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

MARTÍN FRANCISCO FRICKE

ABSTRACT. Someone who believes "I believe it will rain" can easily be mistaken about the rain. But it does not seem likely, and might even be impossible, that he is wrong about the fact that *he believes* that it is going to rain. How can we account for this authority about our own beliefs—the phenomenon known as first person authority? In this paper I examine a type of theory proposed, in distinct forms, by Crispin Wright and Jane Heal for an explanation of our authority. Both authors claim that our second-order beliefs (of the form "I believe that *p*") are constitutive of the first-order beliefs that are self-ascribed in them (beliefs of the form "*p*") and both try to derive the necessity of first person authority from this constitutive relation. My paper analyses and criticizes the two proposals and suggests a non-constitutive, alternative account of first person authority.

KEY WORDS. First person authority, self-knowledge, self-ascription, belief, intentional states, promising, Crispin Wright, Jane Heal, Gareth Evans, Ludwig Wittgenstein.

A menudo, cuando tratamos de averiguar qué es lo que piensan o sienten otras personas, nos equivocamos. Malinterpretamos las señales, ignoramos circunstancias particulares o no entendemos lo que se dice. Sin embargo, la situación es diferente cuando se trata de nosotros mismos. Aunque hay excepciones, en general no necesitamos interpretarnos a nosotros mismos para saber qué es lo que pensamos o sentimos. Parece que, más bien, lo sabemos de una manera inmediata y sin que haya la posibilidad de que nos equivoquemos. Al parecer, 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 que los filósofos llaman "la autoridad de la primera persona"?

Instituto de Investigaciones Filosóficas y Centro Peninsular en Humanidades y Ciencias Sociales, Universidad Nacional Autónoma de México. / martin_fricke@yahoo.com.uk

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 es lo que creemos? ¿Por qué parece improbable, quizá incluso imposible, equivocarse en sinceras afirmaciones 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 afirmar 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 las teorías constitutivas de la autoridad de la primera persona que han ofrecido Crispin Wright y Jane Heal. Según estas teorías, la autoridad se debe al hecho de que nuestras auto-adscripciones de estados intencionales contribuyen a la constitución de estos mismo 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. WRIGHT, SOBRE LA AUTORIDAD DE LA PRIMERA PERSONA

En un artículo sobre la filosofía de la mente, de Wittgenstein, Crispin Wright escribe que

> la autoridad que adscribimos generalmente a las creencias propias 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 intente.
>
> *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 Wittgenstein. Según Wright, Wittgenstein nos enseña que la auto-adscripción de estados intencionales no está basada en una observación de éstos, ni en inferencias. Más bien, según Wittgenstein en la interpretación de Wright, la relación entre auto-adscripciones de estados intencionales y estos estados intencionales mismos es *a priori*. ¿Qué significa eso? Examinemos el ejemplo de auto-adscripciones de

creencia. Supongamos que un sujeto dice "Creo que *p*". Según Wright, esta auto-adscripción 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 auto-adscripción "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, con base en 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 auto-adscripción "Creo que *p*" y la creencia de que *p* es *a priori*, porque hacer la auto-adscripción 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 es detectable.

Sin embargo, Wright no postula que nuestras auto-adscripciones 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 auto-adscripción es falsa. El criterio de la auto-adscripción es un criterio "refutable" (*defeasible*) para la presencia de un estado intencional. Pero las situaciones en las cuales la auto-adscripción resulta falsa son excepciones. "Verdad es la posición por defecto". ("[t]ruth is the default position," Wright 1989a: 633.) Aquí vemos que la teoría de Wright sólo dice que auto-adscripciones de creencias son *parte* de lo que constituye tener las creencias adscritas. En condiciones normales (la posición "por defecto"), la auto-adscripción de una creencia siempre es correcta, porque las condiciones normales junto con la auto-adscripción constituyen que tenemos la creencia adscrita. Pero hay condiciones diferentes en las cuales la auto-adscripción, junto con estas condiciones, no son suficientes para la presencia de la creencia adscrita.

El "juego de lenguaje" que incorpora estas relaciones constitutivas entre auto-adscripciones 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 defecto como verdadera. Es decir, no se necesitan razones para aceptar la verdad de una auto-adscripción. Sólo para rechazarla y declararla como falsa se necesitan razones, porque eso sería una posición diferente de la "por defecto". Entonces, generalmente, es decir, en condiciones normales, se toman las sinceras auto-adscripciones 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 tendrá a resultar en una concepción total de su psicología que elucida más —de hecho, enormemente más— que cualesquier [datos] que podrían ser obtenidos por respetar todos los datos excepto la testimonia 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 entender a otras personas si tomamos sus auto-adscripciones de estados intencionales como verdaderas, que si no tomamos estas auto-adscripciones 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.

2. OBJECIONES CONTRA WRIGHT

En esta sección quiero elaborar dos objeciones contra la propuesta de Wright que todavía no se encuentran en esta forma en la literatura sobre el tema 1. La primera objeción es que la teoría pone demasiado peso en las contingencias de nuestro juego de lenguaje y, así, pierde su poder explicativo. Mi segunda pregunta es si Wright logra caracterizar nuestras auto-adscripciones de creencia como conocimiento. Mi objeción es que hay razones para suponer que un conocimiento no puede ser constitutivo de su propio objeto.

1. 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 auto-adscripción 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$_X$" significa que está lloviendo. (Estoy usando el término técnico "llover$_X$" porque nuestro término ordinario "llover" no obedece esta regla. Pero aparte de esta regla "llover$_X$" significa lo mismo como "llover".) Si nuestro juego de lenguaje contuviera esta regla, podríamos decir que el enunciado "Está lloviendo$_X$" es un criterio para la lluvia. El enunciado es constitutivo de la lluvia. En analogía con las reglas de Wright para auto-adscripciones de estados intencionales, podríamos decir que el enunciado "Está lloviendo$_X$" 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$_X$" 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$_X$" en nuestro juego de lenguaje porque nosotros no podemos determinar cuando sería correcto usarlo. Para poder usar el término "llover$_X$" 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 enuncia "Está lloviendo$_X$", 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 auto-adscripciones 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 auto-adscripciones 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 auto-adscripciones; 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.

2. 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 [mi énfasis], 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 todo conocimiento necesariamente está guiado por el mundo (más específicamente, por su objeto), no al revés. Es decir, los posibles objetos del 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 todavía no tenemos el conocimiento del objeto, este objeto mismo todavía no puede existir. ¿Por qué? Porque se supone que el objeto está constituido (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 esos propios estados intencionales que, según él, son constitutivas de esos 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 pueden 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 deber 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 podamos *saber* que nos duele algo o *saber* qué creemos. Según Wittgenstein, tales afirmaciones no tienen sentido porque lo que llamamos auto-adscripciones 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 crea que *p*. Si estas ideas wittgensteinianas son correctas, entonces claramente nuestras "auto-adscripciones" autoritativas de estados intencionales no expresan *conocimiento*. Una respuesta menos radical en defensa de la teoría de Wright podría mantener que nuestras auto-adscripciones autoritativas sí son adscripciones y reportes, 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 era presente— pero eso es común entre nuestras creencias. Tenemos muchas creencias irracionales, prejuicios que nos formamos como una respuesta racional a nuestro

ambiente. Aunque la teoría de Wright podría evitar así mi objeción, quedaría la siguiente problemática: Parece que, según la teoría, las auto-adscripciones autoritativas de nuestros estados intencionales requieren que no seamos completamente racionales. Lo más racionalmente que formamos nuestras creencias, lo menos posible será 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.

3. HEAL

Jane Heal hace la siguiente interrogante a la teoría de Wright: Si, en condiciones normales, un estado intencional puede ser constituido sólo por la opinión del sujeto que 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. La creencia acerca de la lluvia condiciona las intenciones que formo. Ahora bien, Wright aparentemente dice que el solo hecho de que tengo la opinión de tener tal creencia acerca de la lluvia ya constituye el que tengo la creencia. La simple opinión "Creo que está lloviendo" hace, en condiciones normales, que creo que está lloviendo. ¿Pero tiene sentido suponer que la sola opinión "Creo que está lloviendo" (una opinión no acerca de la lluvia, sino acerca de una creencia mía) 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. Heal trata de dar tal explicación comparando la auto-adscripción de estados intencionales con la práctica de hacer promesas y examinando una propuesta de Gareth Evans sobre la auto-adscripción de creencia. Su discusión es detallada e interesante, y aquí sólo podré esbozarla en grandes rasgos.

Heal distingue dos sentidos de la noción de prometer. En el sentido "natural", prometer algo significa exhibir una "tendencia real" de hacerlo. Un jardinero usa la noción en este sentido cuando dice, "Mis calabazas prometen ser grandes este año" (cf. Heal 2003: 277). En el sentido "personal", prometer algo significa adquirir un compromiso de hacer algo, por ejemplo diciendo "Prometo pagar mis deudas antes del viernes". Los dos sentidos de prometer no son exclusivos: alguien que se compromete a hacer algo por tanto exhibe una tendencia real de hacerlo, el prometer personal tiene un aspecto natural también (pero no necesariamente al revés).

Una persona que sinceramente dice "Prometo hacer *X*", de hecho sí promete, en el sentido personal, hacer *X*. La sincera aserción "Prometo hacer *X*", tiene una autoridad especial. En condiciones normales, es co-

rrecto que el que hace la aserción promete hacer *X*. En cambio, la sincera aserción "*Él* promete hacer *X*", no tiene la misma autoridad; es fácil equivocarse, aunque siendo sincero, en averiguar si otra persona en realidad promete hacer algo o no. Aquí vemos que "Prometo hacer *X*" tiene una autoridad de la primera persona. Ahora bien, la pregunta de Heal es: ¿Cómo es posible tener esta autoridad, dado que el prometer personal también tiene un aspecto natural? Para saber si algo exhibe una tendencia real tenemos que observarlo y las conclusiones que sacamos de nuestras observaciones fácilmente pueden ser equivocadas. Pero nadie tiene que observarse a sí mismo para hacer una sincera promesa como "Prometo pagar antes del viernes". No es necesario averiguar si uno tiene una tendencia real de pagar para hacer la promesa. Además, parece muy poco probable, y quizá imposible, que la persona que sinceramente dice "Prometo pagar antes del viernes", no tenga una tendencia real de pagar y de hecho no prometa aunque sinceramente dice que lo hace [2]. ¿Por qué es eso?

En respuesta, Heal examina el proceso que nos permite llegar a una promesa personal. Según Heal, antes de hacer tal promesa, consideramos si deberíamos o no prometer hacer *X* y en caso de llegar a una conclusión positiva inmediatamente decimos "Prometo hacer *X*". Este proceso de deliberación que nos lleva a la conclusión de que sí deberíamos prometer hacer *X* es *constitutivo* de una tendencia real de hacer *X*. Si decidimos que debemos prometer hacer *X* y actuamos en consecuencia con esta decisión diciendo "Prometo hacer *X*", entonces necesariamente tenemos la tendencia de hacer *X* e hicimos una verdadera promesa (aunque todavía es posible que no lograremos cumplirla). La deliberación (concluida) constituye la tendencia —por lo menos en un ser con "integridad psicológica suficiente para realizar las resoluciones que él se hace" (Heal, 2003: 283). Eso explica la autoridad de la primera persona en nuestras promesas personales. El proceso de deliberación que nos lleva a hacer la aserción "Prometo hacer *X*", es lo que constituye que prometemos y así garantiza que la aserción misma es verdadera.

Heal detecta una estructura similar en las auto-adscripciones de creencia. La noción de creencia, igual que la de prometer, también tiene dos sentidos: el sentido de una creencia "natural", un "estado de un organismo que pueda ser responsivo al mundo en percepción y que guía el comportamiento hacia la satisfacción de deseos" (Heal, 2003: 284f). Animales no reflexivos tienen tales creencias y también nosotros cuando tenemos creencias llamadas "tácitas", "no conscientes" o "inconscientes". No tenemos la autoridad de la primera persona hacia nuestras creencias naturales. Pero también podemos tener creencias en un sentido "personal". Conocemos las creencias personales con la autoridad de la primera persona, sin anteriormente observarnos, sin hacer inferencias y con poca probabilidad

de que nos equivoquemos. Las creencias personales, igual que las naturales, son estados responsivos al mundo en percepción y guían el comportamiento hacia la satisfacción de deseos. ¿Cómo es posible que tales estados sean conocidos con la autoridad de la primera persona, es decir, sin previa observación y sin mucha probabilidad de error? Wright nos dice que la mera opinión de que tenemos una creencia es, en condiciones normales, suficiente para que tengamos la creencia. La objeción de Heal es que la teoría de Wright hace misterioso cómo una creencia constituida por mera opinión puede tener el aspecto natural de un estado objetivo, observable y efectivo en el control de comportamiento. Según Heal, la teoría de Wright no es capaz de elucidar este aspecto, más bien, las creencias de primer orden pierden su aspecto natural y objetivo, y la teoría tiene un "asomo, distintamente idealista".

Heal piensa que no es necesario ser idealista sobre la realidad y eficacia de las creencias si consideramos la analogía de prometer. La promesa personal tiene un aspecto natural también, aunque es adquirido sin previa observación y aunque no existe mucha probabilidad de error en la aserción "Prometo". Eso es posible porque el proceso que nos lleva a hacer una promesa personal es constitutivo del aspecto natural (y objetivo) de la misma. Según Heal, podemos apreciar que algo similar ocurre en la adquisición de creencias personales si consideramos un caso aparentemente descrito por Gareth Evans (cf. Evans 1982: 225), "el caso donde una persona es interrogada, '¿Crees que *p*?' y no tiene una respuesta lista, porque todavía no ha decidido sobre el asunto. Para contestar la pregunta tiene que decidir qué creer [*make up her mind*]" (Heal 2003: 286). Alguien me pregunta: "¿Crees que el próximo Presidente de los Estados Unidos será republicano?" Todavía no he pensado sobre el asunto y no tengo ninguna opinión acerca de si será o no republicano. En respuesta a la pregunta puedo decir: "No he pensado sobre el asunto". Pero también puedo considerar, en este mismo momento, las probabilidades de que el próximo Presidente será o no republicano. Parece que si llego a una conclusión afirmativa ("El próximo Presidente será republicano") puedo afirmar inmediatamente: "Sí, *creo* que el próximo Presidente de los Estados Unidos será republicano". Aparentemente, en este proceso de deliberación, se contestan simultáneamente dos preguntas distintas:

1. ¿De qué partido será el próximo Presidente de los Estados Unidos?
2. ¿Crees que el próximo Presidente de los Estados Unidos será republicano?

Heal toma este fenómeno como indicio de una estructura similar a la del caso de la promesa. En el caso de la promesa, la sincera aserción "Prometo", es al mismo tiempo lo que hace verdadero que la persona tiene una tendencia de hacer lo que promete y, por lo tanto, lo que hace la aserción

verdadera. En el caso de la creencia, según Heal, el proceso de deliberación que nos permite contestar (2) es, al mismo tiempo, un proceso que nos permite contestar (1). Heal concluye que el juicio con que contestamos (2) también constituye un juicio que contesta (1). El proceso de deliberación nos lleva a un juicio que es, a la vez, una respuesta a ambas preguntas:

> Allí tuvimos una acción que era ambas una promesa y una aserción. Aquí tenemos un juicio que es a la vez el juicio 'p' (ya que resulta de una deliberación de este asunto y tiene estas consecuencias) y el juicio 'Creo que p' (ya que tiene esta forma y estas consecuencias también). Así, se trata de un juicio que, al representar el ego de cierta manera, representa el mundo de cierta manera.
>
> *There we had an action which was both a promise and an assertion. Here we have a judgement which is both the judgement 'p' (since it results from deliberation on that issue and has those consequences) and the judgement 'I believe that p' (since it has that form and those consequences too). So it is a judgement which represents the world as being a certain way in representing the self as being a certain way (Heal 2003: 287).*

Según Heal, el juicio expresado en "Creo que *p*" es el inicio de un estado continuo que tiene dos aspectos: el de la creencia "*p*" y el de la creencia "Creo que *p*".

> El suceso del juicio expresado en 'Creo que p' tiene consecuencias de largo plazo. En una mente que es estable y retentiva, será el inicio de un estado continuo que tiene los mismos dos aspectos que el juicio y que, más tarde, puede expresarse igualmente en 'p' y en 'Creo que p'. Este estado continuo, ciertamente una de las cosas que llamamos 'creencia', es, a la vez, una aprehensión del ego y, al mismo tiempo, una aprehensión del mundo. Se sigue que la teoría constitutiva de la autoridad de la primera persona es adecuada para él.
>
> *The event of judgement expressed in 'I believe that p' has long-term consequences. In a mind which is stable and retentive, it will be the onset of a continuing state which has the same two aspects as the judgement and which can later express itself equally in 'p' and 'I believe that p'. This continuing state, surely one of the things which we call 'belief', is both an apprehension of the self and at the same time an apprehension of the world. It follows that the constitutive account of first-person authority is appropriate for it. (Heal 2003: 287).*

Heal arguye que este estado con dos aspectos no sólo existe en el caso en que la pregunta ¿Crees que *p*?, primeramente nos provoca a pensar sobre el asunto porque todavía no tenemos una opinión acerca de él. Según Heal, en otros casos también es plausible que existan estados similares. ¿Por qué? Porque en estos casos asentimos sin duda alguna a ambas preguntas y por eso parece que existe un solo estado que exhibe ambos aspectos de creencia. El ejemplo de Heal consiste en las preguntas: ¿Hay orquídeas en la luna?, y, ¿Crees que haya orquídeas en la luna? Ya que

negamos ambas preguntas con igual facilidad, Heal concluye que aquí también se expresa un mismo estado en los juicios "No hay orquídeas en la luna" y "Creo que no haya orquídeas en la luna[3]". La creencia expresada en el segundo juicio constituye y es idéntico con la creencia expresada en el primer juicio. La creencia de segundo orden es autoritativa, porque es constitutiva de la creencia expresada en el primer juicio, se hace verdadera a sí misma. Heal concluye que la teoría constitutiva también es adecuada para casos donde creencias de primer y segundo orden ya existen antes de que el sujeto se pregunte qué es lo que cree.

4. OBJECIONES CONTRA HEAL

En lo que sigue quiero hacer tres objeciones a la teoría de Heal. La primera repite una objeción que también hice contra la propuesta de Wright; la segunda cuestiona la coherencia de la idea de que un solo estado podría expresarse en dos creencias; y la última pregunta si la teoría como concebida en realidad explica el fenómeno de la autoridad de la primera persona.

1. Heal piensa que nuestras creencias de segundo orden son constitutivas de las creencias de primer orden que son sus objetos. La creencia "Creo que *p*" es parte de lo que constituye la creencia "*p*". Según Heal, la autoridad de nuestras creencias de segundo orden deriva de esta relación de constitución. Si su teoría en realidad tiene esta forma lógica (más adelante desarrollaré unas dudas sobre esta auto-caracterización de su propuesta), entonces está sujeta a la misma crítica que hice arriba contra la teoría de Wright: La propuesta no es compatible con la idea de que nuestras creencias autoritativas de segundo orden sean auto*conocimientos*. Eso es porque un conocimiento no puede ser constitutivo de su propio objeto y, de tal forma, de su propia verdad. Para que un estado sea conocimiento debe ser guiado por el mundo, no al revés. Pero según Heal, las creencias autoritativas no son guiadas por sus objetos; más bien, ellas constituyen sus objetos y, así, hacen que sean verdaderas. En este sentido, las creencias autoritativas "guían", es decir, determinan el mundo de tal manera que sean verdaderas y no al revés. Eso significa que las creencias autoritativas como son concebidas por Heal no pueden ser consideradas como conocimientos.

2. ¿Tiene sentido decir que un juicio "es a la vez el juicio 'p' [...] y el juicio 'Creo que p'" (Heal 2003: 287)? Las proposiciones expresadas en "*p*" y "Creo que *p*" tienen contenidos diferentes, ya que tienen implicaciones diferentes: "*p*", trivialmente, implica que *p*, pero "Creo que *p*" no implica que *p*. Este hecho pone en duda la idea de que un solo juicio podría ser ambos el juicio "*p*" y el juicio "Creo que *p*", porque parece plausible que el contenido de un juicio es parte de lo que individua un juicio de otro. Dos juicios que tienen contenidos diferentes deben ser juicios diferentes. Ya

que los juicios "*p*" y "Creo que *p*" tienen contenidos diferentes no pueden ser idénticos, como afirma Heal.

Un argumento análogo aplica a la aserción de Heal, de que la creencia de primer orden "*p*" debe ser identificada con la creencia "Creo que *p*" (cf. Heal 2003: 285). Dos creencias que tienen implicaciones diferentes (la primera implica que *p*, la segunda no) no pueden tener el mismo contenido y, por lo tanto, no pueden ser idénticos. Si la teoría de Heal requiere que consideremos el juicio (o la creencia) de que *p* como idéntico con el juicio (o la creencia de que creo que *p*), entonces parece que la teoría no puede ser correcta.

¿Es posible interpretar la teoría de Heal de tal manera que se eviten estos problemas? Como vimos, Heal explica la idea central en la siguiente oración: "se trata de un juicio que, al representar el ego de cierta manera, representa el mundo de cierta manera" (Heal 2003: 287). Eso lo podríamos también entender de la siguiente manera: Si juzgo "*p*" y si "*p*" implica "*q*", entonces, quizá, se podría decir que al representar las cosas como siendo *p* también las represento como siendo *q*. ¿Pero podría eso ser lo que Heal quiere decir? Probablemente no, porque, primero, "Creo que *p*" no implica "*p*", ni "*p*" "Creo que *p*"; así, la implicación presupuesta por esta interpretación no existe. Segundo, incluso si juicio "*p*" implica "*q*" no necesariamente se sigue que, al juzgar "*p*" también juzgo "*q*". Considerase: Yo podría juzgar que algo es un círculo sin también ser dispuesto de juzgar que se trata de un conjunto de puntos que están equidistantes de un único punto. Los juicios "*p*" y "*q*" son sucesos mentales distintos, incluso si la verdad de uno implica la verdad del otro.

Una interpretación más prometedora es la que se sugiere a partir del segundo pasaje de Heal que cité arriba: Las creencias "Creo que *p*" y "*p*" son "dos aspectos" de un "estado continuo" que "es, a la vez, una aprehensión del ego y, al mismo tiempo, una aprehensión del mundo" (Heal 2003: 287). El término importante en esta explicación es "aspecto". Un solo estado puede tener diferentes aspectos, los cuales no son idénticos entre ellos. La creencia "*p*" y la creencia "Creo que *p*" son dos creencias distintas y no idénticas, ya que tienen contenidos diferentes. Pero aunque no son idénticos, podría ser que son aspectos distintos de un solo objeto (o estado) —como el color y la textura de una hoja son dos aspectos distintos y no idénticos de un solo objeto. Si formulamos la teoría de Heal usando la noción de aspectos, podemos abandonar la afirmación problemática, según la cual la teoría correcta hace una "identificación [...] entre la creencia de que uno está en cierto estado [intencional] y el estado mismo" (Heal 2003: 285). Más bien, la teoría afirma que la creencia de segundo orden ("Creo que *p*") y la creencia (u otro estado intencional) de primer orden ("*p*") son diferentes aspectos de un solo estado.

Pero, ¿cuál es este solo estado que tiene estos dos aspectos? La persona que tiene las creencias, por ejemplo, tiene ambos aspectos: el aspecto de creer "Creo que *p*", y el aspecto de creer "*p*". Pero la afirmación de que las dos creencias tienen una sola persona, como el mismo sujeto, no parece un descubrimiento interesante y, al parecer, tampoco puede explicar por qué la creencia de segundo orden es una creencia autoritativa. Más plausiblemente, Heal quiere decir que los dos juicios "Creo que *p*" y "*p*" son dos aspectos de un solo juicio y que las creencias correspondientes son dos aspectos de una sola creencia. Probablemente, el contenido del juicio y de la creencia correspondiente sería una simple conjunción de los contenidos de los juicios individuales: "*p* y creo que *p*". El juicio "*p*" es un "aspecto" del juicio "*p* y creo que *p*"; y el juicio "Creo que *p*" también es un "aspecto" del juicio "*p* y creo que *p*". Análogamente, podemos decir que las creencias "*p*" y "Creo que *p*" son "aspectos" de la (sola) creencia "*p* y creo que *p*".

En esta interpretación, la autoridad de la primera persona se debe al hecho de que creencias de segundo orden no vienen "a solas". Más bien, tal creencia siempre es sólo un aspecto de un estado que también tiene, como segundo aspecto, el estado que se auto-adscribe en la creencia de segundo orden. Alguien que cree, "Creo que *p*", en realidad cree "*p* y creo que *p*", por eso la creencia "Creo que *p*" no puede ser falsa.

Si esta es la interpretación más prometedora de Heal, no obstante surge la siguiente pregunta: ¿En qué sentido es correcto hablar de una teoría *constitutiva* de la autoridad de la primera persona? En la nueva interpretación no parece correcto afirmar que la creencia "Creo que *p*" constituye la creencia "*p*". Más bien, la creencia "Creo que *p*" es parte de lo que constituye la creencia "*p* y creo que *p*". Asimismo, la creencia "*p*" *también* es parte de lo que constituye la creencia "*p* y creo que *p*". La creencia de segundo orden y la de primer orden juntas constituyen el solo estado intencional que tiene ambas creencias como aspectos. Parece, entonces, que Heal está equivocada en caracterizar su teoría como una propuesta que comparte con la de Wright su forma lógica. De hecho, su teoría no es constitutiva en el sentido en que se entienden normalmente las "teorías constitutivas" de la autoridad de la primera persona.

3. A la luz de un análisis cuidadoso, mi segunda crítica se convirtió en una crítica meramente terminológica. Mi última crítica es más sustancial. Suponiendo que Heal tiene razón en describir nuestras creencias autoritativas como parte de estados intencionales más complejos, que incluyen las creencias de primer orden que hacen verdaderas a nuestras creencias autoritativas, ¿por qué somos *autoritativos* en formar estos estados intencionales? En otras palabras, ¿por qué somos buenos en juntar creencias de primer y de segundo orden en tales estados intencionales? ¿Por qué somos autoritativos en formar creencias del tipo "*p* y creo que *p*"? Para explicar la autoridad de la primera persona no es suficiente mostrar que, en ciertas

circunstancias, las creencias "*p*" y "Creo que *p*" están unidas en un solo estado de la forma "*p* y creo que *p*". Para explicar la autoridad también tenemos que elucidar por qué somos más exitosos en formar tales estados que en formar opiniones acerca de las creencias de otras personas.

Wright afirma que existen juegos de lenguaje en los cuales los participantes conceden la autoridad de la primera persona. Pero Wright no explica cómo es posible que existan estos juegos de lenguaje (y no otros que conceden la misma autoridad a adscripciones de creencias a otras personas). De manera similar, Heal afirma que existen estados intencionales (de la forma "*p* y creo que *p*") que implican que el sujeto del estado tiene la autoridad de la primera persona. Pero Heal no explica cómo es posible que tengamos estos estados intencionales. El problema es el siguiente: Supóngase que es muy difícil para nosotros formar estados intencionales de la forma "*p* y creo que *p*". Supóngase, además, que es *más difícil* para nosotros formar tales estados que formar opiniones correctas acerca de las creencias de otras personas y que en muchas ocasiones fallamos cuando intentamos formar tales estados. Quizá los estados que formamos cuando fallamos tienen la forma "*p* y creo que *q*". En este caso, Heal tendría razón al decir que si un sujeto tiene un estado intencional de la forma "*p* y creo que *p*" entonces necesariamente tiene una creencia de segundo orden ("creo que *p*") que es correcta. Sin embargo, en este caso, el sujeto no gozaría de la autoridad de la primera persona o, por lo menos, no gozaría de ella por las razones que da Heal. Porque en este caso el sujeto fallaría en muchas ocasiones al formar estados de la forma "*p* y creo que *p*" y formaría con más facilidad opiniones correctas acerca de los estados intencionales de otras personas. El problema con la propuesta de Heal es que, al parecer, nada en su teoría excluye el caso descrito. Me parece que una explicación completa de la autoridad debería ser capaz de excluirlo. En lo que sigue, presentaré una manera de explicar por qué somos buenos en formar creencias de segundo orden y más autoritativos que en formar creencias sobre las creencias de otras personas. Posiblemente, mis ideas también servirían para complementar la propuesta de Heal.

5. CONCLUSIONES Y UNA OBSERVACIÓN DE EVANS

¿Cuáles conclusiones debemos sacar de esta discusión de Wright y Heal? 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 evite las objeciones que he elaborado aquí. Se han propuesto varias versiones de teorías constitutivas distintas de las presentadas aquí. Shoemaker, por ejemplo, defiende una teoría constitutiva de corte funcionalista (cf. Shoemaker 1990); Bilgrami sugiere una teoría constitutiva que relaciona la autoridad con reacciones como el resentimiento (cf. Bilgrami

1998 [4]). Sin embargo, me parece que cualquier teoría constitutiva de la autoridad de la primera persona tiene que responder a la siguiente interrogante: ¿Es la teoría capaz de caracterizar nuestras creencias de segundo orden como auto*conocimiento*? Como expliqué anteriormente, me parece que es dudable que un conocimiento puede ser constitutivo de su propio objeto.

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) [5]? 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 esta 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 remplear uno de ellos en la conclusión. Nadie se sorprende de que generalmente podemos 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 auto-adscripciones correctas de creencia) debería ser la misma que nuestra autoridad para retener los contenidos de premisas y remplearlos en las conclusiones de nuestras inferencias. No hay duda de que es más fácil retener y remplear 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 auto-adscribir las propias creencias [6].

NOTAS

1 La crítica más tajante que se encuentra en la discusión sobre Wright es, en mi opinión, la siguiente, elaborada por Barry Smith: Al parecer, podemos tener estados intencionales como creencias sin *saber* que los tenemos. Wright mismo reconoce que nuestras auto-adscripciones de creencia sólo son suficientes para que tengamos la creencias adscritas, no son necesarias. Wright hace esta calificación para evitar un regreso al infinito. Si todos 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. Pero si podemos tener estados intencionales sin saber que los tenemos, ¿por qué es que en algunos casos los estados dependen de nuestras auto-adscripciones (en el caso cuando tenemos conocimiento de ellos), pero en otros casos no dependen de ellas (en el caso cuando no tenemos conocimiento de ellos)? Como dice Smith: "[E]n 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?" *([I]n 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.))*

2 No sabemos con la autoridad de la primera persona si lograremos cumplir con nuestras promesas. Pero sí sabemos con autoridad si prometemos o no.

3 Un ejemplo con una respuesta afirmativa sería más apropiado para argüir el caso, ya que una respuesta negativa a la pregunta ¿Crees que hay orquídeas en la luna? puede tomar dos formas: (a) "No creo que haya orquídeas en la luna" o (b) "Creo que no haya orquídeas en la luna". Para comprobar que el caso está en conformidad con la teoría de Heal, tenemos que suponer que la pregunta se contesta usando (b) que indica la presencia de una creencia negativa y no la mera ausencia de una creencia como la indica (a).

4 Otra teoría interesante de corte constitutivo es la de Tom Stoneham. Al parecer, Stoneham reversa la relación de constitución, afirmando que nuestras creencias de primer orden pueden ser constitutivas de creencias de segundo orden acerca de ellas. Cf. Stoneham 1998.

5 La explicación de la autoridad que Donald Davidson ha sugerido (cf. Davidson, 1984) también 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, 2007.

6 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 auto-adscripciones 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).

REFERENCIAS

Bilgrami, A. (1998), "Self-knowledge and resentment", in C. Wright, et al. (eds.), *Knowing our Own Minds*. Oxford: Oxford University Press, pp. 207-241.
Davidson, D. (1984), "First person authority", *Dialectica* 38: 101-111 (reimpresión en D. Davidson (2001), *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press, pp. 3-14).
Evans, Gareth (1982), *The Varieties of Reference*. Oxford: Clarendon Press.
Fernández, J. (2003), "Privileged access naturalized", *Philosophical Quarterly* 53: 352-372.
Fricke, M. F. (2007), "Davidson y la autoridad de la primera persona", *Diánoia* 52: 49-76.
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", in J. Heal, *Mind, Reason and Imagination*. Cambridge: Cambridge University Press, pp. 273-288.
Shoemaker, S. (1990), "First-person access", *Philosophical Perspectives: Action Theory and Philosophy of Mind* (*Nous* Supplement) 4: 187-214.
Smith, B. (1998), "On knowing one's own language", in 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* 98: 125-144.
Wittgenstein, L. (1990), "Philosophische Untersuchungen", in L. Wittgenstein, *Werkausgabe*, vol. 1, Frankfurt/Main: Suhrkamp, pp. 225-580.
Wright, C. (1989a), "Wittgenstein's later philosophy of mind. Sensation, privacy and intention", *Journal of Philosophy* 86: 622-634.
Wright, C. (1998b), "Wittgenstein and the central project in theoretical linguistics", in A. George (ed.), *Reflections on Chomsky*. Oxford: Blackwell, pp. 233-264.
Zimmermann, A. (2004), "Unnatural access", *Philosophical Quarterly* 54: 435-438.
Zimmermann, A. (en prensa), "Basic self-knowledge: Answering Peacocke's criticism of constitutivism", *Philosophical Studies*.