dolor", como sugiere la última frase del pasaje. La posibilidad de que alguien pueda usar una u otra formulación no implica que el auto-conocimiento de esa persona sea de la misma clase que el de alguien que sabe que tiene una avería en el coche. Finkelstein imagina a tres filósofos cada uno de los cuales da una respuesta diferente al caso descrito anteriormente sobre la persona que tiene un dolor en la rodilla y cree justificadamente que es más propenso a contraer cierta enfermedad. El primero sostiene que es correcto decir de esa persona que *siente* un dolor, pero que es antinatural y confuso, por no decir un sinsentido, que lo *sabe* o que *no lo sabe*. El segundo filósofo sostiene que esa persona está autorizada para llevar a cabo inferencias basadas en su auto-adscripción de dolor pero que, al no estar epistémicamente justificada, *no sabe* que siente dolor. Finalmente, el tercero sostiene que esa persona "*sí* sabe que siente dolor; *podemos* decir eso acerca de él. El error consiste en pensar que el conocimiento siempre requiere justificación epistémica. Él sabe que tiene dolor, pero saber eso no es como saber que uno tiene termitas en el sótano de su casa" (2003, p. 151-152). Según Finkelstein, ninguno de los tres estaría en desacuerdo acerca de nada que tenga relevancia filosófica, de ahí que su interés no sea el de dar una explicación sobre cómo conocemos nuestros estados mentales sino sobre por qué tenemos autoridad de primera persona.

Una de las principales críticas a Finkelstein está relacionada con esta afirmación acerca de la irrelevancia filosófica en lo referente a la epistemología del autoconocimiento. Según Vega (2011), la cuestión epistémica sí es relevante para el

auto-conocimiento, puesto que en la medida en que el neo-expresivismo se toma en serio la idea de que los avowals son también aserciones, "debería estar comprometido con una descripción epistémica de las auto-adscripciones en primera persona, incluso si esta descripción se aleja de concepciones epistemológicas tradicionales" (2011, p. 49). Para Vega, los tres filósofos del ejemplo anterior sí pueden estar en desacuerdo sobre algo de relevancia filosófica: "sobre si tiene sentido desarrollar una epistemología sustantiva del autoconocimiento y si nuestra epistemología debe unificarse bajo una sola noción de conocimiento" (2011, p. 47). El error de Finkelstein es creer que el único modelo epistémico válido para el autoconocimiento autoritativo es una forma de introspeccionismo (o detectivismo en términos de Finkelstein). En la misma línea, Stroud (2011) sostiene que, a su juicio, los dos primeros filósofos fallan en algo y que es el tercer filósofo quien está en lo correcto, por lo que su respuesta es la que hay que indagar y tratar de comprender. Según Stroud, en esta respuesta hay "involucrado un tipo de conocimiento de lo que pensamos y sentimos y de lo que hacemos que debemos tener si queremos tener uno de los conocimientos más mundanos y puramente proposicionales que muchos filósofos parecen pensar que es el único conocimiento que tenemos", siendo, por tanto, una cuestión de verdadera importancia filosófica (2011, p. 33).

Acerca de esta idea, conviene recordar lo sostenido en el primer bloque sobre la ausencia de la necesidad de elementos conceptuales para el conocimiento de determinados estados mentales. Como he defendido que ocurre en los intercambios intersubjetivos con su madre, el bebé tiene la capacidad de comunicarse de manera expresiva con ella, lo cual implica su habilidad para detectar e intercambiar información expresiva significativa. El bebé es capaz de captar ("saber") si el estado mental que ve en la madre es el mismo estado mental que siente porque, como he argumentado, ver y sentir un estado mental son dos formas distintas de captar ("conocer") una y la misma cosa, ambas provenientes de nuestra capacidad innata para la comunicación expresiva. Ahora bien, el tipo de "saber" o de "conocer" del bebé no requiere, como he sostenido, elementos conceptuales. Por tanto, el conocimiento que el bebé posee no es un "conocimiento epistémico" sino, más bien, un *conocimiento no conceptual,* puesto que para tener conocimiento epistémico sí son necesarios elementos conceptuales, es decir, para poder afirmar de alguien que conoce algo en sentido tradicional (epistémico), este ha de poseer los conceptos necesarios para ello. Tal vez este sea el sentido en el que Finkelstein interpreta a Wittgenstein a la hora de sostener que decir "Sé que me duele" y

"Siento dolor" son dos formas de decir lo mismo, ya que para sentir un dolor no son necesarios elementos conceptuales. Sin embargo, Finkelstein no tiene en cuenta que, una vez introducidos en el lenguaje, una vez que poseemos el concepto de "dolor" (así como otra multitud de conceptos necesarios para ello), la auto-adscripción "Me duele" sí contiene elementos conceptuales y, por tanto, puede ser considerada un conocimiento epistémico, aun cuando su obtención no requiera de una base epistémica.

Al respecto, Vega (2011) sostiene que el modo más plausible de entender por qué las auto-adscripciones pueden ser usadas asertivamente y constituir conocimiento, es porque la auto-adscripción como tal implica una creencia que se expresa en el acto mismo de auto-atribuirse una determinada condición mental. Para Vega, aun cuando admitamos, como hace Finkelstein, que un avowal de, por ejemplo, dolor como "Me duele la cabeza" no se hace sobre ningún tipo de base, es decir, carece de un fundamento epistémico, esto, por si mismo, no significa que el avowal no pueda ser al mismo tiempo una expresión de la creencia de segundo orden de que a uno le duele la cabeza (2011, p. 42). En una línea similar, Bar-On (2004) sostiene que al expresar un avowal se puede afirmar que el sujeto que a-expresa el estado mental en que se encuentra posee una creencia de segundo orden acerca de su propio estado que tiene el mismo contenido que la proposición s-expresada. En otras palabras, podemos atribuir una creencia de segundo orden a un sujeto simplemente en virtud de su expresión de un avowal, es decir, de su emisión intencional de una auto-adscripción de un estado mental presente. De esta manera, según Bar-On "se me puede atribuir la creencia de que, por ejemplo, es hambre lo que siento cuando expreso que tengo hambre, incluso si no he formado una opinión sobre el tipo de estado en el que estoy, siempre y cuando me asigne hambre a mí mismo" (2004, p. 366).

Llegados a este punto es importante señalar algunas diferencias fundamentales entre el enfoque de Finkelstein (así como el de Bar-On) y el enfoque de segunda persona. La primera de estas diferencias radica en la extensión neo-expresivista de su explicación a todos los estados mentales, incluidas las creencias y los deseos. Según el enfoque de segunda persona esto es un error ya que, como acabo de señalar, no todos los estados mentales son del mismo tipo y, consecuentemente, su análisis no tiene por qué (o no puede) ser el mismo. De la idea de que aprendamos a sustituir expresiones naturales por expresiones lingüísticas no se sigue que esto ocurra con todos nuestros estados mentales. Afirmar lo contrario implica obviar la necesidad de elementos conceptuales para el conocimiento de determinados estados mentales. Una cosa son los

estados mentales que no precisan de elementos conceptuales para su conocimiento y que el bebé aprende a sustituir por expresiones lingüísticas (aun cuando no posea los conceptos necesarios para su comprensión), como las emociones básicas y las sensaciones, y otra los que sí requieren de elementos conceptuales, como es el caso de las creencias. Volviendo al ejemplo anterior, la niña sabe que mediante los avowals que sustituyen a las expresiones naturales está haciendo lo mismo que con estas, sabe que tanto al estirarse hacia el osito con la intención de cogerlo como al decir "Quiero mi osito" está expresando el deseo de coger su osito, pero ha de comprender que hay determinados términos auto-adscriptivos que no sustituyen una expresión natural y para cuya comprensión requerirá de la posesión de determinados elementos conceptuales. No es lo mismo que la niña aprenda a sustituir algo que ya posee (las expresiones con las que de manera natural se comunica con su madre) que aprender que sus afirmaciones acerca del mundo contienen un elemento conceptual (la creencia) que está expresando sin saberlo (puesto que aún no conoce el concepto "creencia") y que puede afirmar mediante auto-adscripciones de creencia. Asimismo, tendrá que aprender la distinción entre dos sentidos en los que puede auto-adscribirse una creencia (en los que puede utilizar el término "creo") y que, aun cuando esté también ausente en los análisis neo-expresivistas, según el enfoque de segunda persona resulta fundamental para la comprensión de esta actitud proposicional y, por tanto, de lo que hacemos con ella. Por un lado, está el sentido en el que *afirmamos* (damos por cierto) algo que no conocemos de manera directa, que no hemos comprobado o que no está demostrado. En estos casos lo que hacemos cuando decimos, por ejemplo, "Creo en Dios", es *subrayar* nuestro *total* compromiso con la verdad de la existencia de Dios[49] y, por tanto, nuestro compromiso a actuar como se espera de alguien que cree en Dios (aunque luego no lo hagamos y se pueda cuestionar dicho compromiso). Por otro lado, está el sentido en el que *dudamos* de algo, por ejemplo, cuando decimos "Creo que está lloviendo". En este caso lo que hacemos es *rebajar* nuestro compromiso con la verdad de que, de hecho, esté lloviendo.

A propósito, el análisis expresivista de Frápolli y Villanueva (2012) sobre este tipo de predicados resulta muy esclarecedor. Frápolli y Villanueva defienden un expresivismo mínimo en el campo de la semántica según el cual "una posición puede denominarse "expresivista" —en el sentido mínimo que defendemos— si está basada simplemente en las siguientes características clave del significado de estas expresiones:

[49] Nótese que las oraciones que utilizamos para este tipo de auto-adscripciones "Creo *en p*" pueden sustituirse por "*P existe*".

pueden usarse como funciones de proposiciones, y no se usan para describir cómo es el mundo" (2012, p. 471). Su propuesta está centrada en el subconjunto particular de predicados de segundo orden que pueden usarse para producir proposiciones a partir de otras proposiciones. Según estos autores una teoría del significado puede ser llamada 'expresivista' si cumple con, al menos, dos de las siguientes tesis:

> (1) *Higher-order functions* (HOF). There are natural-language expressions with the following structural properties: they are non-extensional, non-truth-conditional functions of propositions. At least one of the items of the following list can be analyzed along these lines: belief, knowledge, necessity, possibility, good, bad, right, wrong.
> (2) *Non-descriptivism* (ND). These terms are not used to describe the way the world is.
> (3) *Truth-conditional status* (TCS). Expressions containing these terms lack truth conditions, even though they are syntactically correct – they are not "truth-apt".
> (4) *Attitude expressions* (AE). These terms are used to express some attitude *A* towards a particular piece of content (2012, p. 471).

El expresivismo mínimo, en cambio, prescinde de la tesis AE para los predicados de segundo orden, ya que, según estos autores, aun cuando podemos "agregar aspectos cognitivos, epistémicos o emocionales al significado general de estas nociones, como afirma AE […] estos aspectos no pertenecen al núcleo semántico que explica el comportamiento inferencial de los conceptos expresivistas" (2012), p. 475). Asimismo, si bien es cierto que ND indica que las expresiones de segundo orden no son descriptivas, es decir, no se usan para hablar de cómo es el mundo, esto no implica que las expresiones que contienen esos términos carezcan de condiciones de verdad, sino, más bien, que dichos términos son irrelevantes en cuanto a las condiciones de verdad (2012, pp. 477–480). Según Frápolli y Villanueva, la tesis TCS ha de reformularse de la siguiente manera: "*Irrelevancia de las condiciones de verdad* (TCI). Lo predicados de segundo orden no modifican las condiciones de verdad de las expresiones bajo su alcance" (2012, p. 478). Esta tesis permite tanto dirimir el conflicto con el expresivismo "simple", puesto que *no* afirma que las auto-adscripciones de creencia carezcan de condiciones de verdad, como de dar una explicación sobre lo que hacemos con dichos términos, a saber, modificar las condiciones de *evaluación* de las expresiones bajo su alcance. En el caso del predicado que estamos tratando, "Creo", modifica las circunstancias en las que se ha de evaluar el compromiso adquirido con la verdad de las proposiciones bajo su alcance. De esta manera, "Creo" en un avowal de creencia como, por ejemplo, "Creo que está lloviendo", modifica las condiciones en la cuales estamos evaluando el hecho de que esté lloviendo. O bien podemos estar subrayando nuestro

compromiso con la verdad de que está lloviendo, o bien estamos rebajando nuestro compromiso con dicha verdad, mostrando nuestra duda al respecto. Volviendo al ejemplo de la niña, en el caso de las creencias tendrá que comprender que lo que está haciendo al auto-adscribirse la creencia de que *p* puede ser tanto una forma de *subrayar* su compromiso con la verdad de que *p* como afirmar que ese es solo su propio punto de vista, entre otros posibles, que no tiene por qué ser correcto.

Además de la extensión de su análisis a todos los estados mentales, con los problemas y carencias que conlleva, hay otra diferencia fundamental entre los neo-expresivistas y el enfoque de segunda persona: ambos autores obvian (o no dan la importancia que realmente tienen) las interacciones interpersonales en la explicación de su enfoque, aun cuando estén implícitas en el mismo, ya que, como señalé en el apartado anterior, el bebé no puede aprender a sustituir una expresión natural por otra lingüística si no es mediante la interacción con su madre (y, más tarde, con otros adultos, como vimos en el primer bloque). Obviar la importancia de las interacciones conduce a los neo-expresivistas a centrarse en cómo aprendemos a sustituir nuestras expresiones naturales por auto-adscripciones sin tener en cuenta que el proceso de aprendizaje de las auto-adscripciones no está, en manera alguna, separado del aprendizaje de las adscripciones de estados mentales a los demás. Dada la capacidad del bebé para comunicarse de manera expresiva con su madre (que, recordemos, implica su habilidad para detectar e intercambiar información expresiva significativa y de captar si el estado mental que ve en la madre es el mismo estado mental que siente), resulta difícil, por no decir incoherente, pensar que la niña aprende *solo* a sustituir sus expresiones naturales por expresiones lingüísticas sin que, al hacerlo, las reconozca también en el repertorio expresivo de su madre. Del mismo modo que la niña aprende que "Quiero el osito" sustituye a su deseo de alcanzar el osito, aprende, asimismo, que esas mismas palabras dicha por su madre sustituyen el deseo de su madre de alcanzar el osito. La niña aprende *tanto* a sustituir por expresiones lingüísticas sus propias expresiones *como* a reconocerlas en su madre.

Este hecho resulta fundamental para entender el rasgo relacional de la adquisición de las expresiones lingüísticas por parte de la niña así como de las normas sociales que regulan el comportamiento de los miembros de la comunidad. En este sentido, los enfoques neo-expresivistas son, al igual que los agencialistas, enfoques individualistas cuyo objetivo es explicar las características distintivas del

autoconocimiento exclusivamente desde el individuo, sin tener en cuenta la relación con los demás.

En este apartado hemos visto los enfoques neo-expresivistas de Bar-On y Finkelstein. Ambos autores sostienen que la autoridad de primera persona proviene de nuestra capacidad expresiva y del hecho de que podamos expresar nuestros estados mentales mediante auto-adscripciones. Asimismo, ambos afirman que dichas auto-adscripciones poseen condiciones de verdad puesto que son aserciones lingüísticas (s-expresiones en el caso de Bar-On —es decir, expresiones lingüísticas con una estructura semántica— y aserciones pertenecientes a nuestro "juego del lenguaje" en el caso de Finkelstein), por lo que, mediante los avowals, *además* de expresar el estado mental comunicamos algo que puede ser verdadero o falso.

He señalado que Bar-On ofrece una explicación de la transparencia de los estados mentales que subraya o suscribe la captación directa de la condición comunicativa de los estados mentales (aunque Finkelstein comparte una idea similar se niega a usar el término "transparencia" por que la entiende únicamente en el sentido de Evans o Moran). Además, la concepción de Bar-On de los estados mentales como condiciones en las que los sujetos se encuentran y no estados *dentro* de los sujetos y que inferimos mediante la expresión o "vemos" *a través* de ella coincide con la ofrecida por el enfoque de segunda persona. Sin embargo, tanto Bar-On como Finkelstein obvian la condición comunicativa de los estados mentales, aun cuando, como he argumentado, esta está implícita en su explicación.

Asimismo, he suscrito algunas de las críticas de otros autores a ambos enfoques. En el caso de Bar-On he afirmado que, como señala Coliva, su explicación de los avowals falsos resulta problemática puesto que elimina la posibilidad de que un avowal falso sea un avowal según la propia definición de Bar-On. De otra parte, las críticas de Vega y Stroud a Finkelstein señalan la necesidad de un análisis sobre el tipo de conocimiento involucrado en su explicación del conocimiento de uno mismo de sus estados mentales. Al respecto he sostenido que el tipo de conocimiento de los estados mentales mediante los cuales se comunica el bebé con su madre, es un conocimiento no conceptual, es decir, que no requiere de elementos conceptuales y que forma parte de nuestra capacidad para la comunicación expresiva. Sin embargo, he sostenido que, una vez somos usuarios competentes del lenguaje, al auto-adscribirnos un estado mental no puede decirse que no hayan elementos conceptuales involucrados, por lo que las auto-

atribuciones lingüísticas sí pueden ser consideradas como conocimiento epistémico, aun cuando su obtención no requiera de fundamentación epistémica.

En relación con este tema, he afirmado que el individualismo de los neo-expresivistas se refleja en que su explicación de la autoridad de primera persona no tiene en cuenta las interacciones personales, ya que está basada en la asimetría entre primera y tercera persona. Esto hace que los neo-expresivistas obvien el aspecto relacional de la adquisición de las expresiones lingüísticas por parte de la niña, la cual, he sostenido, no está separada en manera alguna del reconocimiento de las mismas en el repertorio expresivo de su madre. Asimismo, he afirmado que de la idea de que aprendemos a sustituir expresiones naturales por expresiones lingüísticas no se sigue que esto sea así para todos los estados mentales, puesto que ni todos los estados mentales son del mismo tipo, como muestra el conocimiento no conceptual de algunos de ellos, ni tienen que tener, por tanto, el mismo tipo de análisis. Al respecto, he señalado el error de los neo-expresivistas de extender su análisis a todos los estados mentales, incluidas las creencias y he ofrecido una interpretación de las mismas en términos del expresivismo mínimo defendido por Frápolli y Villanueva. Según esta interpretación, las creencias son predicados de segundo orden cuya función es la de modificar las condiciones de evaluación de las proposiciones bajo su alcance. En el caso descrito por Bar-On, la niña tendrá que comprender los distintos usos de la auto-adscripción "Creo que p", siendo uno de ellos subrayar su compromiso con la verdad de que *p*, lo cual conlleva *comportarse* como alguien que cree que *p,* y otro rebajar dicho compromiso, lo cual implica su duda al respecto y la *posibilidad de un comportamiento distinto* respecto a *p*. De ahí la importancia fundamental de la comprensión por parte de la niña de esta distinción, y los distintos compromisos que adquiere al usarla. Aun cuando la niña sepa que mediante los avowals que sustituyen a las expresiones naturales está haciendo lo mismo que con estos, ha de comprender que hay determinados términos auto-adscriptivos que no sustituyen las expresiones naturales con las que se comunica con su madre sino que sirven, entre otras cosas, para dar razones de nuestra conducta y de la conducta de los demás y que son necesarios para comprender las normas sociales con las que nos regulamos, como sostendré en el siguiente bloque.

6. Conclusiones

En este bloque he presentado el debate acerca del denominado "problema del autoconocimiento", es decir, acerca de las características especiales distintivas del autoconocimiento: la autoridad de primera persona y la transparencia de los estados mentales. La autoridad está relacionada con el hecho de que, por lo general, nadie está en mejor situación que uno mismo para saber lo que desea, cree, siente, etc. Se nos supone autoritativos respecto a lo que decimos acerca de nosotros mismos, es decir, respecto de nuestros avowals (auto-atribuciones en primera persona de estados mentales presentes), pero no así respecto a lo que decimos acerca de los demás o de nuestros estados corporales (como el peso, la altura, etc.). La transparencia está relacionada con el hecho de que, por lo general, si quiero saber lo que pienso, creo, deseo, etc., no necesito inferirlo de mi comportamiento ni preguntarme a mí mismo, lo sé de una manera directa, no inferencial, transparente. De esta manera, nuestros avowals gozan de una presunción de verdad de la que no gozan nuestras auto-atribuciones de estados corporales ni las atribuciones de estados mentales a los demás.

Para llevar a cabo el análisis del problema, he dividido los diversos enfoques en tres grandes grupos: los introspeccionistas, los agencialistas y los expresivistas. Los introspeccionistas defienden que el autoconocimiento es un hecho empírico que consiste en rastrear, o descubrir, la propia vida mental a través de un mecanismo epistémico similar al de la percepción. Entre los defensores de este enfoque he presentado los que considero más representativos, Armstrong y Lycan, quienes sostienen que los estados mentales son "objetos internos" solo accesibles para el propio sujeto mediante un mecanismo de auto-escaneo interno del cerebro que detecta el estado mental de primer orden, dando como resultado un estado mental de segundo orden (una auto-adscripción). Dicho mecanismo es, de esta manera, la causa de la auto-adscripción y, al igual que los estados mentales que rastrea, se supone material y falible, aunque generalmente confiable. Para estos autores, por tanto, el autoconocimiento es el resultado de un acceso epistémico privilegiado a un estado de cosas psicológico independiente y previo a la auto-adscripción. La autoridad y la transparencia son entendidas *a posteriori*, y de manera contingente, ya que sólo conozco mis estados mentales tras el (confiablemente correcto) escaneo de mi mecanismo epistémico interno.

De otra parte, los agencialistas rechazan el introspeccionismo pues, según ellos, el autoconocimiento no es una cuestión de *acceso* a un contenido, sino que es está relacionado con la acción de un agente. Según estos autores, los estados mentales son algo que hacemos los sujetos en cuanto seres racionales y moralmente responsables. Dentro de este enfoque he distinguido entre la propuesta deliberativa de Moran y las propuestas constitutivistas de Wright y Bilgrami. La diferencia entre ellas radica en que Moran suscribe la transparencia en el sentido de Evans, que el denomina Condición de la Transparencia, como método de autoconocimiento y defiende la agencia deliberativa como garante de la autoridad de primera persona. Los constitutivistas, por su parte, destacan la parte constitutiva de la agencia sosteniendo que, en condiciones normales, la auto-adscripción trae consigo, o constituye, el estado mental auto-atribuido. Para estos autores el autoconocimiento es una condición *a priori* de las reglas que gobiernan nuestro uso de las auto-atribuciones de estados mentales. Según estas reglas, que rigen las prácticas lingüísticas de la comunidad, no hay por qué dudar de las auto-atribuciones psicológicas de un agente que cumple con las condiciones normales de racionalidad (un agente que es sincero, cognitivamente lúcido y alerta, con el repertorio conceptual relevante y sin razón para pensar que podría auto-engañarse), en el caso de Wright, y con las condiciones para la agencia responsable (un sujeto sincero, cognitivamente lúcido y alerta, con el repertorio conceptual relevante y dispuesto —si se diera cuenta de que no estaba cumpliendo con su compromiso— a aceptar las críticas por no haber estado a la altura e intentar hacerlo mejor), en el caso de Bilgrami. De esta forma, Bilgrami va más allá de Wright pues aunque coincide con él en que asumir la autoridad de primera persona debe ser nuestra posición por defecto, para Bilgrami no es solo una precondición de las reglas de nuestro juego del lenguaje respecto a las auto-atribuciones mentales, sino que es la actitud que hemos de tomar los unos con los otros en cuanto agentes responsables. De ahí que sostenga que el autoconocimiento es una condición *a priori* de la agencia responsable, ya que si renunciamos a la intuición de que, generalmente, conocemos los estados mentales que nos auto-atribuimos, renunciamos, asimismo, a nuestro derecho a tratarnos como agentes morales, conscientes y responsables de lo que hacemos (no podemos ser responsables de aquello que desconocemos).

Finamente, he presentado los enfoques neo-expresivistas de Bar-On y Finkelstein, quienes sostienen que la autoridad de primera persona proviene de nuestra capacidad expresiva y del hecho de que podamos expresar nuestros estados mentales

mediante auto-adscripciones. Los neo-expresivistas, al contrario que el expresivismo "simple", afirman que las auto-adscripciones poseen condiciones de verdad puesto que son aserciones lingüísticas, por lo que, mediante los avowals, *además* de expresar el estado mental comunicamos algo que puede ser verdadero o falso.

En cuanto a las críticas a estos enfoques, he expuesto algunas de las que considero más relevantes. En el caso del introspeccionismo he señalado la contingencia del sistema de monitoreo, la posibilidad de que existan personas auto-ciegas cuyo sistema de monitoreo no funciona, el hecho de que el acceso a los estados mentales propios sea exclusivo del individuo, lo cual conlleva la incertidumbre acerca de si los demás también los tienen y la crítica proveniente del externismo, según el cual los significados de nuestras palabras y el contenido de nuestros pensamientos están determinados, en parte, por factores externos a nosotros. En el caso del agencialismo, la crítica de Cassam al racionalismo de Moran, centrada en las actitudes recalcitrantes, muestra que este último realiza un análisis excesivamente exigente para lo que realmente somos los seres humanos, seres cuyas actitudes están lejos de ser completa y absolutamente racionales. Finkelstein, por su parte, le objeta a Wright que su solución al problema del seguimiento de reglas cae en la misma crítica que él señala, ya que, según Finkelstein, las estipulaciones no evitan el regreso al infinito, so pena de dotarlas de contenido intrínseco. Asimismo, según Finkelstein, su "objeción de la responsabilidad" muestra que el constitutivismo no puede dar cuenta de los estados mentales fenoménicos ya que si por el simple hecho de que me auto-adscriba un estado mental de dolor lo constituyo de alguna manera, podría culpárseme por tenerlo, es decir, por habérmelo provocado.

Finalmente, he expuesto algunas de las críticas al neo-expresivismo. En el caso de Bar-On, su explicación de los avowals falsos pero sinceros es, según Coliva, extremadamente problemática, ya que si la causa racional, la presencia del estado mental, es idéntica a lo que lo hace verdadero [*truth-maker*] y el segundo está ausente, también lo está la primera, es decir, si el estado mental no está presente en el sujeto, la auto-adscripción puede, a lo sumo, *parecer* un avowal, pero no ser considerado como tal. Por último, Vega y Stroud critican la carencia de una explicación epistémica de las auto-adscripciones en el enfoque de Finkelstein. Según estos autores, la explicación epistémica resulta fundamental si se pretende sostener que los avowals son, además de expresiones, también aserciones, por lo que, al contrario de lo que afirma Finkelstein, esta es una cuestión de verdadera importancia filosófica. Al respecto he sostenido la

existencia de un tipo de conocimiento no conceptual de determinados estados mentales, a saber, de aquellos con los que se comunica el bebé con su madre y que, consecuentemente no requieren de elementos conceptuales para su conocimiento. Asimismo, he sostenido que al auto-adscribirnos un estado mental no puede decirse que no hayan elementos conceptuales involucrados una vez somos usuarios competentes del lenguaje, por lo que las auto-atribuciones de estados mentales sí pueden ser consideradas como conocimiento epistémico, aun cuando su obtención no requiera de fundamentación epistémica.

En lo referente a la relación de estos enfoques con el enfoque de segunda persona, he afirmado que todos y cada uno de estos enfoques basan su explicación del autoconocimiento en la dicotomía entre primera y tercera persona, lo cual descarta las interacciones personales como relevantes para la explicación del autoconocimiento y sitúan al individuo en el centro de la explicación. He argumentado que ambos presupuestos conllevan una serie de problemas. En el caso del introspeccionismo he afirmado que para los defensores de este enfoque, las relaciones del sujeto con los demás no tienen relevancia ninguna respecto a la transparencia, la autoridad y el autoconocimiento. Por tanto, siendo un enfoque completamente individualista (y, consecuentemente, un enfoque opuesto al de segunda persona) que parte del sujeto como un ser individual y aislado, teniendo en cuenta solo el acceso privado de este a sus estados mentales, conlleva la incertidumbre acerca de si las demás personas también tienen estados mentales al igual que uno mismo.

En el caso de los agencialistas, he señalado que aun cuando sí incluyen las interacciones personales en su enfoque, estas son consideradas exclusivamente desde una perspectiva teórica y en relación al individuo. Wright apela a las prácticas lingüísticas de la comunidad como base de las reglas que constituyen nuestros conceptos psicológicos y determinan el significado de las palabras usadas para expresarlos. Asimismo, entiende la autoridad como una concesión *otorgada* a cualquiera a quien consideremos un sujeto racional, por lo que presupone la autoridad como algo relacionado con los demás miembros de la comunidad. Bilgrami, siguiendo a Strawson, sostiene que nuestra comprensión de la libertad y del autoconocimiento de la agencia responsable provienen de nuestra involucración en interacciones personales. Sin embargo, ambos autores explican las características distintivas del autoconocimiento desde el punto de vista del individuo. Por un lado, la transparencia es vista únicamente como una relación del individuo respecto a sus propios estados mentales, obviando la

posibilidad de otro tipo de transparencia relacionada con nuestra interacción con los demás miembros de la comunidad. Por otro, la autoridad es vista como una característica inalienable del individuo, obviando la posibilidad de que la relación con los demás miembros de la comunidad pueda influir en ella. Al respecto, he suscrito la crítica de Borgoni sobre la concepción idealizada del autoconocimiento de la que adolecen los agencialistas en la que los avowals son realizados en un "entorno esterilizado" que obvia los factores sociales. Casos como los de injusticia epistémica muestran la importancia de la inclusión de las interacciones personales en la explicación del autoconocimiento, ya que revelan que la autoridad de primera persona conlleva un elemento atribucional, de reconocimiento por parte del otro de dicha autoridad. Esta crítica está relacionada con la crítica de Cassam a Moran, según el cual Moran hace un análisis excesivamente racional que no se ajusta a cómo nos comportamos realmente los seres humanos, siendo, a lo sumo, un análisis para un *homo philosophicus* mítico, cuyas creencias y otras actitudes son siempre lo que deberían ser racionalmente, más que para los seres humanos normales, quienes no son ciudadanos epistémicos modelo y cuyas actitudes están lejos de ser como racionalmente deben ser. De otra parte, el individualismo de Moran se refleja, asimismo, en un análisis de la deliberación que no tiene en cuenta el reconocimiento por parte de nuestro interlocutor de las razones en cuanto razones. Al respecto he sostenido que al igual que el lenguaje, las razones no pueden ser privadas, es decir, que lo que cuenta como una razón para el agente deliberativo ha de contar, asimismo, como una razón para su interlocutor, aun cuando este no esté de acuerdo con su contenido.

De otra parte, los agencialistas incluyen un elemento fundamental para el enfoque de segunda persona, un rasgo constitutivo en los seres humanos ya desde muy temprana edad y que está directamente relacionada con las interacciones personales: la normatividad. Sin embargo, estos autores tratan la normatividad también de manera individual, excluyendo en su enfoque la apelación directa la normatividad social, la cual, según el enfoque de segunda persona está íntimamente relacionada con los compromisos que adoptamos como agentes responsables al auto-atribuirnos estados mentales. Estos autores entienden los compromisos como derivados de una decisión del agente, sin tener en cuenta la existencia de otro tipo de compromisos que todos *asumimos* al participar en el "juego del lenguaje" de las auto-atribuciones de estados mentales y que provienen de las normas sociales que regulan el *comportamiento* derivado de dichas auto-atribuciones y que, se supone, todos seguimos. Según el

enfoque de segunda persona, nuestros compromisos como agentes responsables no se limitan a aquellos derivados de una decisión directa, sino que cada una de nuestras auto-atribuciones de estados mentales conlleva un compromiso con su verdad y, por tanto, un compromiso a comportarnos como se espera que se comporte alguien que se auto-atribuye el estado mental en cuestión.

Finalmente, los neo-expresivistas coinciden con algunos de los presupuestos del enfoque de segunda persona en lo relativo tanto a la adquisición del lenguaje de las auto-atribuciones de estados mentales como a las nociones de expresión y transparencia, aunque no dejan de ser enfoques individualistas centrados en el individuo y basados en la dicotomía entre primera y tercera persona. En primer lugar, la noción de estado mental como una condición en la que el sujeto se encuentra y no como un estado dentro del sujeto, coincide con la explicación ofrecida en el primer bloque, según la cual, los estados mentales no son objetos ocultos en el interior del individuo solo accesibles para los demás mediante la realización de inferencias. De otra parte, y en relación con esta misma idea, la transparencia-a-la-condición-mental-del-sujeto defendida por Bar-On subraya la captación directa de la condición comunicativa de los estados mentales, ya que, según esta autora, las expresiones naturales de los estados mentales (y las expresiones lingüísticas que las sustituyen) muestran de manera transparente (directa, no inferencial) tanto la condición en la que el sujeto se encuentra como al sujeto estando en dicha condición.

Asimismo, la explicación de los neo-expresivistas proveniente de las ideas de Wittgenstein acerca de la adquisición del lenguaje de las auto-atribuciones de estados mentales, encaja a la perfección con lo defendido en el primer bloque acerca de la autoridad de la madre respecto al bebé, autoridad que he denominado “autoridad de segunda persona”. Al igual que la madre es una autoridad para el bebé respecto a las situaciones que desconoce (o para las que no tiene una respuesta asignada) lo es respecto a la manera en la que ha de sustituir sus expresiones naturales por expresiones lingüísticas. La niña aprende mediante la autoridad de la madre a expresar de otra manera (lingüística) las expresiones naturales con las que se comunica con ella. Sin embargo, al no dar la importancia a las interacciones personales en su enfoque (aun cuando estén implícitas en su explicación), los neo-expresivistas obvian la importancia del rasgo relacional del aprendizaje por parte de la niña de las auto-adscripciones de estados mentales. Del mismo modo que la niña aprende a sustituir unas expresiones por otras e incluirlas en su repertorio expresivo aprende, asimismo, a reconocerlas en el

repertorio expresivo de la madre. Este hecho, he sostenido, resulta crucial para entender tanto el rasgo relacional de este tipo de aprendizaje como de las normas sociales con las que se regulará una vez haya adquirido la competencias necesarias para ello (como sostendré en el siguiente bloque).

De otra pare, he afirmado que este esquema de aprendizaje no puede extenderse a las actitudes proposicionales como las creencias, ya que, al contrario que las emociones básicas y las sensaciones con las que la niña se comunica con su madre, las primeras requieren de elementos conceptuales para su comprensión. Al respecto, he sostenido que el enfoque expresivista de Frápolli y Villanueva da una explicación, a mi juicio correcta, de lo que hacemos cuando usamos este tipo de auto-atribuciones: modificar las condiciones en las que se han de evaluar las proposiciones bajo su alcance. De esta manera, he sostenido que en el caso de las auto-atribuciones de creencias, podemos tanto subrayar como rebajar nuestro compromiso con la verdad de los predicados bajo su alcance. En el caso descrito por Bar-On, la niña tendrá que comprender los distintos usos de la auto-adscripción "Creo que p", a saber, subrayar su compromiso con la verdad de que *p*, lo cual conlleva *comportarse* como alguien que cree que *p,* y otro rebajar dicho compromiso, lo cual implica su duda al respecto y la *posibilidad de un comportamiento distinto* respecto a *p*.

En el siguiente bloque explicaré la importancia de un enfoque de segunda persona para la comprensión del autoconocimiento. Mostraré cómo el hecho de que los enfoques acerca del autoconocimiento no tengan en cuenta las interacciones conlleva una importante carencia en la comprensión del mismo. Para ello, me centraré en las explicaciones de la adquisición de las normas sociales (*folk-psychology*) en el desarrollo del niño y en lo que al respecto han afirmado los diferentes enfoques que en psicología y en teoría de la mente se han llevado a cabo sobre el asunto.

BLOQUE III
Autoconocimiento y Folk-Psychology

1. Introducción

El fenómeno de la atribución de estados mentales ha sido uno de los temas más debatidos en las últimas décadas tanto en psicología del desarrollo como en filosofía de la mente. El debate, que en la actualidad sigue generando una amplia controversia, se centra en el desarrollo de la adquisición de la denominada Folk-Psychology[50] (en adelante FP), entendida esta como la capacidad para comprender el comportamiento de las demás personas, es decir, como la capacidad para comprender por qué los demás hacen lo que hacen. Tradicionalmente, el estudio del fenómeno se ha centrado en la investigación de la habilidad humana para vernos los unos a los otros como seres 'mentales', es decir, para vernos como dotados de determinados estados cognitivos, afectivos, perceptuales, que supuestamente guían nuestro comportamiento y mediante los cuales predecimos y explicamos nuestra conducta y la conducta de los demás.

Con objeto de explicar nuestra capacidad para entender el comportamiento de los demás, es decir, la FP, la investigación ha estado tradicionalmente dividida en dos versiones contrapuestas de este fenómeno, ambas centradas en la naturaleza de los mecanismos que subyacen al proceso de atribución de estados mentales. De una parte, hay teóricos que defienden la denominada Teoría de la Teoría (en adelante TT), según la cual los seres humanos, a partir de los cuatro años aproximadamente, interpretamos la conducta sobre la base de una teoría formada por constructos o supuestos teóricos sobre estados mentales derivada de la observación de la conducta propia y ajena. Según estos teóricos, las atribuciones de estados mentales son producidas por este tipo de teorización basada en un corpus sistemático de conocimiento que detalla las conexiones entre entradas perceptivas, estados internos y resultados conductuales (Gopnik, 1993, 2003; Gopnik & Meltzoff, 1997; Gopnik et al., 1999; Gopnik & Wellman, 1992, 1994; Meltzoff, 2002; Wellman, 1990). De otra parte, los defensores de la Teoría de la

[50] Esta capacidad ha recibido varios apelativos en los estudios acerca del tema: 'psicología popular' (*folk-psychology*), 'teoría de la mente' *(theory of mind*), 'lectura de mentes' (*mind-reading*), 'simulación de mentes' (*mind-simulating*), entre otros. En este trabajo utilizaré el término 'folk-psychology' para referirme a este fenómeno puesto que engloba *las* capacidades generales necesarias para la auto-atribución de estados mentales, mientras que los otros términos pueden ser entendidos como *una* de las capacidades particulares que explican este fenómeno.

Simulación (en adelante TS) sostienen que nuestra habilidad para atribuir estados mentales no consiste en la aplicación de una teoría acerca de los estados mentales sino que lo que hacemos es utilizar nuestros propios recursos cognitivos para simular o pretender que estamos en la posición del otro y, así, imaginar o generar los estados mentales que atribuimos (Goldman, 1989, 1993, 2000, 2006; Gordon, 1986, 1995, 2007, 2009).

En este bloque desarrollaré una explicación de la FP y del autoconocimiento desde el enfoque de segunda persona, tomando en consideración tanto lo sostenido en el primer bloque acerca del desarrollo ontogenético, como lo sostenido en el segundo acerca del fenómeno del autoconocimiento. Sostendré que el modelo adecuado para describir la adquisición de la FP no es el observacional ni el de la simulación, sino el modelo de *comunicación reguladora mutua*. Bajo este modelo, la FP no es entendida como una estrategia para predecir y explicar el comportamiento, sino como una práctica normativo-regulativa.

En lo referente al fenómeno del autoconocimiento, sostendré una concepción de la transparencia relacionada con la comunicación expresiva y el expresivismo mínimo. Según esta concepción, los estados mentales mediante los que nos comunicamos a través de la comunicación expresiva son transparentes tanto para el sujeto como para la persona con la que estamos interaccionando. De otra parte, las actitudes proposicionales entendidas como predicados de segundo orden son, según esta concepción, necesariamente trasparentes. En cuanto a la autoridad de primera persona, sostendré que está relacionada tanto con el estatus de agente responsable como con el reconocimiento de los demás. Finalmente, sostendré una concepción del autoconocimiento que engloba dos nociones de conocimiento: el conocimiento pre-conceptual, que no requiere de elementos conceptuales, y el conocimiento epistémico (el conocimiento en sentido tradicional), que sí requiere del lenguaje y de conceptos. Según esta noción de autoconocimiento, no conocemos nuestros estados mentales desde la perspectiva de primera o de tercera persona, sino desde las perspectivas de primera, de segunda y de tercera persona. Los tres modos de conocer nuestros estados mentales forman parte del autoconocimiento.

En siguiente apartado haré un breve resumen de los enfoques tradicionales de la FP y mostraré las deficiencias de las que, según el enfoque de segunda persona, adolecen estos enfoques para la comprensión del fenómeno. En el apartado 3 explicaré cómo el enfoque de segunda persona da cuenta de la adquisición de la FP para concluir

en el apartado 4 con la explicación de la transparencia, la autoridad de primera persona y el autoconocimiento. Finalmente, expondré una breves conclusiones del marco presentado en el apartado 5.

2. Teorías individualistas y Folk-Psychology: Teoría de la Teoría y Teoría de la Simulación

Los defensores de la TT sostienen que las atribuciones del estados mentales son producidas por una especie de teorización basada en un corpus de conocimiento sistemático acerca de relaciones causales-inferenciales entre los inputs sensoriales (entradas perceptivas), otros estados mentales y los outputs conductuales (comportamientos) (Gopnik, 1993, 2003; Gopnik & Meltzoff, 1997; Gopnik & Wellman, 1994). La idea central de estos autores es que al atribuir estados mentales a otros y a nosotros mismos empleamos conceptos teóricos de una teoría psicológica de sentido común que llegamos a poseer a partir, aproximadamente, de los cuatro años de edad, cuando somos capaces de comprender la atribución de creencias falsas. Esta afirmación proviene de los experimentos realizados con primates a finales del siglo pasado por Premack & Woodruff, (1978), los cuales pretendían comprobar el nivel de comprensión de la conducta humana que tenían los chimpancés. Los excelentes resultados obtenidos por Premack y Woodruff les llevaron a concluir que los chimpancés poseían una serie de conceptos, informaciones y regularidades estructuradas como una "teoría de la mente" mediante la cual interpretaban la conducta. El éxito de estos experimentos condujo a determinados teóricos a sostener que nuestras habilidades en la comprensión de las interacciones humanas podían ser, asimismo, explicadas apelando a la posesión de una serie de datos e informaciones acerca de la mente y la conducta que toman la forma de una teoría. De esta manera, Wimmer y Perner (1983) crearon el denominado "test de la falsa creencia" (*false belief task,* en adelante FBT) para determinar en qué momento del desarrollo alcanzábamos la posesión de dicha teoría. La elección de este test fue debida a la presuposición de que solo alcanzamos la comprensión de las mentes ajenas, es decir la TT, cuando somos capaces de dominar el concepto de creencia y dominar el concepto de creencia implica comprender que los contenidos representacionales de alguien pueden no coincidir con

cómo es el mundo en realidad. Wimmer y Perner diseñaron el test siguiendo los comentarios de Benett (1978), Dennett (1978) y Harman (1978) al trabajo de Premack y Woodruff (1978), con el que se pretendía investigar "la competencia de los niños para representar la creencia concreta de otra persona que difiere de lo que el sujeto sabe que es verdad" (Wimmer & Perner, 1983, p. 106). El test se diseñó y desarrolló de la siguiente manera: unos niños observan como otro niño, Maxi, coloca chocolate en un armario X. Maxi sale entonces del cuarto y su madre cambia el chocolate del armario X al armario Y. Los niños entonces han de indicar en qué armario buscará el chocolate Maxi cuando vuelva al cuarto. Solo cuando los niños son capaces de representar la creencia falsa de Maxi (de que el chocolate está en el armario X) y diferenciarla de lo que ellos saben que es el caso (que el chocolate está en el armario Y) señalaran correctamente el armario X. Según Wimmer y Perner, "la importancia práctica de representar las creencias erróneas de otra persona está en el uso de esta representación como marco de referencia para interpretar o anticipar las acciones de la otra persona. Es decir, las interpretaciones y las anticipaciones deben *limitarse* al ámbito de las creencias de la otra persona" (1983, p. 106).

Alison Gopnik (Gopnik, 1993, 2003; Gopnik & Wellman, 1994), una de las principales defensoras de este enfoque, sostiene que los niños construyen una explicación coherente y abstracta de la mente, en forma de teoría, que les permite *explicar* y *predecir* fenómenos psicológicos. Para Gopnik, aun cuando esta teoría no es explícita sino implícita, el tipo de estructura cognitiva que conlleva parece compartir muchas características con una teoría científica. Al igual que en las teorías científicas, en la teoría de la mente se postulan entidades no observadas (creencias y deseos) y leyes que las conectan, como el silogismo práctico. Asimismo, permiten la predicción, y cambian como resultado de la falsificación de pruebas. La teoría se adquiere, por tanto, a través de un proceso de formación y cambio de teoría a lo largo del desarrollo ontogenético del niño, llegando a su completa posesión alrededor de los cuatro años de edad, cuando son capaces de pasar el FBT.

En lo referente al autoconocimiento, según Gopnik, no existe diferencia entre nuestras auto-atribuciones y las atribuciones de estados mentales a otros puesto que "la teoría de la mente del niño es igualmente aplicable uno mismo y a los demás" (Gopnik, 1993, p. 333). El autoconocimiento, al contrario de lo que afirma el introspeccionismo, no proviene de una acceso interno a nuestros estados mentales. El hecho de que nos parezca tener un acceso directo y no inferencial a nuestros estados mentales, que nos

parezca que no hacemos ninguna inferencia a la hora de auto-atribuirnos un estado mental, no es sino una ilusión que Gopnik denomina la "ilusión del experto":

> I will suggest a speculative analogy between the illusion of privileged knowledge of our own psychological states and what might be called the illusion of expertise. In the case of expertise, direct and immediate experience may be combined with a long, indirect (and theoretical) cognitive history (Gopnik, 1993, p. 334).

La idea de Gopnik es que pericia e inmediatez van la una con la otra. La apariencia de la percepción directa e inmediata es el resultado de una larga y constante puesta en práctica de la teoría. Así, nuestra destreza en el manejo de la teoría de la mente, sumado al hecho de convivimos constantemente con nosotros mismos, hace que no seamos conscientes de las inferencias que realizamos. De ahí que interpretemos las auto-atribuciones como percepciones directas de nuestros propios estados mentales.

La teoría en pugna con la TT es, como hemos señalado, la TS. Al contrario que los teóricos de la TT, los defensores de la TS sostienen que representamos los estados mentales y los procesos de otros *simulándolos mentalmente*, o generando estados y procesos similares en nosotros mismos. Aún cuando son varias las formas en las que la simulación ha sido interpretada, dando lugar a diferentes propuestas, las principales variantes provienen de Alvin I. Goldman y Robert Gordon (Goldman, 1989, 1993, 2000, 2006; Gordon, 1986, 1995, 2007, 2009). Ambos autores coinciden en sostener el método de simulación para explicar la atribución de estados mentales a otros, sin embargo disienten en la explicación de cómo realizamos dicha simulación. Mientras que para Gordon la simulación no requiere de autoconocimiento, para Goldman es un requisito necesario puesto que, según él, la simulación se realiza a partir del conocimiento de nuestros propios estados mentales.

La explicación de Goldman respecto a la adquisición de los conceptos mentales (necesarios para el autoconocimiento) tiene, según él, "mucho en común con la doctrina clásica del introspeccionismo" (1993, p. 372). Según Goldman, aprendemos primero a aplicarnos conceptos psicológicos a nosotros mismos mediante la identificación introspectiva de nuestros propios estados mentales. A partir de ese conocimiento simulamos estar en la situación del otro simulando o generando los conceptos psicológicos correspondientes en nosotros mismos, para después extenderlos al caso del otro. De esta manera, la simulación requiere del conocimiento de los propios estados mentales 'reales' para poder ser simulados, así como de una comprensión previa de

ciertos conceptos mentales básicos. Goldman sostiene un modelo cuasi-perceptual de la aprehensión de los estados mentales propios similar a la de Lycan, que vimos en el apartado 3.1 del bloque anterior, en el que "el "órgano" de la introspección es la atención, cuya orientación pone a un sujeto en una relación apropiada con un estado determinado (2006, p. 244).

Al contrario que Goldman, Gordon niega que sea necesaria la introspección para simular o atribuir estados mentales. Para Gordon, la simulación no consta de razonamiento teórico sino práctico mediante el cual imaginamos una situación hipotética. Este "razonamiento hipotético-práctico" se realiza mediante la proyección imaginativa de uno mismo en la situación del otro pero sin que tenga lugar el resultado conductual (1986, p. 158). Los procesos de toma de decisión funcionan o*ff-line* generando una 'presunta' decisión y el mecanismo simulador atribuye directamente el estado, o la decisión generada, por medio de lo que Gordon denomina "rutina de ascenso semántico" (2007, p. 151). Según Gordon, la rutina de ascenso semántico es el método que utilizamos, en primer lugar, para auto-atribuirnos estados mentales:

> An *ascent routine* (AR) allows a speaker to self-ascribe a given propositional attitude (PA) by redeploying the process that generates a corresponding *lower level* utterance. Thus, we may report on our beliefs about the weather by reporting (under certain constraints) on the weather. (2007, p. 151)

De esta manera, para auto-atribuirnos una creencia basta con preguntar y responder preguntas de primer orden. Así, para saber si está lloviendo uno no necesita preguntarse "¿creo que está lloviendo?" sino sencillamente "¿está lloviendo?" y dar la respuesta "está lloviendo" (en el caso de que esté lloviendo). Tras ello, uno se involucra en una "rutina de ascenso" en la que añade la construcción "creo" a este resultado. La propuesta de Gordon es, en este sentido, similar a la de la transparencia de Evans. El mismo Gordon señala que Evans "fue el primero en plantear la hipótesis de que las personas usan una estrategia de reutilización para responder preguntas sobre nuestras propias creencias" (2007, p. 153). Para el caso de las atribuciones a otros, el proceso sería el mismo que el mencionado para la auto-atribución, pero añadiendo "cree" (en vez de "creo") al resultado de la simulación.

Como acabamos de ver, los defensores de la TT y de la TS difieren en su explicación de la adquisición y funcionamiento de la FP. Mientras que los primeros sostienen la formación de una teoría de la mente mediante la observación y la

comprensión del comportamiento de los demás, la cual aplicamos tanto a nuestra conducta como a la de los demás, los últimos sostienen que lo que hacemos es ponernos en el sitio del otro para, así, simular y comprender sus estados mentales. Sin embargo, lo verdaderamente relevante para el enfoque de segunda persona que defiendo en este trabajo no son sus divergencias sino los presupuestos que ambos enfoques comparten. Como señalé en el primer bloque, tanto la TT como la TS parten de una concepción individualista del fenómeno en el que el foco de estudio es el sujeto como alguien separado de los demás, entendidos estos como seres *a comprender*, bien sea mediante la aplicación de una teoría de la mente o bien mediante la simulación mental de su situación[51]. Ambos enfoques se basan en una concepción cartesiana de la mente en la que esta se desarrolla de manera paralela a la de los demás, el "yo" es visto en oposición a un "él" o "ella" cuyos estados mentales son entidades abstractas e inobservables, por lo que comprender a los demás implica conocer, entender, inferir sus estados o experiencias internas y privadas. Para estos teóricos, el mecanismo de la FP es predictivo y explicativo, es decir, su objeto es *predecir* y *explicar* la conducta de los demás mediante las actitudes proposicionales subyacentes de creencia y deseo. Mientras que los teóricos de la TT suponen que este proceso se realiza desde la perspectiva del observador, es decir, desde la perspectiva de tercera persona, los teóricos de la TS invierten esta explicación y sostienen que este proceso se lleva a cabo desde la perspectiva del sujeto, es decir, desde la perspectiva de primera persona. En cualquiera de los dos casos la interacción interpersonal presupone una brecha, una separación previa entre uno mismo y los demás que solo es posible salvar mediante operaciones mentales.

En el siguiente apartado desarrollaré una explicación de la adquisición de la FP basada en el modelo de *comunicación reguladora mutua* defendido por el enfoque de segunda persona.

[51] Aun cuando se han propuesto teorías 'híbridas' que toman elementos de ambos enfoques (Carruthers & Smith, 1996; Currie & Ravenscroft, 2002; Nichols & Stich, 2003; Stich & Nichols, 1992, 2003), el presupuesto individualista continúa presente en ellas.

3. Segunda persona y Folk-Psychology

En el apartado anterior hemos visto los dos enfoques en pugna en la explicación de la FP, la TT y la TS. Ambos enfoques parten de un presupuesto individualista en el que el sujeto es visto como separado de los demás, es decir, para estos autores existe una brecha entre el individuo y los demás que en el caso de la TT se salva mediante la creación de una teoría en base a la observación y la inferencia y en el caso de la TS mediante la simulación (o proyección) mental en el otro. Estos enfoques, por tanto, parten de lo que Gallagher (2001) denomina la "suposición mentalista", según la cual "conocer a otra persona es conocer la mente de esa persona" (2001, p. 91). De esta manera, conocer los estados mentales de otra persona implica un elemento metarrepresentacional, ya que los estados mentales son entendidos como entidades abstractas e inobservables solo accesibles mediante representaciones mentales (ya sea mediante una supuesta teoría de la mente o mediante la simulación). La idea es que la mente de los otros es opaca, es decir, no puede ser percibida de manera directa, por lo que este tipo de representaciones resulta imprescindible para comprender a los demás. Esta suposición, como señala Gomila (2002), está directamente relacionada con el énfasis que estos autores ponen en el "interés explicativo o predictivo de las atribuciones [que] acentúa el estatus interno, inobservable, teórico o imaginario, de los estados atribuidos" (2002, p. 133).

Sin embargo, los defensores de la TT y la TS obvian, o no tienen en cuenta, otra forma de comprensión de los demás anterior a la conceptual, una capacidad pre-teórica de los bebés que se hace patente en lo que he denominado *comunicación expresiva*. Como he sostenido en el primer bloque de este trabajo, según el enfoque de segunda persona el modelo adecuado para describir este tipo de interacciones no es ni el teórico-observacional ni el de la simulación, sino el modelo de *comunicación reguladora mutua* en el que los bebés y los adultos son tratados como una sola unidad de estudio. En este enfoque el "yo" no es entendido como opuesto a un "él" o "ella", sino como un "yo-tú" interactivo en el que no existe un desarrollo paralelo de la mente del bebé, sino un desarrollo basado en experiencias interactivas compartidas y regulación de las emociones.

En lo que sigue, desarrollaré una explicación de la FP desde el enfoque de segunda persona basándome tanto en lo defendido en los bloques anteriores de este

trabajo como en los diversos enfoques que, en los últimos años, se han propuesto como alternativas a la presuposición mentalista de la TT y la TS.

Al comienzo de este trabajo dividí las etapas del desarrollo ontogenético basándome en la clasificación de Trevarthen (1979), según la cual el desarrollo ontogenético del bebé puede dividirse en tres etapas: el primer tipo de interacción entre el bebé y su madre corresponde a la etapa de la intersubjetividad primaria, desde las primeras semanas de vida hasta, aproximadamente, los nueve meses de edad, momento en el cual el bebé comienza a participar en triángulos de atención conjunta. La etapa siguiente del desarrollo comprende desde los nueve hasta los 20-24 meses de edad y se corresponde con la intersubjetividad secundaria. En este período los triángulos de atención conjunta comienzan a desarrollarse y volverse más complejos con la introducción del bebé en el lenguaje. Esta introducción al lenguaje da lugar a la tercera etapa del desarrollo, la intersubjetividad terciaria (Bråten & Trevarthen, 2007), momento en el cual los bebés comienzan a involucrarse en intercambios lingüísticos y emocionales mas complejos en los que los conceptos y la abstracción juegan un papel primordial. Según el enfoque de segunda persona, es a partir de esta etapa en la que podemos empezar a hablar tanto de la adquisición de la FP como de autoconocimiento y (auto-)atribuciones mentales.

La pregunta central, llegados a este punto, es: ¿cómo adquiere el niño la FP? Como señalé en el bloque anterior dedicado a los enfoques acerca del autoconocimiento, la introducción al lenguaje de la FP comienza con la sustitución (o acompañamiento) de expresiones naturales por expresiones lingüísticas, no por la observación del otro o por la simulación mental de su situación. Para el bebé la madre no es alguien a quien comprender sino alguien con quien interacciona y regula sus emociones. El intercambio comunicativo es mutuamente manifiesto, tanto el bebé como la madre reconocen estar comunicándose mediante el intercambio de expresiones. Los estados mentales con los que se comunica con su madre no son algo que tenga que inferir sino algo que ve, oye, siente directamente. Este elemento de acceso directo a los estados mentales es ignorado o negado tanto por la TT como por la TS, para quienes, producto de una comprensión de los estados mentales como entidades internas e inobservables, la "teoría de la mente" (FP) es entendida como un mecanismo destinado a predecir y explicar la conducta, un mecanismo que el bebé ha de adquirir mediante la observación o la simulación del otro, es decir, desde la perspectiva de tercera o de primera persona. Sin embargo, como he sostenido en el primer bloque, tanto la

perspectiva de primera persona —la del "yo", la de la autoconciencia y la subjetividad–, como la de tercera persona —la del "él" o "ella", la del distanciamiento y la objetividad—, emergen posterior y simultáneamente en el desarrollo ontogenético y están basadas en, o constituidas por, la perspectiva de segunda persona —la del yo-tú, la de la interacción y la intersubjetividad, la de la comunicación expresiva.

Al respecto, Antoni Gomila (2001, 2002), un férreo defensor de la perspectiva de segunda persona, señala el olvido filosófico a la que esta perspectiva ha sido sometida. Según Gomila, en lo que respecta a la contraposición entre las perspectivas de primera y tercera persona "debería resultar sorprendente que no se plantee siquiera la posibilidad de que esta contraposición no sea en realidad exclusiva y excluyente, y de que sea posible, por tanto, distinguir una perspectiva diferente, genuina, acerca de la mente" (2001, pp. 65–66). La explicación de Gomila de la perspectiva de segunda persona coincide en muchos aspectos con la comunicación expresiva del enfoque de segunda persona defendido en este trabajo: al igual que para el enfoque de segunda persona, para Gomila la perspectiva de segunda persona es la que aparece primero en el desarrollo ontogenético, lo cual tiene sentido desde el punto de vista evolutivo puesto que "dado el largo periodo de dependencia infantil, cuánto antes pueda acceder el bebé al mundo mental, para orientarse en él, mejor" (2002, p. 134). Asimismo, para Gomila, la perspectiva de segunda persona nos ofrece la capacidad de involucrarnos en patrones de interacción intersubjetivos con otros agentes: los estados mentales que captamos a través de la perspectiva de segunda persona "son constitutivamente corporales o bien, a la inversa, las actitudes y configuraciones corporales son también mentales" (2002, p. 134), es decir, son estados expresivos perceptibles de manera directa, "vemos" que alguien está enfadado, alegre, eufórico o triste, no movimientos musculares que interpretemos por analogía, hipótesis o inducción.

Como acabo de señalar, estas afirmaciones coinciden con el enfoque de segunda persona defendido en este trabajo, sin embargo y a diferencia de este, Gomila sostiene una perspectiva de segunda persona basada en la mutua *atribución* de estados mentales. Según Gomila, "mi atribución a otro con quien estoy en relación depende de mi darme cuenta de que él también está atribuyéndome estados mentales en la medida en que esas atribuciones median nuestra interacción" (2002, p. 125). Gomila define este tipo de atribución como una atribución no proposicional, una "atribución implícita de estados mentales" (2001, p. 70; 2002, p. 134) y sostiene que este tipo de atribución está presente ya desde los primeros meses de vida:

> En primer lugar, podemos encontrar otras formas de interacción comunicativa pre-intencional cuya posibilidad implica igualmente la atribución implícita de estados mentales, concebidos de nuevo en términos expresivos, públicos. Esa serie de patrones de interacción se incluyen bajo el rótulo de "referencia social" y abarcan una serie de capacidades que aparecen en torno al año de vida: *la atención visual conjunta*, los proto-declarativos y proto-imperativos mediante el gesto de apuntar, la aceptación de instrucciones o la petición de ayuda (2002, p. 134, cursivas añadidas).

Según Gomila (Gomila & Pérez, 2017), el patrón de interacción que hace posible la atención visual conjunta, involucra un tipo de atribución intencional que no tiene contenido proposicional, "probablemente del tipo 'ella mira eso, yo también'" (p. 291). Gomila (2001) se basa en este punto en el esquema griceano del significado natural, según el cual la tercera condición para el mismo es que el emisor tenga la intención de que el receptor se dé cuenta de que el emisor pretende causar el efecto por medio de la emisión. De esta manera, aun cuando Gomila desvincula el esquema griceano del medio lingüístico y de su formulación en términos de proposiciones, la inclusión de este tipo de atribución implícita hace que su comprensión de la intención comunicativa comparta el mismo esquema de Tomasello que vimos en el primer bloque, a saber, "Tú pretendes que [yo comparta la atención con (X)]" (Tomasello, 1999, p. 102). La explicación de Gomila, por tanto, es susceptible al mismo tipo de crítica señalada por Roessler, según la cual, "interpretar la atención conjunta proto-declarativa como una cuestión de expresar, reconocer y actuar sobre intenciones comunicativas es mentalizar en exceso el fenómeno" (2005, p. 242). De esta manera, Gomila, al igual que Tomasello, comparten la suposición de la existencia de una brecha entre el reconocimiento de que el otro pretende que el bebé comparta la atención con X y la formación de la intención del este de atender a X. Asimismo, aunque el tipo de atribución que defiende Gomila no requiera de elementos proposicionales, la afirmación de que en los casos de atención visual conjunta el bebé atribuye la intención al otro de que este pretende que siga la mirada, conlleva que el bebé "decide, a la luz de sus propios deseos y planes, si aceptar la invitación de compartir la atención" (Roessler, 2005, p. 240), es decir, al igual que, según Gomila, el patrón de interacción que hace posible la atención conjunta involucra un tipo de atribución intencional del tipo "ella mira eso, yo también", puede, asimismo, involucrar un tipo de atribución intencional del tipo, "ella mira eso, yo no", según decida el bebé en relación a sus deseos y planes. Como señalé en el primer bloque, esto

implica considerar al bebé dotado de habilidades racionales que superan con creces las de un bebé de 9 meses.[52]

Según el enfoque de segunda persona que defiendo en este trabajo, "ver" un estado mental no implica necesariamente "atribuir" dicho estado mental, ni siquiera de manera implícita. El bebé no necesita atribuir un estado mental que ya está viendo y, por tanto, captando (o "conociendo"). Más que la idea de una mutua atribución de estados mentales, que como hemos visto resulta problemática, lo que caracteriza la perspectiva de segunda persona es el mutuo intercambio de estados mentales mediante la comunicación expresiva, es decir, más que una perspectiva de segunda persona de la *atribución* mental, lo que defiendo en este trabajo podría denominarse una perspectiva de segunda persona de la *interacción* mental.

Retomando la pregunta central, ¿cómo adquirimos la FP durante la infancia?, he señalado que la introducción al lenguaje de la FP comienza, no por la observación del otro o por la simulación mental de su situación, sino con la sustitución (o acompañamiento) de expresiones naturales por expresiones lingüísticas. En otras palabras, nuestra introducción a la FP no se realiza mediante las perspectivas de primera o de tercera persona sino mediante la perspectiva de segunda persona. La niña no ve a la madre como alguien a quien comprender sino como alguien con quien interacciona y *regula* sus emociones. En esta etapa del desarrollo, mediante la autoridad de segunda persona de la madre, la niña aprende a expresar de otra manera (en este caso lingüística) las expresiones naturales con las que se comunica con ella. He afirmado, asimismo, que este esquema de aprendizaje no puede extenderse a las actitudes proposicionales como las creencias, ya que, al contrario que las emociones básicas y las sensaciones con las que la niña se comunica con su madre, las primeras requieren de elementos conceptuales para su comprensión. La pregunta que se plantea ahora es: ¿*cómo* y *para qué* adquirimos en nuestra infancia las actitudes proposicionales?

La respuesta a esta pregunta está relacionada con la normatividad que, según expuse en el primer bloque, comenzamos a desarrollar a partir de los 12-14 meses, mediante los triángulos de atención conjunta. Como he sostenido en ese bloque, el bebé intercambia y regula sus estados mentales (sus emociones y sensaciones) en la

[52] A la hora de explicar este tipo de atribución implícita, no proposicional, y teniendo en cuenta que una atribución de un estado mental conlleva una posesión mínima, parcial, implícita, del concepto de dicho estado mental (Pérez & Gomila, 2018, p. 94), Gomila sostiene la existencia de conceptos psicológicos básicos que no requieren de habilidades lingüísticas ni, por tanto, de contenido proposicional, y que, según Gomila, el bebé adquiere mediante interacciones interpersonales. Para un desarrollo más extenso de esta idea véase Pérez y Gomila (2018).

interacciones en segunda persona con la madre mediante la comunicación expresiva. Más adelante, cuando los triángulos de atención conjunta entran en juego, el bebé comienza también a regular sus propias impresiones acerca de los objetos externos mediante la autoridad de segunda persona de la madre y a incorporar las respuestas expresivas de esta frente a las situaciones que desconoce (Rochat, 2001; Roessler, 2005). Tras ello, y a través de la internalización de la madre, la autoconciencia social del bebé comienza a desarrollarse, así como su sentido de independencia (Taipale, 2016). En esta etapa del desarrollo, el bebé comienza a entender que, además de la suya, no existe solo la perspectiva de la madre sino también perspectivas alternativas acerca de lo apropiado y lo inapropiado de sus expresiones emocionales (Eilan, 2005). El desarrollo de una conciencia básica de la normatividad en el bebé va más allá de la autoridad de la madre, convirtiéndose en una conciencia de las normas sociales. Este ingreso al mundo social es lo que hace que desarrolle una conciencia básica de la normatividad "social", pasando de la pregunta acerca de cómo debería sentirse frente a una situación determinada, según la respuesta emocional *de la madre,* a la pregunta acerca de cómo esperan *los demás* que se sienta ante dicha situación (Taipale, 2016). La regulación emocional mutua se traduce en patrones de corrección/sanción en los que las actitudes reactivas juegan un papel primordial y son la entrada a la normatividad social, a la regulación con un mundo que está regido por *normas sociales*. El desarrollo posterior de esta conciencia social y la comprensión de estas normas sociales que rigen las interacciones y el uso de las (auto-)adscripciones de estados mentales por parte de la —ya a esta altura del desarrollo— niña, llegará con su introducción en el lenguaje, el cual posibilitará que adquiera la habilidad de usar y entender el repertorio conceptual necesario, es decir, las actitudes proposicionales como las creencias y los deseos que necesitará para dar razones y justificaciones tanto de su comportamiento como del comportamiento de los demás. A partir de este momento, la niña comenzará a regularse en base a las normas sociales y mediante sus interacciones con los demás miembros de la comunidad.

Esta idea de la FP como práctica normativo-regulativa, está relacionada tanto con las ideas de Strawson y Bilgrami que vimos en el bloque anterior, como con las de una serie de autores que sostienen que la comprensión la FP no se basa en descripciones de estados mentales internos, sino en estructuras normativas socialmente compartidas que generan expectativas sobre cómo deben comportarse los agentes (Andrews, 2009, 2012, 2015; Gallagher, 2001; Gallagher & Hutto, 2008; Hutto, 2004, 2008; McGeer,

2001, 2007, 2015; Zawidzki, 2013). Para estos autores, las atribuciones de estados mentales no son necesarias para anticipar la conducta en nuestras interacciones. En palabras de Andrews (2015), "en lugar de confiar en estados mentales ocultos para cerrar la brecha entre la misma circunstancia y un comportamiento diferente, los psicólogos del sentido común (*folk psychologist*) pueden confiar en su conocimiento sobre las normas de comportamiento social" (p. 52). Estos autores, al igual que el enfoque de segunda persona que defiendo en este trabajo y a diferencia de Dennett (1989), quien sostiene la denominada "postura intencional" (*intentional stance*) según la cual la FP es una estrategia para *predecir y explicar* el comportamiento mediante las creencias y deseos que un agente debe tener de acuerdo a las *normas de racionalidad*[53], sostienen que la FP tiene una función *regulativa* que consiste en aprender, enseñar y exhortar a otros a comportarse de acuerdo con las *normas sociales compartidas* que rigen nuestras interacciones. Estas normas sociales compartidas son las que posibilitan nuestra comprensión del comportamiento y nuestra posibilidad de anticiparlo, dado que proveen de estructuras normativas que indican los comportamientos a seguir dada una situación particular. Según Zawidzki (2013) para la adquisición y la internalización y el cumplimiento de las normas, no existe una necesidad aparente de atribución de actitud proposicional. Como señala McGeer (2001), el trabajo es realizado y llevado a cabo en nuestras interacciones, las cuales integran las normas y rutinas que estructuran dichas interacciones (p. 119). De esta manera, el enfoque regulativo rechaza el individualismo de los enfoques tradicionales abogando por una explicación de la FP en términos de relaciones interpersonales:

> The standard view conceptualizes folk-psychology in primarily epistemic terms, as an individually realized method or mechanism for explaining and predicting behaviour by

[53] [T]here is yet another stance or strategy one can adopt: the intentional stance. Here is how it works: first you decide to treat the object whose behavior is to be predicted as a rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and its purpose. Then you figure out what desires it ought to have, on the same considerations, and finally you predict that this rational agent will act to further its goals in the light of its beliefs. A little practical reasoning from the chosen set of beliefs and desires will in many—but not all—instances yield a decision about what the agent ought to do; that is what you predict the agent *will* do (Dennett, 1989, p. 17). Una de las principales críticas a la postura intencional de Dennett radica en esta restricción a las normas racionales. Al igual que el "racionalismo" de Moran, el enfoque de Dennett no tiene en cuenta que los seres humanos no somos ciudadanos epistémicos modelo y mostramos, frecuentemente, fallos de racionalidad. Según Stich (1981) un sistema intencional "es un sistema idealmente racional; cree, quiere y hace exactamente lo que debe, como lo estipula una teoría normativa de la racionalidad. Las personas, por el contrario, no son idealmente racionales, y ahí reside un problema devastador para Dennett" (p. 79), porque si la postura intencional fuese correcta, no podríamos predecir comportamientos irracionales, cuando, de hecho, sí podemos hacerlo. Sin embargo, como señala Fernández Castro (2017a), "[n]uestras estrategias de anticipación están basadas en estructuras normativas relacionadas con personalidades, estatus sociales, reglas sociales, etc. Podemos predecir que el agente se comporte de acuerdo con esos estándares y emitir juicios basados en ellos. Por lo tanto, un curso de conducta que viola las normas racionales no es un contra-ejemplo para la visión regulativa. De hecho, ciertos comportamientos pro-sociales pueden considerarse irracionales" (p. 15).

> way of detecting the underlying mental causes of that behaviour. By contrast, the regulative view conceptualizes folk-psychology in fundamentally interpersonal terms, as a shared "mind-making" practice. It presents folk-psychological competence as primarily involving a capacity to form and regulate our minds in accordance with a rich array of socially shared, and socially maintained, sense-making norms – a capacity that delivers real epistemic benefits, at least with agents that are likewise norm-conforming (McGeer, 2015, p. 267).

Esta tesis del enfoque regulativo que afirma que en nuestras interacciones interpersonales nos regulamos unos a otros de acuerdo a normas sociales compartidas, está relacionada, como señalé anteriormente, con la afirmación de Strawson de que nuestras actitudes reactivas son lo que subyace, lo que justifica nuestras prácticas normativas relacionadas con la responsabilidad que conlleva nuestro estatus de agentes libres y con la extensión de Bilgrami a la agencia humana como normativa en sí misma. Como sugieren numerosos estudios, a partir aproximadamente de los dos años, los niños son capaces de comprender la naturaleza convencional y normativa de algunas prácticas sociales, principalmente mediante el juego de ficción (de simulación/pretensión) y otros juegos gobernados por normas. Los niños no solo aprenden a jugar de acuerdo a las reglas, sino que también comprenden la estructura normativa de dichos juegos, como muestra su aplicación a terceros mediante objeciones, intervenciones o correcciones cuando estos violan dichas reglas (Rakoczy, Brosche, Warneken, & Tomasello, 2009; Rakoczy, Warneken, & Tomasello, 2008). Asimismo, otros estudios muestran que los niños de todas las culturas reaccionan a comportamientos contra-normativos con emociones punitivas como la ira (Edwards, 1987; Sripada & Stich, 2006). Los seres humanos, por tanto, estamos dispuestos por naturaleza a experimentar actitudes reactivas como la aprobación o la gratitud cuando otros cumplen lo que las normas requieren de ellos, o como el resentimiento y la indignación cuando las incumplen. Del mismo modo, experimentamos actitudes auto-reactivas cuando otros nos las muestran, como la vergüenza, el remordimiento, el orgullo. Sin embargo, ante, por ejemplo, el resentimiento o la indignación de los demás, no solo nos sentimos inclinados a sentirnos mal, sino que ofrecemos excusas, justificaciones, explicaciones y, llegado el caso en el que nuestra acción ha sido injustificada, disculpas y una disposición a comportarnos de acuerdo a lo que se espera de alguien que comparte dichas normas. De esta manera, como señala McGeer, "nuestra sensibilidad a las actitudes reactivas muestra que tenemos la disposición a ser regulados por otros para producir pensamientos y acciones que cumplan con las normas" (2015, p. 274).

Llegados a este punto estamos en condiciones de responder a una de las preguntas anteriormente suscitadas: *¿para qué* aprendemos en nuestra infancia el uso de las actitudes proposicionales?, en otras palabras, ¿cuál es el objetivo principal de la adquisición de las actitudes proposicionales en nuestra infancia? Siguiendo el análisis realizado hasta el momento, en el que, de una parte, he afirmado que las (auto)-atribuciones de creencias modifican las circunstancias en las que se ha de evaluar el *compromiso* adquirido con la verdad de las proposiciones bajo su alcance (ver Bloque II, apartado 5) y, de otra, que mediante las (auto-)atribuciones de creencias y deseos justificamos nuestra conducta ante las actitudes reactivas de los otros, podemos afirmar que la principal función de las (auto)-atribuciones de creencia y deseo es la de (auto)-asignar ciertas responsabilidades con el fin de facilitar la exculpación (o censura) de una acción particular. Al respecto, Fernández Castro afirma:

> I maintain that propositional attitude ascriptions are not tools for describing or theorizing about others' psychological states. Instead, they burden agents with merit or responsibilities toward a particular content for the purpose of rationalizing, exculpating, justifying or condemning certain actions. Propositional attitude ascriptions help to arbitrate what is permitted or forbidden in our social situations (Fernández Castro, 2017b, p. 2).

La función principal de la atribución de las actitudes proposicionales de creencia y deseo, por tanto, es una función evaluativa, es decir, justificadora, exculpatoria, condenatoria, de elucidación del comportamiento contra-normativo y de los compromisos adquiridos (Andrews, 2009; Fernández Castro, 2017a; Hutto, 2004; McGeer, 2015; Zawidzki, 2013). Esta afirmación está relacionada con lo defendido en el bloque anterior respecto a los compromisos que adoptamos como agentes responsables cuando nos auto-atribuimos estados mentales. Ser un agente responsable también implica ser un agente que es capaz de ajustar los estados mentales que se auto-adscribe con sus acciones y sus acciones con los estados mentales auto-adscritos. El compromiso, por tanto, puede ser entendido como el compromiso a actuar de acuerdo a lo que decimos de nosotros mismos, es decir, como el compromiso a actuar según se espera que actué alguien que se auto-adscribe determinado estado mental. Asimismo, esta función evaluativa no se restringe a las auto-atribuciones sino también a las auto-atribuciones de estados mentales de los demás. Como he afirmado, cuando la niña aprende a auto-atribuirse un estado mental, aprende, asimismo, a reconocerlo en la auto-atribución de su madre. La niña sabe que "Quiero el osito" dicho por ella expresa su deseo de alcanzar el osito al igual que sabe que esta misma expresión dicha por su

madre expresa el deseo de la madre de alcanzar el osito. El caso de las actitudes proposicionales de creencia no es diferente. Cuando la niña aprende que "creo que va a llover" rebaja su compromiso con la verdad de que va a llover y, por tanto, ofrece una justificación, una razón, para prevenir una futura sanción en caso de que no llueva, aprende, asimismo, a entender esa misma auto-adscripción dicha por su madre de la misma manera. Este hecho, he afirmado, resulta fundamental para entender el rasgo relacional de la adquisición de las expresiones lingüísticas por parte de la niña así como de las normas sociales que regulan el comportamiento de los miembros de la comunidad, puesto que implica la comprensión por parte de la niña de que el lenguaje, al igual que la comunicación expresiva, es una habilidad social compartida con los demás miembros de la comunidad y que las razones y, por tanto, las actitudes proposicionales pueden ser, asimismo atribuidas a los demás para justificar su comportamiento.

Ahora bien, ¿*cómo* aprende la niña el uso de las actitudes proposicionales? Según el enfoque de segunda persona, el aprendizaje de las actitudes proposicionales no se realiza principalmente mediante la observación del comportamiento de los demás ni mediante la simulación de su situación, sino mediante las interacciones con los demás miembros de la comunidad. En el caso de las actitudes proposicionales, a diferencia de los estados mentales con los que la niña se comunica con su madre mediante la comunicación expresiva y que aprende a sustituir (o acompañar) por expresiones lingüísticas, la comprensión de las primeras no puede realizarse de la misma manera que la de estas últimas puesto que no forman parte del repertorio de estados mentales (expresiones) a sustituir y requieren elementos conceptuales para su comprensión. Como he sostenido siguiendo el análisis del expresivismo mínimo de Frápolli y Villanueva (2012), las creencias son predicados de segundo orden cuya función principal es la de modificar las condiciones de *evaluación* de las proposiciones bajo su alcance. En este caso, la niña tendrá que comprender que hay determinados términos auto-adscriptivos que no sustituyen las expresiones naturales con las que se comunica con su madre sino que sirven, entre otras cosas, para dar razones de nuestra conducta y de la conducta de los demás y que son necesarios para regularnos de acuerdo a las normas sociales compartidas. De esta manera, la niña tendrá que comprender los distintos usos de la auto-adscripción "Creo que *p*", siendo uno de ellos subrayar su compromiso con la verdad de que *p*, lo cual conlleva comportarse como alguien que cree que *p,* y otro rebajar dicho compromiso, lo cual implica su duda al respecto y la

posibilidad de un comportamiento distinto respecto a *p*. Asimismo, tendrá que comprender que ese tipo de adscripciones sirven para justificar, razonar, exculpar, condenar determinados comportamientos contra-normativos. ¿Qué circunstancias, por tanto, se requieren para que esto ocurra?

Como he afirmado, según el enfoque de segunda persona que defiendo en este trabajo, la base, el fundamento de este aprendizaje no reside en la observación o en la simulación de los demás, sino en las interacciones de los niños con sus cuidadores, es decir, en las interacciones en segunda persona. Es mediante este tipo de interacciones por las que los niños llegan a comprender y saber usar las actitudes proposicionales. Sin embargo, no todas las interacciones en segunda persona sirven a este propósito. Hutto (2008), y autores como Dunn y Brophy (2005), sostienen que las interacciones en segunda persona que resultan fundamentales para adquirir las actitudes proposicionales que están a la base de la FP, son aquellas en las que se lleva a cabo una comunicación oral a través de narrativas y de historias (cuentos). Hutto (2008) afirma que "debemos reconsiderar las opiniones recibidas sobre el tipo de contexto en el que normalmente opera la FP, tomando en serio la idea de que nuestro punto de partida es la segunda persona" (p. 6). Para Hutto, desprendernos de la idea de que damos sentido a los demás desde la observación de su conducta, es decir, desde la perspectiva de tercera persona, posibilita replantear y reorientar nuestro pensamiento sobre la naturaleza de nuestras expectativas sociales cotidianas y sobre cómo aprendemos a dar razones psicológicas de la conducta de las personas. La explicación y la predicción de las acciones de los demás desde una perspectiva de tercera persona, no solo es ontogenéticamente posterior, sino que es menos fiable y frecuente que nuestros medios intersubjetivos para llegar a comprender a los demás. Hutto sostiene la que denomina "Hipótesis de la Práctica Narrativa", según la cual para que los niños adquieran la capacidad para comprender las acciones intencionales en términos de razones y dominen su aplicación práctica, deben estar expuestos y participar en un tipo particular de práctica narrativa. Según Hutto, para que los niños se familiaricen con la estructura básica de la FP y las posibilidades para ejercerla en la práctica según las normas, aprendiendo así cómo y cuándo usarla, deben estar expuestos a encuentros *directos* con historias sobre personas que actúan por *razones*, a narrativas proporcionadas en *contextos interactivos* por cuidadores responsables, de manera que los niños tengan la oportunidad de *participar* en las narrativas de manera correcta. Para Hutto, en el transcurso de las narrativas los niños se implican en los eventos que se les describen y se mueven de manera emotiva como lo

harían en sus compromisos interpersonales más básicos. Por ello, las narrativas no deben ser entendidas como un asunto pasivo, ya que conllevan una amplia gama de habilidades emotivas e interactivas que ya se encuentran en la infancia, en la intersubjetividad primaria, siendo, por tanto, extensiones naturales de las experiencias tempranas de los niños (2008, pp. 183–186).

Dunn y Brophy destacan, asimismo, la importancia de las interacciones conversacionales de segunda persona en este proceso. Según estos autores, "necesitamos observar la capacidad lingüística de los niños no solo en términos de una habilidad cognitiva o característica individual, sino también en términos de sus experiencias diádicas" (2005, p. 63). Estos autores sostienen que la comprensión de los estados cognitivos surge a través de una comprensión previa de las emociones —lo cual indica que la comunicación expresiva continúa en nuestras interacciones lingüísticas y es parte de lo que posibilita este aprendizaje. Según Dunn y Brophy, en las primeras etapas del desarrollo de la comprensión social, las experiencias emocionales desempeñan un papel de propiciadores del discurso sobre los estados mentales (2005, pp. 64-65). En esta misma línea, los estudios de Bartsch y Wellman (1995) muestran que los niños explican primero las acciones de las personas en términos de emociones y deseos y, más tarde, a través de sus experiencias sociales, incorporan la noción de creencia en su comprensión de por qué las personas se comportan como lo hacen. Para estos autores, no existe una sola transición en los niños de tres a cuatro años en la comprensión de los estados mentales, es decir, no hay solo un cambio de antes a después de una comprensión de la creencia falsa, sino una progresión desde una psicología de la emoción y el deseo a una psicología de la creencia y el deseo (1995, pp. 43-44).

Desde el punto de vista del enfoque de segunda persona, estos estudios muestran la coherencia del análisis sobre la introducción al lenguaje de las (auto-) atribuciones de estados mentales: dado que, como he afirmado, los niños aprenden en primer lugar a sustituir las expresiones naturales de los estados mentales que "conocen" de manera no conceptual, es decir, los deseos y las emociones con las que se comunican con su madre, por expresiones lingüísticas, parece razonable sostener que es mediante estos estados mentales por los que los niños comienzan a explicar las acciones de los demás y que, más tarde, mediante las interacciones en segunda persona objeto de este análisis, introduzcan las actitudes proposicionales de creencia en sus explicaciones.

De otra parte, además de la necesidad de estar expuestos a —y así poder participar en— narrativas proporcionadas en contextos interactivos por cuidadores responsables, necesitamos, asimismo, otro tipo de interacciones en segunda persona. Como afirmé anteriormente, numerosos estudios sugieren que, a partir aproximadamente de los dos años, los niños son capaces de comprender la naturaleza convencional y normativa de algunas prácticas sociales, principalmente mediante el juego de ficción (de simulación/pretensión) y otros juegos gobernados por normas en los que se llevan a cabo actividades cooperativas (Huyder, Nilsen, & Bacso, 2017). En concreto, Fernández Castro (2019) sostiene que los niños deben participar en encuentros sociales en los que se produzcan desviaciones relevantes del comportamiento, de tal manera que sus acciones desencadenen respuestas reactivas o sancionadoras que los participantes desean evitar. Fernández Castro defiende la que denomina "Hipótesis Conversacional Justificativa", según la cual los niños adquieren la capacidad de comprender las acciones de los demás en términos mentales, es decir, de dar razones mediante la atribución de actitudes proposicionales, por medio de la exposición y participación en diferentes situaciones de conversación. Según Fernández Castro, los niños necesitan participar en actividades conjuntas en las que intervienen respuestas regulativas como, por ejemplo, los juegos de simulación cooperativos, las acciones cooperativas, disputas y desacuerdos, etc. Un gran número de experimentos en psicología del desarrollo muestran la plausibilidad de esta hipótesis (Brown, Donelan-McCall, & Dunn, 1996; Dunn & Brown, 1993; Foote & Holmes-Lonergan, 2003; Slomkowski & Dunn, 1992; Tesla & Dunn, 1992). Aún cuando, como señala Fernández Castro, no son evidencia concluyente, los resultados de estos experimentos indican una correlación entre este tipo de contextos y la capacidad de atribuir actitudes proposicionales. Asimismo, otros estudios sugieren que cuando el diseño experimental del FBT no busca probar las capacidades predictivas sino las explicativas o las justificativas los niños son capaces de pasarlo antes de los cuatro años (Bartsch & Wellman, 1989; Robinson & Mitchell, 1995). Como apunta Fernández Castro (2019), los resultados de estos experimentos refuerzan la idea de que "la adquisición del conceptos mentales está vinculada a la racionalización y normalización de las acciones contra-normativas, en lugar de las capacidades anticipatorias" (p. 22), es decir, que la función primaria de las actitudes proposicionales debe relacionarse, como afirmé anteriormente, con fines justificativos o explicativos, lo cual no significa que la atribución de estado mental no pueda servir para propósitos predictivos.

En este apartado he sostenido que el modelo adecuado para describir la adquisición de la FP no es el observacional ni el de la simulación sino el modelo de *comunicación reguladora mutua.* Los niños pasan de la regulación emocional mutua a una regulación con un mundo que está regido por *normas sociales* (lo cual no implica que dejen de regularse emocionalmente en sus interacciones en segunda persona). He rechazado la suposición mentalista desde la que parten los enfoques tradicionales, TT y TS, abogando por una introducción a la FP basada en la comunicación expresiva defendida en los bloques anteriores. He sostenido que nuestra comprensión de la FP no está basada en descripciones de estados mentales internos, sino en estructuras normativas socialmente compartidas que generan expectativas sobre cómo deben comportarse los agentes. La FP es entendida, por tanto, no como una estrategia para predecir y explicar el comportamiento, sino como una práctica normativo-regulativa. Los niños aprenden e interiorizan las normas sociales que rigen nuestras interacciones, nuestro comportamiento, así como las actitudes proposicionales necesarias para entender, justificar, sancionar comportamientos contra-normativos que despiertan actitudes reactivas en ellos, y que son las que nos mueven tanto a auto-regularnos como a regular a los demás y ser regulados por ellos. De esta manera, la respuesta a las preguntas acerca de *cómo* y *para qué* aprendemos el uso de las actitudes proposicionales que están a la base de la FP se haya, por un lado, en las interacciones en segunda persona y, por otro, en nuestra disposición natural a experimentar actitudes reactivas y a regularnos con los demás. Basándome en diferentes estudios y enfoques relacionados con la segunda persona, he sostenido que la repuesta al *para qué*, es decir, la función principal de las actitudes proposicionales de creencia y deseo, es una función evaluativa, es decir, justificadora, exculpatoria, condenatoria, de elucidación del comportamiento contra-normativo y de los compromisos adquiridos. En cuanto al *cómo* aprendemos su uso, he afirmado que es mediante determinados tipos de interacciones en segunda persona, en concreto, mediante interacciones de segunda persona en las que, de una parte, los niños están expuestos a encuentros directos con historias sobre personas que actúan por *razones*, a narrativas proporcionadas en contextos interactivos por cuidadores responsables, de manera que pueden participar en las narrativas de manera correcta y, de otra, a contextos en los que participan en actividades conjuntas en las que intervienen *respuestas regulativas* como, por ejemplo, los juegos de simulación cooperativos, las disputas y los desacuerdos.

En el siguiente apartado, presentaré la explicación de la transparencia, la autoridad de primera persona y el autoconocimiento que ofrece el enfoque de segunda persona que defiendo en este trabajo.

4. Segunda persona y autoconocimiento

En el análisis llevado a cabo en el bloque anterior acerca de los enfoques sobre el autoconocimiento, he sostenido que todos estos enfoques basan su explicación del autoconocimiento en la dicotomía entre primera y tercera persona, lo cual descarta las interacciones personales como relevantes para la explicación del autoconocimiento y sitúan al individuo en el centro de la explicación. Asimismo, aun cuando algunos de ellos incluyen las interacciones en su enfoque, lo hacen desde una perspectiva teórica que obvia los factores sociales. He argumentado que tanto situar al individuo en el centro de la explicación, como descartar las interacciones en segunda persona resulta problemático. En el caso de Moran conduce a una idea del ser humano excesivamente racional, algo muy alejado de lo que, en realidad, somos los seres humanos y cuya actitud deliberativa obvia el reconocimiento por parte del interlocutor. En el caso de Wright y de Bilgrami, resulta en una concepción idealizada del autoconocimiento en el que los avowals son realizados en un "entorno esterilizado" que obvia factores sociales tales como la injusticia epistémica. En el caso de Bar-On y Finkelstein, conlleva la omisión del aspecto relacional del aprendizaje de las expresiones lingüísticas.

En lo que sigue, presentaré una explicación del autoconocimiento basada en el enfoque de segunda persona que defiendo en este trabajo. Para ello, comenzaré con las características especiales distintivas de transparencia y autoridad analizadas por estos autores.

4.1 Transparencia y segunda persona

Una de las rasgos en el que los enfoques acerca del autoconocimiento coinciden es en caracterizar la transparencia de los estados mentales como una forma de conocimiento directo y no inferencial[54]. Sin embargo, a la hora de explicar esta noción, estos enfoques difieren de manera significativa. De una parte, los introspeccionistas sostienen una

[54] Con algunas excepciones como, por ejemplo, Byrne (2005, 2011), como he señalado en la nota a pie 17.

transparencia interna hacia los estados mentales, ya que, para estos autores los estados mentales son entidades internas accesibles mediante un mecanismo de introspección. El agencialismo de Moran, de otra parte, da un "giro" de lo interno a lo externo modificando el sentido en el que la transparencia es entendida, pasando de la transparencia interna hacia los estados mentales a la "transparencia hacia el mundo", es decir, a la idea de que la creencia acerca de un contenido es transparente a la afirmación de dicho contenido. Para los constitutivistas, la transparencia es una característica a priori de la agencia responsable, en el caso de Bilgrami, o de las reglas que gobiernan el uso de las (auto-)adscripciones de estados mentales, en el caso de Wright. Finalmente, Bar-On introduce la noción de "transparencia-a-la-condición-mental-del-sujeto", según la cual los estados mentales que comunican nuestras expresiones naturales son transparentes tanto para nosotros como para la persona con la que estamos interaccionando.

Según el enfoque de segunda persona que sostengo en este trabajo, la trasparencia a los estados mentales no tiene un sentido unívoco puesto que, como he afirmado, no todos los estados mentales son del mismo tipo. De una parte, coincido con Bar-On en suscribir el análisis expresivista wittgensteiniano acerca de las expresiones naturales según el cual estas resultan transparentes, es decir, según el cual las expresiones naturales son captadas de manera directa y no inferencial: "Vemos la emoción — ¿En oposición a qué? — No vemos las contorsiones del rostro y hacemos inferencias de ellas (como el doctor que da un diagnóstico) a la alegría, aflicción o aburrimiento" (Wittgenstein, 1981, §225). Según la concepción de la expresión defendida en este trabajo, las expresiones son las condiciones comunicativas de los estados mentales. Las lágrimas no son un signo o síntoma de tristeza, sino una condición de esta. No hay una brecha ontológica entre los estados mentales y las expresiones. Los gestos faciales, los movimientos corporales, el tono de la voz, son la condición comunicativa de los estados mentales. Durante nuestras interacciones en segunda persona, vemos, oímos, sentimos los estados mentales de los demás. En este punto, como he señalado anteriormente, coincido con el análisis de Gomila (2002) en la afirmación de que los estados mentales que captamos a través de la perspectiva de segunda persona "son constitutivamente corporales o bien, a la inversa, las actitudes y configuraciones corporales son también mentales" (p.134).

De otra parte, sostengo que las actitudes proposicionales, entendidas como predicados de segundo orden cuya función es la de modificar las condiciones de

evaluación de las proposiciones bajo su alcance (Frápolli y Villanueva, 2012), son necesariamente transparentes. No puedo no conocer lo que estoy haciendo cuando hago correctamente un uso pragmático, evaluativo, de la auto-adscripción de creencia porque el mismo hecho de poder hacer un uso evaluativo de manera correcta conlleva conocer las relaciones del predicado de segundo orden con las proposiciones bajo su alcance. En otras palabras, usar de manera correcta una auto-adscripción de una actitud proposicional de creencia conlleva conocer lo que hago al usarla, es decir, conocer el grado de compromiso derivado del tipo de uso que estoy haciendo de ella. Esta afirmación está relacionada tanto con el análisis de Bilgrami como con la concepción regulativa de la FP. Para que consideremos al otro como un agente libre, responsable, como un usuario competente en el uso de la FP, hemos de considerar como una condición previa y necesaria que *conoce* tanto sus practicas relacionadas con la responsabilidad como las actitudes reactivas que las sustentan y que es, asimismo, capaz de justificarlas (Bilgrami, 2006; McGeer 2015). Para ello, es necesario que el agente conozca los distintos usos de las actitudes proposicionales, que sepa, como he señalado, que puede tanto subrayar su compromiso con la verdad de que *p*, lo cual conlleva comportarse como alguien que cree que *p*, como rebajar dicho compromiso, lo cual implica su duda al respecto y la posibilidad de un comportamiento distinto respecto a *p*.

Sin embargo, coincido con Bilgrami en que las actitudes proposicionales entendidas como disposiciones, es decir, como una disposición a realizar una acción, comportamiento, etc., pueden no ser transparentes para el agente, lo cual explica, en parte, determinados comportamientos contradictorios y permite dar cuenta de los casos de autoengaño, como vimos en el bloque anterior.

4.2 Autoridad y segunda persona

Al igual que con la transparencia, la noción de autoridad de primera persona, la presunción de corrección de las auto-atribuciones de estados mentales en primera persona, ha sido entendida de diferentes maneras por los diversos enfoques acerca del autoconocimiento. En el caso de los introspeccionistas y en concordancia con su definición de los estados mentales como objetos internos accesibles mediante la introspección, está relacionada con el *acceso epistémico privilegiado* a estos, puesto que, para estos autores, solo uno mismo tiene el privilegio de poder de acceder a sus propios estados mentales. Los agencialistas, en cambio, sostienen que la autoridad de

primera persona no es una cuestión de acceso privilegiado a objetos internos, sino que está relacionada con la acción de un *agente*, es decir, los estados mentales no son algo que, meramente, observamos sino algo que *hacemos* en tanto que seres racionales y conceptualmente competentes. Por último, para los expresivistas la autoridad de primera persona proviene de la suma de nuestra capacidad para expresar nuestros estados mentales junto con nuestra capacidad para hacerlo de manera lingüística.

Al respecto, he presentado en el bloque anterior varias críticas a estas explicaciones desde el punto de vista del enfoque de segunda persona. A continuación, volveré a algunas de esas críticas relacionándolas con la explicación que ofrece el enfoque de segunda persona acerca de la autoridad.

En el apartado 4.1 del bloque anterior, dedicado a Moran, he afirmado que su análisis de la deliberación refleja el presupuesto individualista del que parte al no tener en cuenta el reconocimiento por parte de nuestro interlocutor de las razones en cuanto razones. Respecto a ello, he sostenido que, al igual que el lenguaje, las razones no pueden ser privadas (en el sentido wittgensteiniano del término), es decir, que lo que cuenta como una razón para el agente deliberativo ha de contar, asimismo, como una razón para su interlocutor, aun cuando este no esté de acuerdo con su contenido. Esta afirmación está relacionada con lo sostenido tanto acerca de la adquisición por parte de los niños de las expresiones lingüísticas con las que sustituyen (o acompañan) las expresiones naturales mediante la autoridad de segunda persona de la madre, como con la adquisición de las actitudes proposicionales que son necesarias para dar razones y justificaciones de su conducta, cuando esta se desvíe de las normas sociales y/o provoque actitudes reactivas en su interlocutor. Como he afirmado anteriormente, según el enfoque de segunda persona, tanto la adquisición de las expresiones lingüísticas con las que los niños sustituyen sus expresiones naturales como el aprendizaje de las actitudes proposicionales tienen una dimensión o aspecto relacional, es decir, cuando son adquiridas o aprendidas por los niños lo son tanto para ellos mismos como para los demás miembros de la comunidad. De esta manera, las auto-atribuciones y las atribuciones de estados mentales a los demás son adquiridas o aprendidas simultáneamente. Del mismo modo que los niños aprenden a auto-atribuirse actitudes proposicionales para dar razones de su propia conducta, aprenden asimismo, a atribuir estados mentales a los demás para dar razones de la conducta de estos. Por tanto, del mismo modo que los niños aprenden a dar razones de su conducta aprenden, asimismo,

a dar razones de la conducta de los demás. Al respecto, Carla Bagnoli (2007) sostiene que la misma estructura de las razones y la justificación es de segunda persona:

> [T]he structure of justification is also second-personal. Reasons are considerations offered to another to justify a course of action or a mental state. My proposal is that we understand the second-personal structure according to a dialogical model. "I shall" is the conclusion of a dialogue that emerges from the recognition that I should account for my actions to others, and that I should demand justifications from them. I put myself under the rational scrutiny of others and demand the same. What I offer as a reason must count as a reason also for others, and what others offer as a reason must be something intelligible to me as a reason (p. 50).

Bagnoli se centra en los compromisos que adquirimos como miembros de una misma comunidad y en nuestra necesidad de reconocimiento mutuo. Para Bagnoli, tanto la comunidad, como sus miembros en tanto que agentes normativos y, por tanto, sus afirmaciones, están fundamentadas en el reconocimiento simultáneo y mutuo. Su propuesta está directamente relacionada con el modelo de *comunicación reguladora mutua* defendido por el enfoque de segunda persona. Como he afirmado, en el modelo que defiendo en este trabajo los bebés y los adultos son tratados como una sola unidad de estudio. En este modelo, al igual que en el de Bagnoli (aún cuando esta no alude ni explica el desarrollo ontogenético), el "yo" no es entendido de manera individual como opuesto a un "él" o "ella", sino como un "yo-tú" interactivo en el que no existe un desarrollo paralelo de la mente, sino un desarrollo basado en experiencias interactivas compartidas. Como he expuesto, esto conlleva una comprensión relacional del aprendizaje de la atribución de estados mentales y de las razones de las que forman parte en la que no existe una brecha entre las auto-atribuciones y las atribuciones de estados mentales a los demás. Por tanto, tampoco existe una brecha a la hora de comprender por qué los demás actúan como lo hacen, puesto que las razones que explican y/o justifican mi comportamiento son iguales a las razones que explican y/o justifican el comportamiento de los demás.

Al igual que con la explicación del aprendizaje de las (auto-)atribuciones de actitudes proposicionales y de las razones, la dimensión relacional de la que parte este modelo hace que su forma de entender la autoridad se distancie de manera significativa del modelo individualista. La diferencia fundamental entre la manera de abordar la cuestión de la autoridad bajo este modelo y bajo el modelo individualista radica en que, bajo este último, la autoridad es entendida como una característica exclusiva del sujeto —por el privilegio epistémico en el acceso a los estados mentales en el caso de los

introspeccionistas, por la capacidad expresiva en el caso de los expresivistas y por la condición de agente responsable en los agencialistas—, mientras que bajo el enfoque de segunda persona la autoridad es entendida de manera interpersonal, es decir, como algo que no depende únicamente del sujeto. Bajo este modelo, la autoridad de primera persona no es entendida como un logro individual, sino como un logro del que somos capaces en la medida en que nos relacionamos con los demás y que depende del reconocimiento. Como señala Borgoni (2018), los casos de injusticia epistémica descritos en el bloque anterior son prueba de que "la autoridad no está garantizada a menos que nuestros iguales las reconozcan en nuestras palabras […] Para ser considerado como una autoridad en un tema dado, uno necesita el reconocimiento de sus iguales de tal estado" (pp. 6–7). Estos casos muestran que nuestra autoridad de primera persona puede ser fundamentalmente socavada desde el exterior a través de prácticas de discriminación. Según Bagnoli, "nuestra autoridad de primera persona es vulnerable a los demás precisamente porque no son simples espectadores. Son nuestros interlocutores" (2007, p. 51). Obviar la necesidad de reconocimiento conduce al ideal racional agencialista, a la concepción idealizada del autoconocimiento, al 'entorno esterilizado' que obvia los factores sociales y que no puede dar cuenta de los casos en los que nuestra autoridad no es reconocida.

Una vez explicada la dimensión relacional de la autoridad de primera persona, veamos la explicación de la autoridad que ofrece el enfoque de segunda persona que defiendo en este trabajo.

Como he argumentado en el primer bloque y señalado anteriormente, el primer sentido de autoridad se da en las relaciones interpersonales entre el bebé y su madre. Desde las primeras semanas de vida, el bebé interacciona con su madre regulando sus emociones a través del intercambio de los estados mentales que comunican sus expresiones naturales. Tras la aparición de los triángulos de atención conjunta, una vez alcanzada la edad de entre doce y catorce meses y como resultado del vínculo psicológico creado entre la madre y el bebé en sus interacciones intersubjetivas a través de la comunicación expresiva y de la confianza emocional derivada del éxito en las prácticas regulativas de sus interacciones diádicas, el bebé comienza a buscar una referencia social en la madre, una valoración emocional de la situación a la que se enfrenta cuando esta es desconocida o cuando no tiene un valor asignado a la misma (Sorce et al., 1985). La respuesta emocional que la madre da al bebé a través de la comunicación expresiva transmite a este último una valoración de la situación que será

la que determine cómo debe afrontar, qué comportamiento ha de seguir, ante dicha situación (Brinck, 2008; Reddy, 2008). La madre es vista, por tanto, como una autoridad, como el referente a seguir, como la persona que sabe cómo debe sentirse y, en consecuencia, comportarse en tales circunstancias (Rochat, 2001; Roessler, 2005). La autoridad que aparece primero en el desarrollo ontogenético del bebé no es, por tanto, la autoridad de primera persona sino la autoridad de segunda persona que el bebé atribuye a su madre.

Ahora bien, ¿cómo alcanza el bebé la autoridad de primera persona?, ¿qué requisitos requiere para que se le reconozca dicha autoridad?

Como ha afirmado anteriormente, en esta etapa del desarrollo en la que surge la autoridad de segunda persona, las relaciones interpersonales asumen un significado normativo. El bebé asimila el valor de la información que obtiene de su madre no solo en sentido descriptivo sino también en sentido normativo, como es muestra de ello que el bebé, a partir de entonces, continúe respondiendo y actuando de la misma manera que aprendió de su madre cuando se enfrente a situaciones similares. Conforme el bebé va interiorizando la respuestas expresivas de su madre, va tornándose cada vez más independiente de su presencia, de la relación diádica con ella, dando lugar a relaciones diádicas adicionales en las que el bebé toma a otros, asimismo, como referentes sociales. El bebé comienza entonces a desarrollar una conciencia básica de la normatividad "social", pasando de la pregunta acerca de cómo debería sentirse frente a una situación determinada según la respuesta emocional *de la madre,* a la pregunta acerca de cómo esperan *los demás* que se sienta ante dicha situación (Taipale, 2016). La regulación emocional mutua se traduce entonces en patrones de corrección/sanción, convirtiéndose en una regulación con un mundo que está regido por *normas sociales* y en los que las actitudes reactivas que tienen lugar en las interacciones personales con los demás, tanto propias como ajenas, juegan un papel primordial.

En esta etapa, como he sostenido en el apartado anterior, la introducción al lenguaje de la que es objeto posibilitará que adquiera la habilidad de usar y entender el repertorio conceptual necesario para comprender las estructuras normativas socialmente compartidas, que son las que generan expectativas sobre cómo deben comportarse los agentes y mediante las cuales nos auto-regulamos, regulamos a los demás y somos regulados por ellos (McGeer, 2015). Para ello, tendrá que aprender el uso de las actitudes proposicionales (es decir, adquirir los conceptos de creencia y deseo) que necesitará, por un lado, para señalar los compromisos que adopta para con sus

afirmaciones, y, por otro, para dar razones y justificaciones cuando estas sean requeridas en pos de la regulación, tanto de su comportamiento como —dada la dimensión relacional de la adquisición de las mismas— del comportamiento de los demás. Dicho proceso, he afirmado, se realiza a través de su participación tanto en el tipo de narrativas descritas por Hutto (2008), como en las actividades conjuntas descritas por Fernández Castro (2019), en las que intervienen respuestas regulativas tales como los juegos de simulación cooperativos, las disputas y los desacuerdos. Tras años de entrenamiento, de participación en dichas narrativas y prácticas, una vez alcanzada la compresión y el dominio de las estructuras normativas, de las normas sociales que regulan nuestro comportamiento y, por tanto, de las actitudes proposicionales necesarias para ello, el (ya a esta altura del desarrollo) niño se haya ya en condiciones de ser considerado un agente con autoridad de primera persona.

Sin embargo, como ha afirmado anteriormente, la autoridad de primera persona depende, asimismo, del reconocimiento de los demás miembros de la comunidad. Además de que la adquisición de la autoridad de primera persona no es un logro solitario, sino un logro del que somos capaces en la medida en que nos relacionamos con los demás, para que un niño sea *reconocido* como un agente competente en el uso de la FP, y, por tanto, con autoridad de primera persona, no basta con que alcance los requisitos conceptuales y adquiera el conocimiento de las normas sociales que rigen el uso de las (auto-)atribuciones de estados mentales. Como señala McGeer (2015), el niño ha de tener (con suficiente frecuencia) éxito tanto en el conocimiento de los estados mentales que se auto-atribuye, como en el de actuar de manera acorde a ellos según establecen las normas sociales compartidas, pues, de lo contrario, "sería considerado un extraño, alguien irracional o quizás simplemente incompetente debido a la inmadurez" (p. 270).

Esta afirmación está relacionada con lo sostenido en el bloque anterior acerca de los compromisos que adquirimos al auto-atribuirnos *cualquier* estado mental. La responsabilidad del agente responsable no se limita a la responsabilidad de cumplir con los compromisos que adquiere intencionadamente, es decir, a los compromisos en el sentido de Bilgrami, sino también a la responsabilidad de actuar como se espera que actúe un usuario competente de la FP, es decir, la responsabilidad de actuar de acuerdo a lo que, según las normas sociales, se espera de alguien que se auto-atribuye dicho estado mental. La autoridad de primera persona, por tanto, es la autoridad del agente responsable, del agente conocedor de las normas sociales, que es capaz de auto-

regularse, ser regulado y regular a los demás de acuerdo a ellas y no está solo relacionada, por tanto, con cómo conocemos nuestros estados mentales sino también con la responsabilidad de hacer coincidir nuestras auto-atribuciones con nuestros actos y con el reconocimiento de la misma por parte de los demás miembros de la comunidad.

Ahora bien, ¿qué explicación del autoconocimiento se deriva de este análisis?, es decir, ¿cómo conocemos nuestros estados mentales según el enfoque de segunda persona? Veámoslo en el siguiente y último apartado.

4.3 Autoconocimiento y segunda persona

Si aceptamos que sin lenguaje, sin conceptos, no son posibles ni las auto-atribuciones ni el autoconocimiento, no podemos hablar de autoconocimiento en el bebé hasta que este adquiera los elementos conceptuales y lingüísticos necesarios para ello. Aunque, como he sostenido anteriormente, el bebé "conoce" tanto sus estados emocionales como los expresados por su madre, como muestra su comportamiento durante la comunicación expresiva, este conocimiento del bebé no es el conocimiento "epistémico" necesario para el autoconocimiento sino, más bien, un *conocimiento no conceptual,* (*pre-reflexivo, pre-conceptual*). Decimos que el bebé "conoce" el estado mental en que se encuentra puesto que "saber (conocer) que está en un estado mental" y "*sentir* un estado mental" son dos formas distintas de decir lo mismo en este contexto (al igual que lo son "ver un estado mental en la expresión de la madre" y "conocer un estado mental comunicado expresivamente por la madre", según el concepto de expresión como condición comunicativa del estado mental defendida en este trabajo). Sin embargo, a diferencia de este tipo de conocimiento pre-conceptual, el conocimiento necesario para el autoconocimiento requiere de elementos conceptuales, es decir, para poder afirmar de alguien que conoce algo, incluido uno mismo, en sentido tradicional (epistémico), este ha de poseer los conceptos necesarios para ello. Como afirma Borgoni (2019), "[s]i el autoconocimiento requiere el concepto del yo, que considero una afirmación indiscutible, y un concepto de orden superior como la creencia, parece seguro decir que los bebés carecen de autoconocimiento a la edad de un año" (pp. 15–16).

Este tipo de conocimiento pre-conceptual del que el bebé goza desde los primeros meses de vida, como he sostenido a lo largo de este trabajo, es el que le permite la entrada al lenguaje de las (auto-)atribuciones de estados mentales. Dicho

aprendizaje, he argumentado, tiene una dimensión relacional, es decir, al aprender a sustituir un expresión natural por una lingüística, el bebé aprende, asimismo, a reconocerla en su madre, lo que muestra que, para el bebé, la madre no es alguien a quien comprender, como suponen los enfoques individualistas, sino alguien con quien interacciona y regula sus emociones a través de la comunicación expresiva. Una vez introducidos en el lenguaje, este conocimiento pre-conceptual, lejos de desaparecer, continúa a lo largo de toda la vida y sigue, de hecho, siendo el modo en el que conocemos tanto nuestros estados mentales, como los expresados de manera natural por los demás. Sin embargo, tras nuestra introducción al lenguaje, una vez que poseemos el concepto de, por ejemplo, "dolor" (así como otra multitud de conceptos necesarios para ello), la auto-adscripción "Me duele" sí contiene los elementos conceptuales necesarios para poder ser considerada un conocimiento epistémico y, por tanto, como auto-conocimiento, aun cuando su obtención no requiera de una base epistémica. De esta manera, cuando digo "me duele" puede decirse de mí que sé que me duele porque sé lo que significan esas palabras y sé lo que estoy haciendo con ellas, a saber, expresar mi dolor y el compromiso a comportarme como se espera que se comporte alguien que se auto-atribuye un dolor.

Ahora bien, esta forma pre-conceptual de conocimiento de los estados mentales que forman parte de la comunicación expresiva, no puede extenderse a los estados mentales que requieren de elementos conceptuales para su conocimiento, como es el caso de las actitudes proposicionales. Como he afirmado en el apartado dedicado a la transparencia, las actitudes proposicionales, entendidas como predicados de segundo orden, son necesariamente transparentes. El hecho mismo de estar usando correctamente la auto-adscripción de la actitud proposicional implica que conozco, tanto los compromisos que estoy adquiriendo al usarla, como su uso dentro de una justificación, una razón, una explicación de mi comportamiento, así como las relaciones conceptuales con las demás actitudes proposicionales y las proposiciones bajo su alcance.

Sin embargo, las actitudes proposicionales entendidas como disposiciones, en el sentido de Bilgrami, no son necesariamente transparentes. Las disposiciones, ni son siempre conocidas ni, cuando lo son, son siempre conocidas desde la perspectiva de primera persona. Al igual que con la adquisición de expresiones lingüísticas para sustituir o acompañar las expresiones naturales, que con el aprendizaje de las actitudes proposicionales y que con la autoridad de primera persona, el autoconocimiento tiene un

aspecto relacional. Además del conocimiento en primera persona de nuestros estados mentales que acabo de presentar existe un tipo de conocimiento resultado de nuestra interacción. Como acabo de afirmar, el conocimiento pre-conceptual que se da a través de la comunicación expresiva no desaparece sino que continúa a lo largo de toda la vida y sigue, de hecho, siendo el modo en el que conocemos tanto nuestros estados mentales, como los expresados de manera natural por los demás. El aspecto relacional de esta afirmación implica, de una parte, que no sólo conocemos nuestros estados mentales de manera no conceptual sino también —a través de la comunicación expresiva— los de los demás y, de otra, que *también* conocemos nuestros estados mentales cuando interaccionamos con los demás y *gracias* a los demás, a sus reacciones emocionales que captamos de manera directa y no inferencial, es decir, que conocemos de manera no conceptual a través de la comunicación expresiva. Como señala Gomila (2002) "podemos conocer nuestras propias mentes a través de nuestro conocimiento de las de otros" (p. 137). Al igual que ocurre con la regulación emocional en la infancia durante la comunicación expresiva, en nuestras interacciones en segunda persona mi reacción emocional desencadenada por el estado mental del otro puede llevar a este a reaccionar a su vez a su reconocimiento de tal reacción, lo que puede afectarme de nuevo mí. La dinámica de intercambio mutuo (dialógico, recíproco) de estados mentales que se da en las interacciones en segunda persona (es decir, la regulación emocional mutua que se da en nuestras interacciones intersubjetivas) "puede dar lugar a que me de cuenta de mi propio estado emocional a través del reconocimiento de la reacción ajena a mi estado. […] Puedo llegar a darme cuenta de mi resentimiento hacia alguien por la amargura de su reacción que tanto me irrita" (Gomila, 2002, p. 137).

Además de servir de entrada a la regulación social al traducirse en patrones de corrección/sanción, la regulación emocional continúa a lo largo de nuestra vida a través de la comunicación expresiva en nuestras interacciones en segunda persona, al igual que lo hace en la infancia. Como señalan Sorce y sus colaboradores "las expresiones faciales de emoción, tono de voz, movimientos corporales, no son meramente respuestas indicativas de estados internos, también son patrones de estímulo que regulan el comportamiento de otros" (Sorce et al., 1985, p. 195). No solo nos conocemos, por tanto, desde la perspectiva de primera persona sino también desde la perspectiva de segunda persona, es decir, mediante nuestra interacción con los demás y gracias a los demás.

Finalmente, el conocimiento de nuestros estados mentales puede, asimismo, tener lugar desde de la perspectiva de tercera persona. Casos, entre otros, como en los que determinamos lo que creemos mediante la observación de nuestro comportamiento, tales como cuando afirmamos "Creo que estoy un poco enojado con ella, dado lo que acabo de hacer", o cuando alguien nos señala un estado mental que ha visto en nuestro comportamiento, una actitud de la que no somos conscientes y que no coincide con lo que decimos de nosotros como, por ejemplo, "Dices que no eres racista pero siempre que se sienta alguien de otra raza a tu lado te levantas y cambias de asiento", o cuando nuestro terapeuta nos convence de que tenemos, por ejemplo, un resentimiento inconsciente hacia nuestro hermano porque creemos que nuestra madre siempre le ha querido más a él, o cuando nos miramos al espejo y nos percatamos de nuestro cansancio porque nuestra cara así lo expresa. Este tipo de conocimiento en tercera persona no es solo posible sino, asimismo, necesario. Como señalé anteriormente, nuestra autoridad de primera persona no solo está relacionada de conocimiento de nuestros estados mentales sino, también, con la responsabilidad de hacer coincidir nuestras auto-atribuciones con nuestros actos. Como señalan Hilan Bensusan y Manuel de Pinedo (2007), en lo que respecta a nuestras creencias "[s]omos responsables ante ellas de una manera que no somos responsables ante nada ajeno a nuestra propia creación. Sin acceso en tercera persona, nuestras creencias no podrían ser corregibles. Sin el acceso de primera persona, no podrían corregirse, ya que nadie sería responsable de ellas para rectificarlas. Cualquier creencia puede ser falsa y puede ser juzgada como tal y rectificada, y esto se debe a que son accesibles a través de estas dos rutas" (p. 38). Los casos en los que los otros regulan nuestro comportamiento, casos en los que nuestras creencias en primera persona no coinciden con nuestros actos o con lo que los demás ven en, o creen de, nosotros, son un ejemplo de ello.

En este apartado, he presentado la explicación que el enfoque de segunda persona ofrece del autoconocimiento. He sostenido que la trasparencia a los estados mentales no tiene un sentido unívoco como defienden los enfoques individualistas. De una parte, los estados mentales con los que nos comunicamos mediante la comunicación expresiva resultan transparentes tanto para el sujeto como para la persona con la que estamos interaccionando. Este tipo de transparencia está relacionada con el análisis wittgensteiniano de las expresiones naturales, según el cual las expresiones naturales comunican el estado mental de manera directa y no inferencial, es decir, transparente y

con el análisis del conocimiento pre-conceptual del que gozamos desde nuestras primeras semanas de vida y mediante el cual "conocemos" o "sentimos" los estados mentales. De otra parte, las actitudes proposicionales, entendidas como predicados de segundo orden son necesariamente transparentes puesto que, el mismo hecho de poder usar correctamente una auto-adscripción de una actitud proposicional de este tipo implica que conozco, tanto los compromisos que estoy adquiriendo al usarla, como su uso dentro de una justificación, de una razón, de una explicación de mi comportamiento, como asimismo, sus relaciones conceptuales con las demás actitudes proposicionales y con las proposiciones bajo su alcance.

En cuanto a la noción de autoridad, he sostenido que la autoridad que aparece primero en el desarrollo ontogenético no es la autoridad de primera persona sino la autoridad de segunda persona que el bebé atribuye a su madre, que es a quien toma como referente social ante las situaciones que desconoce o para las que no tiene un valor asignado. En lo que respecta a la autoridad de primera persona, he sostenido que, al igual que el aprendizaje de las actitudes proposicionales y de las razones (justificaciones, explicaciones), esta ha de ser entendida de forma relacional, es decir, que la autoridad no depende solo del individuo sino también del reconocimiento por parte de los demás miembros de la comunidad. Asimismo, he afirmado que esta autoridad no está solo relacionada con cómo conocemos nuestros estados mentales sino también con la responsabilidad de hacer coincidir nuestras auto-atribuciones con nuestros actos en tanto que agentes responsables conocedores de las normas sociales, y capaces de auto-regularnos, ser regulados y regular a los demás de acuerdo a ellas.

Finalmente, he sostenido una explicación el autoconocimiento que engloba dos nociones de conocimiento: el conocimiento pre-conceptual, que no requiere de elementos conceptuales, y el conocimiento epistémico (el conocimiento en sentido tradicional), que sí requiere del lenguaje y de conceptos. Según esta explicación, el conocimiento necesario para poder hablar de autoconocimiento es el conocimiento epistémico, que requiere de la posesión de elementos conceptuales. De esta manera, aun cuando tenemos un conocimiento pre-conceptual de determinados estados mentales, como muestra el comportamiento de los bebés durante la comunicación expresiva, no podemos hablar de autoconocimiento hasta que adquirimos los elementos conceptuales necesarios para ello. Sin embargo, una vez los adquirimos, las auto-atribuciones de estados mentales provenientes del conocimiento pre-conceptual sí pueden ser consideradas como autoconocimiento. Cuando digo, por ejemplo, "Estoy triste" puede

decirse de mí que sé que estoy triste porque sé lo que significan esas palabras y sé lo que estoy haciendo con ellas, a saber, expresar mi tristeza y el compromiso a comportarme como se espera que se comporte alguien que se auto-atribuye tristeza.

De otra parte, esta concepción del autoconocimiento incluye tanto la perspectiva de primera como la de segunda y la de tercera persona. Además del conocimiento en primera persona de nuestros estados mentales, proveniente del conocimiento no conceptual y de la auto-atribución de actitudes proposicionales (las cuales son conocidas necesariamente como he argumentado), existe un tipo de conocimiento que es resultado de nuestra interacción. Este conocimiento, he sostenido, proviene del intercambio mutuo de estados mentales mediante la comunicación expresiva que tiene lugar en nuestras interacciones de segunda persona. La dinámica de intercambio mutuo (dialógico, recíproco) de estados mentales que se da en estas interacciones permite, en ocasiones, que conozcamos un estado mental propio a través del reconocimiento de la reacción del otro a un estado expresado por nosotros. Por último, el conocimiento de nuestros estados mentales puede también tener lugar desde de la perspectiva de tercera persona, por ejemplo, mediante la observación de nuestro comportamiento o mediante el testimonio de otra persona. Este tipo de conocimiento, he sostenido, resulta necesario para conocer y poder corregir determinados estados mentales que no son accesibles desde la perspectiva de primera persona.

5. Conclusiones

En este bloque he sostenido una explicación de la FP y del autoconocimiento basada en el enfoque de segunda persona que defiendo en este trabajo.

En el segundo apartado he presentado una breve descripción de los enfoques tradicionales y señalado las deficiencias de las que adolecen según el enfoque de segunda persona. He señalado que tanto la TT como la TS parten de una concepción individualista del fenómeno en el que el foco de estudio es el sujeto como alguien separado de los demás, entendidos estos como seres *a comprender*, bien sea mediante la aplicación de una teoría de la mente o bien mediante la simulación mental de su situación. He sostenido que el modelo adecuado para describir la adquisición de la FP no es el observacional ni el de la simulación, sino el modelo de *comunicación*

reguladora mutua. Bajo este modelo, la FP no es entendida como una estrategia para predecir y explicar el comportamiento, sino como una práctica normativo-regulativa.

El tercer apartado ha estado dedicado a presentar la explicación que de la FP ofrece el enfoque de segunda persona. Según este enfoque, la FP no ha de ser entendida como una estrategia para predecir y explicar el comportamiento de los demás sino como una práctica normativo-regulativa. Nuestra comprensión de la FP no está basada en descripciones de estados mentales internos, sino en estructuras normativas socialmente compartidas que generan expectativas sobre cómo deben comportarse los agentes. He sostenido que la respuesta a las preguntas acerca de *cómo* y *para qué* aprendemos el uso de las actitudes proposicionales que están a la base de la FP se haya, tanto en las interacciones en segunda persona, como en nuestra disposición natural a experimentar actitudes reactivas y a regularnos con los demás. La repuesta al *para qué*, es decir, la respuesta a cuál es la función principal para la que adquirimos las actitudes proposicionales de creencia y deseo, proviene de las actitudes reactivas que tienen lugar durante nuestras interacciones interpersonales (en nuestras interacciones en segunda persona). Estas actitudes resultan necesarias en las ocasiones en las que nuestro comportamiento produce una actitud reactiva en el otro y este nos pide dar razones del mismo. De ahí que la principal función de la adquisición de las actitudes proposicionales sea una función evaluativa, es decir, justificadora, exculpatoria, condenatoria, de elucidación del comportamiento contra-normativo y de los compromisos adquiridos. En cuanto al *cómo* aprendemos su uso, he afirmado que es mediante determinados tipos de interacciones en segunda persona, en concreto a interacciones de segunda persona en las que, por un lado, los niños están expuestos a encuentros directos con historias sobre personas que actúan por *razones*, a narrativas proporcionadas en contextos interactivos por cuidadores responsables, de manera que pueden participar en las narrativas de manera correcta y, por otro lado, a contextos en los que participan en actividades conjuntas en las que intervienen *respuestas regulativas* como, por ejemplo, los juegos de simulación cooperativos, las disputas y los desacuerdos.

En el último apartado he presentado la explicación del autoconocimiento según el enfoque de segunda persona. En lo referente a la transparencia, he sostenido que los estados mentales mediante con los que nos comunicamos a través de la comunicación expresiva son transparentes tanto para el sujeto como para la persona con la que estamos interaccionando. Asimismo, ha afirmado que las actitudes proposicionales

entendidas como predicados de segundo orden son, según esta concepción, necesariamente trasparentes. En lo que respecta a la autoridad, en relación con lo defendido en el primer bloque, he sostenido que la primer noción de autoridad que aparece en el desarrollo ontogenético no es la de primera sino la autoridad de segunda persona que el bebé atribuye a su madre. Asimismo, he sostenido que el niño pasa de la autoridad de segunda persona de la madre, de que esta sea su fuente de normatividad, a la autoridad propia al comprender las normas sociales la FP: una vez alcanzada la compresión y el dominio de las estructuras normativas, de las normas sociales que regulan nuestro comportamiento y, por tanto, de las actitudes proposicionales necesarias para ello, el niño puede ser considerado un agente con autoridad de primera persona. Sin embargo, la autoridad de primera persona no tiene solo que ver con cómo conocemos nuestros estados mentales y las normas sociales, sino con la responsabilidad de hacer coincidir nuestras auto-atribuciones con nuestros actos, así como con el reconocimiento de la misma por parte de los demás miembros de la comunidad. Solo uno mismo puede tener autoridad de primera persona pues solo uno mismo es responsable y capaz de hacer que sus auto-atribuciones coincidan con sus actos.

Finalmente, he sostenido una concepción del autoconocimiento que engloba dos nociones de conocimiento: el conocimiento pre-conceptual, que no requiere de elementos conceptuales, y el conocimiento epistémico (el conocimiento en sentido tradicional), que sí requiere del lenguaje y de conceptos. Según esta noción de autoconocimiento, no conocemos nuestros estados mentales desde la perspectiva de primera o de tercera persona, sino desde las perspectivas de primera, de segunda y de tercera persona. El conocimiento de primera persona se da tanto a través del conocimiento pre-conceptual como a través de la auto-atribución de actitudes proposicionales. Por un lado, cuando nos auto-atribuimos un estado mental que conocemos de manera no conceptual como, por ejemplo un dolor, mediante la expresión "Me duele la cabeza", podemos afirmar que sabemos que nos duele la cabeza porque sabemos lo que significan y lo que estamos haciendo con esas palabras, a saber, expresar nuestro dolor y nuestro compromiso a comportarnos como alguien a quien le duele la cabeza. Por otro lado, conocemos nuestras actitudes proposicionales de manera necesaria cuando nos las auto-atribuimos puesto que el hecho mismo de estar usando correctamente una auto-adscripción de la actitud proposicional implica que conozco, tanto los compromisos que estoy adquiriendo al usarla, como su uso dentro de una justificación, una razón, una explicación de mi comportamiento, así como las relaciones

conceptuales con las demás actitudes proposicionales y las proposiciones bajo su alcance.

El conocimiento de segunda persona, he sostenido, tiene su origen en el intercambio mutuo de estados mentales mediante la comunicación expresiva que tiene lugar en nuestras interacciones de segunda persona. Durante la comunicación expresiva que tiene lugar en nuestras interacciones en segunda persona, se crea una dinámica de intercambio mutuo de estados mentales que permite, en ocasiones, que conozcamos un estado mental propio a través del reconocimiento de la reacción del otro al estado expresado por nosotros. Por último, he sostenido que el conocimiento de nuestros estados mentales puede tener lugar, asimismo desde de la perspectiva de tercera persona. Casos, como por ejemplo, cuando nos damos cuenta de que estamos cansados porque vemos la expresión del cansancio en nuestro rostro al mirarnos al espejo, o cuando nuestro terapeuta, amigo, familiar, nos advierte de una actitud que nosotros desconocíamos. Este tipo de conocimiento, he afirmado, es necesario tanto para conocer como para poder corregir los estados mentales que, por algún motivo, no son accesibles desde la perspectiva de primera persona, así como para corregir nuestras auto-atribuciones en caso de no coincidir con nuestros estados mentales.

Con este análisis he tratado de destacar la importancia de un enfoque de segunda persona para la comprensión de la FP y de su relación con la constitución del autoconocimiento.

Conclusiones

En esta tesis he ofrecido una explicación del autoconocimiento desde un enfoque expresivista de segunda persona. Este enfoque parte de la perspectiva de segunda persona y recoge las intuiciones tanto del expresivismo clásico como del neo-expresivismo y del expresivismo semántico.

Según este enfoque, existe una continuidad entre la capacidad para la intersubjetividad en el desarrollo ontogenético, el autoconocimiento y la Folk Psychology. He estructurado la tesis en tres grandes bloques en los que he profundizado en cada uno de estos ámbitos que, aunque tradicionalmente han sido tratados de manera separada, en este trabajo han sido tratados de manera conjunta. De esta manera, he trazado una continuidad entre los estudios en psicología del desarrollo, en epistemología y en filosofía de la mente, mostrando la necesidad de un estudio conjunto a la hora de dar cuenta del fenómeno del autoconocimiento.

He mostrado que existe un presupuesto común, al que he denominado "individualista", desde el que tradicionalmente han partido estos estudios. Adoptar este presupuesto implica asumir una dicotomía exclusiva y excluyente entre la primera y la tercera persona que ignora la existencia de la perspectiva de segunda persona. Este hecho conduce tanto a una visión parcial del fenómeno como a una serie de problemas que, como he mostrado, se dirimen adoptando un enfoque relacional que toma como punto de partida la perspectiva de segunda persona.

El primer bloque de esta tesis ha estado dedicado a los estudios en psicología del desarrollo, por lo que ha contado con la evidencia empírica propia de este campo. He mostrado la importancia de la perspectiva de segunda persona en el estudio del desarrollo ontogenético de los bebés, así como en la constitución de la conciencia de sí mismos y de los demás, es decir, de las perspectivas de primera y tercera persona. He sostenido que el enfoque de segunda persona ofrece una explicación más plausible del desarrollo ontogenético de los bebés que los enfoques alternativos que parten de la dicotomía entre la primera y tercera persona.

He defendido la existencia de una comunicación expresiva mediante la cual los bebés son capaces de conectar y comunicar sus experiencias subjetivas con los demás, es decir, de participar en interacciones intersubjetivas. Para fundamentar esta

comunicación expresiva, he ofrecido una explicación de la noción de expresión compatible con la misma. Según esta noción, los gestos faciales, los movimientos corporales, el tono de la voz, no son un medio a través del cual podemos conocer los estados mentales de los demás sino que son constitutivos de los mismos. Ser constitutivo en este contexto, no significa que las expresiones constituyan los estados mentales ni, al contrario, que los estados mentales constituyan las expresiones. Ser constitutivo significa que las expresiones no están separadas de los estados mentales ni, por tanto, relacionadas con ellos de ninguna manera, sino que son parte de una y la misma cosa. En concreto, he sostenido que las expresiones son la condición comunicativa de los estados mentales.

Según esta concepción de la expresión, durante nuestras interacciones con los demás vemos, oímos, los estados mentales de manera directa, sin necesidad de realizar inferencia alguna de la expresión al estado mental, puesto que la expresión no es algo distinto del estado mental. Esta concepción permite disolver la dicotomía entre la primera y la tercera persona, en lo que respecta al conocimiento de los estados mentales de los demás, al incluir la posibilidad de un acceso directo a los mismos mediante la comunicación expresiva que tiene lugar a través de la perspectiva de segunda persona. Por tanto, he defendido que la intersubjetividad entre los bebés y sus cuidadores tiene lugar durante la comunicación expresiva mediante el intercambio de expresiones, es decir, de estados mentales en su condición comunicativa. De esta manera, la perspectiva de segunda persona puede ser descrita como la perspectiva de la interacción mental, la perspectiva mediante la cual intercambiamos nuestros estados mentales de manera expresiva.

He sostenido, asimismo, que la perspectiva de segunda persona, no solo precede ontogenéticamente a las otras dos perspectivas, sino que también juega un papel constitutivo en la emergencia de las mismas. Finalmente, he presentado un modelo de comunicación reguladora mutua según el cual los bebés conciben a la persona que está a su cuidado como una autoridad —que he denominado autoridad de segunda persona. He defendido que a través de esta autoridad los bebés extienden la regulación emocional que tiene lugar en sus interacciones diádicas, a una regulación normativo-social en sus interacciones triádicas a través de los triángulos de atención conjunta.

El segundo bloque ha estado dedicado a la presentación y crítica de las diferentes propuestas que se han llevado a cabo sobre el fenómeno del autoconocimiento. De acuerdo las similitudes inherentes a estas propuestas, he distinguido tres tipos de enfoques: el introspeccionista, el agencialista y el neo-expresivista. En el primer apartado, dedicado al enfoque introspeccionista, he señalado que la explicación que ofrecen estos autores resulta diametralmente opuesta a la ofrecida por el enfoque de segunda persona. Según los introspeccionistas, las interacciones con los demás no tienen relevancia alguna en el estudio del fenómeno del autoconocimiento. El autoconocimiento proviene, a su entender, de un acceso privilegiado del individuo a sus propios estados mentales. Tras la presentación de estas propuestas, he expuesto las principales críticas a las mismas destacando entre ellas la dificultad para ofrecer una explicación de la existencia de estados mentales en los demás.

En el apartado dedicado a los agencialistas he sostenido que aunque estos autores sí incluyen las interacciones en sus propuestas, lo hacen desde una perspectiva teórica y centrada en el individuo. Este hecho conduce a una comprensión parcial del fenómeno del autoconocimiento y supone un problema a la hora de dar cuenta del comportamiento que, de hecho, tenemos los seres humanos. La comprensión parcial de este fenómeno conlleva la ausencia del elemento relacional en su explicación de las nociones de razón, responsabilidad, compromiso y autoridad, así como la ausencia del tratamiento de la normatividad social en sus enfoques. Por ello, he presentado una explicación relacional de estas nociones que elimina las carencias a las que se enfrenta la explicación parcial de las mismas. Asimismo, he mostrado la necesidad de la inclusión de la normatividad social en la explicación del fenómeno del autoconocimiento.

Finalmente, en el apartado dedicado a los neo-expresivitas he sostenido que el presupuesto individualista conduce, en su caso, a una visión parcial del proceso de adquisición de las (auto-)atribuciones de estados mentales. Estos autores centran su explicación exclusivamente en la capacidad para la auto-atribución de estados mentales, obviando su relación con la capacidad para la atribución de estados mentales a los demás. Por el contrario, de acuerdo con el enfoque expresivista de segunda persona, ambas capacidades se adquieren simultáneamente. Este hecho, he sostenido, resulta crucial para una correcta comprensión de dicho proceso. Asimismo, su explicación

sobre el autoconocimiento se extiende a todos los estados mentales sin distinción. He argumentado que la no diferenciación de los diferentes tipos de estados mentales conduce a una visión errónea del mismo. Finalmente, he suscrito la validez de algunos de los argumentos sobre la transparencia de los estados mentales, en concreto los referentes a la noción de la transparencia-a-la-condición-mental-del-sujeto que, como he afirmado, describe el acceso directo a los estados mentales que tiene lugar mediante la comunicación expresiva.

En el tercer y último bloque he analizado las propuestas de los enfoques tradicionales acerca de la Folk Psychology, así como las propuestas actuales que rechazan la explicación de estos enfoques. En este bloque he ofrecido, asimismo, una explicación del autoconocimiento desde el enfoque expresivista de segunda persona.

He mostrado que, al igual que en los estudios acerca del desarrollo ontogenético y del autoconocimiento, los enfoques tradicionales sobre la Folk Psychology parten de la dicotomía entre primera y tercera persona. En su caso, esta dicotomía ha separado la investigación en dos tipos de teorías contrapuestas, cada uno de ellas fundamentada en uno de los dos polos de la dicotomía. De una parte, los defensores de la Teoría de la Teoría desarrollan su enfoque partiendo de la perspectiva de tercera persona. De otra parte, los defensores de la Teoría de la Simulación lo hacen desde la perspectiva de primera persona. En el apartado dedicado a estas teorías, he sostenido que el mayor problema al que se enfrentan proviene del presupuesto individualista que ambas comparten. Según este presupuesto, los demás son seres a comprender, por lo que la principal función de la Folk Psychology es, para estos autores, la de predecir y explicar la conducta de los demás.

En el siguiente apartado, he analizado los distintos argumentos que contra este supuesto se han llevado a cabo desde algunos de los enfoques actuales, suscribiendo la validez de los mismos y ofreciendo una explicación de la Folk Psychology desde el enfoque de segunda persona. En este enfoque, la Folk Psychology es entendida no como una teoría sino como una práctica normativo-regulativa cuya principal función no es la de predecir y explicar la conducta sino una función regulativa consistente en aprender, enseñar y exhortar a otros a comportarse de acuerdo a las normas sociales compartidas que rigen las interacciones de los miembros de la comunidad.

Para concluir, he ofrecido una explicación del autoconocimiento que recoge lo defendido tanto acerca del desarrollo ontogenético, como acerca de la Folk Psychology.

He sostenido que la transparencia de lo mental no lo es solo hacia los propios estados mentales sino también hacia los estados mentales que los demás nos comunican de manera expresiva durante las interacciones, es decir, mediante la comunicación expresiva que tiene lugar a través de la perspectiva de segunda persona. De acuerdo a esta definición de la transparencia, he sostenido la existencia de dos tipos de conocimiento: el conocimiento epistémico, en sentido tradicional, y el conocimiento pre-conceptual, entendido este último como el conocimiento con el que contamos ya desde nuestras primeras semanas de vida. He afirmado que este tipo de conocimiento, lejos de desaparecer una vez introducidos en el lenguaje y, por tanto, una vez poseemos los conceptos necesarios para ello, continúa presente a lo largo de nuestra vida. En cuanto a la autoridad de primera persona, he defendido una concepción relacional de la misma según la cual esta está vinculada tanto con el conocimiento de nuestros estados mentales, como con la responsabilidad de hacer que nuestras auto-atribuciones de estados mentales coincidan con nuestros actos (con lo que, según las normas sociales y lingüísticas se espera de alguien que se auto-atribuye el estado mental en cuestión), así como con el reconocimiento por parte de los demás miembros de la comunidad. Finalmente, he sostenido que el autoconocimiento incluye no sólo la perspectiva de primera o de tercera persona sino la perspectiva de primera, la de segunda y la de tercera persona.

Conclusions

In this dissertation I have proposed a second-person expressivist approach to self-knowledge that has as its starting point the second-person perspective and gathers the intuitions from classic expressivism as well as from neo-expressivism and semantic expressivism.

According to this approach, there exists a continuity between the capacity for intersubjectivity in the ontogenetic development, self-knowledge and Folk Psychology. I have segmented the dissertation into three large blocks where I have studied in depth each one of these topics. Although they have been traditionally dealt with in a separate manner, in this dissertation I have considered them jointly. Thus, I have drawn a continuity line in the studies by Development Psychology, Epistemology and

Philosophy of Mind, that highlights the necessity for a combined study when accounting for the phenomenon of self-knowledge.

I have shown that there is a common assumption —which I have called "individualistic"— whence all traditional approaches on this phenomenon have taken as their starting point. Taking on this assumption implies adopting a unique and excluding dichotomy between the first and the third person that ignores the existence of the second-person perspective. This leads us to a partial vision of the phenomenon as well as to a series of problems that, I have shown, are settled by adopting an approach that takes as its starting point the second-person perspective.

The first block of this dissertation has been devoted to the studies in Development Psychology; and so, it has made use of the empirical evidence from this field. I have shown the relevance of the second-person perspective both to the ontogenetic development of infants and to the constitution of their self-awareness and the awareness of others, i.e., of the first- and third-person perspectives. I have argued that the second-person approach allows for a more plausible understanding of infants' development during this period than the alternative approaches that assume the dichotomy between the first and the third person.

I have argued for the existence of an expressive communication through which infants are able to connect and communicate their subjective experiences with others, i.e., to participate in intersubjective interactions. In order to ground this expressive communication, I have brought forward and explanation of the notion of expression compatible with it. According to this conception, facial gestures, body movements, and tone of voice are not means through which we can know the mental states of others but they are constitutive of them. To be constitutive in this context does not mean that expressions constitute mental states or that mental states constitute expressions. To be constitutive means that the expressions are neither separated from the mental states nor are they related to them in any way, since they are part of one and the same thing. Precisely, I have argued that the expressions are the communicative condition of mental states.

According to this conception of expression, during our interactions with the others we see, hear, mental states directly, without needing to make any inferences from the expression to the mental state, since the expression is not something different from the mental state. This conception allows solving the dichotomy between the first and the

third person in regards to the knowledge of others' mental states, when including the possibility of a direct access to them through the communicative expression that occurs by means of second person perspective. Then, I have defended that the intersubjectivity among infants and their caregivers happens during expressive communication through the exchange of expression, i.e., of mental states in their communicative condition. Thus, second-person perspective could be described as the perspective of mental interaction, the perspective by which we exchange our mental states in an expressive manner.

I have argued that the second-person perspective not only precedes the ontogenetic development of the other two perspectives, but that it also plays a constitutive role in their emergence. Finally, I have introduced a mutual communicative regulation model according to which infants conceive their caregivers as an authority– which I have called second-person authority. I have sustained that it is through this authority that infants extend the emotional regulation that takes place within their dyadic interactions over a socio-normative regulation in their triadic interactions by manner of the joint-attention triangles.

The second block has been devoted to presenting and criticizing the different proposals that have been brought forward regarding self-knowledge. Following the inherent similarity within these proposals, I have distinguished three different approaches: introspectionist, agencialist and neo-expressivist. In the first section, dealing with introspectionism, I have pointed out that the explanation that these authors present opposes radically the one offered by the second-person approach. Introspectionists affirm that interactions with others have no relevance in the study of self-knowledge. Self-knowledge arises, they argue, from a privileged and exclusive access to their inner mental states that individuals have. Once their proposals were unfolded, I have detailed their primary critiques highlighting among them the difficulty to produce an explanation of the existence of others' mental states.

In the section dealing with agentialism, I have argued that, although these authors do include interactions in their proposals, they do it from a theoretical perspective and centred in the individual. This leads to a partial understanding of the phenomenon and it has serious problems when it comes to accounting for the behaviour that, in fact, we humans display. This partial understanding of the phenomenon implies the absence of the relational element in their explanation of the notions of reason,

responsibility, commitment and authority, as well as the lack of a treatment of social normativity in their approaches. Because of this I presented a relational explanation of these notions that resolves the shortcomings that the partial explanation of them tackles. Furthermore, I have shown the necessity of including social normativity within the explanation of self-knowledge.

Finally, in the section dealing with neo-expressivism, I have maintained that the individualist assumption leads, in their case, to a partial vision of both the process of learning the self-attributions of mental states and the notion of expression. Regarding the first question, these authors focus their explanation exclusively on the capacity for self-attribution of mental states, not taking into consideration its connection with the capacity for attributing mental states to others. In contrast, according to the second-person expressivist approach, both capacities are learned simultaneously. This fact, I have contended, happens to be crucial for a proper understanding of such a process. Regarding the notion of expression, in the neo-expressivist approach, it appears severed from the notion of mental state, which leads to its understanding as a means through which we know mental states. Moreover, their explanation of self-knowledge comprises all mental states without distinction. I have contended that both the separation of the notions of expression and mental state and the lack of differentiation between different mental states, results in an inadequate understanding of them. Finally, I have agreed with some of the arguments about transparency of mental states, particularly about transparency-to-the-subject's-mental-condition.

I have analyzed the proposal of the traditional approaches on Folk Psychology in the third and last block, as well as the recent proposals that reject the explanation of these approaches. In this block I have also provided an explanation of self-knowledge from the second-person expressivist approach.

Just as the studies on ontogenetic development and self-knowledge, the traditional approaches on Folk Psychology are based on the dichotomy between first- and third person. This assumption has split the research into two types of opposing theories, each one grounded in one of the poles of the dichotomy. On one side, the supporters of the Theory Theory construct their approach from the third-person perspective. On the other side, the supporters of the Simulation Theory embrace the first-person perspective. In the section dealing with these theories, I have argued that the major problem they faces arises from the individualist assumption that both share.

According to this assumption, the others are beings to be understood. For these authors, then, the primary function of Folk Psychology becomes predicting and explaining the behaviour of others.

In the next section I have studied the diverse arguments held by some current approaches, subscribing their validity and providing a second-person approach to Folk Psychology. This approach understands Folk Psychology, not as a theory, but as normative-regulative practice whose primary function is not predicting and explaining behaviour, but a regulative function consisting of learning, teaching and exhorting others to behave according to the shared social norms that rule the interactions of the community members.

To conclude, I have provided an explanation of self-knowledge that gathers what has been defended both on the ontogenetic development and Folk Psychology. I have maintained that the transparency of the mental is not only towards one's own mental states, but also towards the mental states that others communicate in an expressive manner during interactions, i.e., through the expressive communication that occurs by means of the second-person perspective. Following this definition of transparency, I have argued for the existence of two kinds of knowledge: epistemic knowledge, in a traditional sense, and the pre-conceptual knowledge, this last one understood as the knowledge we have from the first weeks of life. I have defended that this type of knowledge, far from disappearing once introduced within language and, therefore, once we possess the necessary concepts for it, stays present through life. Regarding first-person authority, I have argued for a relational conception of it, according to which it is linked both with the knowledge of our mental states and the responsibility to put in line our self-attributions with our deeds (with what is expected from someone who self-attributes the mental state in question, according to social and linguistic norms), as well as with the recognition by other members of the community. Finally, I have argued that self-knowledge includes not only first- or third-person perspectives but first-person, second-person and third-person perspectives.

Bibliografía
References

Adamson, L. B., & Bakeman, R. (1991). The development of shared attention during infancy. *Annals of Child Development*, *8*, 1–41.

Amano, S., Kezuka, E., & Yamamoto, A. (2004). Infant shifting attention from an adult's face to an adult's hand: A precursor of joint attention. *Infant Behavior and Development*, *27*(1), 64–80. https://doi.org/10.1016/j.infbeh.2003.06.005

Andrews, K. (2009). Understanding norms without a theory of mind. *Inquiry*, *52*(5), 433–448.

Andrews, K. (2012). *Do apes read minds? Toward a new folk psychology*. Cambridge: MIT Press.

Andrews, K. (2015). The folk psychological spiral: Explanation, regulation, and language. *Southern Journal of Philosophy*, *53*, 50–67.

Armstrong, D. M. (1968). *A materialist theory of the mind*. London: Routledge.

Armstrong, D. M. (1981). *The nature of mind and other essays*. Ithaca, NY: Cornell University Press.

Austin J. L. (1962). *How to do things with words*. Oxford: Oxford University Press.

Austin, J. L. (1970). *Philosophical papers*. Oxford: Clarendon Press.

Ayer, A. J. (1952). *Language, truth, and logic*. New York: Dover.

Bagnoli, C. (2007). The authority of reflection. *Theoria*, *58*, 42–52.

Bar-On, D. (2004). *Speaking my mind: Expression and self-knowledge*. Oxford: Clarendon Press.

Bar-On, D. (2009). First-person authority: Dualism, constitutivism, and neo-expressivism. *Erkenntnis*, *71*(1), 53–71. https://doi.org/10.1007/s10670-009-9173-y

Bar-On, D. (2011). Neo-expressivism: Avowals' security and privileged self-knowledge. In A. Hatzimoysis (Ed.), *Self-knowledge* (pp. 189–201). Oxford: Oxford University Press.

Bar-On, D. (2013). Origins of meaning: Must we "go gricean"? *Mind and Language*, *28*(3), 342–375. https://doi.org/10.1111/mila.12021

Bar-On, D. (2015). Transparency, expression and self-knowledge. *Philosophical Explorations*, *18*(2), 134–152.

Bar-On, D., & Long, D. C. (2001). Avowals and first-person knowledge. *Philosophy and Phenomenological Research*, *62*, 311–335.

Bar-On, D., & Long, D. C. (2003). Knowing selves : Expression, truth, and knowledge. Retrieved September 10, 2018, from http://philosophy.sites.unc.edu/files/2013/10/Bar-On_Long2003_KnowingSelves_wRef.pdf

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge. MA: MIT Press/Bradford Books.

Barresi, J., & Moore, C. (1996). Intentional relations and social understanding. *Behavioral and Brain Sciences*, *19*(01), 107–122. https://doi.org/10.1017/S0140525X00041790

Bartsch, K., & Wellman, H. M. (1989). Young children's attribution of action to beliefs and desires. *Child Development*, *60*(4), 946–964.

Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. New York:

Oxford University Press.
Bates, E., Benigni, L., Bretherton, I., Camaioni, L., & Volterra, V. (1979). *The emergence of symbols: Cognition and communication in infancy*. New York: Academic Press Inc.
Bates, E., Camaioni, L., & Volterra, V. (1976). Sensoriomotor performatives. In E. Bates (Ed.), *Language and context: The acquisition of pragmatics* (pp. 49–71). New York: Academic Press Inc.
Bateson, M. C. (1975). Mother-infant exchanges: The epigenesis of conversational interaction. In D. Aaronson & R. W. Rieber (Eds.), *Developmental psycholinguistics and communication disorders. Annals of the New York Academy of Sciences* (Vol. 263, pp. 101–113). New York: New York Academy of Science.
Bateson, M. C. (1979). The epigenesis of conversational interaction: A personal account of research development. In M. Bullowa (Ed.), *Before speech: The beginning of human communication* (pp. 63–77). London: Cambridge University Press.
Batki, A., Baron-Cohen, S., Wheelwright, S., Connellan, J., & Ahluwalia, J. (2000). Is there an innate gaze module? Evidence from human neonates. *Infant Behavior and Development*, *23*(2), 223–229. https://doi.org/10.1016/S0163-6383(01)00037-6
Bennett, J. (1978). Commentary on three papers about animal cognition. *Behavioral and Brain Sciences*, *1*(4), 556–560.
Bensusan, H., & Pinedo, M. de. (2007). When my own beliefs are not first-personal enough. *Theoria. Revista de Teoría, Historia y Fundamentos de La Ciencia*, *22*(1), 35–41.
Bilgrami, A. (1998). Self-knowledge and resentment 1. In C. Wright, B. Smith, & C. Macdonald (Eds.), *Knowing our own minds: Essays in self-knowledge* (pp. 207–242). Oxford: Oxford University Press.
Bilgrami, A. (2006). *Self-knowledge and resentment*. Cambridge. MA: Harvard University Press.
Bilgrami, A. (2012). The unique status of self-knowledge. In A. Coliva (Ed.), *The self and self-knowledge* (pp. 263–278). Oxford: Oxford University Press.
Blackburn, S. (1993). *Essays in Quasi-realism*. Oxford: Oxford University Press.
Blackburn, S. (1998). *Ruling passions*. Oxford: Clarendon Press.
Borgoni, C. (2018). Authority and attribution: the case of epistemic injustice in self-knowledge. *Philosophia*, 1–9. https://doi.org/10.1007/s11406-018-0002-x
Borgoni, C. (2019). *First-person authority: persons, expressions and communication*.
Brandom, R. (1994). *Making it explicit. Reasoning, representing, and discursive commitment*. Cambridge. MA: Harvard University Press.
Bråten, S. (2007). *On being moved. From mirror neurons to empathy*. (S. Bråten, Ed.). Amsterdam/Philadelphia: John Benjamins Publishing.
Bråten, S., & Trevarthen, C. (2007). Prologue: From infant intersubjectivity and participant movements to simulation and conversation in cultural common sense. In S. Bråten (Ed.), *On being moved: From mirror neurons to empathy* (pp. 21–34). Amsterdam/Philadelphia: John Benjamin Publishing Company.
Brazelton, T. B., Koslowski, B., & Main, M. (1974). The origins of recipriocity: The early mother-infant interaction. In M. Lewis & L. Sosenblum (Eds.), *The effects of the infant on its caregiver* (pp. 49–76). New York: Wiley.
Bretherton, I., McNew, S., & Beeghly-Smith, M. (1981). Early person knowledge as expressed in gestural and verbal communication: When do infants acquire a "theory of mind"? In M. E. Lamb & L. R. Sherrod (Eds.), *Infant social cognition: Empirical and theoretical considerations* (pp. 333–373). Amsterdam/Philadelphia: Lawrence Erlbaum Associates Ltd.

Brinck, I. (2004). Joint attention, triangulation and radical interpretation: A problem and its solution. *Dialectica*, *58*(2), 179–206. https://doi.org/10.1111/j.1746-8361.2004.tb00296.x

Brinck, I. (2008). The role of intersubjectivity in the development of intentional communication. In J. Zlatev, T. P. Racine, C. Sinha, & E. Itkonen (Eds.), *The shared mind: Perspectives on intersubjectivity* (pp. 39–66). Amsterdam: John Benjamins Publishing.

Brown, J. N., Donelan-McCall, N., & Dunn, J. (1996). Why talk about mental states? The significance of children's conversations with friends, siblings, and mothers. *Child Development*, *67*(3), 836–849.

Bruner, J. (1977). Early social interaction and language acquisition. In H. R. Schaffer (Ed.), *Studies in mother–infant interaction* (pp. 271–289). New York: Academic Press Inc.

Bruner, J. (1995). From joint attention to the meeting of minds. In C. Moore & P. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 1–14). Hillsdale, NJ: Lawrence Erlbaum Associates Ltd.

Bruner, J., & Scaife, M. (1975). The capacity for joint visual attention in the infant. *Nature*, *253*(5489), 265–266. https://doi.org/10.1038/253265a0

Buber, M. (1958). *I and Thou*. (R. G. Smith, Trans.) (2nd edn). Edinburgh: T. & T. Clark.

Burge, T. (1979). Individualism and the Mental. *Midwest Studies in Philosophy*, *4*(1), 73–121. https://doi.org/10.1111/j.1475-4975.1979.tb00374.x

Bushnell, I. W. R., Sai, F., & Mullin, J. T. (1989). Neonatal recognition of the mother's face. *British Journal of Developmental Psychology*, *7*(1), 3–15. https://doi.org/10.1111/j.2044-835X.1989.tb00784.x

Byrne, A. (2005). Introspection. *Philosophical Topics*, *33*(1), 79–104.

Byrne, A. (2011). Transparency, belief, intention. *Aristotelian Society Supplementary Volume*, *85*(1), 201–221. https://doi.org/10.1111/j.1467-8349.2011.00203.x

Carnap, R. (1935). Philosophy and logical syntax.

Carruthers, P., & Smith, P. K. (1996). *Theories of Theories of Mind*. Cambridge: Cambridge University Press.

Caskey, M., Stephens, B., Tucker, R., & Vohr, B. (2011). Importance of parent talk on the development of preterm Infant vocalizations. *Pediatrics*, *128*(5), 910–916. https://doi.org/10.1542/peds.2011-0609

Chrisman, M. (2007). From epistemic contextualism to epistemic expressivism. *Philosophical Studies*, *135*, 225–254.

Chrisman, M. (2012). Epistemic expressivism. *Philosophy Compass*, *7*(2), 118–126.

Cohn, J. F., & Tronick, E. (1983). Three-month-old infants' reaction to simulated maternal depression. *Child Development*, *54*(1), 185–193. https://doi.org/10.1111/j.1467-8624.1983.tb00348.x

Currie, G., & Ravenscroft, I. (2002). *Recreative minds: Imagination in philosophy and psychology*. Oxford: Oxford University Press.

Darwall, S. L. (2006). *The second-person standpoint*. Cambridge: Harvard University Press.

Delafield-Butt, J. T., & Gangopadhyay, N. (2013). Sensorimotor intentionality: The origins of intentionality in prospective agent action. *Developmental Review*, *33*(4), 399–425. https://doi.org/10.1016/j.dr.2013.09.001

Delafield-Butt, J. T., & Trevarthen, C. (2015). The ontogenesis of narrative: from moving to meaning. *Frontiers in Psychology*, *6*, 1157. https://doi.org/10.3389/fpsyg.2015.01157

Dennett, D. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences*, *1*(4), 568–570.
Dennett, D. (1989). *The intentional stance*. Cambridge. MA: MIT Press.
Descartes, R. (1977). *Meditaciones metafisicas. Con objeciones y respuestas*. (V. Peña, Ed. & Trans.), *Alfaguara*. Madrid.
Dunn, J., & Brophy, M. (2005). Communication, relationships and individual differences in children's understanding of mind. In J. W. Astington & J. Baird (Eds.), *Why language matters for theory of mind* (pp. 50–69). Oxford, UK: Oxford University Press.
Dunn, J., & Brown, J. R. (1993). Early conversations about causality: Content, pragmatics and developmental change. *British Journal of Developmental Psychology*, *11*(2), 107–123.
Edwards, C. P. (1987). Culture and the construction of moral values: A comparative ethnography of moral encounters in two cultural settings. In J. Kagan & S. Lamb (Eds.), *The emergence of morality in young children* (pp. 123–151). Chicago: University of Chicago Press.
Eilan, N. (2005). Joint attention, communication and mind. In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds* (pp. 1–33). Oxford: Oxford University Press.
Eilan, N. (2017). Knowing and understanding other minds: On the role of communication. In *Paper presented at Other Minds, Other Wills*. University of Chicago. Chicago, IL.
Evans, G. (1982). *The varieties of reference*. Oxford: Oxford University Press.
Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in hun ians from birth. *Proceedings of the National Academy of Sciences*, *14*(99), 9602–9605. https://doi.org/10.1073/pnas.
Fernández Castro, V. (2017a). Regulation, normativity and Folk Psychology. *Topoi*.
Fernández Castro, V. (2017b). *Talking the way to other minds: Assessment, conversation and Folk Psychology. (Tesis doctoral)*. Universidad de Granada, Granada (España).
Fernández Castro, V. (2019). Justification, conversation, and Folk Psychology. *Theoria. An International Journal for Theory, History and Foundations of Science*, *34*(1), 75–91.
Ferrer, J. (2014). El papel de la segunda persona en la constitución del autoconocimiento. *Daimon. Daimon Revista Internacional de Filosofía*, *62*, 71–86.
Finkelstein, D. H. (1999). On the distinction between conscious and unconscious states of mind. *American Philosophical Quarterly*, *36*(2), 79–100.
Finkelstein, D. H. (2003). *Expression and the inner*. Cambridge. MA: Harvard University Press.
Finkelstein, D. H. (2010). Expression and avowal. In K. D. Jolley (Ed.), *Wittgenstein: Key concepts* (pp. 185–198). London and New York: Routledge.
Finkelstein, D. H. (2012). From transparency to expressivism. In J. Conant & G. Abel (Eds.), *Rethinking Epistemology* (Vol. 2, pp. 101–118). Berlin/Boston: De Gruyter.
Fogel, A. (2011). *Infant development: A topical approach*. Cornwall-on-Hudson, NY: Sloan Publishing.
Foote, R. C., & Holmes-Lonergan, H. A. (2003). Sibling conflict and theory of mind. *British Journal of Developmental Psychology*, *21*(1), 45–58. https://doi.org/10.1348/026151003321164618
Franco, F. (2005). Infant pointing: Harlequin, servant of two masters. In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and*

other minds (pp. 129–164). Oxford: Oxford University Press.
Frápolli, M. J., & Villanueva, N. (2012). Minimal expressivism. *Dialectica*, *66*(4), 471–487.
Frege, G. (1953). *The foundations of arithmetic*. (J. L. Austin, Trans.). Evanston, Ill: Northwestern University Press.
Freud, S. (1911). Formulations on the two principles of mental functioning. In J. Strachey (Ed.), *Standard edition of the complete psychological works of Sigmund Freud, Volume XII (1911-1913): The case of Schreber, papers on technique and other works* (pp. 215–226). London: Hogarth Press and the Institute of Psycho-Analysis.
Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198237907.001.0001
Gallagher, S. (2001). The practice of mind: Theory, simulation or primary interaction. *Journal of Consciousness Studies*, *8*(5–7), 82–108.
Gallagher, S., & Hutto, D. (2008). Understanding others through primary interaction and narrative practice. In J. Zlatev, T. Racine, C. Sinha, & E. Itkonen (Eds.), *The Shared Mind: Perspectives on Intersubjectivity* (pp. 17–38). Amsterdam: John Benjamins.
Gibbard, A. (2003). *Thinking how to live*. Cambridge. MA: Harvard University Press.
Gibbard, A. (2012). *Meaning and normativity*. Oxford: Oxford University Press.
Goldman, A. (1989). Interpretation psychologized. *Mind & Language*, *4*(3), 161–185. https://doi.org/10.1111/j.1468-0017.1989.tb00249.x
Goldman, A. (1993). The psychology of folk psychology. *Behavioral and Brain Sciences*, *16*(1), 15–28.
Goldman, A. (2000). The mentalizing folk. In D. Sperber (Ed.), *Metarepresentations*. Oxford: Oxford University Press.
Goldman, A. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford: Oxford University Press.
Gómez, J.-C. (2005). Joint attention and the notion of subject: Insights from apes, normal children, and children with autism. In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds* (pp. 65–84). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199245635.003.0004
Gomila, A. (2001). La perspectiva de segunda persona: mecanismos mentales de intersubjetividad. *Contrastes, Vol. Suplementario 6. Monográfico Sobre Filosofía Actual de La Mente*.
Gomila, A. (2002). La perspectiva de la segunda persona de la atribución mental. *Azafea*, *4*, 123–138.
Gomila, A., & Pérez, D. (2017). Lo que la segunda persona no es. In D. Pérez & D. Lawer (Eds.), *La segunda persona y las emociones* (pp. 275–297). Buenos Aires: Editorial SADAF.
Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences*, *16*(01), 1. https://doi.org/10.1017/S0140525X00028636
Gopnik, A. (2003). The theory theory as an alternative to the innateness hypothesis. In L. Antony & N. Hornstein (Eds.), *Chomsky and his critics* (pp. 238–254). Oxford: Blackwell. https://doi.org/10.1002/9780470690024.ch10
Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories*. Cambridge. MA: MIT Press.

Gopnik, A., Meltzoff, A. N., & Kuhl, P. K. (1999). *The scientist In the crib: minds, brains, and how children learn*. New York: William Morrow & Co.
Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really Is a theory. *Mind & Language*, *7*(1–2), 145–171. https://doi.org/10.1111/j.1468-0017.1992.tb00202.x
Gopnik, A., & Wellman, H. M. (1994). The Theory Theory. In L. Hirschfield & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 257–293). New York: Cambridge University Press.
Gordon, R. M. (1986). Folk psychology as simulation. *Mind & Language*, *1*(2), 158–171. https://doi.org/10.1111/j.1468-0017.1986.tb00324.x
Gordon, R. M. (1995). Simulation without Introspection or Inference from me to you. In *Mental simulation: Evaluations and applications* (pp. 53–67).
Gordon, R. M. (2007). Ascent routines for propositional attitudes. *Synthese*, *159*(2), 151–165. https://doi.org/10.1007/s11229-007-9202-9
Gordon, R. M. (2009). Folk psychology and mental simulation. Retrieved from https://plato.stanford.edu/archives/fall2009/entries/folkpsych-simulation/
Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., & Parlato-Oliveira, E. (2015). Early development of turn-taking in vocal interaction between mothers and infants. *Frontiers in Psychology*, *6*, 1167. https://doi.org/10.3389/fpsyg.2015.01167
Gratier, M., & Trevarthen, C. (2008). Musical narratives and motives for culture in mother-infant vocal interaction. *Journal of Consciousness Studies*, *15*, 122–158.
Hains, S. M. J., & Muir, D. W. (1996). Infant sensitivity to adult eye direction. *Child Development*, *67*(5), 1940–1951. https://doi.org/10.1111/j.1467-8624.1996.tb01836.x
Harman, G. (1978). Studying the chimpanzee's Theory of Mind. *Behavioral and Brain Sciences*, *1*(4), 576–577.
Haviland, J. M., & Lelwica, M. (1987). The induced affect response: 10-week-old infants' responses to three emotion expressions. *Developmental Psychology*, *23*(1), 97–104. https://doi.org/10.1037/0012-1649.23.1.97
Hobson, R. P. (1993). *Autism and the development of mind*. Hove, England: Lawrence Erlbaum.
Hobson, R. P. (2002). *The cradle of thought*. London: Macmillan.
Hobson, R. P. (2005). What puts the jointness into joint attention? In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds* (pp. 185–204). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199245635.003.0009
Hopkins, R. (2009). Objectivity and realism in aesthetics. In S. Davies, K. M. Higgins, & R. Hopkins (Eds.), *A Companion to Aesthetics* (pp. 444–449). Oxford: Blackwell.
Hume, D. (1739). A Treatise of Human Nature. Retrieved August 28, 2018, from http://www.gutenberg.org/ebooks/4705
Hutto, D. (2004). The limits of spectatorial folk psychology. *Mind & Language*, *19*(5), 548–573.
Hutto, D. D. (2008). *Folk psychological narratives: The sociocultural basis of understanding reasons*. Cambridge. MA; London England: MIT Press.
Huyder, V., Nilsen, E., & Bacso, S. (2017). The relationship between children's executive functioning, theory of mind, and verbal skills with their own and others' behaviour in a cooperative context: Changes in relations from early to middle school-age. *Infant and Child Development*, *26*(6).

Jacobsen, R. (1996). Wittgenstein on self-knowledge and self-expression. *The Philosophical Quarterly*, *46*(182), 12–30.
Johnson, M. H., & Morton, J. (1991). *Biology and cognitive development: The case of face recognition*. Oxford: Blackwell Publishing Ltd.
Kaye, K. (1982). *The mental and social life of babies. How parents create persons*. Chicago: The University of Chicago Press.
Kokkinaki, T. (2010). Inter-subjectivity during free infant–father "protoconversation" and within-"protoconversation" pauses. *Early Child Development and Care*, *180*(1–2), 87–106. https://doi.org/10.1080/03004430903414737
Kokkinaki, T., & Kugiumutzakis, G. (2000). Basic aspects of vocal imitation in infant-parent interaction during the first 6 months. *Journal of Reproductive and Infant Psychology*, *18*(3), 173–187. https://doi.org/10.1080/713683042
Kripke, S. (1982). *Wittgenstein on rules and private language*. Cambridge. MA: Harvard University Press.
Kugiumutzakis, G. (1985). *The origin, development and function of the early infant imitation*. University of Uppsala, Sweden.
Kugiumutzakis, G. (1993). Intersubjective vocal imitation in early mother-infant interaction. In J. Nadel & L. Camaioni (Eds.), *New perspectives in early communicative development* (pp. 23–47). London: Routledge.
Kugiumutzakis, G. (1996). Le développement de l'imitation précoce de modèles faciaux et vocaux. *Enfance*, *49*(1), 21–25. https://doi.org/10.3406/enfan.1996.2980
Kugiumutzakis, G. (1998). Neonatal imitation in the intersubjective companion space. In S. Bråten (Ed.), *Intersubjective communication and emotion in early ontogeny* (pp. 63–88). Cambridge: Cambridge University Press.
Leekam, S. (2005). Why do children with autism have a joint attention impairment? In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds* (pp. 205–229). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199245635.003.0010
Lycan, W. G. (1987). *Consciousness*. Cambridge. MA: MIT Press.
Lycan, W. G. (1996). *Consciousness and experience*. Cambridge. MA: MIT Press.
Mahler, M. S. (1952). On Child Psychosis and Schizophrenia. *The Psychoanalytic Study of the Child*, *7*(1), 286–305. https://doi.org/10.1080/00797308.1952.11823164
Mahler, M. S., Pine, F., & Bergman, A. (1975). *The psychological birth of the human infant: symbiosis and individuation*. New York: Basic Books.
Malloch, S. N. (1999). Mothers and infants and communicative musicality. *Musicae Scientiae, Special Issue, 1999–2000*, 29–57.
Malloch, S. N., & Trevarthen, C. (2009). Musicality: Communicating the vitality and interests of life. In S. Malloch & Trevarthen (Eds.), *Communicative musicality: Exploring the basis of human companionship* (pp. 1–11). Oxford: Oxford University Press.
McDowell, J. (1982). *Wittgenstein on rules and private language*. Oxford: Blackwell.
McDowell, J. (1984). Wittgenstein on following a rule. *Synthese*, *58*, 325–363.
McDowell, J. (1991). Intentionality and Interiority in Wittgenstein. *Puhl*, 148–169.
McDowell, J. (1994). *Mind and world.* Cambridge. MA: Harvard University Press.
McDowell, J. (1998). Response to Crispin Wrigh. In C. Wright, B. C. Smith, & C. Macdonald (Eds.), *Knowing our own minds* (pp. 47–62). Oxford: The Clarendon Press.
McGeer, V. (2001). Psycho-practice, psycho-theory and the contrastive case of autism. How practices of mind become second-nature. *Journal of Consciousness Studies*, *8*(5–6), 109–132.

McGeer, V. (2007). The regulative dimension of folk psychology. In D. D. Hutto & M. Ratcliffe (Eds.), *Folk Psychology Re-Assessed* (pp. 137–156). Kluver/Springer. https://doi.org/10.1007/978-1-4020-5558-4_8

McGeer, V. (2015). Mind-making practices: the social infrastructure of self-knowing agency and responsibility. *Philosophical Explorations*, *18*(2), 259–281. https://doi.org/10.1080/13869795.2015.1032331

Meltzoff, A. N. (1985). The roots of social and cognitive development: Models of man's original nature. In T. M. Field & N. A. Fox (Eds.), *Social perception in infants* (pp. 1–30). Ablex Publishing Corporation.

Meltzoff, A. N. (1990). Foundations for developing a concept of self: The role of imitation in relating self to other and the value of social mirroring, social modeling, and self practice in infancy. In D. Cicchetti & M. Beeghly (Eds.), *The self in transition: Infancy to childhood* (pp. 139–164). Chicago, IL: University of Chicago Press.

Meltzoff, A. N. (2002). Elements of a developmental theory of imitation. In A. N. Meltzoff & W. Prinz (Eds.), *The imitative mind: Development, evolution and brain bases* (pp. 19–41). New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511489969.002

Meltzoff, A. N. (2007). The "like me" framework for recognizing and becoming an intentional agent. *Acta Psychologica*, *124*(1), 26–43. https://doi.org/10.1016/j.actpsy.2006.09.005

Meltzoff, A. N., & Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Development*, *54*(3), 702–709. https://doi.org/10.1111/j.1467-8624.1983.tb00496.x

Meltzoff, A. N., & Moore, M. K. (1989). Imitation in newborn infants: Exploring the range of gestures imitated and the underlying mechanisms. *Developmental Psychology*, *25*(6), 954–962. https://doi.org/10.1037/0012-1649.25.6.954

Meltzoff, A. N., & Moore, M. K. (1998). Infant intersubjectivity: Broadening the dialogue to include imitation, identity and intention. In S. Bråten (Ed.), *Intersubjective communication and emotion in early ontogeny* (pp. 47–62). Cambridge, UK: Cambridge University Press.

Moore, D. G., Hobson, R. P., & Lee, A. (1997). Components of person perception: An investigation with autistic, non-autistic retarded and typically developing children and adolescents. *British Journal of Developmental Psychology*, *15*(4), 401–423. https://doi.org/10.1111/j.2044-835X.1997.tb00738.x

Moore, G. E. (1903). *Principia ethica*. Cambridge: Cambridge University Press.

Morales, M., Mundy, P., Delgado, C. E. F., Yale, M., Messinger, D., Neal, R., & Schwartz, H. K. (2000). Responding to joint attention across the 6-through 24-month age period and early language acquisition. *Journal of Applied Developmental Psychology*, *21*(3), 283–298. https://doi.org/Doi 10.1016/S0193-3973(99)00040-4

Moran, R. A. (2001). *Authority and estrangement: An essay on self-knowledge*. Princeton: Princeton University Press.

Moran, R. A. (2003). Responses to O'Brien and Shoemaker. *European Journal of Philosophy*, *11*(3), 402–419.

Murray, L. (1980). *The young sensitivities and expressive capacities of infants in communication with their mothers*. PhD thesis, University of Edinburgh.

Murray, L., & Trevarthen, C. (1985). Emotional regulation of interactions between two-months-old and their mothers. In T. M. Field & N. A. Fox (Eds.), *Social perception in infants* (pp. 177–198). Norwood, NJ: Ablex.

Nadel, J., Carchon, I., Kervella, C., Marcelli, D., & Reserbat-Plantey, D. (1999). Expectancies for social contingency in 2-month-olds. *Developmental Science*, *2*(2), 164–173. https://doi.org/10.1111/1467-7687.00065

Nagy, E., & Molnar, P. (1994). Homo imitans or homo provocans? In search of the mechanism of inborn social competence. *Nternational Journal of Psychophysiology*, *18*(2), 128.

Nagy, E., & Molnar, P. (2004). Homo imitants or homo provocans? Human imprinting model of neonatal imitations. *Infant Behavior and Development*, *27*(1), 54–63. https://doi.org/10.1016/j.infbeh.2003.06.004

Nichols, S., & Stich, S. (2003). *Mindreading: An integrated account of pretence, self-awareness, and understanding of other minds*. Oxford: Oxford University Press.

Pérez, D., & Gomila, A. (2018). La atribución mental y la segunda persona. In T. Balmaceda & K. Pedace (Eds.), *Temas de Filosofía de la Mente: Atribución mental* (pp. 69–98). Editorial SADAF.

Piaget, J. (1952). *The origins of intelligence in children*. (M. Cook, Trans.). New York: International Universities Press.

Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a Theory of Mind? *Behavioral and Brain Sciences*, *1*(4), 515–526.

Price, H. (2011). *Naturalism without mirrors*. Oxford and New York: Oxford University Press.

Prinz, J. (2004). The Emotional Basis of Aesthetic Judgment. Houston: American Society for Aesthetics.

Putnam, H. (1975). The meaning of 'meaning.' In *Mind, language and reality. Philosophical Papers* (Vol. 2, pp. 215–271). Cambridge: Cambridge University Press.

Rakoczy, H., Brosche, N., Warneken, F., & Tomasello, M. (2009). Young children's understanding of the context relativity of normative rules in conventional games. *British Journal of Developmental Psychology*, *27*, 445–456.

Rakoczy, H., Warneken, F., & Tomasello, M. (2008). The sources of normativity: Young children's awareness of the normative structure of games. *Developmental Psychology*, *44*(3), 875–881.

Reddy, V. (2003). On being the object of attention: Implications for self-other consciousness. *Trends in Cognitive Sciences*, *7*(9), 397–402. https://doi.org/10.1016/S1364-6613(03)00191-8

Reddy, V. (2005). Before the "third element": Understanding attention to self. In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: communication and other mnds* (pp. 85–109). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199245635.003.0005

Reddy, V. (2008). *How infants know minds*. Cambridge: Harvard University Press.

Reddy, V. (2010). Engaging minds in the first year: The developing awareness of attention and intention. In J. G. Bremner & T. D. Wachs (Eds.), *The Wiley-Blackwell handbook of infant development* (2nd ed., pp. 365–393). Oxford, UK: Wiley-Blackwell. https://doi.org/10.1002/9781444327564.ch12

Reddy, V., & Trevarthen, C. (2004). What we learn about babies from engaging with their emotions. *Zero to Three*, *24*(3), 9–15.

Robinson, E. J., & Mitchell, P. (1995). Masking of children's early understanding of the representational mind: Backwards explanation versus prediction. *Child Development*, *66*(4), 1022–1039. https://doi.org/10.1111/j.1467-8624.1995.tb00920.x

Rochat, P. R. (2001). Social contingency detection and infant development. *Bulletin of the Menninger Clinic*, *65*(3), 347–360. https://doi.org/10.1521/bumc.65.3.347.19847

Roessler, J. (2005). Joint attention and the problem of other minds. In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds* (pp. 230–259). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199245635.003.0011

Rommetveit, R. (1974). *On message structure: A framework for the study of language and communication*. London: Wiley.

Ryle, G. (1949). *The concept of mind*. London: Hutchinson & Co.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36*(4), 393–414. https://doi.org/10.1017/S0140525X12000660

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology*, *18*(9), 668–671. https://doi.org/10.1016/j.cub.2008.03.059

Shoemaker, S. (1968). Self-reference and self-awareness. *Journal of Philosophy*, *65*(19), 555–567.

Shoemaker, S. (1994). Self-knowledge and «inner sense». *Philosophy and Phenomenological Research*, *54*, 249–314.

Slomkowski, C. L., & Dunn, J. (1992). Arguments and relationships within the family: Differences in young children's disputes with mother and sibling. *Arguments and Relationships within the Family: Differences in Young Children's Disputes with Mother and Sibling*, *28*(5), 919–924.

Sorce, J. F., Emde, R. N., Campos, J. J., & Klinnert, M. D. (1985). Maternal emotional signaling: Its effect on the visual cliff behavior of 1-year-olds. *Developmental Psychology*, *21*(1), 195–200. https://doi.org/10.1037/0012-1649.21.1.195

Sripada, C., & Stich, S. (2006). A framework for the psychology of norms. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind: Culture and cognition* (pp. 280–301). New York: Oxford University Press.

Stevenson, C. L. (1937). The emotive meaning of ethical terms. *Mind*, *46*, 14–31.

Stevenson, C. L. (1944). *Ethics and language*. New Haven: Yale University Press.

Stich, S. (1981). Dennett on intentional systems. In *Collected Papers. Volume !. Mind and Language 1972-2010* (pp. 71–90). Oxford: Oxford University Press.

Stich, S., & Nichols, S. (1992). Folk psychology: Simulation or tacit theory? *Mind and Language*, *7*, 35–71.

Stich, S., & Nichols, S. (2003). Folk Psychology. In S. Stich & T. Warfields (Eds.), *The blackwell guide of philosophy of mind* (pp. 235–255). Oxford: Blackwell.

Strawson, P. F. (1949). Truth. *Analysis*, *9*, 83–97.

Strawson, P. F. (1974). Freedom and resentment. In *Freedom and resentment and other essays* (pp. 1–28). London and New York: Routledge.

Stroud, B. (2011). Feelings and the ascription of feelings. *Teorema*, *30*(3), 25–33.

Taipale, J. (2016). Self-regulation and beyond: Affect regulation and the infant–caregiver dyad. *Frontiers in Psychology*, *7*, 1–13. https://doi.org/10.3389/fpsyg.2016.00889

Tesla, C., & Dunn, J. (1992). Getting along or getting your own way: the development of young children's use of argument in conflicts with mother and sibling. *Social Development*, *1*(2).

Tomasello, M. (1993). On the interpersonal origins of self-concept. In U. Neisser (Ed.), *The perceived self. Ecological and interpersonal sources of self knowledge* (pp.

174–184). New York: Cambridge University Press.
Tomasello, M. (1999). *The cultural origins of human cognition*. *The cultural origins of human cognition*. Cambridge. MA: Harvard University Press.
Tomasello, M. (2008). *Origins of human communication*. *MIT Press*. Cambridge. MA; London England: MIT Press.
Tomasello, M., & Jeffrey Farrar, M. (1986). Joint attention and early language. *Child Development*, *57*(6), 1454–1463. https://doi.org/10.1111/j.1467-8624.1986.tb00470.x
Trevarthen, C. (1974). Conversations with a two-month-old. *New Scientist*, *2nd May*, 230–235.
Trevarthen, C. (1977). Descriptive analyses of infant communicative behavior. In H. R. Schaffer (Ed.), *Studies in mother-infant interaction: The Loch Lomond symposium* (pp. 227–269). London: Academic Press.
Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. In M. Bullowa (Ed.), *Before speech: The beginning of interpersonal communication* (pp. 321–347). Cambridge: Cambridge University Press.
Trevarthen, C. (1982). The primary motives for cooperative understanding. In G. Butterworth & P. Light (Eds.), *Social cognition: Studies of the development of understanding* (pp. 77–109). Brighton, UK: Harvester Press.
Trevarthen, C. (1993a). The function of emotions in early infant communication and development. In J. Nadel & L. Camaioni (Eds.), *New perspectives in early communicative development* (pp. 48–81). London: Routledge.
Trevarthen, C. (1993b). The self born in intersubjectivity: The psychology of an infant communicating. In U. Neisser (Ed.), *The perceived self. Ecological and interpersonal sources of self knowledge* (pp. 121–173). New York: Cambridge University Press.
Trevarthen, C. (1998). The concept and foundations of infant intersubjectivity. In S. Bråten (Ed.), *Intersubjective communication and emotion in early ontogeny* (pp. 15–46). Cambridge: Cambridge University Press.
Trevarthen, C. (1999). Musicality and the intrinsic motive pulse: Evidence from human psychobiology and infant communication. *Musicae Scientiae, Special Issue, 1999–2000. Rhythms, Musical Narrative, and the Origins of Human Communication.*, 155–215.
Trevarthen, C. (2002). Making sense of infants making sense. *Intellectica 34*(1), 161–188.
Trevarthen, C. (2008). The musical art of infant conversation: Narrating in the time of sympathetic experience, without rational interpretation, before words. *Musicae Scientiae*, *12*(1_suppl), 15–46. https://doi.org/10.1177/1029864908012001021
Trevarthen, C., & Aitken, K. J. (2001). Infant intersubjectivity: Research, theory, and clinical applications. *Journal of Child Psychology and Psychiatry and Allied Disciplines*. https://doi.org/10.1017/S0021963001006552
Trevarthen, C., Delafield-Butt, J. T., & Schögler, B. (2011). Psychobiology of musical gesture: Innate rhythm, harmony and melody in movements of narration. In A. Gritten & E. King (Eds.), *Music and gesture ii* (pp. 11–43). Aldershot: Ashgate.
Trevarthen, C., & Hubley, P. (1978). Secondary intersubjectivity: Confidence, confiding and acts of meaning in the first year. In A. Lock (Ed.), *Action, gesture and symbol: The emergence of language* (pp. 183–229). London: Academic Press.
Trevarthen, C., & Reddy, V. (2007). Consciousness in Infants. In M. Velmans & S. Schneider (Eds.), *The Blackwell Companion to Consciousness* (pp. 41–57).

Oxford: Blackwell. https://doi.org/10.1002/9780470751466.ch4
Tronick, E. (1989). Emotions and emotional communication in infants. *American Psychologist*, *44*(2), 112–126.
Tronick, E., Als, H., Adamson, L., Wise, S., & Brazelton, T. B. (1978). The infant's response to entrapment between contradictory messages in face-to-face interaction. *Journal of the American Academy of Child Psychiatry*, *17*(1), 1–13. https://doi.org/10.1016/S0002-7138(09)62273-1
van der Meer, A. L., van der Weel, F. R., & Lee, D. N. (1995). The functional-significance of arm movements in neonates. *Science*, *267*(5198), 693–695.
Vega, J. (2011). Self-knowledge as knowldge? *Teorema*, *30*(3), 35–49.
Velleman, J. D. (2009). *How we get along*. Cambridge: Cambridge University Press.
von Hofsten, C. (1982). Eye-hand coordination in the newborn. *Developmental Psychology*, *18*(3), 450–461. https://doi.org/10.1037/0012-1649.18.3.450
Vygotsky, L. S. (1962). *Thought and language*. (E. Hanfmann & G. Vakar, Trans.). Cambridge: MIT Press.
Vygotsky, L. S. (1977). The development of higher psychological functions. *Soviet Psychology*, *15*(3), 60–73. https://doi.org/10.2753/RPO1061-0405150360
Wellman, H. M. (1990). *The Child's theory of mind*. Cambridge, MA: MIT Press.
Werner, H., & Kaplan, B. (1963). *Symbol formation: An organismic-developmental approach to language*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and the constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103–128.
Wittgenstein, L. (1953). *Philosophical Investigations*. (G. E. M. Anscombe, Trans.). Oxford: Wiley-Blackwell.
Wittgenstein, L. (1958). *The blue and brown books*. Oxford: Basil Blackwell.
Wittgenstein, L. (1981). *Zettel*. (G. E. M. Anscombe & G. H. von Wright, Eds., G. E. M. Anscombe, Trans.). Berkeley: University of California Press.
Wolff, P. H. (1987). *The development of behavioral states and the expression of emotions in early infancy: New proposals for investigation*. Chicago: University of Chicago Press.
Woodward, A. L. (2005). Infants understanding of joint attention. In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: communication and other minds* (pp. 110–128). Oxford: Oxford University Press.
Wright, C. (1984). Kripke' s account of argument against private language. In *Rails to infinity* (pp. 91–115). Cambridge. MA: Harvard University Press.
Wright, C. (1986). Rule-following, meaning and constructivism. In *Rails to infinity* (pp. 53–80). Cambridge. MA: Harvard University Press.
Wright, C. (1987). On making up one's mind: Wittgenstein on intention. In *Rails to infinity* (pp. 116–142). Cambridge. MA: Harvard University Press.
Wright, C. (1989a). Excerpts from a critical study of Colin McGinn' s Wittgenstein on meaning. In *Rails to infinity* (pp. 143–169). Cambridge. MA: Harvard University Press.
Wright, C. (1989b). Wittgenstein's rule-following considerations and the central project of theoretical linguistics. In *Rails to infinity* (p. 170.214). Cambridge. MA: Harvard University Press.
Wright, C. (1991). Wittgenstein's later philosophy of mind: Sensation, privacy and intention. In *Rails to infinity* (pp. 291–318). Cambridge. MA: Harvard University Press.
Wright, C. (1996a). The problem of self-knowledge (I). In *Rails to infinity* (pp. 319–

344). Cambridge. MA: Harvard University Press.
Wright, C. (1996b). The problem of self-knowledge (II). In *Rails to infinity* (pp. 345–373). Cambridge. MA: Harvard University Press.
Wright, C. (1998). Self-knowledge: The wittgensteinian legacy. In C. Wright, B. C. Smith, & C. Macdonald (Eds.), *Knowing our own minds* (pp. 13–45). Oxford: Oxford University Press.
Wright, C. (2001a). On Mind and world. In *Rails to infinity* (pp. 444–464). Cambridge. MA: Harvard University Press.
Wright, C. (2001b). *Rails to infinity*. Cambridge. MA; London. England: Harvard University Press.
Wright, C. (2015). Self-knowledge: the reality of privileged access. In S. C. Goldberg (Ed.), *Externalism, Self-Knowledge, and Skepticism* (pp. 49–74). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781107478152.004
Wright, C., Smith, B. C., & Macdonald, C. (1998). *Knowing our own minds: Essays in self-knowledge*. Oxford University Press.
Zawidzki, T. W. (2013). *Mindshaping: A new framework for understanding human social cognition*. Cambridge: Cambridge MIT Press, A Bradford Book.