

el saber filosófico

# T Ó P I C O S

Jorge Martínez Contreras • Aura Ponce de León

C O O R D I N A D O R E S



Asociación Filosófica de México



# TÓPICOS DEL SABER FILOSÓFICO

VOLUMEN III

*coordinado por*  
**JORGE MARTÍNEZ CONTRERAS**  
**AURA PONCE DE LEÓN**

*colaboradores*

FRANCISCO DE JESÚS ÁNGELES CERÓN \* PHYLLIS M. CORREA  
JUAN GRANADOS VALDÉZ \* JULIÁN HERNÁNDEZ CASTELANO  
JUAN CARLOS MORENO ROMO \* HORACIO CERUTTI GULDBERG  
LUIS MIGUEL GALLARDO SALAZAR \* CARLOS LENKERSDORF SCHMIDT  
BERTOLD BERNREUTER \* HEINZ KRUMPEL \* MAURICIO BEUCHOT  
FERNANDO CALOCA AYALA \* NAPOLEÓN CONDE GAXIOLA  
CARLOS EMILIO GENDE \* JOSÉ MARÍA GONZÁLEZ GARCÍA  
FRANCISCO HILARIO HERNÁNDEZ RESÉNDIZ \* JULIETA LIZAOLA  
MARTHA ILIA NÁJERA CORONADO \* PALOMA OLIVARES  
MARÍA ROSA PALAZÓN MAYORAL \* SONIA TORRES ORNELAS  
MARÍA DEL CARMEN VALVERDE VALDÉS \* JOSÉ FRANCISCO ZÚÑIGA GARCÍA  
JOSÉ ALFREDO ARAUJO MARTÍNEZ \* JAVIER CORONA FERNÁNDEZ  
RODOLFO CORTÉS DEL MORAL \* JOSÉ MENDÍVIL MACÍAS VALADEZ  
AURELIANO ORTEGA ESQUIVEL \* MARTÍN FRANCISCO FRICKE  
ROSARIO HERRERA GUIDO \* RAÚL NAVARRETE JACOBO  
MARÍA CRISTINA RÍOS \* SABINE KNABENSCHUH DE PORTA  
MARÍA DEL CARMEN GARCÍA AGUILAR \* SANDRA ANCHONDO PAVÓN  
ALINA MENDOZA CANTÚ \* JORGE RENDÓN ALARCÓN  
WERNER GABRIEL \* MANUEL VELÁZQUEZ MEJÍA  
MARÍA TERESA RODRÍGUEZ GONZÁLEZ



**XXI**  
siglo  
veintiuno  
editores



**siglo xxi editores, s.a. de c.v.**

CERRO DEL AGUA 248, ROMERO DE TERREROS, 04310, MÉXICO, D.F.

**siglo xxi editores, s.a.**

TUCUMÁN 1621, 7° N, C1050AAG, BUENOS AIRES, ARGENTINA

**siglo xxi de españa editores, s.a.**

MENÉNDEZ PIDAL 3 BIS, 28036, MADRID, ESPAÑA

B52

S33

2007 *El saber filosófico* / coordinado por Jorge Martínez

Contreras y Aura Ponce de León ; colaboradores

Francisco de Jesús Ángeles Cerón ... [et al.]. — México : Siglo XXI :

Asociación Filosófica de México, A.C., 2007

3 v. — (Filosofía)

Contenido: v. 1. Antiguo y moderno. — v. 2. Sociedad y ciencia. — v. 3. Tópicos del saber filosófico

ISBN-13: 978-968-23-2693-6 (obra completa)

ISBN-13: 978-968-23-2694-3 (v. 1)

ISBN-13: 978-968-23-2695-0 (v. 2)

ISBN-13: 978-968-23-2696-7 (v. 3)

1. Filosofía. I. Martínez Contreras, Jorge, ed.

II. Ponce de León, Aura, ed. III. Ángeles Cerón, Francisco de Jesús, colab. IV. Ser.

primera edición, 2007

© siglo xxi editores, s.a. de c.v.

en coedición con la asociación filosófica de méxico, a.c.

isbn-13: 978-968-23-2693-6 (obra completa)

isbn-13: 978-968-23-2696-7 (volumen 3)

derechos reservados conforme a la ley  
impreso y hecho en méxico

WRIGHT Y LA AUTORIDAD DE LA PRIMERA PERSONA:  
PROBLEMAS DE TEORÍAS CONSTITUTIVAS DE LA AUTORIDAD

MARTÍN FRANCISCO FRICKE

Instituto de Investigaciones Filosóficas  
(Unidad Académica de Ciencias Sociales y Humanidades  
UACSHUM - UNAM/ Mérida, Yucatán)

Es fácil equivocarse en sus creencias acerca de qué piensan o sienten *otras* personas. Pero ¿es posible no tener un dolor si uno piensa que uno lo tiene? ¿Es posible creer, "Creo que va a llover", sin creer que va a llover? Parece que cuando se trata de los *propios* pensamientos y sensaciones, tenemos más autoridad en nuestro conocimiento que cuando se trata de los pensamientos y sensaciones de otras personas. ¿Cómo se explica este fenómeno, lo que los filósofos llaman la "autoridad de la primera persona"?

Podemos distinguir entre el conocimiento que tenemos de las propias sensaciones ("Tengo un dolor", "Tengo sed", "Tengo la sensación de ver algo rojo") y el conocimiento que tenemos de las propias actitudes proposicionales, también llamadas estados intencionales ("Creo que va a llover", "Quiero que se vaya", "Temo que ya se acabó tu tiempo"). En este trabajo me interesa principalmente el segundo tipo de conocimiento y en el área de éste me enfoco en el conocimiento de las propias creencias. ¿Por qué tenemos una autoridad especial cuando sinceramente afirmamos qué creemos? ¿Por qué parece improbable, quizás incluso imposible, equivocarse en afirmaciones sinceras como "Creo que todavía hay leche en el refrigerador" o "Creo que X ganará las elecciones presidenciales"? Es posible que nos equivoquemos sobre la leche o acerca de las elecciones, pero ¿también es posible que nos equivoquemos en decir que *creemos* que todavía hay leche o que X ganará las elecciones? Mi pregunta en este trabajo es por qué eso no parece probable e incluso no parece posible.

En lo que sigue, examinaré la teoría constitutiva de la autoridad de la primera persona que ha ofrecido Crispin Wright. Según esta teoría, la autoridad se debe al hecho de que nuestras autoadscripciones de estados intencionales contribuyen a la constitución de estos mismos estados. Trataré de demostrar que este tipo de explicación tiene problemas graves. Como alternativa sugeriré una explicación epistemológica basada en una observación de Gareth Evans.

## 1. LA LÓGICA DE RELACIONES CONSTITUTIVAS

Antes de entrar en la discusión, quiero hacer algunas observaciones generales sobre relaciones constitutivas:

Primero, la siguiente afirmación parece plausible: Cuando afirmamos que X constituye Y podemos concluir que la existencia de X es *suficiente* para la existencia

de *Y*. Es decir, siempre que se da *X*, también se da *Y*, ya que *X* constituye *Y*. Sin embargo, cuando afirmamos que *X* constituye *Y* no podemos concluir que la existencia de *X* es *necesaria* para la existencia de *Y*. Es posible que haya diferentes maneras de constituir una cosa. Por ejemplo, ciertas piedras y otros materiales ordenados de cierta manera constituyen mi casa. Pero la existencia de estas piedras, etc. no es necesaria para la existencia de mi casa. De hecho, es probable que en futuras reparaciones de mi casa se remplazarán partes de mi casa con otras piedras, etcétera. Entonces, en el futuro otras piedras, etcétera, constituirán mi casa.

Segundo, aunque lo que dije en mi primera observación parece plausible, a menudo, cuando se dice que *X* es *constitutivo* de *Y*, eso sólo indica que *X* es *parte* de lo que constituye *Y*. En este caso es posible que la sola existencia de *X* no sea suficiente para la existencia de *Y*. Por ejemplo, ciertas piedras son constitutivas de mi casa, pero estas piedras son sólo parte de lo que constituye mi casa junto con cemento, vidrios, etc. Si eso es correcto, entonces el hecho de que *X* sea constitutivo de *Y*, no implica que *X*, sólo, es lo que constituye *Y*, sino solamente que *X* es parte de lo que constituye *Y*.

Tercero, parece que la relación de constitución implica más que estas simples relaciones lógicas. Si *X* es suficiente para *Y*, todavía no se sigue que *X* constituye *Y*. Si la bandera de Buckingham Palace está arriba eso es un indicador suficiente para la presencia de la reina en su palacio. Pero obviamente el hecho de que la bandera esté arriba no *constituye* la presencia de la reina en su palacio. ¿Qué más necesitamos además de que *X* sea suficiente, para que se trate de una relación constitutiva? Tal vez cuando decimos que *X* constituye *Y* indicamos que *Y*, además de ser necesitado por *X*, también *consiste* en *X*.

## 2. RELACIONES CONSTITUTIVAS Y LA AUTORIDAD DE LA PRIMERA PERSONA

En la discusión sobre la autoridad de la primera persona se han propuesto ambas afirmaciones: que nuestras creencias de segundo orden son constitutivas de nuestras creencias de primer orden y que nuestras creencias de primer orden son constitutivas de las de segundo orden.<sup>1</sup> La teoría de Crispin Wright defiende la primera de estas afirmaciones y la generaliza a todas las autoadscripciones de estados intencionales. ¿Cómo se usa tal aseveración para explicar la autoridad de la primera persona? Si nuestras creencias de segundo orden ["Creo que *p*"] constituyen nuestras creen-

<sup>1</sup> Akeel Bilgrami incluso defiende una posición según la cual ambas ideas son verdaderas. Cf. A. Bilgrami: "Self-Knowledge and Resentment", C. Wright *et al.* (ed.): *Knowing Our Own Minds*, Oxford 1998, pp. 207-241. Sin embargo, yo acabo de argüir que cuando decimos que *X* constituye *Y* indicamos que *Y*, además de ser necesitado por *X*, también consiste en *X*. Pero ¿es posible que *X* consista en *Y* y *Y* en *X*? Me parece que generalmente tal afirmación no tiene mucho sentido. (Compárese: El bosque consiste en, y es constituido por cierto número de árboles en cierta distribución en cierta área. Pero cierto número de árboles, etc. no consisten en, ni son constituidos por un bosque.) El único caso en que, tal vez, tiene sentido afirmar que *X* constituye *Y* y al revés, es el caso de una identidad entre *X* y *Y*. (Compárese: Esta silla constituye esta silla y esta silla consiste en esta silla.)

cias de primer orden ["*p*"], entonces no es sorprendente que estas creencias siempre son correctas: Si tenemos una creencia "Creo que *p*", esta creencia es correcta porque es lo que constituye que también tenemos la creencia "*p*". Es decir, si tenemos la creencia "Creo que *p*", necesariamente también tenemos la creencia "*p*", porque la existencia de la primera es una condición suficiente para la existencia de la segunda. Eso significa que la creencia "Creo que *p*" necesariamente es correcta. La creencia "se hace verdadera" a sí misma.

La situación no es tan clara cuando afirmamos que nuestras creencias de segundo orden son constitutivas de nuestras creencias de primer orden, pero que sólo son *parte* de lo que constituye las creencias de primer orden. En este caso la ocurrencia de una creencia de segundo orden en sí no "se hace verdadera" a sí misma. Más bien, para que sea verdadera una creencia de segundo orden es necesario que otras condiciones también sean cumplidas, porque la creencia de segundo orden sólo puede constituir la creencia de primer orden junto con otras condiciones. Entonces, ¿cómo podría la afirmación de que nuestras creencias de segundo orden son parte de lo que constituye nuestras creencias de primer orden ayudarnos a explicar la autoridad de la primera persona? Nos podría ayudar para tal explicación si fuera el caso que las condiciones adicionales que deben ser cumplidas para que exista la creencia de primer orden *generalmente* se cumplen. Supóngase que la creencia "Creo que *p*" (de una persona) junto con ciertas condiciones C constituye la creencia "*p*" (en la misma persona). Ahora, si sabemos que generalmente se da C podemos concluir que generalmente la creencia "Creo que *p*" (de cualquier persona) es correcta porque esta creencia junto con C constituye la creencia "*p*" (en la misma persona) y eso significa que la creencia "Creo que *p*" es verdadera. Ésta es la estructura de la teoría de Wright.

### 3. WRIGHT SOBRE LA AUTORIDAD DE LA PRIMERA PERSONA

Después de estas observaciones generales, déjenme presentar los detalles de la teoría de Crispin Wright. Según él,

[...] la autoridad que adscribimos generalmente a las propias creencias [...] de un sujeto sobre sus estados intencionales es un *principio constitutivo*: algo que [...] entra primitivamente en las condiciones de identificación de lo que un sujeto cree, espera e intenta.

(the authority standardly granted to a subject's own beliefs [...] about his intentional states is a constitutive principle: something that [...] enters primitively into the conditions of identification of what a subject believes, hopes, and intends) (Wright, 1989a:632).

Las razones que Wright tiene para esta aseveración provienen de las *Investigaciones filosóficas* de Ludwig Wittgenstein. Según Wright, Wittgenstein nos enseña que la autoadcripción de estados intencionales no está basada en una observa-

ción de nuestros estados intencionales, ni en inferencia. Más bien, según Wittgenstein en la interpretación de Wright, la relación entre autoadscripciones de estados intencionales y estos estados intencionales mismos es *a priori*. ¿Qué significa eso? Examinemos el ejemplo de autoadscripciones de creencia. Supongamos que un sujeto dice, "Creo que *p*". Según Wright, esta autoadscripción es un criterio para la identificación de la creencia de que *p* en el sujeto. Es decir, normalmente cuando un sujeto dice "Creo que *p*" eso es un indicador suficiente para concluir que el sujeto cree que *p*. Eso es así no porque la creencia de que *p* tiende a *causar* la autoadscripción "Creo que *p*", sino porque estar dispuesto a decir sinceramente, "Creo que *p*", es parte de lo *significa* creer que *p*. (En cambio, cuando afirmo, sobre la base de una observación, "El libro está en la mesa", esta afirmación no es parte de lo que significa que el libro está en la mesa.) La relación entre la autoadscripción "Creo que *p*" y la creencia de que *p* es *a priori* porque hacer la autoadscripción es parte de lo que significa tener la creencia de que *p*. Como dice Wright también, la sincera afirmación "Creo que *p*" no *refleja* la extensión del concepto "mi creencia de que *p*", sino la *determina* (cf. Wright, 1989b). No está basada en un logro cognitivo, en la detección de algo, sino es lo que hace que algo sea detectable.

Sin embargo, Wright no postula que nuestras autoadscripciones son infalibles. Puede haber situaciones en las cuales es más coherente con el resto de lo que un hablante hace y dice suponer que su autoadscripción es falsa. El criterio de la autoadscripción es un criterio "refutable" ("defeasible") para la presencia de un estado intencional. Pero las situaciones en las cuales la autoadscripción resulta falsa son excepciones. "Verdad es la posición de default" ("*[t]ruth is the default position*", Wright, 1989a:633). Aquí vemos que la teoría de Wright sólo dice que autoadscripciones de creencias son *parte* de lo que constituye tener las creencias adscritas. En condiciones normales (la posición "default"), la autoadscripción de una creencia siempre es correcta, porque las condiciones normales junto con la autoadscripción constituyen que tenemos la creencia adscrita. Pero hay condiciones diferentes en las cuales la autoadscripción junto con estas condiciones no son suficientes para la presencia de la creencia adscrita.

El "juego de lenguaje" que incorpora estas relaciones constitutivas entre autoadscripciones de estados intencionales y los estados intencionales mismos tiene la siguiente forma: Si un sujeto se adscribe sinceramente un estado intencional esta adscripción se acepta por default como verdadera. Es decir, no se necesitan razones para aceptar la verdad de una autoadscripción. Sólo para rechazarla y declararla como falsa se necesitan razones, porque eso sería una posición diferente del "default". Entonces, generalmente, es decir, en condiciones normales, se toman las sinceras autoadscripciones como correctas. En este juego de lenguaje, la autoridad de la primera persona es algo *concedido* por los otros participantes del juego.

Wright nota —y eso me parece importante— que la existencia de tal juego de lenguaje depende de ciertas "contingencias profundas":

El éxito de un juego de lenguaje que funcionaría de esta manera dependería de ciertas contingencias profundas. Dependería, por ejemplo, de la contingencia de que tomar las auto-concepciones de otras personas en serio [...] casi siempre tenderá a resultar en una concepción total de su psicología que elucida más —de hecho, enormemente más— que cualesquiera [datos] que podrían ser obtenidos por respetar todos los datos excepto el testimonio del sujeto sobre sí mismo. Y eso, a su vez, depende de la contingencia de que somos, cada uno de nosotros, permanentemente pero —según la concepción propuesta— subcognitivamente movidos a opiniones acerca de nuestros propios estados intencionales, [opiniones] que en realidad harán un buen servicio para otros en su intento de entendernos.

*([T]he success of a language game that worked this way would depend on certain deep contingencies. It would depend, for instance, on the contingency that taking the self-conceptions of others seriously [...] will almost always tend to result in an overall picture of their psychology which is more illuminating—as it happens, enormously more illuminating— than anything which might be gleaned by respecting all the data except the subject's self-testimony. And that in turn rests on the contingency that we are, each of us, ceaselessly but —on the proposed conception— subcognitively moved to opinions concerning our own intentional states which will indeed give good service to others in their attempt to understand us).*  
(Wright, 1989a:632.)

En esta cita vemos dos “contingencias profundas”: 1) Es más fácil de entender a otras personas si tomamos sus autoadscripciones de estados intencionales como verdaderas que si no tomamos estas autoadscripciones en cuenta y 2) permanentemente, pero sin razones (*i. e.*, subcognitivamente), “nos vienen” opiniones sobre nuestros propios estados intencionales que generalmente son correctas. Si estas dos condiciones no se cumplen, el juego de lenguaje que nos otorga la autoridad de la primera persona no funciona.

#### 4. OBJECIONES

Barry Smith nota —correctamente, en mi opinión— que la teoría de Wright no permite que las opiniones que un sujeto tiene acerca de sus propios estados intencionales sean confirmadas por otras observaciones. Si estas opiniones (*i. e.*, las sinceras autoadscripciones de estados intencionales) no concuerdan con el resto de su comportamiento (lingüístico y no lingüístico), pueden ser rechazadas como falsas por un intérprete. Pero si el resto de su comportamiento está de acuerdo con sus sinceras autoadscripciones eso no *confirma* las autoadscripciones, ya que ellas son verdaderas por default. Pero, según Smith, eso no concuerda con nuestras experiencias ordinarias. Parece que, en la vida diaria, sí puede pasar, por ejemplo, que las autoadscripciones de una persona nos parecen correctas en vista de lo que ya pensábamos sobre la persona desde antes. Pero eso no es compatible con la teoría de Wright:

No hay lugar, según esta teoría [de Wright] para la idea de que lo que los sujetos dicen sobre sí mismos puede confirmar, o ser confirmado, por cómo los habíamos entendido hasta ahora. Similarmente, a menudo pensamos que lo que las personas dicen sobre sí mismas suena verdadero, dado lo que sabemos acerca de ellas. Pero las autoadscripciones de las personas no pueden recibir una confirmación independiente por lo que otros están preparados para adscribirles, porque los hechos acerca de los estados intencionales en los cuales un sujeto está no son decididos independientemente de las opiniones [que estas personas tienen] acerca de eso. [...] Lo que estamos dispuestos a pensar acerca de alguien no puede confirmar lo que él dice sobre sí mismo, porque [lo que dice sobre sí mismo] es parte de la evidencia [para lo que deberíamos pensar acerca de él].

*[T]here is no scope on this account [of Wright's] for the idea that what subjects say about themselves can confirm, or be confirmed by, how we had understood them up until now. Similarly, we often take what people say about themselves to ring true, given what we know about them. But people's self-ascriptions cannot be given independent confirmation by what others are prepared to ascribe to them, because the facts about which mental states subjects are in is not something settled independently of their opinions about this. [...] What we are inclined to think about someone cannot confirm, because it includes as evidence, what they say about themselves) (Smith,1998:413).*

Otra crítica de Smith es que Wright no explica por qué en algunos casos nuestros estados intencionales son constituidos por las opiniones que tenemos acerca de ellos, mientras en otros casos no. Esta pregunta surge porque parece posible tener un estado intencional sin *saber* que uno lo tiene. Obviamente, en este caso, el estado intencional no puede ser constituido por nuestra opinión acerca de él, la suposición es que no tenemos tal opinión. Wright debería responder que estados intencionales son constituidos de diferentes maneras. Como vimos antes, eso es compatible con la lógica de la relación constitutiva. Pero tal vez Smith tiene razón en exigir una explicación de por qué, según Wright, nuestros estados intencionales (aunque sí tienen el mismo contenido) se constituyen de diferentes maneras en diferentes ocasiones:<sup>2</sup>

Es posible tener estados intencionales sin tener conocimiento de ellos. Como Wright reconoce, nuestros juicios acerca de nuestros propios estados intencionales sólo son suficientes (por default) para que estemos en estos estados intencionales, no necesarios. Así, en casos donde no tenemos autoconocimiento, la presencia en nosotros de los estados intencionales no se debe a nuestros juicios acerca de ellos. ¿Por qué, entonces, debería ser que en casos donde *sí* conozco mi propia mente mis juicios ayudan a determinar, o tal vez constituyen, en cuáles estados estoy? ¿Por qué es la presencia de estados intencionales dependiente de mi juicio en un caso, pero no en el otro?

<sup>2</sup> Nótese que esta explicación también es necesaria para evitar un regreso al infinito. Si todos los estados intencionales fueron constituidos por la opinión de que los tenemos, entonces esta opinión misma —que es un estado intencional— también debería ser constituida por tal opinión, etc. Para evitar este regreso, Wright tiene que justificar que algunos estados intencionales (por lo menos las opiniones constitutivas de otros estados intencionales) no son constituidos por opiniones acerca de ellos.

([I]t is possible to have intentional states without having knowledge of them. As Wright acknowledges, our judgements regarding our own intentional states are only sufficient (by default) for us to be in those intentional states, not necessary. So in cases where we lack self-knowledge, the presence in us of intentional states owes nothing to our judgements about them. Why, then, should it be that in cases where I do know my own mind, my judgements help to determine, or perhaps constitute, which states I am in? Why is the presence of intentional states dependent on my judgement in one case, but not in the other?) (Smith, 1998:413f.).

Jane Heal hace la siguiente crítica: Si, en condiciones normales, un estado intencional puede ser constituido sólo por la opinión del sujeto de que él está en este estado intencional, entonces ¿cómo es posible que este estado intencional tenga “eficacia real” (cf. Heal, 2003:286)? Si creo que está lloviendo, salgo con paraguas. ¿Pero tiene sentido suponer que la sola opinión “Creo que está lloviendo” me hace llevar el paraguas? Parece que aquí falta una explicación de cómo estados intencionales constituidos por opiniones del sujeto pueden tener una verdadera eficacia.

En lo que sigue, quiero elaborar dos críticas de la teoría de Wright que no son mencionadas por Smith o Heal.

Mi primera duda es si una estrategia lingüística como la que persigue Wright es capaz de explicar por qué tenemos la autoridad de la primera persona. Wright analiza nuestro “juego de lenguaje” e identifica la autoridad de la primera persona como algo que es parte de nuestro juego. El juego es definido por sus reglas y algunas de sus reglas dicen que, en condiciones normales, hacer una sincera autoadscripción de un estado intencional significa estar en este estado intencional y por eso tal adscripción necesariamente es verdadera. Pero me parece que esta afirmación explica muy poco. Eso podemos ver si examinamos el siguiente caso imaginario. El juego de lenguaje podría contener esta regla: En condiciones normales, *hacer* el sincero enunciado “Está lloviendo<sub>x</sub>” significa que está lloviendo. (Estoy usando el término técnico “llover<sub>x</sub>” porque nuestro término ordinario “llover” no obedece esta regla. Pero aparte de esta regla “llover<sub>x</sub>” significa lo mismo que “llover”). Si nuestro juego de lenguaje contuviera esta regla, podríamos decir que el enunciado “Está lloviendo<sub>x</sub>” es un criterio para la lluvia. El enunciado es constitutivo de la lluvia. En analogía con las reglas de Wright para autoadscripciones de estados intencionales, podríamos decir que el enunciado “Está lloviendo<sub>x</sub>” no *refleja* la extensión del concepto “lloviendo”, sino la *determina*. La consecuencia de esta regla es que en el juego de lenguaje que estamos considerando normalmente el enunciado “Está lloviendo<sub>x</sub>” es verdadero. Verdad es la posición de default para este enunciado.

Nuestro juego de lenguaje no contiene la regla descrita. ¿Por qué no? La razón, en términos generales, es muy simple. No tiene sentido tener el término “llover<sub>x</sub>” en nuestro juego de lenguaje porque nosotros no podemos determinar cuándo sería correcto usarlo. Para poder usar el término “llover<sub>x</sub>” necesitaríamos una capacidad que en condiciones normales debería ser infalible en detectar la lluvia. Sólo así podría ser verdad que, en condiciones normales, el hecho de que alguien enun-

cia, "Está lloviendo<sub>x</sub>", sería parte de lo que significa que está lloviendo. Lo que vemos aquí es que hay una explicación importante de por qué tenemos el juego de lenguaje que tenemos. No podemos tener cualquier juego de lenguaje sino sólo tales juegos que somos capaces de realizar.

Ahora bien, ¿por qué somos capaces de realizar el juego de lenguaje que contiene las reglas que Wright describe para el uso de autoadscripciones de estados intencionales, pero no somos capaces de realizar el juego de lenguaje que contiene reglas análogas acerca de ciertos enunciados sobre la lluvia? La respuesta debe estar en las "contingencias profundas" que menciona Wright. Nuestro juego de lenguaje no funcionaría si no fuéramos "movidos subcognitivamente a opiniones correctas" sobre nuestros propios estados intencionales. No existen "contingencias profundas" que nos hacen capaces de realizar juegos de lenguaje con enunciados autoritativos sobre la lluvia, o sobre los estados intencionales de otras personas. Aquí vemos que las "contingencias profundas" juegan un papel explicativo muy importante en la teoría de Wright. Podemos imaginarnos muchos juegos de lenguaje: juegos que confieren una autoridad especial a nuestras autoadscripciones de estados intencionales, como el que describe Wright; juegos que confieren una autoridad similar a nuestras adscripciones de estados intencionales a otras personas, pero ninguna autoridad a autoadscripciones; juegos que confieren una autoridad especial a enunciados sobre la lluvia, la economía, etc. La descripción de nuestro juego de lenguaje en sí explica muy poco si queremos saber por qué tenemos la autoridad de la primera persona. Para una explicación satisfactoria debemos saber por qué somos capaces de *realizar* el juego que de hecho jugamos. ¿Cómo es posible que se den las "contingencias profundas" que describe Wright? Parece que, en la teoría de Wright, falta una respuesta a esta pregunta para que la teoría sea completa.

Mi segundo punto crítico es que las opiniones sobre los estados intencionales —como Wright las describe— no pueden calificar como *conocimiento* de sí mismo. Wright claramente quiere elucidar "nuestro *conocimiento* de nuestras propias intenciones y de nuestros estados intencionales en general" ("our *knowledge* of our own intentions and of our intentional states in general", Wright, 1989a:630 (las cursivas son mías), cf. también *ibid.*, 631, 633). Según Wright el conocimiento expresado en "Creo que *p*" (o "Temo que *p*", "Quiero que *p*" etc.) es constitutivo de la creencia "*p*" (o del temor que *p*, de la intención de que *p*, etc.). Pero ¿es posible que un conocimiento sea constitutivo de su propio objeto?

He aquí un argumento en contra de esa posibilidad: Me parece que el conocimiento necesariamente es guiado por el mundo (más específicamente, por su objeto), no al revés. Es decir, los posibles objetos de conocimiento deciden qué conocimiento podemos tener. La existencia del conocimiento no decide (no es lo que determina) qué objetos existen. Ahora, supongamos que un conocimiento sea constitutivo de su propio objeto, como la teoría de Wright arguye. Una buena pregunta frente a esta suposición es: ¿Cómo se puede adquirir este conocimiento? Mientras que todavía no tenemos el conocimiento del objeto, este objeto mismo todavía no puede existir. ¿Por qué? Porque se supone que el objeto está constitui-

do (por lo menos en parte) por el conocimiento de él. Así, si el conocimiento todavía no existe, tampoco su objeto puede existir, ya que su existencia depende del conocimiento de él. Pero eso significa que la adquisición del conocimiento no puede ser provocada o guiada en algún sentido por el objeto del conocimiento. Porque durante la adquisición del conocimiento este conocimiento todavía no existe y por lo tanto tampoco existe el objeto que supuestamente se constituye gracias al conocimiento. Pero si la adquisición del conocimiento no está guiada por el objeto del conocimiento, entonces en realidad no se puede tratar de conocimiento. La adquisición de cualquier conocimiento debe ser guiada por el mundo, el objeto del conocimiento. Eso no es posible si el conocimiento es constitutivo de su propio objeto. (El supuesto "conocimiento" se produce cuando uno quiera, no cuando su objeto está presente y justifica la adquisición del conocimiento.) Por eso tal conocimiento que es constitutivo de su propio objeto no existe.

Si eso es correcto, entonces Wright se equivoca en caracterizar las opiniones acerca de los propios estados intencionales que, según él, son constitutivas de los mismos estados como conocimiento. Si estas opiniones son constitutivas de sus propios objetos (los estados intencionales), entonces no se puede tratar de conocimiento, porque no es posible que formemos tales opiniones en respuesta a la presencia de sus propios objetos. Sus objetos todavía no existen cuando formamos las opiniones. Ya que las opiniones son constitutivas de sus propios objetos, es necesario que las opiniones estén presentes para que sus objetos estén presentes. Eso significa que las opiniones deben formarse independientemente, y de hecho antes, de que sus objetos puedan estar presentes. Pero si las opiniones se forman antes de que estén presentes sus objetos, entonces la formación de estas opiniones no puede ser guiada por los objetos. Más bien, la formación de las opiniones entonces debe ser lo que "guía" el mundo (provoca la presencia de sus objetos). Eso significa que no se puede tratar de conocimiento, porque para ser un conocimiento una opinión debe ser guiada por el mundo, no al revés.

La conclusión es que si tenemos un autoconocimiento autoritativo de los propios estados intencionales Wright no logra explicarlo. Además, Wright se equivoca en suponer que su teoría contesta la pregunta de cómo es posible tener conocimiento de los propios estados intencionales (cf. Wright, 1989a:630). Esta objeción no es fatal. En respuesta, Wright simplemente podría abandonar el término "conocimiento" y sólo hablar de opiniones (o creencias) autoritativas. Nótese que Wittgenstein, la inspiración para la teoría de Wright, negaba que podemos *saber* que nos duele algo o *saber* qué creemos. Según Wittgenstein, tales afirmaciones no tienen sentido porque lo que llamamos autoadscripciones de estados mentales en realidad no son reportes sobre nosotros mismos, de hecho tampoco son *adscripciones* (Wittgenstein, 1990:§246). Según la interpretación expresivista de Wittgenstein, un enunciado como "Creo que *p*" en realidad sólo expresa que *p*, no que el sujeto cree que *p*. Si estas ideas wittgensteinianas son correctas, entonces claramente nuestras "autoadscripciones" autoritativas de estados intencionales no expresan *conocimiento*. Una respuesta menos radical en defensa de la teoría de Wright, podría mantener que nuestras autoadscripciones autoritativas sí son adscripciones y re-

portes, pero que no expresan conocimiento sino sólo opiniones o creencias. Entonces probablemente habría algo de irracionalidad en formar estas opiniones —porque se formaron aunque su objeto no estaba presente— pero eso es común entre nuestras creencias. Tenemos muchas creencias irracionales, prejuicios que nos formamos como una respuesta racional a nuestro ambiente. Aunque así la teoría de Wright podría evitar mi objeción, quedaría la siguiente problemática: Parece que según la teoría las autoadscripciones autoritativas de nuestros estados intencionales requieren que no seamos completamente racionales. Entre más racionalmente formamos nuestras creencias, menos es posible que formemos opiniones autoritativas del tipo que Wright describe. Esta consecuencia de su teoría parece un poco extraña y en contra de la idea de que una persona racional no tiene menos, sino más posibilidad de conocerse a sí misma y de tener (por lo menos) opiniones autoritativas acerca de sus propios estados intencionales.

##### 5. CONCLUSIONES Y UNA OBSERVACIÓN DE EVANS

¿Cuáles conclusiones debemos sacar de esta discusión de la teoría de Wright? Una posibilidad es la de mantener la idea de una teoría *constitutiva* de la autoridad de la primera persona y tratar de modificar la teoría de tal manera que evita las objeciones que he elaborado aquí. Ésa es la estrategia de Jane Heal. Ella defiende la idea constitutiva, pero trata de caracterizar los estados intencionales que nos adscribimos como algo más sustancial que en la teoría de Wright. (Heal se enfoca en el caso de autoadscripciones de creencias.) Pero Heal está de acuerdo con Wright en tanto que considera nuestras creencias de segundo orden (que se expresan en autoadscripciones) como constitutivas de las creencias adscritas. Por eso su teoría tiene que responder, igual que la de Wright, a la objeción recién discutida: ¿Es la teoría capaz de caracterizar nuestras creencias de segundo orden como *autoconocimiento*? Como expliqué anteriormente, me parece que es dudoso que un conocimiento pueda ser constitutivo de su propio objeto. Otra alternativa, diferente de las de Wright y Heal, es la de modificar la estructura de la teoría. Tom Stoneham, por ejemplo, arguye que no es el caso que las creencias de segundo orden son constitutivas de las de primer orden sino al revés. Creencias de primer orden necesariamente son constitutivas de cualquier creencia de segundo orden y por eso cualquier creencia de segundo orden necesariamente es correcta (cf. Stoneham, 1998). Aunque la teoría de Stoneham tiene sus propios problemas, ciertamente merece más discusión.

Sin embargo, para concluir, quiero sugerir que investiguemos una teoría no constitutiva de la autoridad de la primera persona. Una teoría no constitutiva, entre otras, es la que investiga la *epistemología* de nuestro autoconocimiento (suponiendo que sí tenemos tal conocimiento). Aquí la pregunta central es, ¿cómo *adquirimos* nuestro conocimiento de las propias creencias (y de los demás estados

intencionales)?<sup>3</sup> La idea es que la epistemología del autoconocimiento puede revelar por qué disfrutamos de la autoridad de la primera persona. Un buen punto de partida para tal teoría es, en mi opinión, la siguiente observación de Gareth Evans:

Si alguien me pregunta "¿Piensas que va a haber una tercera guerra mundial?", para responderle, tengo que atender precisamente a los mismos fenómenos exteriores que a los que atendería si estuviera contestando a la pregunta "¿Habrá una tercera guerra mundial?" Me pongo en una posición [apropiada] para contestar a la pregunta de si creo que  $p$  al iniciar aquella operación (cualquiera que ésta sea) para contestar a la pregunta de si  $p$ .

*(If someone asks me "Do you think there is going to be a third world war?", I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question "Will there be a third world war?" I get myself in a position to answer the question whether I believe that  $p$  by putting into operation whatever procedure I have for answering the question whether  $p$ ) (Evans, 1982:225).*

Evans nota aquí que, al contestar la pregunta, "¿Crees que  $p$ ?", no examinamos, en un acto de introspección, las propias creencias, sino consideramos la pregunta de primer orden, "¿Es verdad que  $p$ ?" y, en caso de que la respuesta sea positiva, procedemos directamente a contestar "Sí, creo que  $p$ ". La cita entonces describe una manera de formar verdaderas creencias de segundo orden. Si descubro que es verdadero que  $p$ , entonces también es verdadero que yo creo que  $p$ . Para formar la creencia de segundo orden sólo necesito retener el contenido de lo que he descubierto y prefijarlo con "creo que ...". El resultado es una verdadera creencia de segundo orden: "Creo que  $p$ ". Entonces, para formar verdaderas creencias de segundo orden sólo tenemos que retener el contenido de nuestras creencias de primer orden e integrarlo en creencias de segundo orden. ¿Tenemos una autoridad especial en hacer eso? Una manera de argüir que sí es apuntar una analogía con inferencias. Para inferir de la primera premisa " $p \rightarrow q$ " y de la segunda premisa " $p$ " nuestra conclusión " $q$ " tenemos que retener los contenidos de ambas premisas y remplazar uno de ellos en la conclusión. Nadie se sorprende de que generalmente podamos lograr eso sin problemas. (Por lo menos en inferencias simples, no tendemos a equivocarnos por no acordarnos bien de los contenidos de nuestras inferencias. En general, nuestras falacias se deben a otras fallas, por ejemplo a la idea de que afirmar el antecedente constituye una inferencia válida.) Me parece que nuestra autoridad para formar verdaderas creencias de segundo orden (y hacer autoadscripciones correctas de creencia) debería ser la misma que nuestra autoridad para retener los contenidos de premisas y remplazarlos en las conclusiones de nuestras inferencias. No hay du-

<sup>3</sup> La explicación de la autoridad que Donald Davidson ha sugerido (cf. Davidson, 1984) también se enfoca en la pregunta de cómo adquirimos nuestro conocimiento de los propios estados mentales. Pero no investiga cómo en realidad adquirimos este conocimiento sino cómo *podríamos* obtenerlo. Eso provoca una inquietud acerca de cuál es el estatus de tal teoría, a veces llamada "trascendental". Véase al respecto Fricke (en dictamen).

da de que es más fácil retener y reemplazar contenidos de esta manera que observar el comportamiento de otras personas y sacar conclusiones acerca de sus creencias. En consecuencia tenemos una autoridad especial en autoadscribir las propias creencias.<sup>4</sup>

## BIBLIOGRAFÍA

- Bilgrami, A. (1998), "Self-knowledge and resentment", C. Wright *et al.* (eds.), *Knowing our own minds*, 207-241, Oxford, Oxford University Press.
- Davidson, D. (1984), "First person authority", *Dialectica*, vol. 38, 101-111 (reimpresión en D. Davidson (2001), *Subjective, intersubjective, objective*, 3-14, Oxford, Clarendon Press.
- Fernández, J. (2003), "Privileged access naturalized", *Philosophical Quarterly*, vol. 53, pp. 352-372.
- Fricke, M. F. (en dictamen), "Davidson y la autoridad de la primera persona", *Diánoia*.
- Gallois, A. (1996), *The world without, the mind within: An essay on first-person authority*, Cambridge, Cambridge University Press.
- Heal, J. (2003), "On first-person authority", J. Heal, *Mind, reason and imagination*, Cambridge, pp. 273-288.
- Smith, B. (1998), "On knowing one's own language", C. Wright *et al.* (eds.), *Knowing our own minds*, Oxford, Oxford University Press, pp. 391-428.
- Stoneham, T. (1998), "On believing that I am thinking", *Proceedings of the Aristotelian Society*, vol. 98, pp. 125-144.
- Wittgenstein, L. (1990), "Philosophische Untersuchungen", L. Wittgenstein, *Werkausgabe*, Frankfurt/Main, Suhrkamp, vol. 1, pp. 225-580.
- Wright, C. (1989a), "Wittgenstein's later philosophy of mind. Sensation, privacy and intention", *Journal of Philosophy*, vol. 86, pp. 622-634.
- , (1998b), "Wittgenstein and the central project in theoretical linguistics", A. George (ed.), *Reflections on Chomsky*, Oxford, Blackwell, pp. 233-264.
- Zimmermann, A. (2004), "Unnatural access", *Philosophical Quarterly*, vol. 54, pp. 435-438.
- Zimmermann, A., "Basic self-knowledge: Answering Peacocke's criticism of constitutivism", *Philosophical Studies* (en prensa).

<sup>4</sup> Una defensa de una teoría epistemológica como la que he esbozado aquí requiere mucho más detalle. Teorías similares (pero con un enfoque en la justificación de nuestras autoadscripciones autoritativas) se encuentran en Gallois, 1996, y Fernández, 2003. Véase Zimmermann, 2004, para una crítica de Fernández, 2003, y Zimmermann (en prensa) para una defensa de una teoría constitutiva (aunque de forma diferente de Wright).