Revista Argentina de Ciencias del Comportamiento, 2011, Vol. 3, N°1, 34-53

ISSN 1852-4206
www.psyche.unc.edu.ar/racc

RACC
Revista Argentina de Ciencias del Comportamiento

Revista Argentina de Ciencias del Comportamiento (RACC)

Autoconocimiento y Atribución de Estados Mentales en Teoría de la Mente

Skidelsky, Liza[*a]

[a] Universidad de Buenos Aires/CONICET, Buenos Aires, Argentina

Intencionalidad y Conciencia: Abordajes Recientes

Resumen

Muchos filósofos consideran que el fenómeno del autoconocimiento refleja la particularidad de que podemos saber lo que pensamos, creemos, deseamos, de una manera distinta a la manera en que conocemos los estados mentales de los otros. Esta es la tesis de la asimetría entre la primera y la tercera persona. En la literatura epistemológica se han ofrecido diversas propuestas para dar cuenta de esta asimetría. Sin embargo, a diferencia de la compatibilidad esperable entre ámbitos adyacentes, la literatura relacionada con la atribución y auto-atribución de estados mentales o, en general, lo que se suele llamar Teoría de la Mente, o bien no parece rescatar esta asimetría o bien los intentos por rescatarla le quitan un rol primordial a las distintas propuestas de Teoría de la Mente. En este trabajo se intentará mostrar esto en dos partes. En primer lugar, se abordará cómo ha sido defendida, en general, la tesis de la asimetría en la literatura epistemológica sobre el autoconocimiento. El objetivo de este apartado es ofrecer una geografía de las distintas propuestas. En segundo lugar, se analizará a grandes rasgos dos enfoques de Teoría de la Mente, la teoría-teoría y la teoría de la simulación, con el objetivo de mostrar por qué no dan lugar a la tesis de la asimetría, y explicitar algunas de las consecuencias que se desprenderían del intento de conciliar estos enfoques de la Teoría de la Mente con las propuestas epistemológicas que defienden la tesis de la asimetría.

Palabras Claves:
Autoconocimiento; Auto-Atribución; Teoría-Teoría; Teoría de la Simulación; Intencionalidad.

Abstract

Self-knowledge and attribution of mental states in Theory of Mind. Many philosophers consider that self-knowledge reflects the particularity that we can know what we think, believe, desire, in a different way in which we know the mental states of other people. This is the claim of an asymmetry between first and third person. Several approaches han been offered in the epistemological literature in order to account for this asymmetry. Nonetheless, unlike the expected compatibility between adjacent fields, the literature related to the attribution and self-attribution of mental states or, in general, what is called Theory of Mind, does not seem either to preserve this asymmetry or the attempt to preserve it undermines the fundamental role of the different Theory of Mind proposals. This paper will show this in two parts. Firstly, it addresses how the asymmetry thesis han been defended in the epistemological literature. The aim of this section is to offer a geography of the different approaches. Secondly, two proposals in Theory of Mind, the theory theory and simulation theory, will be evaluated in order to show why they do not account for the asymmetry thesis, and some of the consequences that would be gather from the attempt to conciliate these Theory of Mind proposals with the epistemological approaches that defend the asymmetry thesis will be analized.

Key Words:
Self-Knowledge; Self-Attribution; Theory Theory; Simulation Theory; Intentionality.

Recibido el 31 de Enero de 2011; Recibido la revisión el 1 de Marzo de 2011; Aceptado el 27 de Marzo de 2011

1. Introducción

Muchos filósofos consideran que el fenómeno del autoconocimiento refleja la particularidad de que podemos saber lo que pensamos, creemos, deseamos, de una manera distinta a la manera en que conocemos los estados mentales de los otros. Esta es la tesis de la asimetría entre la primera y la tercera persona. En la literatura epistemológica se han ofrecido diversas propuestas para dar cuenta de esta asimetría. Sin embargo, a diferencia de la compatibilidad esperable entre ámbitos adyacentes, la literatura relacionada con la atribución y auto-atribución de estados mentales o, en general, lo que se suele llamar Teoría de la Mente (TM), o bien no parece rescatar esta asimetría o bien los intentos por rescatarla le quitan un rol primordial a las distintas propuestas de TM. En este trabajo me ocuparé de mostrar esto en dos partes. En primer lugar, abordaré

* Enviar correspondencia a: Dra. Skidelsky, Liza
E-mail: lskidelsky@filo.uba.ar

cómo ha sido defendida, en general, la tesis de la asimetría en la literatura epistemológica sobre el autoconocimiento. El objetivo de este apartado es ofrecer una geografía de las distintas propuestas. En segundo lugar, analizaré a grandes rasgos dos enfoques de TM, la teoría-teoría (TT) y la teoría de la simulación (TS), con el objetivo de mostrar por qué no dan lugar a la tesis de la asimetría, y explicitar algunas de las consecuencias que se desprenderían del intento de conciliar estos enfoques de la TM con las propuestas epistemológicas que defienden la tesis de la asimetría.

2. La tesis de la asimetría y los enfoques epistemológicos

El fenómeno del autoconocimiento refleja la particularidad de que, al menos, los humanos podemos saber lo que pensamos, creemos, deseamos, sentimos y demás, de una manera distinta a la manera en que conocemos los estados mentales de los otros. Conocemos los estados mentales de los otros, básicamente, infiriéndolos a partir de su conducta. Sin embargo, en nuestro propio caso y en la mayoría de las veces, no necesitamos ni es relevante (aun cuando esté disponible) la evidencia de nuestra conducta. Así, podemos tener conocimiento de nuestros estados mentales pasados, presentes, futuros e hipotéticos. Podemos conocer nuestros estados mentales pasados (lo que creímos, deseamos, sentimos, etc.) cuando los recordamos y predecir nuestros estados mentales futuros cuando analizamos qué conducta seguir o qué decisión debemos tomar en situaciones futuras particulares. También podemos tener autoconocimiento acerca de estados mentales hipotéticos que tendríamos si tomáramos cierta decisión o tal curso de acción. Sin embargo, en estos casos no está claro que haya una manera especial de acceso a nuestros propios estados mentales de modo que se manifieste alguna asimetría entre la primera y la tercera persona. Por ejemplo, en la auto-atribución de estados mentales recientes parecen intervenir componentes "confabulatorios" para dar falsas explicaciones de la propia conducta (Nisbett & Ross, 1980). Asimismo, parece ser que el recuerdo de nuestros estados mentales pasados "se reconstruye" en función de nuestra condición actual (Levine, Prohaska, Burgess, Rice & Laulhere 2001; McFarland & Ross, 1987). Ambos procesos requieren procedimientos inferenciales, y así no habría asimetría.

Los casos paradigmáticos de autoconocimiento en los que se manifiesta claramente la asimetría son acerca de estados mentales *ocurrentes*. La clase de estados mentales ocurrentes abarca desde dolores hasta juicios. Dado que los primeros poseen ciertas propiedades "cualitativas" que no parecen poseer los estados de pensamiento, creencia, y en general, los estados intencionales, considero que merecen un tratamiento aparte (aunque se mencionarán a lo largo del trabajo). En consecuencia, a los fines de lo que me interesa mostrar aquí, consideraré que el fenómeno del autoconocimiento se acota al conocimiento de los estados intencionales ocurrentes que se manifiesta en juicios auto-atributivos o auto-atribuciones. Asimismo, asumiré, junto a la tradición filosófica, que las adscripciones, de primera persona del presente del indicativo, de estados mentales ocurrentes constituyen autoconocimiento, en cualquier sentido en que se sostenga que algo es conocimiento (por ejemplo, en tanto creencia verdadera justificada o creencia verdadera formada a partir de un mecanismo fiable). Asumiré, junto con los enfoques epistemológicos y de TM de los que me ocuparé, que el fenómeno del autoconocimiento puede describirse como una relación entre estados mentales de primer y segundo orden, que pueden expresarse, en el lenguaje natural, en actitudes proposicionales, i.e. "creo que *p*", y auto-adscripciones de actitudes proposicionales, i.e. "creo que creo que *p*", respectivamente. Finalmente, asumiré que los enfoques epistemológicos y de TM de los que me ocuparé son realistas acerca de los estados intencionales (en cualquier sentido en que se sostenga que estos estados existen).

Así, si se cree que hay algo especial en el autoconocimiento –en el sentido de que podemos saber lo que pensamos, creemos, deseamos, y demás, de una manera distinta de como conocemos los estados mentales de los otros–, la cuestión de la que hay que dar cuenta es, entonces, en qué radica lo especial del autoconocimiento, es decir, en qué radica esta asimetría entre la primera y la tercera persona. Las posiciones que intentan dar cuenta de esta asimetría pueden clasificarse en aquellas que postulan un *acceso epistémico privilegiado* a nuestros propios estados mentales y aquellas que consideran que el individuo posee una *autoridad especial de primera persona* con respecto a las auto-adscripciones (Skidelsky, 2008a). Según la primera perspectiva, el acceso epistémico puede tomar la forma de cierta seguridad epistémica (Descartes, 1641), o de algún método especial, como el "mirar hacia adentro" o introspección (Armstrong, 1981: Lycan, 1987, 1996) o el "mirar hacia fuera" o modelo de la transparencia (Dretske, 1994; Evans, 1982). Entre las posturas que sostienen que poseemos una autoridad de primera persona, se pueden ubicar a las perspectivas constitutivas, como la conceptual-pragmática (Shoemaker, 1994; Wittgenstein, 1988; Wright, 1998) y

del compromiso (Bilgrami, 1998; Moran, 1997), y al expresivismo (Bar-On 2004; Bar-On & Long 2001; Wittgenstein 1980, 1988).

Según uno de los enfoques constitutivos, la perspectiva conceptual-pragmática, la autoridad especial de primera persona con respecto a las auto-adscripciones de estados mentales radica en una cuestión relacionada con la comprensión de los conceptos involucrados en la práctica de la auto-atribución. La idea básica es que la práctica de la auto-atribución está parcialmente constituida por la inmunidad al desafío de la autoridad de la primera persona. Las auto-atribuciones poseen una presunción de verdad cuya negación por parte de un interlocutor, en contextos normales (i.e. cuando el auto-atributor no es insincero o insano), sería irrazonable o impropia, en la medida en que el interlocutor no estaría comprendiendo cómo funciona la práctica de la auto-atribución. En este sentido, la autoridad de primera persona consiste en una cuestión conceptual (o pragmática) acerca de cómo están constituidas nuestras auto-atribuciones (o las prácticas auto-atributivas). Este enfoque está sujeto, fundamentalmente, a dos tipos de críticas (sobre las que volveré más adelante): aquellas que sostienen que falla en justificar o explicar lo constitutivo de nuestras prácticas de auto-atribución y aquellas que sostienen que no captura el hecho de que el autoconocimiento es, en parte, un logro cognitivo.

El otro enfoque constitutivo, la perspectiva del compromiso, considera que la autoridad especial de primera persona se basa en que las auto-atribuciones implican un compromiso, por parte del individuo, con las actitudes que se auto-atribuye, a saber, el de actuar sobre la base de la verdad de las creencias que se auto-atribuye (Moran, 1997) o el de llevar a acciones o conclusiones que están sujetas, en la agencialidad responsable, a actitudes evaluativas (Bilgrami, 1998). En este sentido, la autoridad de primera persona es una condición necesaria para la posibilidad de ciertos rasgos básicos de los humanos: la agencialidad responsable –en el caso de Bilgrami– y la formación racional de creencias que guían la acción –en el caso de Moran. Según Bilgrami (1998), el teórico constitutivo está comprometido con el bicondicional que afirma que si uno tiene la creencia de que *p*, uno *debe* creer que cree que *p*, y viceversa. El autoconocimiento es una condición necesaria de la agencialidad responsable en el sentido de que si se está en un estado intencional individuado en virtud de que puede llevar a acciones (o conclusiones) que pueden ser objeto de actitudes evaluativas (por ejemplo, de crítica o culpa), el estar en ese estado lo compromete a uno a creer que se lo tiene (y viceversa). Así, bajo la condición de la agencialidad responsable, no es posible conceptualmente la ruptura de la conexión entre los estados de primer y segundo orden. El enfoque constitutivo también ha recibido críticas, por dar sólo un ejemplo, véase Brueckner (2003).

La otra postura mencionada que sostiene que poseemos una autoridad especial de primera persona es el expresivismo que considera que la autoridad especial de primera persona radica en que las auto-atribuciones tienen el tipo de autoridad que tienen otros modos de auto-expresión, como los gritos o sonrisas, que expresan aspectos de la psicología de los individuos. En particular, la línea interpretativa más ortodoxa le atribuye a Wittgenstein (1980, 1988) esta perspectiva (Malcolm, 1954; Strawson 1954). El expresivismo ha sido fuertemente criticado dado que parece borrar, o volver muy insustancial, la autoridad especial de primera persona (Wright, 1998). Puesto que las auto-atribuciones no serían aserciones genuinas y, en este sentido, no tendrían contenido proposicional, los individuos no podrían describir, de manera verdadera o falsa, sus propios estados mentales. Esta restricción contrasta fuertemente con la posibilidad que tienen los otros de atribuirnos, de manera correcta o incorrecta, estados mentales (aunque véase el neo-expresivismo de Bar-On & Long, 2001, y Bar-On, 2004 para un intento de superar esta objeción).

En el otro espectro de posiciones, dejando a un lado la bastante vapuleada de la infalibilidad epistémica cartesiana, se encuentran los enfoques de acceso especial. El enfoque de "mirar hacia adentro" sostiene que estamos, habitualmente, en una mejor posición que otras personas para conocer lo que pensamos, dado que tenemos un método especial para acceder a nuestros propios estados mentales que consiste en una especie de percepción interna (Armstrong, 1981; Lycan, 1987, 1996). Según esta perspectiva, accedemos a nuestros propios estados mentales porque tenemos un mecanismo cognitivo de monitoreo que produce estados de nivel superior acerca de estados de nivel inferior. Así, aquello que es especial en el autoconocimiento no sólo es epistémico sino empírico. Este enfoque ha sido criticado, por las posturas constitutivas, en base a que falla en hacer justicia a la asimetría no-contingente entre el autoconocimiento y el conocimiento de los otros. Por otro lado, los enfoques que postulan una especie de percepción externa o "mirar hacia afuera" plantean que el autoconocimiento se obtiene "mirando a través" de las creencias hacia la evidencia externa (Dretske, 1994; Evans, 1982). Así, para determinar si uno cree que *p*, hay que examinar la evidencia

relevante, en el mundo exterior, para afirmar la verdad de *p*. Estas posturas han recibido críticas tales como que sólo son útiles para dar cuenta de la formación de estados mentales (para lo cual, seguramente, se requiere considerar la evidencia del entorno), pero no se aplican a los casos comunes de autoconocimiento de estados mentales (Martin, 1998).

Ambas perspectivas, las de acceso especial y las constitutivas, consideran que el autoconocimiento consiste en algún tipo de conexión entre estados mentales de primer orden y de segundo orden, o auto-atribuciones, y se diferencian en el tipo de conexión que postulan entre ambos órdenes. Mientras que los enfoques epistémicos de acceso especial consideran que el autoconocimiento es un hecho *empírico* –en el sentido de que habría una conexión *causal* entre los estados mentales de primer orden y los de orden superior–, los enfoques constitutivos no-epistémicos (o parcialmente epistémicos, como los de Shoemaker, 1994 o Peacocke, 1998) consideran que el autoconocimiento no es, básicamente, un hecho empírico sino *conceptual* –en el sentido de que habría una conexión *a priori e*ntre los estados mentales de primer orden y los de orden superior o las auto-atribuciones. Así, estos últimos enfoques parecen tener la consecuencia de que el autoconocimiento, al no estar basado, al menos, en nada empírico, se torna "insustancial" o no es un logro cognitivo.

Se puede entender que algo es un logro cognitivo, al menos, de dos maneras (Fricker, 1998). Una manera es entenderlo como el producto de algo que la persona hace (o intenta hacer), de un esfuerzo de algún tipo para lograr ese conocimiento, por ejemplo, atender cuidadosamente, focalizar la atención, concentrarse, hacer una observación o inferencia. En este sentido, nuestro conocimiento empírico siempre es un logro cognitivo. Esto se opone, en el caso del autoconocimiento, a la idea de simplemente encontrarse a uno mismo teniendo una creencia o verse inclinado a tener una creencia. La otra manera es entender el logro cognitivo como teniendo una capacidad de rastreo de estados mentales de nivel inferior por estados mentales de nivel superior, ontológicamente diferentes, de manera que el mecanismo aprehende hechos independientes. Esto se opone a la idea de que habría algún tipo de conexión conceptual o *a priori* entre ambos tipos de estados mentales.

El enfoque de acceso especial de "mirar hacia adentro" se basa en la idea de que el autoconocimiento es un logro cognitivo en el segundo de los sentidos mencionados. Poseemos un mecanismo de escaneo y monitoreo cerebral del flujo de la información cognitiva que resulta en estados conscientes de orden superior acerca de estados de orden inferior. Según Armstrong, la introspección consiste en un "proceso de auto-escaneo en el cerebro" (1968, p. 324) que procede en un nivel cognitivo inferior (hoy podría decirse, en el nivel subpersonal) y es rápido, constante y simple, esto es, completamente no-inferencial (similar a la detección de presión en nuestra espalda). Esto no excluye, por supuesto, que pueda haber actos de introspección deliberada (Armstrong, 1999). Según Lycan (1987, p. 72): "Estar consciente, de manera introspectivamente activa, de que P es tener un escáner interno funcionando que opera sobre algún estado que es en sí mismo psicológico y entrega información acerca de ese estado a la propia unidad de control ejecutiva". En estos procesos intervienen mecanismos atencionales, de manera que, a diferencia de Armstrong, el escaneo requeriría recursos cognitivos más demandantes (Lycan, 1996). Desde ambas perspectivas, los estados de orden superior son el producto de procesos que operarían sobre estados de primer orden. Ambos tipos de estados son ontológicamente distintos, de manera que la conexión entre ambos es causal y contingente. Sin embargo, el acceso a los estados de orden inferior no es inferencial. Para tener autoconocimiento, el individuo no tiene que acceder a las conexiones causales; el mecanismo, simplemente, dado un estado de nivel inferior produciría estados de nivel superior (i.e., entre los efectos causales típicos de las actitudes proposicionales estaría el de producir típicamente un estado de actitud proposicional de orden superior). Esto constituye un logro cognitivo en el sentido de que saber que se cree que *p* es producto de un mecanismo causal fiable de monitoreo interno. Así, la fiabilidad de la autoridad de primera persona no es una cuestión basada en principios normativos de nuestra práctica de atribución y auto-atribución (como en el caso de la perspectiva conceptual-pragmática) o en una conexión necesaria (como en el caso de la perspectiva del compromiso) que relaciona las auto-atribuciones con los estados mentales de manera *a priori*, sino que la explicación es empírica y completamente naturalista: hay un mecanismo cognitivo fiable que produce el acceso a nuestros estados mentales.

Se podría pensar que los enfoques de acceso especial y los constitutivos son compatibles y posiblemente complementarios. El enfoque de acceso especial parece concernir a los estados mentales de autoconocimiento en sí mismos, mientras que el enfoque constitutivo parece concernir a las atribuciones (lingüísticas) de autoconocimiento. Es más, alguien podría decir que uno explica al otro. Es precisamente

porque los estados de autoconocimiento tienen esta naturaleza peculiar –caracterizable en términos de acceso privilegiado– que las atribuciones de autoconocimiento se comportan como lo hacen –caracterizables en términos de autoridad. Por razones que desarrollé en otro lugar considero que ambas posturas son irreconciliables (Skidelsky, 2008a). Esto supone que ambos tipos de enfoque están en contienda. Dicho brevemente, si bien se puede pensar que los enfoques de acceso especial se ocupan de los juicios auto-atributivos mientras que los defensores de los enfoques constitutivos se ocupan de las auto-atribuciones, tomando en cuenta que ambos enfoques son realistas acerca de los estados mentales, estas últimas pueden considerarse como expresiones (sinceras) de los primeros. De modo que ambos enfoques intentan dar cuenta de ambos, los juicios auto-atributivos y las auto-atribuciones (Fricker, 1998). Además, la misma práctica filosófica sobre el problema del autoconocimiento no ofrece indicios, al menos que yo sepa, de que ambos contrincantes se acusen mutuamente de "cambio de tema". En apoyo de esto, una defensora de una postura conciliatoria afirma lo siguiente: "Las teorías de acceso especial y las teorías del artefacto de la gramática son explicaciones alternativas vindicadoras del fenómeno mínimo. Dan explicaciones alternativas de la fiabilidad de nuestras auto-adscripciones de estados mentales, no-inferidas psicológicamente" (Fricker, 1998, p. 160). También la siguiente expresión de un defensor del enfoque constitutivo apoya esta idea: "*El asunto es este*: dada la tensión intolerable entre el modelo perceptivo y el ideal constitutivo, sólo uno de ellos puede servir en nuestro enfoque del autoconocimiento. De manera que, ¿cuál de ellos debe ser?" (Bilgrami, 1998, p. 208).

Ahora bien, luego de esta geografía de respuestas ofrecidas al problema del autoconocimiento, me interesa ver a continuación si los dos enfoques contemporáneos más desarrollados y sistemáticos en TM, la TT y la TS, logran acomodar la asimetría apelando a alguna variante de los enfoques epistemológicos presentados, y si esto es así qué consecuencias trae para las respectivas teorías de la mente. Tal como veremos, los enfoques de TT y TS apelan efectivamente a alguno/s de los enfoques mencionados, de allí que estos se hayan presentado con cierto detalle. También los enfoques de TT y TS se abordarán con cierto detalle de manera que se puedan apreciar claramente las relaciones entre estos enfoques y las propuestas epistemológicas. En este sentido, este trabajo también pretende ofrecer una perspectiva panorámica del estado de la cuestión sobre el tema del autoconocimiento y la auto-atribución en TM. En lo que sigue hablaré en términos de "autoridad de primera persona", ya sea que se base en cuestiones empíricas o conceptuales, como aquello en lo que se asienta la asimetría entre la primera y la tercera persona.

3. La tesis de la asimetría y la teoría de la mente

Los seres humanos solemos describir, explicar y predecir la conducta de los otros y la nuestra misma apelando a estados mentales. Así, si queremos explicar la conducta de alguien que cruza hacia la otra vereda cuando ve venir un perro, seguramente apelaremos a su creencia de que viene un perro y a su deseo de evadirlo, probablemente, por su temor a los perros. De la misma manera, apelando a estados mentales, podemos predecir la conducta de evitar a los perros. Se dice que este tipo de explicaciones son posibles gracias a una psicología de sentido común, o TM, que todos tenemos y usamos para facilitar nuestra comunicación, comprensión mutua y cooperación social. El estudio de la TM se ocupa, principalmente, de dar cuenta de cómo atribuimos estados mentales a los otros y a nosotros mismos en el curso de describir, explicar y predecir la conducta. Esto es, se ocupa de "las habilidades y recursos que la gente utiliza rutinariamente para la anticipación, explicación, y coordinación social del comportamiento" (Gordon, 2009, p. 1).

Según Goldman (2000a), el abordaje de la TM abarca el estudio de la comprensión ordinaria de los estados mentales y el uso de conceptos mentales, de manera que habría, básicamente tres preguntas que contestar: cómo comprendemos ordinariamente los estados mentales, cómo se atribuyen y auto-atribuyen estados mentales, y cómo se adquieren los conceptos de estados mentales y las habilidades para aplicarlos. Dados estos objetivos, Goldman (2000a) considera que el estudio de la TM puede verse como una rama de la epistemología descriptiva, en tanto que la epistemología se ocupa, en parte, de cómo se forman las creencias y la TM involucra la formación de creencias acerca de estados mentales. En lo que sigue me ocuparé particularmente de la segunda pregunta puesto que es allí donde se plantea la cuestión epistemológica de la asimetría entre la primera y la tercera persona respecto de la atribución de estados mentales. Si no se desea sostener una posición de dependencia de la TM de la epistemología como la de Goldman, al menos, no se puede negar que habría un ámbito de estudio de la TM que se solaparía con la epistemología del autoconocimiento y es la cuestión de la auto-atribución de estados mentales.

La idea de autoconocimiento, como mencioné en

§1, consiste en que simplemente "conocemos" (signifique esto lo que sea que signifique "conocer", por ejemplo, sea la concepción tradicional tripartita de creencia verdadera justificada o creencia verdadera formada de manera fiable) nuestros propios estados mentales. De manera que cuando tenemos, por ejemplo, una creencia (consciente), estamos en posición de juzgar que tenemos esa creencia (con ese contenido). En §2 hemos visto distintos enfoques epistemológicos que intentan dar cuenta de esta autoridad de primera persona en el acceso/conocimiento de los propios estados mentales. En estos enfoques no se resalta la diferencia entre la auto-atribución y el autoconocimiento. Así, "la auto-atribución es un tema prominente en filosofía, discutido comúnmente bajo el título de "autoconocimiento"" (Goldman, 2006, p. 223). Tampoco en TM, hasta donde sé, se explicita esta distinción (véase tan sólo como muestra Carruthers 1996a, en donde ambos términos se toman indistintamente). Más allá de que hay características que corresponden a alguno de los términos y no al otro, la dependencia entre ambos fenómenos es clara: en principio, no parece ser posible auto-atribuirse un estado mental sin saber que lo tenemos ni decir que conocemos nuestro propio estado mental sin, de alguna manera, auto-atribuírnoslo.

Veamos, entonces, las respuestas que ha recibido la pregunta acerca de cómo se atribuyen y auto-atribuyen estados mentales en TM, atendiendo particularmente a este último fenómeno. Estas pueden agruparse, a grandes trazos, en aquellas que apelan básicamente a un cuerpo de conocimiento que posee el atributor o, como se suele decir, están "guiadas por teoría", como la TT, o aquellas que apelan a mecanismos que pone en uso el atributor o, como se suele decir, están "guiadas por procesos", como la TS (Goldman, 1989). Obviamente, ésta no es una distinción tan tajante puesto que dentro de los enfoques tradicionales de TT y TS, la mayoría de las teorías aceptan aspectos del otro enfoque. Así, la versión de TT de Carruthers (1996a) acepta la "simulación dentro de una teoría", y la versión simulacionista de Gordon (1992) trata a las generalizaciones de la psicología *folk* en el contexto de una simulación, por dar sólo algunos ejemplos. Sin embargo, estos enfoques mantienen ya sea la teoría o la habilidad de simulación como lo fundamental en la atribución de estados mentales, y en la explicación y predicción de la conducta. Por supuesto que hay otros enfoques disponibles de TM (como la teoría de la interacción de Gallagher, 2001, 2005 o la hipótesis de la práctica narrativa de Hutto, 2008), pero me interesa aquí evaluar las respuestas ofrecidas por los dos enfoques tradicionales. Puesto que las teorías de ambos enfoques varían en ciertos aspectos según el autor, me ocuparé de las propuestas que suelen considerarse, en la literatura sobre el tema, como las más desarrolladas y representativas de TT y TS.

3.1. La teoría-teoría

A grandes rasgos, la TT sostiene que al atribuir estados mentales a otros y a nosotros mismos empleamos conceptos teóricos (no-observacionales) que forman parte de (o se definen a partir de) una teoría psicológica de sentido común. Dependiendo de las distintas versiones de TT, esta teoría *folk* puede adoptar la forma de un conjunto de generalizaciones o leyes (lo que se suele denominar "funcionalismo de sentido común", Lewis, 1972) o la forma de cualquier otra teoría científica (Churchland, 1988; Gopnik, 1993; Gopnik & Meltzoff, 1997; Gopnik & Wellman, 1994). La atribución se realiza por un proceso inferencial a partir de la observación de la conducta y de los eventos del entorno, i.e. reconociendo el rol causal-explicativo de las atribuciones de acuerdo con las generalizaciones de la teoría. Las distintas versiones de la TT difieren según se proponga que la teoría se adquiere a través de un proceso de formación y cambio de teoría (Gopnik, 1996), transmisión cultural (Churchland, 1988), o desarrollo innato de un mecanismo de dominio específico a partir del desencadenamiento del entorno (Leslie, 1987, 1994; Leslie & German, 1995). La atribución y auto-atribución involucra el dominio de generalizaciones teóricas acerca de relaciones causales-inferenciales entre los estímulos sensoriales, otros estados mentales y las respuestas conductuales. Así, las personas utilizan un cuerpo de conocimiento acerca de la psicología que puede ser considerado como una teoría sobre ese dominio.

Según Lewis, la teoría se formula así: "Junte todas las perogrulladas que pueda pensar en relación a las relaciones causales entre los estados mentales, los estímulos sensoriales y las respuestas motoras…Sólo incluya perogrulladas que son conocimiento común entre nosotros –todos las saben, todos saben que los otros las saben, y así en más" (1972, p. 256). La forma general de estas perogrulladas o, dicho de otro modo, las típicas generalizaciones causales-inferenciales de la teoría psicológica *folk*, consiste en: "Cuando alguien está en una tal y cual combinación de estados mentales y recibe estímulos sensoriales del tipo tal y cual, tiende con tal y cual probabilidad a ser causado por ello mismo a pasar a tales y cuales estados mentales y producir tales y cuales respuestas" (p. 256). Otros filósofos como Churchland (1988, pp. 58-59) han dado ejemplos específicos de estas perogrulladas: "Las personas que

quieren que P y creen que Q será suficiente para dar lugar a P, y no tienen deseos conflictivos o estrategias preferidas, intentarán dar lugar a Q" o más específico aun: "Las personas con dolor tienden a aliviar ese dolor". Así, si queremos, por ejemplo, explicar tanto la conducta de alguien que está tomando un remedio como la nuestra misma en esa situación, dadas ciertas condiciones de trasfondo y la (auto-)adscripción de ciertos estados mentales, subsumimos ese caso en alguna/s de las generalizaciones de la teoría, por ejemplo, que las personas con dolor tienden a aliviarlo. La mayoría de las versiones de TT permiten que este uso de la teoría psicológica *folk* sea inconsciente. Incluso, el atributor no necesita estar al tanto de que posee esta teoría (la teoría es poseída a la manera, quizá, del conocimiento lingüístico tácito que el enfoque chomskiano plantea, cf. Davies & Stone, 1995; Stich & Nichols, 1995).

Una característica primordial de la TT es que es un planteo desde el punto de vista de la tercera persona. La atribución se realiza básicamente en función de observación e inferencia. Puesto que se aplica teoría tanto para la atribución a otros como para la auto-atribución, no habría así una asimetría entre la primera y la tercera persona. Ambos tipos de atribuciones son inferenciales en tanto que están mediadas por la teoría psicológica de sentido común que todos poseeríamos. El aparente acceso privilegiado a nuestros propios estados que parecemos percibir como un dato se explica, en verdad, por medio de una analogía entre esta ilusión del acceso privilegiado y la ilusión del experto (Gopnik, 1993). En el caso del experto ocurre el fenómeno de la experiencia directa e inmediata como producto de una larga y constante puesta en práctica de una teoría. De la misma manera, nuestra pericia en las mentes y el comportamiento de los otros es muy buena y aun mayor en nuestros propios casos debido a que convivimos con nosotros. Dados los efectos de la especialización registrados en la percepción (casos de jugadores de ajedrez que reportan "ver" que un rey aislado es vulnerable en vez de "calcular", Chase & Simon, 1973; De Groot, 1978), no somos conscientes de las inferencias que realizamos, e interpretamos las experiencias cargadas de teoría como percepciones directas de nuestros propios estados psicológicos.

Así, según esta versión estricta de TT, es evidente que este enfoque no reconoce la asimetría entre el acceso a los propios estados mentales y los de los otros (o mejor dicho, la reconoce como una mera ilusión). En este sentido, una consecuencia peculiar es que se encuentra en la misma posición que los enfoques conductistas que consideran que no hay diferencia entre estos tipos de atribuciones (Ryle, 1949). Esto es una consecuencia peculiar porque se supone que los defensores de la TT son críticos del conductismo sosteniendo, en general, posturas funcionalistas en relación a la naturaleza de la mente. Sin embargo, ya sea que se considere que la mente es una caja negra y en este sentido nada que posea cumple un rol explicativo en la atribución de estados mentales, o que la mente posee algo, en este caso, teoría, y ésta tiene un rol explicativo en la atribución, ambas perspectivas no recogen la asimetría que la mayoría de los epistemólogos postulan como una característica fundamental del fenómeno del autoconocimiento.

Tampoco en la TT parecería haber lugar para ningún tipo de enfoque naturalista disponible del autoconocimiento como los de acceso especial. No habría lugar para un procedimiento directo no-inferencial de "mirar hacia adentro", como el mecanismo de auto-monitoreo (porque, en principio, los estados mentales son entidades abstractas inobservables), ni tendría sentido otro procedimiento inferencial, pero de "mirar hacia afuera". Lo único que se requiere para tener acceso a nuestros propios estados mentales, según una defensa estricta de la TT, es teoría. Sin embargo, un notorio defensor de la TT, como Carruthers (1996a), apela tanto a un enfoque de acceso especial, la introspección, como al enfoque de "mirar hacia afuera", que hemos visto en §2, en tanto estrategias que permitirían el acceso a nuestros propios estados mentales. Si bien Carruthers (2009, 2010) defiende actualmente una versión más bien estricta (en el sentido de ser casi completamente inferencial), vale la pena analizar uno de los pocos y más desarrollados intentos en TT por acomodar la asimetría.

Carruthers (1996a) propone que nuestras creencias acerca de nuestros estados *ocurrentes* se obtienen de manera intuitiva y no-inferencial (esto último es en el sentido de excluir inferencias conscientes de nivel personal), siendo así una especie de conocimiento cuasi-perceptivo. De modo que tenemos acceso a nuestros estados ocurrentes a través de la introspección de cualidades distintivas de esos estados. Para el caso de estados mentales *permanentes*, primero se requeriría un ascenso semántico para hacer ocurrente ese estado y luego así poder realizar introspección. El ascenso semántico consiste en lo que hemos visto en §2 como el enfoque de "mirar hacia afuera". Usando un ejemplo de Carruthers (1996a), para determinar si creo que la desforestación mundial será terrible, me hago la pregunta de primer orden de si es el caso de que la desforestación mundial será terrible. Si me veo inclinado a contestar afirmativamente, entonces realizo

un ascenso semántico de manera que estoy en condiciones de emitir un juicio ocurrente de segundo orden de que creo que la desforestación mundial será terrible. Y es de este último juicio ocurrente del cual tengo un conocimiento cuasi-perceptual. Así, para el caso de los estados mentales ocurrentes sólo se aplica introspección y para el caso de los estados mentales permanentes se aplica el ascenso semántico y luego la introspección. Aparentemente, nada de esto hay que hacer para el caso de la atribución de estados mentales a otros, para el cual sólo se utiliza la teoría psicológica *folk*.

Ahora bien, hay dos problemas con esta propuesta. La primera se relaciona con aquello que se dice que se introspecta, y la segunda se relaciona con la cuestión de en qué medida esta propuesta se inscribe en la TT. Respecto del primer problema, Carruthers (1996a) sostiene que la teoría psicológica *folk* no es suficiente para individuar estados mentales ocurrentes dado que lo que ofrece es conocimiento teórico general del tipo de las generalizaciones que hemos visto más arriba, pero no conocimiento teórico específico relacionado con qué puede estar inclinada a hacer gente con creencias y deseos con contenidos específicos. Lo que propone Carruthers es que individuamos juicios porque podemos introspectar la ocurrencia de sus vehículos lingüísticos (más específicamente, del lenguaje natural). Así, cuando ocurre el pensamiento de que la desforestación mundial será terrible, sé inmediatamente lo que acabo de pensar porque reconozco el vehículo de ese pensamiento: la forma de las palabras (del lenguaje natural) empleadas. Esta idea de que se puede hacer introspección, en el nivel personal, de los vehículos subpersonales de los pensamientos está desarrollada, en especial, en Carruthers (1996b, 1998).

Lo único que diré aquí, puesto que me he ocupado de criticar en detalle esta idea de que se puede hacer introspección de los vehículos de los estados mentales en Skidelsky (2009), es que las hipótesis acerca de los vehículos de los estados mentales son hipótesis acerca de la arquitectura o maquinaria cognitiva. El conocimiento del funcionamiento de la arquitectura cognitiva requiere observación controlada y experimentación; esto significa que el conocimiento que brindan estas hipótesis requiere investigación empírica. Pero como en el nivel personal no se acceden a los formatos de las representaciones mentales, no se puede querer decir que el vehículo de los pensamientos es el lenguaje natural, sino tan sólo que "nos parece" que es el lenguaje natural. En este caso, accederíamos a los aspectos fenoménicos (de nivel personal) asociados al lenguaje natural (i.e., que "nos parece" oír oraciones en lenguaje natural, en el mismo sentido en que nos parece, por ejemplo, ver imágenes en nuestro pensamiento visual). Esto, por supuesto, no dice absolutamente nada acerca de los vehículos de esos pensamientos en el nivel de la maquinaria cognitiva. Así como para el caso de las imágenes se ha propuesto que a nivel subpersonal los vehículos podrían ser descripciones (Pylyshyn, 2002), y para el caso de los conceptos se ha propuesto que los vehículos serían estados perceptivos (Prinz, 2002), la cuestión de los vehículos de los pensamientos (ocurrentes) no puede dirimirse en un plano (pura o meramente) personal, por mera introspección.

Carruthers (2002) sostiene específicamente que los vehículos serían representaciones de la forma lógica y la forma fonológica de las expresiones lingüísticas, productos de la facultad del lenguaje. Sin embargo, no hay ningún sentido en que podamos tener acceso introspectivo a estas formas. Por ejemplo, Fodor (1998, p.65) afirma: "Lo más cercano que podría ser pensar en inglés sería pensar en alguna regimentación del inglés libre de ambigüedad (quizá en fórmulas de lo que Chomsky llama "FL"). (...) Quizá, por ejemplo, lo que hay en su cabeza cuando piensa *que todos aman a alguien*, según la interpretación en la que "todos" tiene alcance amplio, es "todo$_x$ algún$_y$ (x ama a y)". Esta (…) es la clase de estructura lingüística correcta para ser el vehículo de un pensamiento. Pero (dilema) seguramente no es la clase de estructura lingüística que es dada a la introspección de alguien; si lo fuera, no hubiéramos necesitado que Frege nos enseñe sobre variables ligadas". Machery (2005), en su crítica a esta idea de Carruthers, sostiene lo mismo cuando formula la "tesis de la ceguera de la introspección" que afirma que el hecho introspectivo no puede ser evidencia de que nuestros pensamientos ocurrentes se expresan en lenguaje natural porque no tenemos acceso a la estructura sintáctica de los vehículos simbólicos de nuestros pensamientos y es esta propiedad, justamente, la marca de los vehículos lingüísticos.

Ahora bien, el problema anterior se relaciona con aquellas propiedades de los estados mentales que Carruthers (1996a y b, 1998, 2002) sostiene que se introspectan. Si, como pienso, Carruthers está equivocado, igualmente se podrían ofrecer otras propiedades menos problemáticas para la introspección de manera de salvaguardar este modo de acceso a nuestros propios estados mentales. Si este fuera el caso, de todas formas su propuesta estaría sujeta al segundo problema mencionado. Carruthers afirma que si bien el proceso de adquirir autoconocimiento puede lograrse sin que los principios de la psicología *folk* sean

accesibles al sujeto, no obstante lo que se reconoce cuando se reconoce que se está en un estado mental con un contenido particular es un estado que tiene una caracterización o rol psicológico-*folk* particular. Esto es, el conocimiento de que estoy en un estado mental en particular implica teoría.

Como veremos en §3.2.2, la versión de TS de Goldman para la auto-atribución también incorpora a la introspección, de manera que ésta puede adosarse tanto a la TT como a la TS y, en este sentido, no parece ser una propuesta propia de alguna de estas teorías, sino un complemento que se añade debido a las fallas explicativas de estas teorías para la auto-atribución (quizá, por la necesidad de respetar la asimetría epistemológica). Como también veremos más adelante, para la introspección propuesta por Goldman no hace falta teoría, es decir, no hace falta reconocer mi propio estado mental como teniendo un rol psicológico-*folk* particular. Esto quiere decir que la afirmación de Carruthers no parece estar bien motivada. No se ve por qué se requiere reconocimiento de un rol psicológico-*folk* para reconocer nuestros propios estados mentales (es más, según Goldman 2006, la introspección no puede detectar propiedades funcionales). Si esto es así, la explicación carrutherseana de la auto-atribución ciertamente respeta la asimetría (porque conocemos nuestros propios estados ocurrentes de manera directa por introspección, un método que sólo se aplica a los propios estados mentales, mientras que atribuimos estados mentales a los otros por medio de inferencias a partir de una teoría *folk*), pero al costo de no ser una explicación estrictamente en términos de TT o que le otorgue al cuerpo de conocimiento de la psicología de sentido común un rol central o fundamental.

Ahora bien, mientras que versiones de la TT, como la anterior, que intentan complementarse con enfoques naturalistas disponibles de la asimetría, parecen quitarle un rol fundamental a la TT en la auto-adscripción, parecería que la TT podría, en principio, ser compatible con un enfoque constitutivo, no-naturalista, como el conceptual-pragmático o el de compromiso, introducidos en §2, sin dejar de tener un rol fundamental. Así, aunque apliquemos teoría para atribuirnos nuestros propios estados mentales, esto no excluiría que haya algo especial en el autoconocimiento. Lo especial del autoconocimiento sería la autoridad de primera persona, y ésta radicaría, según el enfoque conceptual-pragmático, en que las auto-atribuciones tienen una presunción de verdad. Tal como vimos en §2, esto significa que en contextos conversacionales normales, cuando los hablantes no son insinceros o insanos, sería irrazonable o impropio que un interlocutor niegue la auto-atribución de un individuo; si lo hace, eso quiere decir que no comprende cómo funcionan las afirmaciones de este tipo. Así, la autoridad de primera persona radicaría en que la práctica de la atribución y la auto-atribución está, parcialmente, constituida por una presunción de verdad, esto es, por la inmunidad de las auto-atribuciones al desafío de los otros. En la práctica de la atribución, tratamos al auto-atributor como la autoridad por *default* de sus propios estados mentales.

Wright (1998) considera que para dar cuenta de la autoridad de la primera persona no basta con afirmar que éste es un rasgo constitutivo del discurso psicológico que utilizamos en la práctica de la atribución de estados mentales o decir, solamente, que la asimetría entre la primera y la tercera persona es una cuestión que pertenece a la gramática del juego de lenguaje de la psicología de sentido común. Esto, simplemente, convierte este rasgo de la práctica en un rasgo primitivo, cuando en verdad requiere una explicación. Wright considera que es legítimo preguntar: ¿qué es lo que hace que las auto-atribuciones sean confiables o verdaderas por *default*? Así, Wright (1989) sostiene que la autoridad que se concede a las creencias propias de un sujeto (o los *avowals* expresados, i.e. las adscripciones, de primera persona del presente del indicativo, de estados mentales ocurrentes) acerca de sus estados intencionales es un *principio constitutivo* que participa en las condiciones de identificación de lo que un sujeto cree. Esta autoridad de las auto-adscripciones al ser, en parte, determinativa de la identidad de un estado mental, funciona como indicativo de que el individuo está en ese estado. La idea es, entonces, que si uno juzga que tiene el estado mental *M* con un contenido que *p*, entonces uno tiene ese estado mental.

Esta "perspectiva por *default*" o concepción minimalista del autoconocimiento, según la cual la autoridad de la primera persona está construida en nuestra concepción de en qué consiste poseer estados mentales, ofrece una solución normativa *a priori* al problema del autoconocimiento, puesto que el principio normativo de la presunción de verdad se basa en una conexión *a priori* entre los estados mentales y las auto-atribuciones. La cuestión radicaría, entonces, en si los defensores de la TT estarían dispuestos a aceptar elementos normativos en su enfoque de la auto-atribución (que serían constitutivos de las condiciones de identidad de los estados mentales y que, aparentemente, no lo serían para el caso de la atribución a otros) y, con ello, opacar o ceder el rol fundamental de la TT en la explicación del autoconocimiento (este

mismo punto valdría también para el caso del enfoque del compromiso) o si, como vimos, la asimetría simplemente radica en una ilusión. Tal como la literatura naturalista sobre la TT muestra, ni las versiones estrictas (como la de Gopnik, 1993) ni las moderadas (como la de Carruthers, 1996a) incorporan elementos normativos en el sentido especificado. Y eso es comprensible porque, tal como sostengo en otro lugar, considerar que hay elementos normativos constitutivos de las condiciones de identidad de los estados mentales es incompatible con posturas naturalistas, como las abordadas, cuyo naturalismo consiste, justamente, en que no hay elementos normativos constitutivos de las condiciones de identidad de los estados mentales (Skidelsky, 2008a).

3.2. Teoría de la simulación

A diferencia de la TT, la TS sostiene que nuestra habilidad de *mindreading* no consiste en una capacidad de teorizar acerca de los estados mentales sino de simularlos. Más allá de las diferencias entre las distintas versiones de la TS, ésta sostiene que el individuo utiliza sus propios recursos cognitivos (en especial, motivacionales, emocionales y de razonamiento práctico) para atribuir estados mentales imaginando o pretendiendo que está en la posición del otro y así generando los estados mentales que se atribuirán al otro. De esta manera, representamos los estados mentales de otros por medio de su simulación mental o generando estados similares en nosotros. Así, a diferencia de la TT que adopta el punto de vista de la tercera persona, la TS adopta el punto de vista de la primera persona. En principio, entonces, a diferencia de las versiones propiamente de TT, para las cuales todo caso de auto-atribución es un caso más entre otros, para la TS todo caso de atribución a otros sería un caso de auto-atribución, i.e. a "uno mismo como otro". En lo que sigue consideraré dos de las versiones más desarrolladas y citadas de la literatura sobre TS, la de Robert Gordon y la de Alvin Goldman. Veremos no sólo que ambas versiones no proveen un enfoque de la auto-atribución basado, en términos estrictos, en la simulación, sino que tampoco las estrategias propuestas para lidiar con este problema parecen dar lugar a una asimetría epistemológica sustancial. Comencemos con la versión de Gordon.

3.2.1. La teoría de la simulación de Gordon

Gordon (1986) propone, en contra de la TT, que las declaraciones de intención inmediata, por ejemplo, "Me serviré ahora café", no parecen poder predecirse por razonamiento nomológico, i.e., por medio de inferencias a partir de premisas teóricas acerca de estados mentales y leyes o generalizaciones de la forma que hemos visto en §3.1. Lo que utilizamos es razonamiento práctico. Al simular razonamientos prácticos, podemos extender nuestras capacidades de auto-predicción a situaciones hipotéticas. Como en un juego de pretensión, la idea es preguntarse qué haría si una situación hipotética fuera actual, y la respuesta sería también una afirmación de intención inmediata. De esta manera nos involucramos en una simulación práctica, i.e., una decisión simulada de qué hacer pero sin el output conductual. Más aún, podemos extender esta simulación práctica para predecir la conducta de los otros. Como en el caso de la auto-predicción, aquí también está involucrada una toma de decisión, pero con un procedimiento de "ponerse en los zapatos del otro", i.e. proyectarse en la situación del otro haciendo "ajustes para diferencias relevantes" (1986, p. 63). Este parece ser el caso de los jugadores de ajedrez que reportan visualizar el tablero desde el lado oponente, tomando las piezas del oponente como propias y pretendiendo que las razones para la acción se han modificado en función de la situación del oponente y así poder predecir qué jugada hará éste.

Así, se deben realizar los cambios imaginativos requeridos para predecir lo que haría *el otro* en sus zapatos). Esto es, hay una "diferencia entre simular a *uno mismo* en la situación de O y simular a *O* en la situación de O" (Gordon, 1995, p. 55). Esto último involucra desde cambios de perspectiva espacio-temporal, pasando por ajustes en valores, temperamento, educación, etc. hasta cambios imaginativos de roles institucionales sobre la base de evidencia de la conducta pasada y presente del otro. Hay que tener en cuenta que el modo por *default* de la simulación no involucra ningún ajuste, y es el modo en el que funciona habitualmente, esto es, automáticamente proyectamos en los otros nuestras propias creencias y conocimiento acerca del entorno. Esto es lo que Gordon (1995) denomina "proyección total". Sin embargo, bajo ciertas circunstancias la proyección total no es confiable y entonces deben hacerse ajustes comenzando con "transportarse a uno mismo en la imaginación a la ubicación espacial o temporal del otro" (Gordon, 1995, p. 102). Esto lleva a modificar nuestro mapa egocéntrico (i.e. el mapa mental en el que las cosas están representadas en relación con uno mismo aquí y ahora), lo cual hace que el pronombre "yo", utilizado en el contexto de una simulación, refiera al otro y no a mí (si esto no fuera así, imaginaríamos situaciones falsas). Así, en la pregunta: ¿qué haré yo ahora?, "yo" y "ahora" no tienen su referencia habitual, i.e. no refieren al simulador y su circunstancia temporal, sino que refieren al otro y a las

condiciones temporales imaginadas.

De este modo, la simulación no consiste en utilizarse a uno mismo como modelo de otro individuo y, por ende, no consiste en una inferencia analógica implícita desde mí al otro. El argumento tradicional por analogía requiere que uno reconozca primero sus propios estados mentales (sean estos reales o imaginados) para luego inferir que el otro se encuentra en estados similares (Mill, 1865). Esto es, se tiene primero que saber de qué tipo de estado mental se trata e identificar su contenido para luego poder atribuirlo al otro. Generalmente, se postula que este autoconocimiento es por medio de introspección o monitoreo de ciertas propiedades de los estados mentales (y que esto requiere poseer los conceptos de los diferentes estados mentales adscriptos). Como veremos en §3.2.2., la versión simulacionista de Goldman adopta esta concepción. En cambio, la proyección imaginativa o el cambio egocéntrico no requiere, como hemos visto, una inferencia analógica. Si no requiere esto último, tampoco requiere introspección. Gordon (1995) sostiene que si bien el método que él propone identifica estados mentales desde la primera persona, dado que el cambio egocéntrico permite transformarnos en otras primeras personas, no es un método limitado a "una persona" en el sentido de que no es un método recognoscitivo.

Recapitulemos hasta aquí cómo sería el procedimiento general empleado en la simulación. La versión de TS de Gordon se basa en la noción de "identificación imaginativa". Según Gordon, el atributor re-centra su mapa egocéntrico cognitivo en el otro de manera que el pronombre "yo" refiere en la imaginación al individuo hacia el cual se re-centró el mapa cognitivo, y el "ahora" y "aquí" refieren al tiempo y lugar de la situación imaginada (del mismo modo que los actores se convierten en sus personajes mientras actúan). De esta manera, se realiza una transformación imaginativa en otro. Luego, los procesos de toma de decisión funcionan *off-line* generando una decisión (pretendida) y el mecanismo simulador atribuye directamente el estado, o la decisión generada, por medio de la rutina de ascenso semántico. Gordon (1995, 1996) adopta esta rutina para dar cuenta de la atribución de los estados propios y de otros, por la cual se obtiene la respuesta a una pregunta acerca de la propia condición mental respondiendo una pregunta que no es acerca de uno mismo o sus propios estados mentales. Así, Gordon adopta un acceso especial epistémico de "mirar hacia afuera", que hemos visto en §2, al estilo del que defiende Evans (1982) para las creencias.

Evans afirma: "al hacer una auto-adscripción de creencia, los propios ojos están, por así decirlo, u ocasionalmente de manera literal, dirigidos hacia afuera –hacia el mundo. Si alguien me pregunta '¿Pensás que va a haber una tercera guerra mundial?', debo atender, al contestarle, precisamente al mismo fenómeno externo al que atendería si estuviera contestando la pregunta '¿Habrá una tercera guerra mundial?' (…) cuando se está en la posición de afirmar que *p*, se está ipso facto en la posición de afirmar 'Creo que *p*'" (1982, pp. 225-6). Tal como vimos en §2 y §3.1, la idea es que dirigimos nuestra atención a los rasgos del mundo que representa el estado mental y no al estado mental mismo. Lo mismo hacemos, según Gordon, para el caso de la atribución a otros. Adscribir a O la creencia de que *p* no es más que afirmar que *p* en el contexto de una simulación de O. Así, la atribución a otros no sería más que un caso de auto-atribución a "uno mismo como otro" (en el contexto de la simulación) por medio de rutinas de ascenso.

Ahora bien, tal como mencioné en §2, las posturas de "mirar hacia afuera" o modelo de la transparencia han recibido críticas tales como que sólo son útiles para dar cuenta de la formación de estados mentales (para lo cual, seguramente, se requiere considerar la evidencia del entorno) o para decidir qué creer o desear, pero no se aplican a los casos comunes de autoconocimiento de estados mentales preexistentes (Goldman, 2000a; Martin, 1998). Por otro lado, gran parte de nuestro autoconocimiento no parece estar disponible a partir de juicios acerca del mundo exterior. Por ejemplo, sabemos qué estamos imaginando o qué decidimos hacer o si estamos contentos sin que tengamos que atender al mundo exterior para saber que estamos en esos estados (Gertler, 2008). Goldman (2000a, 2000b, 2006) discute el enfoque de Gordon argumentando que sólo se aplica a un subconjunto de estados mentales, esto es, las creencias, pero no queda claro cómo se extiende este enfoque al resto de las actitudes y a las sensaciones. Si en una panadería me pregunto, por ejemplo, si *quiero* dos medialunas del mostrador, ¿por cuál pregunta de "mirar hacia afuera" la sustituyo? Si es por ¿parecen apetitosas las dos medialunas del mostrador?, no parece haber una manera de decir si lucen apetitosas *para mí* sin consultar (de manera tácita) un estado interno mío. Por otro lado, si me pregunto si *esperé* que el equipo E ganara ayer, contestar la pregunta de nivel inferior de si el equipo E ganó ayer no ayuda a determinar la respuesta a la pregunta de nivel superior.

Tampoco, según Goldman, queda claro por cuáles preguntas se sustituyen las siguientes: ¿me estoy preguntando ahora si p?, ¿estoy recordando ahora p?,

¿estoy imaginando ahora que p? Tampoco queda claro que haya una pregunta distintiva para cada tipo de actitud. Esto lleva a pensar que este enfoque no es adecuado para el autoconocimiento de actitudes. Puedo preguntarme si *quiero* dos de las medialunas del mostrador o si *creo* que hay dos medialunas en el mostrador. En ambos casos dirijo mi atención al mostrador, pero allí no está la información que necesito saber para distinguir entre las dos preguntas. Asimismo, tendría que haber alguna conexión entre la información relevante para responder a preguntas acerca de estados mentales pasados y presentes. Pero no es cierto que las primeras se contesten apelando a información del mundo externo. Si me pregunto si estuve imaginando que *p* no necesito recabar información del exterior, sólo recordar mi estado mental interno en ese momento. Finalmente, suponiendo que "contestar afirmativamente a la pregunta de si p" significa que uno *juzga* (i.e. cree ocurrentemente) que la respuesta es "p", ¿cómo se llega a la respuesta (afirmativa) de si creo/pienso que *p*?, el paso siguiente parece ser la determinación que uno juzgó que *p*. Pero aquí nos encontramos nuevamente con el problema original.

En su respuesta a la críticas recibidas, Gordon (2007) sostiene que el método de ascenso semántico puede aplicarse a otros estados mentales, no sólo a las creencias. Esto se logra adosando a esta rutina una variante del enfoque expresivista, mencionado en §2. Así, para contestar a la pregunta acerca de qué gusto de helado quiero, no pienso acerca de mi deseo sino acerca de algo en el mundo, esto es, los diferentes gustos disponibles. Y expreso mi actitud de manera auto-adscriptiva porque tengo competencia desde chica (aun antes de tener el concepto "x quiere/desea que p") en anteponer a mis expresiones de deseo la emisión de la forma "quiero que p". Lo mismo vale para los temores, miedos, intenciones, etc. Gordon presenta ahora la rutina de ascenso en términos mecanicistas. Así, la cuestión no es tanto que la rutina se realiza en respuesta a ciertas preguntas, no es algo que hacemos *nosotros*, sino que cuando intento atribuirme un estado mental, el cerebro reutiliza el proceso que usa al generar una emisión correspondiente no-adscriptiva de nivel inferior.

El modelo para la rutina sería "el esquema "*Yo φ que p*" > "Yo φ que p" en donde la primera ocurrencia en cursivas de "*Yo φ que p*" representa el uso no-adscriptivo [expresivo] de la forma y la segunda ocurrencia representa el uso adscriptivo, la auto-adscripción explícita" (Gordon, 2007, p. 160). Así, la misma oración usada para expresar la actitud proposicional también se usa para la auto-adscripción. De esta manera, para cada actitud proposicional habría una rutina de ascenso correspondiente que generaría una emisión distintiva acerca del mundo. Como se mencionó, la rutina procede en ausencia de los conceptos relevantes aunque, por supuesto, para que haya una genuina auto-adscripción se debe tener competencia conceptual (i.e., se debe tener dominio de la semántica de la auto-adscripción de deseos, creencias, etc.). Finalmente, no habría tal regreso en la explicación porque cuando se emite, por ejemplo, "llueve", uno expresa la creencia de que llueve. De modo que para generar la emisión "yo creo que llueve", no se requiere el juicio posterior de que uno juzga que está lloviendo, sólo se requiere una rutina mecánica que prefija la oración con "yo creo".

Como queda claro en este detallado desarrollo, Gordon acude a posturas epistemológicas tradicionales, introducidas en §2, i.e. el enfoque de acceso especial de "mirar hacia afuera" y una variante del expresivismo, como complementos de su TS para dar cuenta de la auto-atribución (y también de la atribución a otros). Ahora bien, más allá de las críticas mencionadas y las posibles respuestas satisfactorias o no, este enfoque no parece capturar la asimetría dado que toda atribución mental es esencialmente auto-atribución, ya sea directamente vía la rutina de ascenso (más expresivismo) o vía uno mismo como otro en una simulación con rutina de ascenso (más expresivismo). De manera que, tanto para atribuir estados mentales a los otros como a nosotros mismos utilizamos un modelo perceptivo inferencial que desplaza la atención desde el estado mental a los objetos del mundo (más los prefijos que permiten diferenciar actitudes). Así, no parece haber nada sustantivo respecto de la autoridad de primera persona. Se podría pensar en adosar alguna otra perspectiva epistemológica que dé lugar a la asimetría. Sin embargo, en este enfoque no sólo no hay lugar para un modelo perceptivo de "mirar hacia adentro" para la primera persona, sino que tampoco acoplarlo sería compatible porque, justamente, todo el punto de la rutina de ascenso es su oposición a cualquier método introspectivo recognoscitivo. De manera que las únicas opciones disponibles para lograr una asimetría sustantiva parecerían ser las no-naturalistas. Así, lo que se obtendría sería un (mega-)enfoque de simulación más uno de "mirar hacia afuera" más el expresivismo, y todo esto complementado con algún enfoque constitutivo conceptual-pragmático y/o del compromiso.

Sin embargo, no sólo no parece claro cómo funcionaría un (mega-)enfoque así, en términos de coherencia interna entre las tesis fundamentales de las

distintas perspectivas, sino que la viabilidad de la TS resulta hasta cierto punto sospechosa si hay que complementarla con tantos y diversos enfoques epistemológicos para que pueda dar cuenta de la asimetría. En este último sentido (y sin la necesidad de incluir ninguna perspectiva constitutiva), en el enfoque de Gordon no queda claro el rol de la TS, siendo que una rutina de ascenso más el expresivismo podrían dar cuenta tanto de la auto-adscripción como de la adscripción a otros. No queda claro no sólo lo que aportaría la simulación en la explicación de la auto-adscripción de estados mentales (puesto que, aparentemente, no se requiere), sino que tampoco quedaría claro su rol para la adscripción a otros, siendo que sin simulación igualmente se podría dar cuenta de este fenómeno a través del enfoque epistemológico de la rutina de ascenso más los prefijos, e.g., "yo creo" y "él cree" o "yo deseo" y "él desea", que aprendemos desde chicos (si estos fueran métodos viables). En este sentido, no parece que haga falta ninguna identificación imaginativa para dirigir la atención hacia el mundo (y utilizar los prefijos aprendidos en cuestión) para la adscripción de estados mentales a otros. Así, en el enfoque simulacionista de Gordon, el rol de la simulación queda desdibujado frente al rol de los complementos epistemológicos que se le adosan. Parecería que todo el peso explicativo podría recaer en estos últimos sin necesidad de ningún elemento simulacionista.

Por otro lado, y más allá de la cuestión de si la simulación juega o no un rol primordial en la explicación de los fenómenos en cuestión (en particular, en la auto-atribución), ninguno de los complementos epistemológicos adosados, por sí mismos o en conjunción, permite rescatar una asimetría (sustantiva). La rutina de ascenso semántico no parece establecer ninguna asimetría entre la primera y la tercera persona puesto que es un proceso inferencial que funciona de la misma manera, en el enfoque de Gordon, tanto para atribuir estados mentales a los demás como a uno mismo. Asimismo, la versión expresivista que se le adosa sólo cumple el rol de diferenciar actitudes, y funciona de la misma manera tanto para el caso de uno mismo como para la atribución a otros, i.e. se prefija "yo creo", "yo deseo", etc. a cualquier emisión (lo que cambia en la identificación imaginativa, recuérdese, es la referencia de "yo"). Veamos, entonces, si la versión de Goldman de TS logra acomodar la asimetría y al mismo tiempo mantener un rol explicativo fundamental para la simulación.

3.2.2. La teoría de la simulación de Goldman

Según la versión de TS de Goldman, la simulación consiste en "imitar", "copiar, "re-experimentar" (Shanton & Goldman, 2010), "replicar" o "emular" (Goldman, 2000b) la actividad mental del target cuya mente se quiere "leer". La simulación juega un rol importante en la atribución de tercera persona, pero esto no quiere decir que se use "siempre", en el sentido de ser el único método. Según Goldman (1989, 2006), la simulación es el procedimiento por *default*, puesto que es el más básico y espontáneo (a veces, se hacen atribuciones basándose sólo en información adquirida inductivamente). En este sentido, la versión de Goldman de TS es una moderada o débil. Prototípicamente, el proceso de simulación tiene tres etapas (Shanton & Goldman, 2010, Goldman & Shanton, en prensa). Antes de describir estas etapas, conviene aclarar que Goldman distingue entre simulación de nivel inferior y simulación de nivel superior (Goldman 2006, Shanton & Goldman 2010). La segunda, a diferencia de la primera, es más compleja computacionalmente y suele involucrar actitudes proposicionales. Un ejemplo de la primera es la atribución de estados emocionales sobre la base de expresiones faciales. Un ejemplo de la segunda es la atribución de decisiones. El tipo de simulación que interesa aquí es del segundo tipo.

En la rutina de simulación, en primer lugar, se crea un conjunto de estados en uno mismo, por medio de la imaginación, que corresponden a los estados iniciales del *target*. Esto es, se pretende tener los deseos y creencias del sujeto de la atribución (relevantes a la situación atributiva en cuestión), que el atributor considera (por evidencia independiente) que el *target* posee. Esto es ponerse "en el lugar del otro" o "desde su perspectiva". Por supuesto, esta etapa requiere que el atributor posea información acerca de los estados iniciales del sujeto de la atribución y que estos estados (así como los del output de la rutina de simulación) sean etiquetados como los del *target*. La segunda etapa consiste en alimentar con estos inputs a uno de los sistemas operativos de la mente (por ejemplo, al mecanismo de toma de decisiones, si de lo que se trata es de dar como output una decisión que otro tomaría) para que dé lugar, de manera *off-line*, a otros estados como outputs. Finalmente, el atributor "lee" o detecta el estado-output y lo proyecta en o lo atribuye al *target*. Según Goldman (2000b), la esencia de la simulación mental de otros es el intento deliberado de reproducir en uno una secuencia de eventos que ocurre (o se piensa que ocurre) en otro. Esto supone que los estados mentales pretendidos tienen algún tipo de similitud, homología o semejanza con sus contrapartidas naturales o no-pretendidas (siendo una cuestión empírica en

cuáles respectos). Y supone también que el atributor pone (o intenta poner) en cuarentena o inhibe sus propios estados mentales durante la simulación (la evidencia de que un proceso es susceptible de parcialidad egocéntrica parece ir a favor de que ese proceso involucra simulación, cf. Shanton & Goldman, 2010).

Con este procedimiento, la TS puede dar cuenta de la atribución y auto-atribución de estados pasados, presentes e hipotéticos, pero no es un enfoque plausible de la auto-atribución de estados ocurrentes, que es el fenómeno que interesa aquí. Goldman mismo advierte que la atribución de los propios estados mentales ocurrentes no parece requerir el procedimiento descripto y, tal como se adelantó más arriba en la discusión de TT, propone que la auto-adscripción se basa en el método de la introspección. Así, la auto-atribución de estados ocurrentes sería por introspección como método por *default* (otros métodos serían la inferencia o confabulación que suponen auto-interpretación, de manera que habría un método dual de autoconocimiento). De este modo, Goldman (1993) adoptaría la idea de que el conocimiento de los propios estados mentales no sólo es un logro cognitivo (o no puede ser "cognitivamente insustancial", sino que adopta de manera explícita una variante del enfoque de acceso especial, mencionado en §2, de "mirar hacia adentro" o introspección, por el cual las creencias de segundo orden son el resultado de procesos causales que rastrean estados de primer orden.

Según la versión introspeccionista o de auto-monitoreo de Goldman (1993, 2000a, 2000b, 2006), tenemos un acceso directo, no-inferencial, y recognoscitivo a nuestros estados mentales ocurrentes. Este acceso no significa que estos sean siempre identificables o discriminables, sino que *pueden potencialmente* identificarse (Goldman, 2000b). La atención (automática o voluntaria), al igual que en la percepción externa, juega un rol importante en la introspección en el sentido de que la facilita, actuando como un "órgano orientador…que pone al sujeto en una relación apropiada con un estado candidato" (Goldman, 2006, p. 244). El sistema cognitivo tiene un acceso introspectivo cuasi-perceptivo (en el sentido de ser recognoscitivo y estar gobernado, como se mencionó, por la atención) a ciertas propiedades distinguibles de los estados mentales, que son categoriales (no-disposicionales) y no-relacionales (en el sentido, de no ser masivamente relacionales como las propiedades de rol funcional). En general, el output de la introspección es una clasificación de los estados mentales según el tipo del estado-caso (por ejemplo, creencia, dolor, sensación de calor), el contenido del estado, y la fuerza o intensidad del estado. Respecto del input causal del mecanismo que permite la clasificación mencionada, en un principio, Goldman (1993) sostuvo que se tiene acceso a las propiedades fenoménicas de los estados mentales, luego se mantuvo neutral acerca de si las propiedades son o no fenoménicas (Goldman, 2000a), para finalmente afirmar que las propiedades que detecta el mecanismo no pueden ser las funcionales (dado su carácter disposicional y relacional), ni las fenoménicas (dado que no está claro que todos los tipos de actitudes tengan rasgos fenoménicos distintivos), ni las representacionales/intencionales (dado que lo que distingue tipos de estados con el mismo contenido o distintas intensidades no es una cuestión adicional acerca del contenido intencional), sino las neuronales (Goldman, 2006).

En este sentido, habría propiedades neuronales distintivas que detecta el mecanismo introspectivo, de manera de dar lugar a la clasificación de representaciones sensoriales en términos de estados mentales. Así, "[u]n nivel alto de activación en una clase de células genera la clasificación introspectiva "dolor" (o "dolor agudo"), un nivel de activación alto en una clase diferente de células genera la clasificación introspectiva "cosquillas", y así en más" (Goldman, 2006, p. 252). Este sería uno de los procesos que operaría en la introspección. A grandes rasgos, habría tres procesos distintos: reconocimiento, reutilización y traducción. El primero consistiría en el ya descripto de tomar rasgos de los estados neurales realizadores de los estados mentales como inputs y dar lugar como output a la clasificación del tipo de estado mental y la intensidad de ese estado. Otro proceso de reutilización o traducción toma el componente que porta el contenido del estado de creencia como input y da como output una representación de ese componente (i.e. el contenido del estado de primer orden es replicado en el de segundo orden). Se requiere un proceso adicional, que no queda claro en los escritos de Goldman si formaría parte del mecanismo de introspección, que combine la representación cuasi-perceptiva "creencia" con la representación del ítem que porta el contenido.

Tal como hemos visto, estos procesos de introspección tienen lugar en la etapa final de la rutina de simulación cuando el atributor "lee" o detecta el estado-output, de manera que para la atribución a otros también es esencial que el sistema simulador reconozca sus propios estados mentales "pretendidos". Para ello se requiere que el sistema acceda introspectivamente a estos estados, y lo haga en posesión de los conceptos mentales relevantes para poder clasificarlos y porque el

producto final de la simulación es siempre la formación de una creencia acerca de un estado mental. De manera que, ya sea que la atribución sea a uno mismo o a otros, el sistema simulador debe reconocer los estados propios, esto es, los estados del atributor, para seleccionar un tipo de estado y un contenido específicos (además de una intensidad particular). Por eso, la "lectura" de los propios estados ocurrentes ocupa un lugar primordial en la atribución tanto de primera como de tercera persona y, en este sentido, la atribución a los otros es parasitaria de la auto-atribución (Goldman, 1993).

Ahora bien, son muy conocidas las críticas que ha recibido la introspección como método para el autoconocimiento, de manera que no las voy a repasar aquí (para un panorama, véase Schwitzgebel, 2010). Lo único que quisiera decir respecto del mecanismo de monitoreo propuesto por Goldman es que no queda clara su naturaleza. No queda claro si es un mecanismo completamente subpersonal de procesamiento de la información cognitiva o un método que opera también a nivel personal (para la distinción personal-subpersonal, véase Skidelsky, 2006). Cuando se dice que la introspección (como etapa final de la simulación o por ella misma para el caso de la auto-adscripción) tiene como fin la formación de una creencia acerca de un estado mental, parece que es un procedimiento que, al menos, tiene que transcurrir, en parte, en el nivel personal, siendo que las creencias son los estados paradigmáticos de nivel personal. Más aún, las creencias están constituidas por conceptos, de manera que si el output de la introspección son clasificaciones conceptuales (no superficiales) de los estados mentales, parece que eso no puede producirlo un mecanismo puramente subpersonal, sino que sería el sujeto (en tanto sistema global) el que culminaría realizando esas clasificaciones.

Este es un problema que no sólo atañe a la versión introspeccionista de Goldman, sino también a toda propuesta de acceso especial de "mirar hacia adentro" que propone un mecanismo de auto-monitoreo para la formación de creencias de segundo orden. Siendo que en la literatura, hay cierto consenso en que los mecanismos subpersonales de procesamiento de la información no poseen estados con contenido conceptual (cf. Bermúdez & Cahen, 2011), al menos Goldman debe alguna explicación de cuál es el sistema que realiza la clasificación conceptual: si un mecanismo completamente subpersonal de auto-monitoreo o el sujeto. Si es lo primero, entonces debería ofrecer algún enfoque de la naturaleza del mecanismo de auto-monitoreo en relación a la posesión de conceptos. Por otro lado, cuando se sostiene que se puede acceder a propiedades neuronales, está claro que la introspección está operando en el nivel subpersonal porque, en parte por lo dicho para el caso de la versión de TT de Carruthers (1996a), la persona no accede introspectivamente a las propiedades subpersonales, en este caso neuronales, de sus estados y procesos mentales (o no mentales). Más bien, habría que pensar en un mecanismo subpersonal que detecta input neuronal. Quizá, Goldman está pensando en un procedimiento que está en ambos niveles, pero entonces no queda claro en qué consiste el mecanismo introspectivo, esto es, cuáles serían los niveles intermedios entre el input neuronal subpersonal y el output ¿subpersonal/personal? de la formación de creencias sobre estados mentales.

Más allá de las críticas tradicionales a la introspección y la oscuridad acerca del mecanismo introspectivo, en la versión de TS de Goldman, a pesar de que la atribución siempre es desde la perspectiva de la primera persona, parecería haber una asimetría entre la atribución de estados mentales a los otros, que sería por medio de la simulación que incluye como último paso el auto-monitoreo o introspección, y a uno mismo, que sería sólo por medio del auto-monitoreo. Así, parecería que esta versión de la TS da cabida a la asimetría y lo haría adhiriendo a una versión particular del modelo perceptivo de "mirar hacia adentro". No obstante, puesto que el auto-monitoreo se requiere tanto para la atribución de primera como de tercera persona, no parece haber una asimetría esencial o sustantiva en la manera en que conocemos nuestros propios estados y los de los otros: accedemos a ambos a través de la introspección de nuestros estados, en un caso "reales" y en el otro, "simulados".

Se podría también pensar, como para los otros enfoques de TM analizados, en adosar otras perspectivas epistemológicas de manera de volver más sustancial a la asimetría. Por supuesto que complementar este enfoque con el de "mirar hacia afuera" no parece tener mucho sentido siendo que, tal como vimos, son enfoques de acceso especial completamente opuestos, en el sentido de que este último se contrapone a cualquier proceso introspectivo. Por otro lado, las perspectivas constitutivas no parecen congeniar apropiadamente con el enfoque naturalista de Goldman. Parece un poco forzado, tal como mencioné en §3.1, congeniar posiciones naturalistas con las constitutivas. No parece conducente intentar congeniar un enfoque naturalista, que propone como aspecto crucial un mecanismo causal-empírico de auto-monitoreo, con un enfoque constitutivo-apriorístico que

postula una conexión necesaria entre los estados mentales de primer y segundo orden. Aun más, hasta resulta incompatible este enfoque empírico con posturas que sostienen, como hace el enfoque de compromiso, que bajo la condición de agencialidad responsable, no es posible conceptualmente la ruptura de la conexión, en el sentido de que aun en mundos posibles donde hubiera una falla del mecanismo causal, igualmente habría autoconocimiento y agencialidad responsable. Es en este sentido que en §2 remarqué que las teorías de acceso especial, en particular las de "mirar hacia adentro", y las constitutivas están en disputa y que sólo una de ellas puede servir para dar cuenta del autoconocimiento (véase la cita de Bilgrami, 1998, p. 208 en ese mismo apartado). De manera que no sólo alguno de los dos enfoques parece de más y, en este sentido, superfluo, sino que habría dificultades casi insalvables para esa compatibilización (que, como mencioné, he desarrollado en Skidelsky, 2008a).

Tomando en cuenta ambas versiones de TS analizadas, quisiera remarcar dos cuestiones. La primera es que queda claro que las propuestas de TS no son suficientes por sí mismas para dar cuenta de la auto-atribución de estados mentales ni, incluso, de la atribución a otros. Como hemos visto, mientras que la versión de Goldman requiere el complemento del enfoque epistémico-recognoscitivo de la introspección no sólo para la auto-atribución sino que también para la atribución a otros, el enfoque de Gordon requiere complementarse con una versión epistémica no-recognoscitiva de ascenso semántico (más expresivismo) también tanto para la atribución de estados a nosotros mismos como a otros. En este sentido, no sólo no queda claro qué le aporta la TS, en términos de poder explicativo, a los enfoques epistemológicos ya existentes de la introspección y del ascenso semántico, sino que estos últimos, en la medida en que puedan sostenerse por sí mismos, parecen quitarle a la simulación un rol primordial en la adscripción y la auto-adscripción. Es más, en la explicación de la auto-adscripción, que es lo que interesa aquí, la simulación no tendría rol alguno, según Goldman. Y aunque en la perspectiva de Gordon no queda muy claro, tal como vimos, aparentemente tampoco lo tendría. Es por ello que mencioné al comienzo de §3.2 que ambas versiones no proveen un enfoque de la auto-atribución basado, en términos estrictos, en la simulación.

En segundo lugar, ninguna de ambas versiones logra dar cuenta de la asimetría. Como hemos visto, mientras que en el enfoque de Gordon ambos tipos de atribuciones utilizan un procedimiento perceptivo inferencial (el ascenso semántico), en la versión de Goldman para ambos tipos de atribuciones se utiliza un modelo perceptivo no-inferencial (de introspección). Recordemos que la tesis de la asimetría, tal como se mencionó al principio de §2 y como se la entiende tradicionalmente, parece decir más que el hecho de que hay una diferencia en la manera en que accedemos a nuestros estados y la manera en que accedemos a los estados de los otros. Parece suponer que el acceso en nuestro caso no es inferencial, mientras que en el caso de los otros lo es (Bilgrami, 2010; Gertler, 2008; Shoemaker, 2010). Ninguna de las versiones de TS analizadas parece respetar esto.

4. Conclusiones

Asumiendo una perspectiva integradora del conocimiento científico-filosófico, se esperaría que en áreas adyacentes, como TM y autoconocimiento, pueda, al menos, haber canales de diálogo de manera de ofrecer un enfoque coherente de ciertos fenómenos, en este caso, la asimetría entre la primera y la tercera persona respecto del conocimiento o atribución de estados mentales. Sin embargo, hemos visto, por un lado, que uno de los enfoques más difundidos de TM, esto es la TT, o bien no parece dar lugar a este fenómeno (en su versión estricta) o bien los intentos por rescatarlo le quitan a la TT un rol primordial (en las versiones moderadas). Y esto parece ser producto de que tanto la atribución a otros como a uno mismo se basa en la perspectiva inferencial de tercera persona. Por otro lado, si bien los enfoques más conocidos de TS que hemos visto apelan a variantes de respuestas ofrecidas a favor de la tesis de la asimetría en teoría del conocimiento, o bien la propia TS parece superflua o habría algo respecto de la propuesta misma de simulación que en principio parecería bloquear una asimetría sustantiva. La idea de que tanto la atribución a otros como a uno mismo se basa en la perspectiva de la primera persona parece llevar a una simetría que en el enfoque de Gordon se refleja en que ambos tipos de atribuciones utilizan un procedimiento perceptivo inferencial, mientras que en la versión de Goldman, en ambos tipos de atribuciones interviene un procedimiento perceptivo no-inferencial. Finalmente, las posibles opciones epistemológicas para respetar la asimetría o marcar una asimetría más sustantiva, no parecen promisorias dado que, como hemos visto, serían problemáticas o incompatibles o superfluas.

Por supuesto que una apreciación más global y, por ello, más adecuada de la cuestión hubiera requerido el análisis de evidencia empírica respecto de las posibles predicciones que se desprenden, tanto de las propuestas

empíricas en TM como de aquellas filosóficas pero también de naturaleza empírica, en torno del fenómeno de la atribución y auto-atribución de estados mentales. Así, es esperable que la predicción de aquellas teorías filosóficas o psicológicas que sostienen que no hay tal asimetría sea que los niños desarrollan las habilidades de auto-atribución y atribución de manera simultánea, mientras que las teorías que intentan dar lugar a la asimetría predicen que la auto-adscripción tendría que desarrollarse antes que la capacidad de adscribir estados mentales a otros. Igualmente, aquellos que están familiarizados con la literatura experimental sobre este tema acordarán probablemente en que no parece haber acuerdo respecto de esta cuestión. Como suele ocurrir en algunas áreas en ciencia, hay evidencia empírica que respalda ambas predicciones. Y tal como suele desarrollarse la práctica científica, habrá que esperar que surja más evidencia concordante.

Suponiendo, entonces, que mi diagnóstico fuera correcto y, con ello, que en el ámbito cognitivo de TM la tesis de la asimetría no parece tener algún lugar prominente, ¿no sería más adecuado abandonar la tesis de la asimetría?, ¿por qué persistir con lo que quizá no sea más que un resabio de la epistemología? Responder esto implica tomar partido respecto de la cuestión más general acerca de las relaciones entre la filosofía y la ciencia, en este caso, la ciencia cognitiva. Según Goldman (1992), el filósofo puede tener el rol de aportador, crítico metodológico o consumidor en relación con las ciencias. El filósofo de la mente aportador contribuye al desarrollo de la ciencia cognitiva creando herramientas intelectuales, identificando tópicos de su propio ámbito para investigar (actitudes proposicionales, referencia, etc.) y ofreciendo una fundamentación conceptual. El crítico metodológico practica filosofía especial de las ciencias, i.e. se ocupa de la legitimidad de los constructos teóricos, de su interpretación en términos realistas o instrumentalistas, de los distintos niveles de teorización en una disciplina, etc. El filósofo consumidor hace un uso filosófico directo de los resultados científicos. Este uso puede adoptar distintas formas que van desde el traspaso de los datos científicos a la reflexión filosófica (en la medida en que ciertas tesis no podrían plantearse sin la ayuda de estos resultados) hasta la utilización de esos resultados empíricos para confirmar o refutar tesis filosóficas particulares, e incluso de un modo más audaz, plantear hipótesis filosóficas que puedan generar modelos susceptibles de contrastación empírica.

Por supuesto que estas tres actitudes no son excluyentes ni incompatibles entre sí. Pero frente a posturas que consideran que el filósofo sólo debe ser aportador, mis simpatías naturalistas, tal como desarrollé en otro lugar, se inclinan hacia los otros roles (Skidelsky, 2008b). Esto parecería sugerir que ante situaciones en las cuales la ciencia cognitiva parece arrojar resultados negativos respecto de alguna tesis filosófica, habría que abandonarla. Sin embargo, considero que los roles de crítico metodológico y filósofo consumidor sugieren otra cosa. A mi entender, estas actitudes reflejan la tesis naturalista de la continuidad entre la filosofía y la ciencia. Las hipótesis filosóficas y científicas están en igualdad de condiciones en el entramado general del conocimiento. De manera que no considero a la ciencia cognitiva como la piedra de toque de las afirmaciones filosóficas ni, consecuentemente, a la filosofía como la piedra de toque de la ciencia. En todo caso, me parece que la continuidad supone un ida y vuelta en el cual puede darse que respecto de ciertas cuestiones tengamos que abandonar algunas tesis filosóficas o, igualmente, algunas científicas. En el caso que nos incumbe, creo que la tesis de la asimetría merece seguir siendo investigada porque la autoridad de primera persona es fundamental, en particular, para cuestiones relacionadas con la agencialidad y el compromiso moral. La autoridad de primera persona parece ser una condición necesaria de la agencialidad y, con ello, de la posibilidad de ser responsables de los propios actos.

Agradecimientos

Una versión anterior de este trabajo fue presentada en el XV Congreso Nacional de Filosofía-AFRA (Buenos Aires, 2010). Agradezco los comentarios de Pablo Rychter (en especial, el que dio lugar a la observación final de la sección 2), Ángeles Eraña (en particular, el que dio lugar a la observación en torno a las relaciones entre la filosofía y la ciencia cognitiva) y Fernanda Velázquez. También agradezco las sugerencias de los evaluadores anónimos de la revista. El trabajo se ha beneficiado con el apoyo financiero de los proyectos de investigación UBACyT 20020090200322 (2010-2012) y PIP-CONICET 2531 (2009-2011).

Referencias

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

Armstrong, D. (1981). *The Nature of Mind and Other Essays*. Ithaca, NY: Cornell University Press.

Armstrong, D. (1999). *The mind-body problem.* Boulder, CO: Westview.

Bar-On, D. (2004). *Speaking My Mind. Expressionism and Self-Knowledge*. Oxford: Clarendon Press.

Bar-On, D. & Long, D. (2001). Avowals and First Person

Privilege. *Philosophy and Phenomenological Research*, *62*, pp. 311-335.

Bermúdez, J. L. & Cahen, A. (2011). Nonconceptual mental content. In E. Zalta (Ed.) *Stanford Encyclopedia of Philosophy,* (Spring 2011 Edition), recuperado en URL = <http://plato.stanford.edu/archives/spr2011/entries/content-nonconceptual/>.

Bilgrami, A. (1998). Self-Knowledge and Resentment. En C. Wright, B. Smith & C. Macdonald (Eds.) *Knowing Our Own Minds* (pp. 107-142). Oxford: Oxford University Press.

Bilgrami, A. (2010). Other minds. En J. *Dancy*, E.*Sosa* & M. *Steup* (Eds.) *A* Companion *to Epistemology* (pp. 566-571). Oxford: Blackwell.

Brueckner, A. (2003). Two Transcendental Arguments Concerning Self-Knowledge. En S. Nuccetelli (Ed.) *New Essays on Semantic Externalism and Self-Knowledge* (pp. 185-200). Cambridge, MA: MIT Press.

Carruthers, P. (1996a). Simulation and Self-Knowledge: A Defence of the Theory-Theory. En P. *Carruthers* & P. Smith (Eds.), *Theories of Theories of Mind* (pp. 22-38). Cambridge: Cambridge University Press.

Carruthers, P. (1996b). *Language, Thought and Consciousness.* Cambridge, MA: Cambridge University Press.

Carruthers, P. (1998). Conscious Thinking: Language or Elimination. *Mind & Language,* 13, 323-342. Reimpreso en P. Carruthers (2005). *Consciousness: Essays from a Higher-Order Perspective*. Oxford: Clarendon Press.

Carruthers, P. (2002). The Cognitive Function of Language. *Behavioral and Brain Sciences, 25*, 657-674.

Carruthers, P. (*2009*). Simulation and the First-Person. *Philosophical Studies*, *144*, 467-475.

Carruthers, P. (*2010*). Introspection: Divided and Partly Eliminated. *Philosophy and Phenomenological Research, 80*, 76-111.

Churchland, P. M. (1988). Folk Psychology and the Explanation of Human Behavior. *Proceedings of the Aristotelian Society, 62*, 209-21.

Chase, W. G., & *Simon*, H. A. (*1973*). Perception in chess. *Cognitive Psychology, 4*, 55-81.

Davies, M. & Stone, T. (1995). Introduction. En M. Davies & T. Stone (Eds.) *Folk Psychology. The Theory of Mind Debate* (pp. 1-44). Oxford: Blackwell.

De Groot, A. (1978). *Thought and choice in chess.* The Hague: Mouton Publishers, 2nd edition.

Descartes, R. (1641). *Meditaciones Metafísicas.* En E. de Olazo y T. Zwanck (Comps.) (1980). *Obras escogidas.* Buenos Aires: Editorial Charcas.

Dretske, F. (1994). Introspection. *Proceedings of the Aristotelian Society*, *94*, 263-278.

Evans, G. (1982). *The Varieties of Reference.* Oxford: Oxford University Press.

Fodor. J. (1998). Do We Think in Mentalese: Remarks on Some Arguments of Peter Carruthers. En *In Critical Condition. Polemical Essays on Cognitive Science and the Philosophy of Mind* (pp. 63-74). Cambridge, MA: MIT Press.

Fricker, E. (1998). Self-Knowledge: Special Access versus Artefact of Grammar-A Dichotomy Rejected. En C. Wright, B. Smith & C. Macdonald (Eds.), *Knowing Our Own Minds* (pp. 155-106) Oxford: Oxford University Press.

Gallagher, S. (2001). The practice of mind: Theory, simulation, or interaction? *Journal of Consciousness Studies*, *8*, 83-107.

Gallagher, S. (2005). *How the body shapes the mind.* Oxford: Oxford University Press.

Gertler, B. (2008). Self-Knowledge. En *The Stanford Encyclopedia of Philosophy (Winter 2008 Edition)*, E. N. Zalta (Ed.), recuperado en: URL = <http://plato.stanford.edu/archives/win2008/entries/self-knowledge/>.

Goldman, A. (1989). Interpretation Psychologized. *Mind and Language*, *4*, 161–185. Reimpreso en in M. Davies & T. Stone (Eds.) (1995). *Folk Psychology: The Theory of Mind Debate*. Oxford: Blackwell Publishers.

Goldman, A. (1992). *Liaisons. Philosophy Meets Cognitive Science.* Cambridge, MA: MIT Press.

Goldman, A. (1993). The Psychology of Folk Psychology. *Behavioral and Brain Sciences, 16,* 15-28. Reimpreso en A. Goldman (Ed.) *Readings in Philosophy and Cognitive Science*. Cambridge, MA: MIT Press.

Goldman, A. (2000a). Folk Psychology and Mental Concepts. *Protosociology, 14*, 4-25.

Goldman, A. (2000b). The mentalizing folk. En D. Sperber (Ed.) *Metarepresentations* (pp. 171-196). Oxford: Oxford University Press.

Goldman, A. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading.* Oxford: Oxford University Press.

Goldman, A. & Shanton, K. (en prensa). The case for simulation theory. En A. Leslie & T. German (Eds.), *Handbook of 'Theory of Mind'*. Mahwah, NJ: Erlbaum.

Gordon, R. M. (*1986). Folk Psychology as Simulation. Mind and Language, 1*, 158-171. Reimpreso en M. Davies & T. Stone (Eds.) (1995) *Folk Psychology. The Theory of Mind Debate*. Oxford: Blackwell.

Gordon, R. M. (1992). The simulation theory: Objections and misconceptions. *Mind and Language 7*, 11-34. Reimpreso en M. Davies & T. Stone (Eds.) (1995) *Folk Psychology. The Theory of Mind Debate*. Oxford: Blackwell.

Gordon, R. M. (1995). *Simulation* without *Introspection* or Inference From Me to You. En M. Davies & T. Stone (Eds.), *Mental Simulation* (pp. 53-67). Oxford: Blackwell.

Gordon, R. M. (1996). '*Radical*' *Simulationism.* En P. Carruthers & P. Smith (Eds.) *Theories of Theories of Mind* (pp.11-21). Cambridge: Cambridge University Press.

Gordon, R. M. (2007). Ascent routines for propositional attitudes. *Synthese, 159*, 151-165.

Gordon, R. M. (2009). Folk Psychology as Mental Simulation. En E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Fall 2009 Edition)*, recuperado en: URL = <http://plato.stanford.edu/archives/fall2009/entries/folkpsych-simulation/>.

Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Brain and Behavioral Sciences*, *16*, 1-14.

Gopnik, A. (1996). The scientist as child. *Philosophy of Science,* 63, 485-514.

Gopnik, A. & Wellman, H. (1994). The Theory Theory. En L. Hirschfield & S. Gelman (Eds.) *Mapping the Mind: Domain Specificity in Cognition and Culture* (pp. 257–93). New York: Cambridge University Press.

Gopnik, A. & Meltzoff A. (1997). *Words, Thoughts and Theories.* Cambridge, MA: MIT Press.

Hutto, D. (2008). *Folk-psychological narratives.* Cambridge, MA: MIT Press.

Leslie, A. (1987). Pretense and representation: The origins of "theory of mind". *Psychological Review, 94*, 412-426.

Leslie, A. (1994). Pretending and believing: Issues in the theory of ToMM. *Cognition, 50*, 211-238.

Leslie, A & German, T. (1995). Knowledge and ability in "theory of mind": One-eyed overview of a debate. En M. Davies & T. Stone (Eds.), *Mental* Simulation (pp. 123-150). Oxford: Blackwell.

Levine, L. J., Prohaska, V., Burgess, S. L., Rice, J. A, & Laulhere, T. (2001). Remembering emotions: The role of current appraisals. *Cognition and Emotion*, *15*, 393-417.

Lewis, D. (1972). Psychophysical and Theoretical Identifications. *Australasian Journal of Philosophy, 50*, 249-58.

Lycan, W. (1987). *Consciousness.* Cambridge, MA: MIT Press.

Lycan, W. (1996). *Consciousness and Experience.* Cambridge, MA: MIT Press.

Machery, E. (2005). You Don't Know How You Think: Introspection and Language of Thought. *British Journal of Philosophy of Science, 56*, 469-485.

Malcolm, N. (1954). Wittgenstein`s Philosophical Investigations*. Philosophical Review*, *53*, 530-559.

Martin, M. (1998). An Eye Directed Outward. En C. Wright, B. Smith & C. Macdonald (Eds.) *Knowing Our Own Minds* (pp. 99-122). Oxford, Oxford University Press.

McFarland, C., & Ross, M. (1987). The relation between current impressions and memories of self and dating partners. *Personality and Social Psychology Bulletin, 13*, 228-238.

Mill, J. S (1865). *An Examination Of Sir William Hamilton's Philosophy And Of The Principal Philosophical Questions Discussed In His Writings.* Vol. 1. Boston: W. V. Spencer Publication.

Moran, R. (1997). Self-Knowledge: Discovery, Resolution, and Undoing. *European Journal of Philosophy, 5*, 141-161.

Nisbett, R. & Ross, L. (1980). *Human Inference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Peacocke, C. (1998). Conscious Attitudes, Attention, and Self-Knowledge. En C. Wright, B. Smith & C. Macdonald (Eds.) *Knowing Our Own Minds* (pp. 63-98). Oxford: Oxford University Press.

Prinz, J. (2002). *Furnishing the Mind. Concepts and Their Perceptual Basis.* Cambridge, MA: MIT Press.

Pylyshyn, Z. (2002). Mental Imagery: In search of a theory. *Behavioral and Brain Sciences 25*, 157-182.

Ryle, G. (1949). *The Concept of Mind.* New York: Philosophical Library.

Schwitzgebel, E. (2010). Introspection. En E. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy (Fall 2010 Edition)*, recuperado en: URL = <http://plato.stanford.edu/archives/fall2010/entries/introspection/>.

Shanton, K. & Goldman, A. (2010). Simulation theory. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*, 527-38.

Shoemaker, S. (1994). Self-Knowledge and 'Inner Sense'. *Philosophy and Phenomenological Research*, *54*, 249-314.

Shoemaker, S. (2010). Self-knowledge and self-identity. En J. *Dancy*, E.*Sosa* & M. *Steup* (Eds.) *A* Companion *to Epistemology* (pp. 724-726). Oxford: Blackwell.

Skidelsky, L. (2006). Personal-*Subpersonal*: The Problems of Inter-level Relations. *Protosociology, Special Issue: Compositionality, Concepts and Representations II: New Problems in Cognitive Science*, *22*, 120-139.

Skidelsky, L. (2008a). Autoconocimiento: conexión constitutiva vs. logro cognitivo. En A. Gianella, M. C. González, y N. Stigol (Comps.) *Pensamiento, representaciones, conciencia: Nuevas reflexiones* (pp. 231-264). Buenos Aires: Alianza.

Skidelsky, L. (2008b). Filosofía de la Mente y Ciencia Cognitiva: la cuestión de cómo concebir y practicar sus relaciones. En D. Pérez y L. Fernández Moreno (Comps.) *Cuestiones filosóficas* (pp. 263-284). Buenos Aires: Catálogos.

Skidelsky, L. (2009). La versión débil de la hipótesis del pensamiento en lenguaje natural. *Theoria 24*, 83-104.

Stich, S. & Nichols, S. (1995). Folk Psychology: Simulation or Tacit Theory? En M. Davies & T. Stone (Eds.) *Folk Psychology. The Theory of Mind Debate* (pp. 123-158). Oxford: Blackwell.

Strawson, P. F. (1954). Critical Notice: Philosophical Investigations", *Mind*, *63*, 70-99.

Wittgenstein, L. (1980). *Remarks On The Philosophy of Psychology.* Oxford: Basil Blackwell.

Wittgenstein, L. (1988). *Investigaciones filosóficas.* México: UNAM.

Wright, C. (1989). Wittgenstein's Later Philosophy of Mind: Sensation, Privacy, and Intention. *Journal of Philosophy*, *86*, 622-634.

Wright, C. (1998). Self-Knowledge: The Wittgensteinian Legacy. En C. Wright, B. Smith & C. Macdonald (Eds.)

Knowing Our Own Mind (pp. 13-46). Oxford: Clarendon Press.