

SIMPOSIO

CUALIDADES SECUNDARIAS Y AUTOCONOCIMIENTO*

JULIA VERGARA

Locke (1690) adopta del físico Robert Boyle y de otros científicos de su época la terminología, el problema y la distinción entre cualidades primarias y secundarias. Además de adoptarlos, Locke ofrece argumentos en su favor. Las cualidades primarias, forma, tamaño, posición, número, movimiento y también solidez y textura, son aquellas cualidades que pertenecen a los cuerpos y permanecen inalterables en ellos no obstante los cambios y divisiones que pudieran sufrir. Al ser propiedades intrínsecas (no relacionales) constituyen el fundamento de las diversas potencias de los cuerpos para producir en nosotros ideas. Locke las considera propiedades reales porque piensa que se encuentran verdaderamente en el mundo, completamente fuera de nosotros, siendo irrelevante para el caso que se las perciba o no.

Las cualidades secundarias, colores, sonidos, sensaciones térmicas y sabores, tienen en cambio una naturaleza mixta; por un lado, son potencias o poderes de los objetos que al unirse a las cualidades primarias producen en nosotros las ideas correspondientes de, por ejemplo, rojo, amargo, agudo. Pero, por el otro, los colores y sabores, las sensaciones de frío y calor propiamente dichos no se encuentran, ontológicamente hablando, en las cosas mismas sino que se encuentran en nuestra sensibilidad o en nuestra experiencia al percibirlos. Debido a que se trata de propiedades que atribuimos irreflexivamente a los cuerpos cuando en realidad no están completamente en ellos surge la vieja idea de considerar a tales atribuciones como "irrealistas". Sin embargo, ello parece excesivo ya que la ciencia moderna ha sustituido los colores por radiaciones electromagnéticas explicando con ello el proceso de acuerdo al cual en la superficie de los cuerpos tales radiaciones se absorben y reflejan de forma variable, y señalando que en seres como nosotros, dotados de un cierto aparato perceptual y de la aptitud del lenguaje, se producen diversos estados discriminativos que eventualmente se expresan en juicios verbales acerca de los colores del mundo.

* Este trabajo fue presentado en las Segundas Jornadas de Filosofía de la Mente y la Psicología organizadas por los proyectos UBACYT FI-156 (UBA), SECYT (UNC) y CIUNSA (UNSA) en Vaquerías, Córdoba, en julio de 1996. Agradezco especialmente a Diana Pérez, Carolina Scotto y Nora Stigol por sus comentarios. En 1995 asistí a un seminario en la Universidad de Columbia que realizó Akeel Bilgrami sobre el problema del autoconocimiento. El marco de esas discusiones dio lugar a este artículo al que la lectura y objeciones críticas de Eduardo Rabossi permitieron su elaboración final. Hacia ambos, mi renovada gratitud.

Pero el paradigma de claridad y objetividad que persigue la ciencia moderna requiere la exclusión de descripciones en términos de propiedades subjetivas, de propiedades del objeto en tanto es experimentado. Así, la relevancia de la distinción entre cualidades primarias y cualidades secundarias proviene del hecho de que las cualidades secundarias son inútiles desde un punto de vista científico. Lo son, tanto si se desea dar cuenta de las interacciones causales que se dan entre las cosas como si lo que se quiere explicar es la percepción que tenemos de ellas. Se ha atribuido a Locke sostener que en nuestro pensar ordinario cometemos el error sistemático de atribuir colores, sonidos y sabores a las cosas materiales confundiendo con ello unas cualidades con otras. No obstante esta interpretación, la distinción Locke-Boyle coincide en gran medida con la que Aristóteles hiciera entre “sensibles comunes” y “sensibles especiales” y la que anteriormente hicieran los atomistas griegos. Fue Demócrito quien sostuvo que lo dulce y lo amargo, el frío y el calor y la gama entera de los colores existen sólo en forma convencional; en realidad, únicamente existen los átomos y el vacío. En nuestra época la distinción primario/secundario ha sido defendida por J. L. Mackie (1976) y revalorizada por J. Bennett (1965), aunque en versiones diferentes. También puede introducirse como un análisis de lo que ordinariamente decimos y juzgamos y es esto lo que ha hecho Crispin Wright (1989) en una refinada versión del antirrealismo acerca de ciertas formas de oraciones que adscriben estados mentales. Charles Taylor (1977) recurre también a la distinción, esta vez en relación con la naturaleza del autoconocimiento en general y con las emociones en particular.

En las secciones que siguen comenzaré por el análisis propuesto por Wright para los juicios en primera persona, la estructura lógica de las autoadcripciones, y la analogía de los mismos con los juicios acerca de cualidades secundarias. Luego examinaré el modelo perceptual (observacional) del autoconocimiento tanto en la versión clásica como en su variante contemporánea para enseguida considerar su modelo alternativo, el modelo inferencial que aconseja G. Ryle. En oposición a ambos enfoques, defenderé la versión que parece haber sugerido Wittgenstein, que suele denominarse tesis de la constitutividad. La idea básica es que el autoconocimiento no es un tipo de conocimiento si se piensa en los paradigmas anteriores (observacional, inferencial) sino la forma misma de la intencionalidad. En lo que hace a la justificación de la autoridad especial de las personas sobre sus propios pensamientos, seguiré la propuesta de Wright aunque introduciendo una variante en la misma. Hacia el final del trabajo y sin detenerme demasiado en ello, mencionaré el conflicto entre el autoconocimiento y ciertas versiones externalistas.

I

Crispin Wright (1989) propone un test. El test consiste en confrontar los mejores juicios, aquellos que se realizan en condiciones cognitivas ideales (C), y su verdad. Pasar el test requiere que ambos términos de la relación sean de algún modo independientes, es decir, que la verdad del juicio no dependa de ningún privilegio cognitivo. Por el contrario, los juicios que no logran pasarlo son aquellos en los que la distancia entre ser verdadero y ser el mejor prácticamente no existe, es decir, que su verdad la constituye lo que juzgamos que es cuando operamos bajo condiciones cognitivas ideales.

Ocurre que los juicios acerca de cualidades secundarias no logran pasar el test dado que, cuando, por ejemplo, decimos “Esto es rojo”, sólo nuestro propio juicio (bajo C) y ninguna otra cosa más lo hace verdadero. Esto simplemente obedece a que en la naturaleza de una propiedad secundaria como el rojo está el “parecer rojo”, a diferencia de lo que ocurre con las cualidades primarias.

Tenemos entonces un contraste entre i) juicios en los que las mejores opiniones *determinan* la extensión del predicado “verdadero” y ii) juicios que en el mejor de los casos *reflejan* una extensión independientemente determinada. Veamos dos ejemplos. Pensemos en una superficie roja y en el juicio “X es rojo”. Por condiciones cognitivas ideales (C) se entiende en este caso buena vista, buena luz, proximidad con el objeto, observación atenta y además que la persona sea competente con el concepto de rojo. Establecidas esas restricciones, para determinar la extensión de un concepto de color debe ser verdadero *a priori* que el concepto se aplique cuando juzgamos que se aplica.

Consideremos ahora una cualidad primaria, la forma, y el juicio correspondiente “X tiene forma de pera”. Aquí las condiciones cognitivas ideales son diferentes de las anteriores. La luz no necesita ser perfecta, el sujeto debe hacer varias observaciones y no una sola, y tercero, para que sea verdadero *a priori* que la mejor opinión de un agente en condiciones cognitivas ideales sea la correcta, debemos asegurarnos de que no existan cambios en la forma de x durante el período de observación, a lo que también puede agregarse que la mejor opinión requeriría de varios observadores para ser efectivamente la mejor. En realidad el punto que interesa mostrar es que la diferencia entre ambos tipos de juicios reside en que en el primer caso (el color) la apariencia visual (bajo C) es suficiente, mientras que en el segundo no lo es. Desde ya que el test no está diseñado para establecer la distinción de manera definitiva ni se pretende que la misma sea generalizable a otras instancias de la distinción primario/secundario. Lo que Wright propone es más bien un marco para la discusión en esos términos, además de dar crédito a la idea de que hay una distinción que vale la pena trazar. El punto cobra especial relevancia porque

los juicios visuales acerca de colores son una mezcla interesante de subjetividad y objetividad. Ni los colores existen en las cosas tal como los vemos ni los vemos como los vemos independientemente de las cosas. De modo que establecer cuál es la mejor opinión acerca de un color es algo razonablemente objetivo, algo que nadie discute ya que la mayoría de las personas suelen en general coincidir. Quiero decir, cuando digo por ejemplo que la bandera es roja no estoy haciendo una apreciación personal como si dijera "Fulano me cae bien" o "Mi *qualia* actual es inefable", simplemente digo que cualquier persona con una visión normal y en condiciones normales vería lo mismo que yo veo. Es en este sentido que las cualidades secundarias no son subjetivas. Pero el hecho de que existan juicios en los que nuestras mejores opiniones determinen su extensión y que el mismo tipo de análisis pueda aplicarse a las autoadcripciones de estados psicológicos como creencias, deseos o emociones los vuelve particularmente atractivos para la legitimación epistémica del autococonocimiento. Los juicios en primera persona, como aquellos que se refieren a colores, son "determinantes de la extensión", es decir, no son juicios que reflejen una extensión independientemente determinada, como ocurre con aquellos referidos a cualidades primarias.

Para mostrar esto último se deben explicitar las condiciones que aseguren *a priori* que los sujetos creen tener un cierto estado intencional si y sólo si lo tienen. ¿Qué impedimentos tendrían que presentarse para que ocurra que si creo que creo que p no sea verdad que creo que p , y si creo que p tampoco lo sea que creo que creo que p ?

Suele decirse, y Wright lo dice, que un impedimento posible es que opere solapadamente alguna forma de autoengaño. El sujeto puede ser incapaz de traer (o conservar) en su conciencia "la verdadera intención" que lo guía, puede creer que actúa por A y no por B. También puede engañarse confundiendo su deseo con lo que en realidad piensa (*wishful thinking*). Pero en ambos casos hay un "motivo" para que la creencia falte y en ciertas circunstancias esto mismo podría tener una explicación fisiológica o farmacológica, luego de un shock, por ejemplo. De todos modos podemos pensar que si se establece un conjunto apropiado de condiciones ideales (C) podríamos asegurar que el bicondicional en cuestión es verdadero.

Además de autoengañarse el sujeto puede no estar atento o puede no contar con los recursos conceptuales pertinentes. Uno podría pensar que si se establecen con todo detalle estas tres condiciones (prohibición de autoengaño, estar atento, contar con el concepto pertinente) el bicondicional quedaría asegurado: los sujetos creen tener un cierto estado psicológico si y sólo si lo tienen.

Wright encuentra que la dificultad mayor consiste en presentar una especificación sustantiva del autoengaño. Porque el autoengaño cubre cual-

quier condición motivada¹ que pueda conducir al sujeto a ignorar su propia creencia, pero necesitamos admitir también condiciones no-motivadas (por ejemplo, las inducidas químicamente). De modo que pareciera que si pretendemos dar una especificación positiva del autoengaño, tendríamos que reformular la condición diciendo que el sujeto debe encontrarse “libre de cualquier cosa que pueda impedir la capacidad de certificar sus propias intenciones” y esto, por supuesto, vuelve insustancial la condición. Los esfuerzos de Wright para formular de manera no trivial la condición que prohíbe el autoengaño a fin de asegurar la verdad *a priori* del bicondicional son varios, pero no me detendré en ellos aquí; digo solamente que luego de batallar con la sustantividad de la condición en cuestión Wright debilita la exigencia del bicondicional cambiando “verdadero *a priori*” por “creíble o razonable *a priori*”. Planteemos la pregunta ahora en estos términos: ¿por qué es razonable creer que si *x* cuenta con el concepto pertinente, si está atento, y además, si no existe evidencia contraria, *x* creerá que tiene la creencia de *p* si y sólo si la tiene? ¿No es esto razonable *a priori*? Personalmente creo que sí. Creo, siguiendo a Wright, que los juicios psicológicos en primera persona, las autoadscripciones, son juicios determinantes de la extensión del mismo modo que lo son los juicios acerca de propiedades secundarias. Es decir que no hay ninguna otra cosa que los haga verdaderos que no sea, con las condiciones correspondientes, nuestra mejor opinión. En suma, nuestra mejor opinión bajo *C* constituye su verdad. Volveremos luego sobre esto.

Antes de continuar quisiera decir algo más acerca de la distinción primario/secundario respecto de los compromisos que la distinción supone. Este trabajo no es una exégesis de Locke, quien por otra parte propuso diferentes versiones de la distinción, las que a su vez dieron lugar a diferentes lecturas. Por tal motivo no voy a detenerme en la discusión acerca de si la teoría de Locke efectivamente se inscribe en el marco de una teoría representacional de la percepción. Mackie y otros creen que sí, Bennett y otros creen que no, y yo por mi parte suscribo a la versión de Bennett cuya defensa de la distinción es la menos comprometida. No se requieren “sense data” ni “representaciones”, como tampoco se requiere aceptar la distinción extrínseco/intrínseco ni suscribir a la dicotomía apariencia/realidad. El contraste de cualidades que establece Bennett concierne sólo al modo en que las cosas interactúan.

En segundo lugar considero que la distinción es interesante no sólo desde un punto de vista científico sino también filosófico. Es decir que no es indispensable apelar a la microestructura de la superficie de los cuerpos y a las longitudes de onda que la luz refleja para admitir que las formas, a diferencia de los colores, interactúan en el mundo de una cierta manera y que esas mane-

¹ Cf. Davidson (1982) y (1985b).

ras de interactuar no son las que de hecho se dan en el caso de los colores, sonidos, sabores. Una cosa, por ejemplo, no puede detener la caída de otra sin dejar de tocarla ni un cubo puede rodar sobre una superficie plana. O puede darse como ejemplo que la marca de un disco circular sobre la cera será siempre circular o que en un recipiente más grande podemos poner uno más chico y así sucesivamente. El punto es que estas conexiones obvias y familiares de interacción causal simplemente no se dan entre las cualidades secundarias.

Por otra parte quisiera agregar que considero a los colores reales y no ilusorios, aun cuando acepto que la realidad que los contiene no sea del todo independiente de los modos subjetivos de percepción. Vale la pena recordar que, ante la duda escéptica que recurre a la alucinación o al sueño, Locke responde en favor del carácter verídico de “el sabor del ajeno, el olor de una rosa, el calor del fuego y la blancura de una página” (*Essay*, IV, ii, 14, xi, 2). También es cierto que las cualidades secundarias son relativas y subjetivas en el sentido de que los propios rasgos experienciales son constitutivos de su presencia. Pero no debe confundirse esto con lo que denota el término filosófico “qualia”, si hubiera algo así, pues éstas son propiedades no de los objetos de la experiencia sino de la experiencia misma: propiedades intrínsecas. Pensar entonces que sólo las cualidades primarias son reales significa olvidar o negar que el mundo –el mundo en el que estamos– está estrechamente ligado a nuestra manera de concebirlo. O puede que signifique no distinguir las nociones de verdadero y objetivo ligadas a lo real. O acaso considerar que sólo son reales las propiedades que la última teoría científica acepta. Naturalmente que algunas de las formulaciones de Locke invitan a estos equívocos ya que su afirmación de que las cualidades secundarias “no están en las cosas mismas” llevó a pensar que entonces eran ideas y siendo ideas, como razonó Berkeley, no pueden estar sino en la mente. Pero esta interpretación sesgada no es lo que Locke entendía, porque del hecho de que no se encuentren completamente en el objeto no se sigue que estén sólo en la mente. Más bien están en la mente porque el objeto tiene ciertos poderes para causarlas, como dice Locke más de una vez. De modo que creo que las cualidades secundarias se caracterizan por los siguientes tres rasgos: son disposicionales, son relacionales y requieren de una mente capaz de percibirlas. Blackburn (1993) dice que es un hecho científico que son supervenientes. Smart y Putnam ponen esto en duda. Yo francamente no lo sé. No entiendo bien si significa que no pueden variar sin que varíen al mismo tiempo las cualidades primarias o si significa otra cosa. Tampoco sé si son reducibles o no. Locke creyó que lo eran. Y Kripke, siguiendo a Locke, las cree reducibles a esencias reales. Sin perjuicio de ello, creo que la distinción primario/secundario es legítima y filosóficamente útil ya que su aceptación no exige asumir ninguno de los compromisos mencionados.

II

Hasta aquí hemos visto que las autoadscripciones comparten las características de los juicios acerca de propiedades secundarias. En ambos casos las propiedades involucradas dependen de la experiencia.²

Preguntemos ahora qué tipo de conocimiento, si lo es, es el autoconocimiento y en tal caso qué modelo epistémico le corresponde. Y si no lo es, veamos cómo hacemos para explicar que normalmente conocemos nuestros pensamientos y deseos propios. Tendríamos que dar cuenta de la “autoridad especial” que sentimos tener sobre ellos, o al menos justificar la fuerte intuición que al respecto tenemos. Pienso que deberíamos comenzar con Hume.

Hume (1739) negó que en nuestra conciencia podamos encontrar al yo en tanto sujeto de experiencia, observando que sólo podíamos encontrar percepciones particulares. La negación de Hume, a la que suscribieron Lichtenberg, Wittgenstein y Anscombe, ha sido ampliamente discutida y ha tenido implicaciones tanto en lo que concierne a la naturaleza del autoconocimiento como al tema puntual de la autorreferencia. Pero Hume dijo también que amar, odiar, sentir o pensar no son otra cosa que “percibir”, lo cual permite agregar que percibir, como “aprehender” o “estar en contacto con”, son tipos de relaciones.

De acuerdo con lo que se denomina el “modelo perceptual”, el autoconocimiento es presentado como una forma de observación interna. Una observación que se concibe en la línea de la percepción sensorial pero que difiere de ella en tanto concierne únicamente al “sentido interno” y no a los sentidos externos. Es cierto que puede hacerse una distinción y convenir con Hume en que el yo consigue eludir sistemáticamente dicha mirada, pero de acuerdo al modelo perceptual no podemos decir lo mismo de los destinatarios de la relación, sus objetos. Desde la época clásica hasta nuestros días este lugar ha sido ocupado alternativamente por “imágenes”, “ideas”, “datos de los sentidos”, “contenidos” (y más conflictivamente, “proposiciones”), cuyo privilegio epistémico procedía del puesto de honor que se les había asignado. Recordemos que tanto para el cartesianismo como para el empirismo clásico

² Charles Taylor (1977) traza también un paralelo con los juicios acerca de cualidades secundarias pero esta vez en relación con la estructura lógica de las emociones. Sostiene que experimentar una emoción es experimentar nuestra situación personal de una cierta manera y que en tal interpretación nos constituimos en los seres que somos. En su análisis, las emociones humanas tienen propiedades que sólo pueden existir en un mundo en donde existen seres con experiencias subjetivas, seres para quienes la situación personal no les es de ningún modo indiferente. Taylor señala que en los juicios acerca de cualidades secundarias y en aquellos que conciernen a las emociones, las propiedades involucradas soportan una dependencia esencial con la experiencia subjetiva (*subject-referring properties*) y que en ambos casos se da una mezcla de subjetividad y objetividad.

una de las cualidades que se les atribuyen a tales objetos (si fuesen para Descartes objetos) es su “transparencia”, cualidad que los vuelve inmunes a todo posible error y que convierte a nuestros juicios acerca de ellos en incorregibles o infalibles, justificándose así la autoridad “especial” de la primera persona sobre sus pensamientos. Por otra parte, el modelo perceptual u observacional se encuentra fuertemente arraigado en las metáforas visuales del lenguaje corriente, de modo que si no se diese por sentado que algo efectivamente es percibido en la introspección, es posible que el problema de Hume no se hubiese presentado. Y por supuesto, si se rechaza el modelo observacional, el problema de Hume, la elusividad del yo, también desaparece.

Por distintas y buenas razones,³ la idea de que la mente se encuentra poblada de objetos ha sido dejada a un lado. Alternativamente se ha propuesto un enfoque monádico de las experiencias mentales en el que las creencias (y demás actitudes proposicionales) “se adjetivan” en el sujeto. De ese modo se ha eliminando la idea de la relación con un objeto (lo cual no significa negar el carácter relacional de la “atribución” de creencia) aunque esto por sí solo no implica renunciar a la idea de que el autoconocimiento sea una forma de observación interna. Aun eligiendo la concepción monádica pero conservando el modelo perceptual, el sujeto mismo, el yo, no resulta eliminado, pues si se permiten experiencias en el sujeto se admite que “alguien” las experimenta (alguien a quien se observa), con lo cual la negación de Hume se revierte de modo tal que ahora ante la mirada introspectiva sólo encontraríamos al yo en una u otra de sus modificaciones. Esto parece reducir al absurdo la idea misma de observación o convenir con Hume en que ante la mirada interna sólo encontramos distintas percepciones. Pero entonces, volvemos a los objetos.

Más aun, consideremos algunas de las cosas que en términos generales están en juego en la percepción. La percepción supone un espacio de cosas en el cual resulta posible recabar información identificatoria sobre algún particular. Supone también la posibilidad del error, el hecho de codificar en la operación una información falsa. Si esto es básico a la percepción, no lo es sin embargo a la introspección, al menos en lo que concierne al yo, pues nadie sensatamente se pregunta si será él quien camina o si será algún otro; ni tampoco si es a él a quien le duele su cabeza o quien cree que está lloviendo. De modo que no es cierto que la introspección nos provee información identificatoria del yo ni buena ni mala, como la que provee algún ítem que identificamos en la percepción normal. Pero supongamos que se decida evitar el problema de Hume. Alguien puede sostener que nuestra mente se encuentra poblada de imágenes y que esto es tan obvio como lo es el hecho de que pode-

³ Cf. Rorty (1979) y Davidson (1985a y 1991).

mos producirlas a voluntad. Podemos por ejemplo cerrar los ojos e imaginar un gato, podemos imaginarlo en la alfombra, podemos tener esa experiencia o imaginar que la tenemos. Sin embargo, no hay algo como una imagen del gato en la alfombra. Lo que tenemos es la “sensación” de ver un gato en la alfombra y una sensación es algo distinto de una observación. Esta última presupone a la primera pero no a la inversa. De modo que no puede hablarse de recabar información identificatoria como ocurre en el caso de la percepción externa.

Por otra parte, de acuerdo al modelo perceptual uno debería tener algún autoconocimiento previo que nos garantice que al observarnos internamente no fuéramos a encontrarnos con la sorpresa de que “allí” se encuentra otra persona. Pero si debo suponer que todo autoconocimiento es observacional, tal conocimiento previo también habría de serlo, con lo cual sería imposible evitar el regreso; en algún momento debo postular un conocimiento de mí que no implique la observación, en cuyo caso el modelo perceptual mismo se desvanece.⁴

Otro elemento que incluye la percepción es la idea de referencia demostrativa. No voy a detenerme en los argumentos que muestran que ello no ocurre en el autoconocimiento, mencionaré solamente lo que sucede en el caso de un estado mental como el dolor. Tener un dolor y sentirlo son conceptual y experiencialmente una y la misma cosa. Sentir un dolor consiste en saber que se lo tiene y no hay en ello algo “adicional”, no hay dolores que puedan no ser sentidos, al menos en tanto la palabra “dolor” signifique lo que en este mundo significa. Tal es, según creo, la gramática lógica del dolor.

Asimismo, en la percepción habitual cabe tanto la identificación correcta como la incorrecta, cosa que no cabe en la introspección. Si creo por ejemplo que el comandante Marcos es un líder revolucionario y sé que lo creo (aunque no agregue nada) no por ello he identificado a un objeto, en este caso el contenido de mi creencia, al cual le presto mi atención trayéndolo a la conciencia. Como hemos dicho antes, varios filósofos han argumentado en contra del enfoque relacional-objetual de la creencia, y yo por mi parte suscribo a tales críticas (por ejemplo, Davidson, 1985, 1990). Pero si alguien quisiera defenderlo postulando además el modelo perceptual del autoconocimiento debería admitir o bien la transparencia del objeto en cuestión y con ello la infalibilidad, o bien que el error es posible, esto es, en nuestro ejemplo, que creo que creo que Clinton y no Marcos es un líder revolucionario o que dudo acerca de si tengo un dolor o no lo tengo. Y alternativamente también, con objetos o sin ellos, que creo que siento sueño, hambre o sed pero que tal vez

⁴ Para los problemas que presenta el modelo perceptual, cf. Shoemaker (1986, 1994).

no sienta nada de eso, o como dijimos antes, que dudo de si seré yo misma quien siente y cree todas esas cosas. Por lo que vemos, el modelo perceptual (versión objetos) tiene serios inconvenientes.

D. Armstrong (1968) presenta una versión distinta del mismo. En lugar de decir que la introspección es “contacto directo” (*acquaintance*) con objetos internos, habla de algo más abstracto. Habla de adquirir información acerca de estados que, de hecho o potencialmente, causan nuestra conducta. Esta es una variante bastante más sofisticada que la anterior ya que incluye no sólo estados de conciencia vívidos sino estados disposicionales. Pero aun con esta variante el paralelo que traza Armstrong con la percepción externa sigue siendo el eje de su enfoque. La introspección no es sino percepción interna. De modo que si se acepta que no existe ninguna diferencia importante entre ambas percepciones, se está aceptando que lo que se aplica en un caso se aplica en el otro. Y así reconoce Armstrong que de existir alguna diferencia entre las creencias acerca del mundo externo y las que tenemos acerca de nosotros mismos, no sería tanto porque estén en juego diferentes modelos epistémicos sino meramente porque las cosas que conocemos son distintas. Por otra parte, tomar literalmente el paralelo entre los dos tipos de percepción le permite a Armstrong rechazar los rasgos cartesianos del autoconocimiento. Me refiero a lo siguiente. Cuando tenemos un conocimiento perceptual de un objeto cualquiera, percibimos sólo algunas de sus partes mientras que otras permanecen ocultas. Esta limitación que generalmente se debe a la perspectiva, la luz, etc., es lo que entre otras cosas permite dar cuenta de los errores perceptuales. En cambio, para Descartes y los empiristas no hay un paralelo posible con las percepciones internas ya que éstas, como dice Hume, “son lo que parecen y parecen lo que son”, es decir que conocemos todas sus propiedades, todas sus facetas, haciendo con ello al error imposible.

Pero para Armstrong el error es y debe ser posible. Es posible porque la información que adquirimos sobre nuestros estados internos (vívidos o disposicionales) es parcial, como lo es la información externa y por ello nunca llegamos a captarlos en su totalidad. Esto último obedece a que las disposiciones se describen en términos potenciales, es decir con una lista no exhaustiva de conductas posibles. Con esta variante del modelo clásico (tesis de la incorregibilidad) Armstrong evita el compromiso de postular estados que operen como intermediarios epistémicos y que conviertan a nuestra mente en la fuente de toda certeza posible (acceso privilegiado).

El ataque de Armstrong al cartesianismo es crucial para su teoría ya que, como él mismo reconoce, “si la incorregibilidad y el acceso privilegiado fuesen verdaderos mi teoría sería falsa” (p. 102). Y vemos que sería falsa por lo siguiente: si los procesos y estados mentales están ligados esencialmente a la conducta (por definición de estado mental la conexión es no-contingente) y

si esos estados son estados físicos del cerebro (materialismo del estado central) entonces la introspección es también un proceso físico, un proceso de autoobservación. Pero si lo es, es lógicamente posible que este mecanismo (como cualquier otro mecanismo) arroje resultados falsos. De lo contrario, si el conocimiento introspectivo fuese incorregible, concluye Armstrong, la teoría materialista del estado central, su teoría, sería falsa. Dicho en términos dramáticos, la vida de una concepción implica la muerte de la otra. Armstrong lo dice también de otra manera. Si la doctrina de la incorregibilidad, transparencia, acceso privilegiado, fuese verdadera, habría una brecha entre los enunciados perceptuales acerca del mundo externo y los enunciados en primera persona, las autoadscripciones, para las cuales nuestra posición sería lógicamente privilegiada; pero si tal brecha existiese, su teoría sería falsa.

Estas dos razones muestran hasta qué punto es crucial para Armstrong oponerse a lo que él caracteriza como el modelo cartesiano de la introspección. Y para ello ofrece otros dos argumentos o más bien uno solo que consta de dos partes: a) cualquier cosa que podamos decir, por breve que fuese, requiere un determinado tiempo. Cuando decimos por ejemplo "Ahora tengo dolor", el "ahora" indica la porción de tiempo durante la cual el juicio ha sido pronunciado. De modo que la indubitabilidad (incorregibilidad) alcanza sólo al instante de la proferencia. Pero una vez que éste ha transcurrido sólo puedo confiar en mi memoria y esperar el instante futuro. Pero entonces existe la posibilidad del error; luego la indubitabilidad lógica del enunciado "Ahora tengo dolor" no puede ser tal. b) Si los estados mentales yacen en el pasado, el hecho de que seamos conscientes de haberlos tenido (por próximos que fuesen) presupone una "existencia distinta" de aquellos. Se sigue entonces la posibilidad lógica de que lo que creemos saber conscientemente sea falso. A continuación Armstrong conjetura cómo deben razonar quienes se oponen a sus argumentos. Dice aproximadamente así: quien defienda la tesis de la incorregibilidad/indubitabilidad tiene que sostener que mi conciencia del dolor y mi dolor no tienen una existencia distinta. Pero para decir esto último, que se trata de un solo y único estado, los defensores de esta idea tienen que estar pensando que del mismo modo que la forma de un cuerpo no tiene una existencia distinta de su tamaño tampoco la tienen el color y la extensión ni el dolor y la sensación de dolor. Para Armstrong, en cambio, la conexión no debe ser conceptual: "debe haber una distinción absoluta entre el observador y lo observado" (p. 107).

Lo que parece ser común a las variantes del modelo perceptual del autoconocimiento es lo siguiente: los objetos o estados que se perciben existen independientemente de que sean percibidos; de modo que se permite que existan con independencia de que existan seres con capacidad para percibirlos (seres con creencias pero sin autoconocimiento). Podría decirse que esta

afirmación puede leerse tanto en un sentido metafísico como en un sentido lógico o conceptual. En este caso tener un dolor sería lógicamente independiente de sentirlo. Es justamente para oponerme a esto último que comencé este trabajo mencionando las cualidades secundarias, con la idea de introducir a continuación el análisis de la estructura lógica de las autoadscripciones que hace Wright. En ambos tipos de juicios, como dijimos antes, las propiedades involucradas dependen de la experiencia (juicios determinantes de la extensión) y esto es algo que va en contra de lo que afirma el modelo perceptual que, como decimos, parece insostenible.

Pero antes de continuar quisiera referirme a otro modelo, alternativo del anterior, presentado por Gilbert Ryle en el capítulo VI de *El concepto de lo mental*. La idea de Ryle es que no existe ninguna diferencia entre la naturaleza de la evidencia que utilizamos para determinar lo que los demás sienten y piensan y la que utilizamos respecto de nuestra propia persona. Ryle rechaza la explicación del autoconocimiento en términos de observación interna señalando que tal concepción es la consecuencia natural del enfoque cartesiano con su fantasmal idea de teatro privado, sus nociones de claridad, distinción, transparencia, y consecuentemente con ello, la idea epistémica de un "acceso privilegiado". Por atractivo que todo esto pueda resultar, dice Ryle, nada semejante existe en nosotros. No existe el teatro ni existen sus actores y sólo cometiendo un grueso error categorial se puede llegar a hablar en esos términos. No obstante esto último debemos dar cuenta del hecho de que normalmente sabemos lo que pensamos y Ryle lo explica de este modo. Comienzo con uno de sus ejemplos. Si vemos que una persona se jacta constantemente de sus logros, si se burla de los defectos ajenos, si es proclive a frecuentar personas importantes y a colocarse en el centro de las reuniones sociales, podemos inferir legítimamente que se trata de una persona vanidosa. No de un modo diferente (aunque con mayor benevolencia) descubro mis propios rasgos de carácter. Observo mi comportamiento en diversas ocasiones, presto atención a mis dichos y respuestas, consulto otra vez mi diario personal y a partir de ello extraigo la conclusión correspondiente. En un caso y en otro se trata de un procedimiento inductivo. Un procedimiento no muy diferente del que aplica el hombre de campo cuando intenta conocer y predecir los cambios del clima o del que aplica el médico para formarse una idea global de cómo es su paciente. Por supuesto que el rigor del método estadístico queda descartado pero puede decirse que en términos generales tales inferencias funcionan. Por otra parte, Ryle reconoce que algunas inducciones son más sencillas que otras y que ciertas personas están mejor dotadas al respecto; pero lo mismo podemos decir acerca de nosotros. No sólo es difícil detectar a un impostor, a veces también es difícil saber por qué hacemos lo que hacemos, con el agravante de que podemos autoengañarnos completamente,

cosa que en general los demás advierten con total claridad. De modo que, según Ryle, la gran diferencia práctica entre ambas formas de conocimiento reside en la situación en que nos encontramos unos y otros, situación que no es asimétrica respecto de las ventajas que tendríamos cuando se trata de nosotros mismos. Depende de los casos. Podemos notar que los ejemplos que elabora Ryle son casos típicos de rasgos disposicionales de carácter y que se requiere explicar el conocimiento de estados puntuales, vívidos, por decirlo así. Por ejemplo, que me descubro habiendo captado una broma o me doy cuenta de que dije lo que dije sólo por desprecio o advierto que me invade una ira repentina. Lo advierto instantáneamente (me pasa a mí). ¿Me doy cuenta de lo que le pasa al otro de la misma manera? Aparentemente puedo saberlo si advierto que ante la broma sonríe, si dice algo negativo de quien acaba de abandonarlo o si lo veo acalorado, gritando y con una mueca de disgusto en su cara. ¿Se dirá entonces que no hay un proceso inductivo, que simplemente observo la conducta ajena en un caso y que en el otro me observo a mí? No exactamente, responde Ryle. La prueba de que alguien tiene sentido del humor o se siente despechado o es fácilmente irritable, la proporcionan sus actos y sus reacciones, y una vez que detectamos esas cualidades duraderas, explicamos la acción particular. Es cierto que puede haber expresiones y acciones que hablan por sí solas, pero por lo general conocernos y conocer a los otros requiere un proceso de inferencia inductivo.

En esta sección hemos visto que el modelo perceptual que defiende Armstrong pone en un pie de igualdad la percepción externa y la interna de modo que las ideas de incorregibilidad e infalibilidad deben dejarse a un lado. El modelo de Ryle también las descarta asimilando ambas formas de conocimiento a un único proceso inferencial. Podríamos preguntar entonces: ¿qué queda de la intuición de la que partimos, de la idea de que creemos saber mejor de nosotros mismos de lo que cualquiera pueda llegar a saberlo?

III

Quienes han aceptado la idea de que el autoconocimiento es un tipo de conocimiento apelan para su justificación a uno de los dos modelos posibles: el perceptual (observacional) o el inferencial. La forma más elaborada del primero, la representada por Armstrong, ha sido fuertemente criticada entre otros por Chisholm (1969), Shoemaker (1986, 1994) y Burge (1988). La otra línea defendida por Ryle sostiene que el autoconocimiento no difiere en nada del que tenemos respecto de las restantes personas y se lo justifica apelando a una serie de inferencias muy rápidas. Esta propuesta también ha sido criticada, entre otros, por Davidson (1987).

En oposición a ambos modelos se ubica la tesis que parece haber surgido con Wittgenstein (1953) y que luego fue adoptada por Davidson y Burge, aunque a mi juicio adquiere una forma particularmente interesante en el análisis que realiza Crispin Wright. La idea es negar que el autoconocimiento sea un tipo de conocimiento en el sentido de los paradigmas mencionados; esto es, negar que haya algo como una observación interna por la cual adquirimos información que resulta ser un “logro cognitivo” y negar que se trate de inferencias. Por logro cognitivo se entiende el que proporciona el conocimiento empírico basado o bien en la observación o bien en inferencias efectuadas a partir de ella. La idea, en cambio, es que el autoconocimiento es la forma misma de nuestra intencionalidad, que es constitutivo de ella así como la sensación de dolor es constitutiva de los estados no intencionales como el dolor. Wittgenstein observa: “No puede para nada decirse de mí (excepto quizá como broma) que *conozco* que tengo dolor. ¿Qué habría de significar esto excepto quizá que *tengo* dolor?” (IF, 246).

Naturalmente que si aceptamos esta tesis o bien decimos que la constitutividad es un hecho primitivo o bien intentamos explicarlo y/o justificarlo. Vimos que en los juicios acerca de propiedades secundarias la instanciación de un color se analiza en términos de la experiencia que produce. Se dice que está en la naturaleza del rojo (tesis disposicional) el “parecer” rojo, de modo tal que mi experiencia de rojo es constitutiva de la presencia de la cualidad en cuestión; y se dice que esto mismo no se aplica a las cualidades primarias respecto de las cuales contamos con un criterio independiente para juzgar si la cualidad se halla o no instanciada. Análogamente decimos que un sujeto cree tener un cierto estado intencional si y sólo si lo tiene, y que esto se justifica, como dice Wright, porque en las autoadscripciones y en los juicios acerca de propiedades secundarias, la extensión (lo que hace verdadero al predicado) depende de nuestra mejor opinión, lo que equivale a decir que el autoconocimiento no es un logro cognitivo. De modo que el carácter directo y la autoridad de la primera persona no se basan en nada empírico.

Claro que decir que el autoconocimiento es la forma misma de nuestra intencionalidad significa que está en la naturaleza del pensamiento el que tengamos conciencia del mismo y, por supuesto, otro tanto ocurre con el resto de las actitudes proposicionales, deseos, planes, expectativas, etc. Decimos de todos ellos que son “nuestros”, que nos pertenecen íntimamente, que son privados, aunque ciertamente ello es posible sólo a partir del lenguaje y la conciencia (además de la percepción del mundo externo). Es cierto también que habitualmente nos referimos a creencias y deseos inconscientes, a cosas que se encuentran en nuestra mente y guían nuestra conducta, pese al hecho de que ignoramos su presencia. Pero ciertamente, por un lado, son eso, creencias o deseos *inconscientes*, y por el otro, tiene sentido que hablemos de este modo

sólo en razón de que habitualmente conocemos lo que sentimos y pensamos, y es en relación con esto último que hablamos de aquellos. Desde ya que lo que sabemos lo sabemos estando despiertos y en estado normal y no cuando dormimos o cuando estamos bajo una droga o anestesia. Si preguntamos “¿Sabe fulano que yo vine?” (y fulano está durmiendo), lo que queremos preguntar es si fulano supo, estando despierto, que yo haya venido y no si sabe que yo vine en ese mismo momento en que no sabe nada porque está durmiendo. Por lo tanto, el autoconocimiento es conocimiento consciente susceptible de ser pensado y expresado por medio del lenguaje y esto último es crucial pues aquellos organismos menos evolucionados que nosotros (y también los recién nacidos y los sordomudos antes del aprendizaje de un código de signos), aun exhibiendo diversas formas de conciencia (conciencia del dolor, del hambre, del peligro) no llegan a tener un conocimiento articulado de sí. El autoconocimiento requiere entonces la conciencia pero no a la inversa.

Ahora bien, cuando Descartes establece el *cogito*, da a continuación dos pasos adicionales. Primero presenta a la mente como la base epistémica, como lo único que conocemos de manera indubitable, lo paradigmáticamente claro y distinto. Pero a partir de ello da un paso ulterior, a saber, que la mente es una sustancia distinta. De allí que en relación con el primero las ideas de “transparencia” y “acceso privilegiado” se conviertan en las ideas emblemáticas de dicho modelo. Posteriormente, cuando se ataca la concepción cartesiana se lo hace atacando ambas nociones. Existieron además quienes prefirieron desechar de plano a la mente misma, como por ejemplo los conductistas. Davidson (1985), sin embargo, y antes Rorty (1979), apuntan su crítica a toda postura filosófica que postule en la mente cualquier tipo de intermedio epistémico (sense-data, representaciones, contenidos) declarando a tal postura “un mito” acerca de nuestra subjetividad. Y mostrando también cómo la idea de certeza indubitable dependía de aquellos. Armstrong y Ryle, en cambio, critican el cartesianismo centrándose en las ideas mencionadas (transparencia, acceso privilegiado) y en sus nociones asociadas de “incorregibilidad” e “infalibilidad”. La crítica puede entenderse como el resultado de una fuerte reacción anticartesiana, pero ciertamente impide reconocer la simple verdad del *cogito*, el carácter directo y la certeza del mismo, arrastrando con ello la imposibilidad de dar cuenta de la autoridad especial que sentimos sobre nuestros pensamientos propios. Si por ejemplo en este momento creo que tengo un árbol frente a mí, puede ser que me equivoque al confundir otra cosa con un árbol, puede incluso que no haya nada delante mío, pero no hay duda de que no me equivoqué al pensar que creo, y sé que lo creo, que hay un árbol frente a mí. Más aun, mi conciencia de saber lo que creo no habría sido tal si no hubiese llevado a cabo ese pensamiento. (Cabe agregar —contra Descartes— que para tener dicho pensamiento debemos contar con los con-

ceptos pertinentes y ello requiere, si se es externalista, de la relación causal con el mundo externo.) Otra cosa distinta es pretender derivar consecuencias metafísicas de ello y concluir en la doctrina de las dos sustancias. Pero si no se pretende tal derivación ni tampoco se quiere colocar a la mente como la base epistémica de toda verdad y de toda certeza posible (si se rechaza la idea de “objetos” con los cuales mantenemos una relación única), entonces se puede considerar que tales pensamientos o juicios (*cogito-like judgements*) son trivialmente verdaderos ya que su verdad está constituida (análogamente a las cualidades secundarias) por el pensamiento o el juicio mismo (por mi “mejor opinión”, como dice Wright).

Burge (1988), por su parte, habla de juicios que se autoverifican (*self-verifying judgements*) considerando que una condición necesaria de los mismos consiste en que realizarlos o simplemente pensarlos los vuelve verdaderos. ¿Pero cuál es el alcance de dicha tesis? Burge mismo los reduce a una clase mínima a la que denomina “basic self-knowledge”. Wright la amplía un poco más acotándolos por medio de las tres salvedades que integran, como vimos antes, la condición C: prohibición de autoengaño, atención y necesidad de contar con el concepto pertinente. Mi propuesta es extender todavía más el dominio conservando las dos restricciones últimas (contar con el concepto y estar atento) pero excluyendo la prohibición de autoengaño. La razón de ello es que el autoengaño requiere una creencia (un motivo inconsciente) que justamente no llega a ser consciente y por lo mismo permanece oculta, ignorada. También hay otras maneras en que ignoramos ciertos sucesos de nuestra mente. Una de ellas se advierte en el fenómeno que se denomina “vista ciega” (*blindsight*): los sujetos reportan no saber conscientemente nada (no tener autoconocimiento) de la figura que se exhibe ante ellos no obstante lo cual sistemáticamente aciertan acerca de qué figura se trata. Por supuesto que el autoengaño requiere para su comprensión de una teoría psicológica pero también resulta un problema filosófico si se lo entiende como una forma de irracionalidad (Davidson, 1985b). Desde mi perspectiva, sin embargo, las formas de autoengaño no debieran entrar en conflicto con la tesis de la constitutividad.⁵ Dijimos antes que la idea es hacer coextensivo el autoconocimiento con las creencias y deseos *conscientes* (actuales o disposicionales). Esto tal vez parezca poner en un pie de igualdad a la tesis de la constitutividad con la que se denomina “tesis de la autointimidad” (*self-intimating thesis*) (Shoemaker, 1990, 1994) y que afirma que pertenece a la naturaleza misma de los estados

⁵ El punto requiere un desarrollo que no puedo llevar a cabo aquí. Además de los dos trabajos de Davidson mencionados antes, cf. *The Multiple Self*, comp. por J. Elster (Cambridge University Press), 1986, y *Perspectives on Self-Deception*, en McLaughlin and Rorty (comps.), Berkeley, University of California Press, 1990.

intencionales y no intencionales que tenerlos conduce a la creencia y al conocimiento de que se los tiene. Sin embargo, considero que son tesis distintas; o al menos que para hacerlas equivalentes debiéramos reformular esta última diciendo que pertenece a la naturaleza misma de los estados intencionales y no intencionales “conscientes” que tenerlos conduce a la creencia y al conocimiento de que se los tiene. De lo contrario tendría sentido decir que la naturaleza del pensamiento es inconsciente y no que dicha noción es relativa a, y que depende de, la noción de conciencia.

Por otra parte se ha criticado la idea de que los juicios en primera persona no representen un logro cognitivo, pues se dice que de ser así se los estaría condenando a un rol epistémico insustancial. Boghossian (1989) observa que no se esperaría de algo cognitivamente insustancial que pueda estar sujeto al tipo de dirección que muestra el fenómeno de la atención. Pareciera que uno puede decidir cuánta atención dirige sobre sus propios pensamientos del mismo modo que decide cuánta atención les presta a los objetos del campo visual. También observa Boghossian que si se admite que el autoconocimiento es falible —que podemos reconstruir mal nuestros pensamientos— se debe admitir la diferencia que existe entre estar en una posición epistémicamente favorable y no estarlo.

Creo que respecto de cierto tipo de *autoadscripciones* (“Tengo dolor”, “Ahora estoy aquí”), somos obviamente infalibles e incorregibles; y también creo que respecto de nuestros estados intencionales vívidos y disposicionales somos en general inmunes al error.⁶ No creo sin embargo lo mismo de ciertos informes referidos al pasado aunque la ausencia de fidelidad en este caso depende del trabajo de la memoria, y ciertamente, la memoria es falible. Es interesante notar que ciertos trabajos actuales sobre la conciencia (me refiero a los trabajos de F. Crick [1994] y C. Koch) dan cuenta de la misma a través de los siguientes dos rasgos: atención y memoria de corto plazo. Recordemos que la tesis de la constitutividad que defiende menciona estados intencionales y no intencionales conscientes. De modo que creo que las objeciones de Boghossian podrían en este sentido responderse. Ahora bien, si se dice que lo de Crick y Koch se encuentra en el ámbito de la explicación científica y no en el ámbito de una justificación filosófica del autoconocimiento, entonces creo que es aquí donde el análisis de Wright resulta particularmente interesante: la verdad de los juicios en primera persona queda lógicamente autoverificada tan pronto como se piensan o formulan, y esto justifica la autoridad especial

⁶ “Infalible” significa que no podemos equivocarnos respecto de la creencia. “Incorregible”, que ninguna evidencia independiente puede corregir a la nuestra. Ambas nociones aluden a nuestra propia autoridad epistémica dejando a un lado la posibilidad de que de hecho algunas de tales creencias (“tengo un árbol frente a mí”) sean falsas.

que sentimos sobre ellos. Por otra parte, la tesis alude no sólo a estados vívidos, actuales, sino también a estados intencionales disposicionales. Se podría preguntar, por ejemplo, de una creencia disposicional cualquiera: ¿Sabe x su nombre, su estado civil y la ciudad en la que vive aunque no esté pensando en ello? La respuesta es que lo sabe y la prueba es que si se le formula la pregunta responde sin dudar.

Volvamos a la objeción de Boghossian de que la tesis de la constitutividad conduce a una epistemología insustancial. Akeel Bilgrami (1990) ha ofrecido la siguiente respuesta. El autoconocimiento no puede fundamentarse en consideraciones epistemológicas porque no es un conocimiento observacional ni tampoco inferencial. Pero afirmar que el autoconocimiento es constitutivo de la intencionalidad no es explicarlo ni justificarlo. Por otra parte Bilgrami considera que la posición de Wright es insuficiente. De modo que su propuesta consiste en excluir al autoconocimiento del ámbito epistemológico y hacerlo depender del ámbito de la responsabilidad moral, empleando una estrategia similar a la que Strawson utilizara al justificar las atribuciones de libertad dentro del contexto normativo o evaluativo. La idea es que el autoconocimiento es constitutivo de la intencionalidad porque existe una conexión *a priori* entre la responsabilidad moral que les imputamos a los agentes en nuestras actitudes reactivas (reproche, ofensa, gratitud) y el conocimiento que ellos mismos tienen de los estados intencionales que dieron lugar a su conducta.⁷ Por su parte Davidson ha dicho muy poco al respecto, salvo haber esgrimido como justificación una razón trascendental. La idea es que no podríamos hacer lo que de hecho hacemos —interpretarnos unos a otros— si no diéramos por sentado que las personas conocen su propia vida mental. Así Davidson presenta al autoconocimiento como condición trascendental de la interpretación.

Mencionaré por último un problema que aún no hemos considerado. Suele decirse que el externalismo plantea *prima facie* un problema al autococonocimiento, a la autoridad de la primera persona. Desde el experimento mental de la Tierra Gemela (Putnam, 1975) y los trabajos de Kripke (1972, 1979) se extrae la consecuencia de que los protagonistas de las respectivas historias de ficción “significan” o “quieren decir” cosas que ellos mismos desconocen. Así, por ejemplo, el personaje ficticio de Kripke, Pierre, dice en realidad “Londres” cuando dice “London” (porque la referencia es la misma) y los Oscar de Putnam, aún en el año 1750, dicen cosas distintas (porque la referencia de “agua” es distinta). Por ello, para dar cuenta de estas anomalías, se apeló a la noción de contenido amplio (*wide content*), y con el fin de restaurar la autori-

⁷ Una versión más elaborada se encuentra en Bilgrami, *Self-Knowledge and Resentment y Self-Knowledge and Intentionality*, ambos en prensa.

dad de la primera persona se apeló a la noción de contenido estrecho (*narrow content*). De una forma u otra la amenaza al autoconocimiento, siendo externalista, se tornaba inevitable. Si el pensamiento depende de factores externos y si uno desconoce los mismos, no conoce lo que cree conocer cuando piensa y habla, al menos en el caso de cierto tipo de palabras. La discusión, en suma, se plantea en términos de la compatibilidad o incompatibilidad de la tesis externalista con la autoridad de la primera persona (Boghossian, 1989).

Considero que la polémica ha surgido con la teoría causal de la referencia y que depende de ella (Bilgrami, 1991), razón por la cual un externalismo formulado en esos términos socava efectivamente el autoconocimiento. Creo también que la postura externalista de Burge basada en consideraciones normativas (cuando digo “tengo artritis”, digo lo correcto porque digo lo que en mi comunidad tal palabra significa) produce el mismo resultado que el que producen las teorías causales, de modo que la bifurcación del contenido (*narrow/wide*) como recurso ante la amenaza al autoconocimiento se vuelve ineludible. Pero tal vez yo no comprenda acabadamente a Burge —quien niega ambas cosas— cuando sostiene que “podemos conocer lo que nuestros eventos mentales son y sin embargo no conocer los hechos generales relevantes acerca de las condiciones de individuación de tales eventos” (Burge, 1988). Creo sin embargo que el problema de la compatibilidad o incompatibilidad en cuestión depende de cuál sea la tesis externalista que se quiera defender, aunque sé que este punto requeriría un desarrollo que no puedo llevar a cabo aquí. Simplemente digo que cierto tipo de externalismo amenaza el autoconocimiento y lleva a la bifurcación del contenido, y que otro tipo de externalismo, como el de Davidson por ejemplo, no lo hace.

De todos modos considero que un examen más acabado del autoconocimiento debería mostrar que no podríamos contar con los conceptos con que contamos si el mundo externo no fuese lo que es. Recordemos que contar con el concepto adecuado era la tercera de las restricciones impuestas por Wright. Por eso creo que una manera de rechazar el escenario cartesiano en el aspecto en que debe ser rechazado consistiría en mostrar que la idea de autoconocimiento depende del dominio del lenguaje, principalmente del dominio del concepto de primera persona, y que éste a su vez depende de la experiencia perceptual. Resta entonces argumentar en favor de esta línea y en favor de un externalismo diferente del que defienden Kripke, Putnam y Burge.

BIBLIOGRAFIA

Armstrong, D. [1968], *A Materialist Theory of Mind*, Routledge & Kegan Paul, Londres.

- Bennett, J. [1971], *Locke, Berkeley, Hume: temas centrales*, México, UNAM.
- Bilgrami, A. [1990], "Self-Knowledge and Intentionality", conferencia leída en SADAF, agosto de 1990.
- [1992], *Belief and Meaning*, Cambridge, Mass., Blackwell.
- , "Self-Knowledge and Resentment" (en prensa).
- , *Self-Knowledge and Intentionality* (en prensa).
- Blackburn, S. [1993], *Essays in Quasi-Realism*, Oxford University Press.
- Boghossian, P. [1989], "Content and Self-Knowledge", *Philosophical Topics*, vol. XVII.
- Burge, T. [1988], "Individualism and Self-Knowledge", en *Self-Knowledge*, comp. por Quassim Cassam, Oxford University Press, 1994.
- Crick, F. [1994], *La búsqueda científica del alma*, Madrid, Debate.
- Chisholm, R. [1969], "On the Observability of the Self", en *Self-Knowledge*, comp. por Quassim Cassam, Oxford University Press, 1994.
- Davidson, D. [1982], "Paradoxes of Irrationality", *Philosophical Essays on Freud*, R. Wollheim y J. Hopkins (comps.), Nueva York y Cambridge, Cambridge University Press.
- [1985a], "El mito de lo subjetivo".
- [1985b], "Engaño y división".
- [1987], "El conocimiento de la propia mente". En *Mente, mundo y acción*, Barcelona, Paidós, 1992.
- [1991], "What is Present to the Mind", *Philosophical Issues, I. Consciousness*.
- Descartes, R. [1640], *Meditations on First Philosophy*, Hackett Publishing Company, 1993.
- Hume, D. [1739], *Treatise of Human Nature*, ed. de L. A. Selby-Bigge, Oxford University Press, 1958.
- Kripke, S. [1972], *El nombrar y la necesidad*, México, UNAM, 1985.
- [1979], "A Puzzle about Belief", en *Meaning and Use*, comp. por A. Margalit, Dordrecht; Reidel.
- Locke, J. [1690], *An Essay Concerning Human Understanding*, The Open University, 1983.
- Mackie, J. L. [1976], *Problemas en torno a Locke*, México, UNAM, 1988.
- McGinn, C. [1983], *The Subjective View: Secondary Qualities and Indexical Thoughts*, Oxford, Clarendon Press.
- Putnam, H., "The Meaning of 'Meaning'", *Mind, Language and Reality, Philosophical Papers*, vol. 2, Cambridge University Press.
- Rorty, R. [1979], *Philosophy and the Mirror of Nature*, Princeton University Press.
- Ryle, G. [1949], *El concepto de lo mental*, Buenos Aires, Paidós, 1967.
- Shoemaker, S. [1986], "Introspection and the Self", en *Self-Knowledge*,

- comp. por Quassim Cassam, Oxford University Press, 1994.
- Shoemaker, S. [1990], "First-person Access", *Philosophical Perspectives*, 4. *Action Theory and Philosophy of Mind*.
- [1994], "Self-Knowledge and Inner Sense", *Philosophy and Phenomenological Research*, vol. LIV, N° 2.
- Taylor, C. [1977], *Human Agency and Language, Philosophical Papers*, Cambridge University Press.
- Wright, C. [1989], "Wittgenstein Rule-following Considerations and the Central Project of Theoretical Linguistics", en *Reflections on Chomsky*, comp. por A. George, Oxford, Basil Blackwell.
- Wittgenstein, L. [1953], *Philosophical Investigations*, Oxford, Basil Blackwell.

Julia Vergara
UNIVERSIDAD DE BUENOS AIRES
jvergara@einstein.com.ar

ABSTRACT

Self-knowledge as a sort of knowledge can be justified through two kinds of paradigm: observational and inferential. This paper criticizes both epistemic lines and suggests instead that first-person statements exhibit the same logical structure as statements about secondary qualities, as proposed by Crispin Wright. The paper introduces a variant to this proposal: self-deception should not be seen as threatening the biconditional that expresses that self-knowledge is constitutive of intentionality. This is the constitutive thesis.