

LA FUERZA DEL PC (PRINCIPIO DE CARIDAD). CIRCUNSTANCIA, HABILIDAD Y CONTENIDO EN LA EXPLICACION DE LA ACCION

FERNANDO BRONCANO

Del PC al PRI

En este trabajo vamos a abordar dos preguntas acerca del Principio de Caridad. La primera, ¿es posible acomodar el PC en una perspectiva naturalista? La segunda, ¿afectará la noción de contenido al *status* del PC? La respuesta para ambas cuestiones es única: una noción de contenido no cartesiana, no dependiente de la concepción simbólica de lo mental, apoya una versión naturalista del Principio de Caridad.

Permítaseme comenzar aclarando en qué sentido hablo de naturalismo: el naturalismo es un programa de investigación metafísico que entraña una tesis ontológica y una tesis epistemológica. Esta es la tesis ontológica:

(1) Las entidades y propiedades admisibles en una ontología deben ser compatibles con el cierre causal de la naturaleza por las propiedades físicas.

(1) se puede reformular también diciendo que toda entidad o propiedad admisible debe obedecer un principio de sobreveniencia sobre propiedades físicas. El principio epistémico es el siguiente:

(2) La filosofía es continua con la investigación científica. Sus tesis se juzgan por métodos similares a las teorías de alto nivel teórico.

El Principio de Caridad de Quine, para la traducción radical, y el de Davidson y Dennett, para la interpretación radical, tiene dificultades para cumplir estas condiciones naturalistas. Por una parte es un principio intrínsecamente normativo, ya que indica que *debemos* conceder *racionalidad* a un sistema para inteligir su conducta como intencional (donde los dos términos en cursiva son términos intrínsecamente normativos). En tanto que principio normativo, y por serlo, tiene dificultades para encajar en una ontología naturalista a menos que indiquemos la manera de hacer compatible su fuerza

normativa con su carácter empírico. Por otra parte es un principio trascendental que establece las condiciones *a priori* de posibilidad de toda interpretación. Por esta segunda razón tiene problemas con la segunda tesis epistemológica, por cuanto una visión científica del mundo no acepta fácilmente la existencia de principios trascendentales que no sean de carácter lógico (y quizá ni siquiera éstos). Mi conjetura es, sin embargo, que podemos cumplir los dos requisitos y todavía mantener el PC como un principio metodológicamente respetable.

En la naturalización del PC hay dos cuestiones diferentes. En primer lugar está la cuestión del alcance y fuerza de un principio de racionalidad: ¿hay que asumir un principio de racionalidad perfecta? ¿Bastaría, por el contrario, con un principio de racionalidad mínima? ¿La perspectiva de diseño perfecto implica que las capacidades de razonamiento implican una lógica perfecta?. Estas cuestiones son importantes para el programa de naturalización del contenido, sin embargo las dejaré a un lado para centrarme en una pregunta más general acerca de la naturaleza ontológica y epistémica del principio de caridad, independientemente de cuál sea la fuerza que le concedamos. Brevemente, la tesis que voy a defender es la siguiente: una solución naturalista al PC, concebido como norma para toda interpretación, implica el compromiso de expresar su fuerza normativa en términos de cuestiones de hecho y, por consiguiente, implica un compromiso ontológico. En cuanto a su fuerza normativa, se fundamenta en que tiene una función propia en cierta clase de interacciones: aquellas que involucran *interacciones intencionales* entre seres iguales, es decir, aquellas que involucran la captación de intenciones en el medio como parte del contenido de la propia acción.¹ Esta es la razón por la que me veré obligado a entrar en discusiones sobre la naturaleza del contenido. Argumentaré en favor de una noción de contenido intencional desde una posición no eliminativista, no instrumentalista, pero tampoco comprometida con el hiperrealismo representacional. Esta perspectiva, es mi argumento, hace mucho más fácil de tragar el carácter naturalista del PC. Por cierto, siguiendo la recomendación de Antoni Gomila, creo que esta refundación del PC debería conllevar un cambio de nombre: será mucho más adecuado que lo llamemos el PRI (*Principio de Racionalidad en la Interpretación*).

El PRI tal como ha sido considerado en la literatura cumple dos funciones. La primera es la de proporcionar una condición suficiente de intencionalidad:

(3) Un sistema es intencional si y sólo si es racional.

¹ Entenderé por contenido de la acción el contenido de los estados mentales del sujeto que la produce, de forma que no distinguiré entre contenido y sentido de una acción.

Esta función convierte el PRI en un principio de alto nivel que permite generalizar los postulados de la psicología intencional desde el provincianismo de especie o de cultura a una psicología sin restricciones espaciotemporales. Si en cualquier momento o lugar, bajo cualquier apariencia física, un objeto muestra conducta racional, entonces tiene intenciones. Y viceversa.

Mientras que la primera función establece una simetría ontológica entre intenciones y racionalidad de acción, la segunda función se asienta, por el contrario, sobre la existencia de una asimetría epistémica en la adscripción de ambas propiedades a un sistema: así, mientras que la racionalidad es abierta, o se predica de las obras como la fe católica, las intenciones pertenecen a la parte oculta del funcionamiento del sistema, puesto que dependen de un estado mental que, al menos en apariencia, solamente se puede conocer en primera persona. El PC funcionaría como puente para la adscripción de intenciones en la explicación y comprensión de la conducta: en el acto de explicación o de comprensión de la acción hacemos un argumento retroductivo o abductivo de esta forma:

(4) Dada una acción observable A del agente S, en las circunstancias M, decimos que A tendría sentido si, dadas las circunstancias M, si el agente fuera racional y si tuviera el deseo de O y la creencia de que A lleva a O, S tendría la intención I de hacer A.²

O en una reconstrucción más formal:

(5) (i) El agente S es racional (PC)

(ii) S desea O en M

(iii) S sabe que A produce O

(iv) S sabe hacer A

(v) El que (i)-(iv) expliquen A es una buena razón para aceptar (i)-(iv)

Este esquema retroductivo convierte al PRI (de ahora en adelante lo llamaremos por su nuevo nombre) en un instrumento metodológico de investigación para adscribir deseos y creencias a los sujetos, dada su conducta. Es,

² No es necesario comprometerse con la existencia de intenciones como algo separado del complejo deseos/creencias, pero en este caso asumiremos esta diferencia.

antes que otra cosa, un principio que establece restricciones en las cláusulas *ceteris paribus* que rodean las generalizaciones de la psicología intencional. Por ejemplo:

(6) (i) La gente con sed, *ceteris paribus*, tiende a buscar bebida en su medio.

(ii) La gente racional, *ceteris paribus*, tiende a buscar las cervezas en su frigorífico antes que ordenarlas por teléfono al bar más caro de la ciudad.

(iii) etc.

De esta manera, en ausencia de otras informaciones, si observamos que Marcelo se dirige al frigorífico y toma una cerveza, activamos una pequeña teoría, cuya instanciación en la circunstancia concreta se parece mucho a la abducción anterior. El resultado es que ya podemos decir que hemos comprendido la acción, o que la hemos explicado, o ambas cosas.

El PRI es, pues, una preconditione normativa para la intelección de la conducta ajena, porque es una preconditione de comprensión de cualquier tipo de conducta intencional. Exige que adscribamos al agente cuya conducta tratamos de explicar una capacidad de creencias y deseos similar a la nuestra en sus mismas condiciones, así como una capacidad de extracción de inferencias válidas similar a la que tenemos nosotros como agentes intérpretes. Una vez que aceptamos el PRI, ya es lo mismo explicar una acción que comprenderla. Para los defensores de la explicación intencional, la asignación de creencias y deseos racionales al sujeto agente se convierte en el último sustrato metodológico que podemos alcanzar en el nivel intencional. Más allá será otra cosa, pero no explicación de la acción intencional. Quine ha desarrollado esta perspectiva para la traducción radical, Davidson para la interpretación radical de acciones y Dennett ha universalizado la perspectiva para la explicación de la conducta de sistemas complejos en los que la explicación física y la explicación funcional sean insuficientes. La progresiva generalización de su campo de aplicación ha convertido así al PRI en un principio fundamental de cualquier ciencia o teoría que emplee contenidos representacionales como entidades teóricas. Su utilidad práctica está, por lo demás, sobradamente demostrada en nuestras prácticas cotidianas de comprensión y predicción de la conducta ajena.³

³ Dennett, 1991, insiste en la prioridad funcional de su capacidad predictiva:

Pero esta omnipresencia y casi omnipotencia del PRI es también la fuente de las sospechas sobre su naturaleza. Una primera forma de entender su naturaleza podría ser como un presupuesto metodológico similar a otros muchos principios que funcionan en las ciencias. Por ejemplo, el principio de inercia es un presupuesto de la segunda ley de Newton, aunque no observemos empíricamente sistemas inerciales en la naturaleza. La ley de inercia se formula (al menos en la física clásica) para sistemas ideales libres de fuerzas que no son, por lo mismo, observables empíricamente. Para quienes están preocupados por la fundamentación científica de la psicología y de las ciencias sociales en general, esta forma de considerarlo resulta bastante aceptable y quizá no tendría que plantear estos escrúpulos naturalistas que alego. Pero el PRI no es sólo ni principalmente un principio metodológico: es una precondition de las prácticas mutuas interpretativas de los agentes intencionales que viven en un contexto social en el que la acción cooperativa y social es la regla. Es en este contexto en el que el PRI adquiere su fuerza ontológica y es aquí donde se plantea su relación profunda con una teoría del contenido. En la versión metodológica, no tendría que haber ninguna relación entre el PRI y la noción de contenido. Pero la relación sí existe cuando consideramos el PRI como un grupo de constricciones sobre los procesos de desarrollo cognitivo que conducen a la capacidad de mutua intelección e incluso manipulación mutua de las representaciones. La naturaleza de esas constricciones debe fundarse, éste será mi argumento, sobre la naturaleza del contenido y de las relaciones que éste tiene con el medio. ¿Por qué no nos conformamos con el *status* que ya tenía de principio metodológico?, ¿no son respetables ciertos principios similares en todas las ciencias? Al fin y al cabo otras ciencias también usan cláusulas *ceteris paribus*, pensemos por citar el ejemplo que usa Dennett, el principio de optimalidad en las ciencias biológicas. De acuerdo a este principio, en ausencia de otras consideraciones comenzamos a investigar un órgano sospechando que su forma se adecua a una función. No habría ningún inconveniente en considerar de esta manera el PRI, pero reparemos en que el principio de optimalidad es tan válido como la teoría evolucionista que lo sostiene. El PRI no puede apoyarse solamente en la teoría evolucionista, *pace* Millikan, porque hay procesos de aprendizaje que sostienen el PRI que no tienen origen evolutivo, sino social, pero si vamos a dar al PRI un estatuto parecido al principio de optimalidad que se aplica a la relación entre la estructura y función de los seres vivos que resultan de la adaptación,

aunque sirve para muchas otras cosas, como por ejemplo para enamorarse, para reconstruir nuestra vida y escribir un diario, etc., estos servicios los cumple precisamente a causa de su aceptable capacidad como útil predictivo.

deberíamos también fundamentarlo en teorías de desarrollo psicológico de los sistemas cognitivos. De acuerdo, ésta es la opción que vamos a defender, pero no olvidemos que la fuerza metodológica deriva de su *status* ontológico, y no a la inversa.

La teoría de la acción es neutra en apariencia respecto de una teoría de la naturaleza del contenido. Casi todos los postulados de Dennett o Davidson podrían mantenerse, *mutatis mutandis*, si resultara cierta una teoría del rol conceptual o una teoría de tipo fodoriano de la dependencia asimétrica, o una teoría del funcionalismo biológico, o...; sin embargo, si nos tomamos el PRI como una tesis empírica que dice algo acerca de la naturaleza y desarrollo de ciertos sistemas cognitivos, a saber, de aquellos que son capaces de interacción social, la naturaleza del contenido, como explicación de la acción intencional, comienza a ser relevante para una comprensión de tal naturaleza. La función biológica del PRI podría haber sido la de *construir contextos relevantes para el mantenimiento de la interacción social*. Nuestros ancestros homínidos vivieron en un medio prioritariamente social. La importancia del medio social para la supervivencia individual explica el desarrollo de métodos rápidos de resolución de problemas sociales que dieron origen a las constricciones que agrupamos bajo el nombre de PRI. Una interacción solamente es una interacción social si conlleva una mutua adscripción de contenidos y, lo que es más importante, manipulación de contenidos. Si nos preguntamos, por ejemplo, cómo pudieron los depredadores carnívoros desarrollar manadas con jerarquías sociales, seguramente tendremos que contar una larga historia sobre feromonas, y sobre expresiones corporales que indican estados emotivos internos y muchas otras cosas sobre la función de las jerarquías para desarrollar mejores métodos de caza, etc. Pero si nos preguntamos cómo los protohomínidos desarrollaron, a partir de manadas con jerarquías sociales, un sistema de aprendizaje cultural, un sistema de convenciones y un sistema de enseñanza para mantener estas convenciones, etc, tendremos que contar una historia sobre la naturaleza de la adscripción de contenidos que nos llevará necesariamente del contenido al PRI y del PRI al contenido.

Mi argumentación no será sin embargo una argumentación demostrativa; comenzaré, mas bien, mostrando algunas características de la acción que nos llevan a cuestionar nuevas formas de pensar el contenido. A saber, mi propuesta es: una concepción situada de las acciones hace plausible el abandono de la concepción simbólica del contenido y, con ella, de la dualidad entre un aspecto interno, *narrow*, especializado en la producción causal, y un aspecto externo, *wide*, especializado en la comprensión pública. El abandono de esta dualidad viene favorecida por la existencia de nuevas vías programáticas abiertas por la nueva IA conexionista y por la robótica. A su vez, una perspectiva del contenido

como un sistema de patrones estables preconceptuales nos permite avanzar una primera aproximación a la naturaleza del PRI: apareció como un sistema de habilidades en la manipulación de los contenidos de otros sistemas cognitivos presentes en el medio. El PRI, en nuestra propuesta, es una restricción a un tipo especial de facultad cognitiva y práctica: la de detectar intenciones en el medio.

El carácter situado y circunstancial de la acción

Comenzaremos por examinar algunas de las consecuencias que tiene el concebir las acciones intencionales desde un prisma externalista. La característica más importante en la que deseo insistir es lo que la escuela californiana (Barwise, Perry, D. Israel entre otros) ha llamado carácter *situado* de la acción. Con esta denominación señalamos el hecho de que la acción intencional se produce en circunstancias o situaciones concretas (el escenario de la acción) que contribuyen a individualizar la acción y, por consiguiente, son relevantes para su explicación y comprensión. La naturaleza situada o circunstancial implica que, además de involucrar preferencias y creencias, la acción tiene que haber sido particularizada por la situación concreta de producción. Esta particularización exige una forma especial de representar la acción en que aparece una relación indicativa con el medio, como señalan Israel *et al.* 1992. En este trabajo se insiste en la diferencia que existe entre acciones tipo y caso cuando adoptamos esta perspectiva circunstancial. Así, desde la perspectiva de Bruto la acción-tipo determinada como <matar, César> no es la misma que la acción-caso <matar, César, en idus de marzo, con la espada, de frente>, dado que en el modo de llevarla a cabo está implicado algo más que la mera relación de concreción tipo-caso. La segunda descripción exige un modo de acción que sería el correlato de un modo de presentación que, en el caso de la representación conceptual, funciona como un particularizador del contenido. Pero los modos de acción son algo menos trivial de lo que su nombre pudiera sugerir puesto que, si es un tópico que todas las acciones se realizan de un modo u otro, el que de hecho se realice de un modo u otro depende de al menos tres factores: el primero es la *noción concreta* que uno se haga de la acción en el momento de su ejecución, el segundo es la *habilidad* o capacidad para llevarla a cabo bajo ese modo de acción y el tercero la *situación o contexto de realización*. Tomando a Bruto como sujeto de la acción <matar, César>, ésta se representa para él bajo un modo indicativo idiosincrático. Podría ser, pongamos por caso, que Bruto confundiera a César con Marco Antonio a causa de su atolondramiento, o porque las condiciones de visibilidad no eran las idóneas. En tal caso, la acción realizada no sería la acción intentada a menos que existiera una colaboración

del medio y de las capacidades del organismo. Supongamos que se han dado todas las condiciones especificadas por una teoría de tipo davidsoniano: la intención primitiva, la preferencia y la proactividad son las correctas, la posición la adecuada, etc., pero Bruto confunde a César con Marco Antonio, la razón causaría entonces una acción que no es la acción intentada. Si tomásemos el camino inverso, desde el mundo o la circunstancia a la mente que planifica la acción, nos encontraríamos con un problema similar, a saber, a menos que especifiquemos con precisión el contexto no conseguiremos determinar el modo de llevar a cabo la acción, es decir, sólo la relación directa con el contexto nos individualiza el modo de acción tal como se presenta ante el agente.

En más de un sentido este modo de entender las acciones está relacionado con un análisis similar de las acciones lingüísticas que llamamos preferencias, pues, al fin y al cabo, las preferencias son acciones situadas que tienen una intención comunicativa. Como cualquier acción situada aprovecha de una manera oportunista lo que ofrece la circunstancia en orden a lograr el objetivo. Por ejemplo, las convenciones lingüísticas, también cualquier otro tipo de convenciones culturales y regularidades, la posición y situación del oyente en el contexto y cualesquiera elementos prominentes del contexto que se ofrezcan como medios para alcanzar el éxito comunicativo.

Supongamos que deseo una cerveza y deseo que el camarero sepa que deseo una cerveza y además deseo que lo llegue a saber por medio de mi acción intencional, y no como resultado no pretendido de mi apariencia de persona sedienta. Podría tal vez intentarlo de estas formas, dependiendo de cuál sea la situación en la que se produce la petición:

- (i) Camarero, otra, por favor
- (ii) Chico, ¿se te han acabado las cervezas?
- (iii) Reparo en que me está mirando a los ojos y levanto ostentosamente mi jarra vacía
- (iv) Me dirijo a la máquina de discos, selecciono la canción "Que me pongan una copa y otra más" y, mirando a los ojos al camarero, le señalo la máquina con el pulgar.

Todos estos modos de acción pueden ser comprendidos de la misma forma, es decir, puede adscribirseles el mismo contenido-tipo, pero está claro que la forma de explotar la situación para conseguir el resultado buscado han sido muy diferentes. La identidad de la acción tipo *<pedir, cerveza>* se conserva de (i)

a (iv), pero necesitamos algo más para individualizar las cuatro acciones, si lo que pretendemos es explicarlas como eventos causados por representaciones. Necesitamos representar el *modo de acción* así como las relaciones del agente con el escenario de la acción.

Se nos dirá que lo que estamos haciendo es cambiar desde una concepción representacionalista de la acción a una concepción intencionalista de origen griceano. Es cierto, pero esto no es todo, ni tampoco lo más importante, porque el problema de cómo el agente explota la situación para conseguir su propósito, y el problema de cómo un observador potencial explota la situación para conseguir comprender la acción y adscribirle algún sentido no han sido solucionados en la teoría griceana.

Nuestro primer paso, antes de volver sobre cómo un potencial intérprete aplicaría un principio de caridad, ha sido el indicar que nuestro intérprete necesariamente tendrá *que situarse* adecuadamente en el escenario de la acción para conseguir explotar la información contenida en el hecho físico de la acción⁴ porque, simétricamente, el agente sólo llegará a realizar la acción si es capaz de individualizarla correctamente dada la situación, lo que también a él le exige situarse adecuadamente. Entiendo por situarse, el colocarse en una perspectiva acertada, que capacite para explotar la información ofrecida en el escenario de la acción, bien sea para realizarla, bien sea para explicarla.

El paso que hemos dado nos llevará, sin embargo, a una dificultad de más ardua solución. Porque al adoptar una concepción situada de la acción nos hemos comprometido aparentemente con una concepción externalista del contenido de la acción, y uno de los problemas que tiene esta concepción es que amenaza directamente cualquier Principio de Caridad (Bilgrami, 1992, pp. 34-49). Supongamos que Pedro, quien ha olvidado el inglés de su niñez que aprendió en Londres como hijo de un emigrante, decide firmar un manifiesto contra el descuido de Londres basándose en su reciente experiencia de que Londres es una ciudad fea. Sin embargo, su experiencia de la niñez le dice que "London" es una bella ciudad, de manera que cuando le presentan el manifiesto

⁴ En el análisis quineano de la traducción radical encontramos un problema similar: el antropólogo radical deberá comenzar por identificar, entre los patrones de estimulación nerviosa en los que aparecen los aborígenes emitiendo oraciones de caso, cuáles son acciones lingüísticas. Pero Quine también da por resuelto este problema, lo que, si consideramos preferencias como las anteriores, deja de estar claro. Pues la frontera entre cuándo estamos explotando la composicionalidad funcional del sistema lingüístico y cuándo estamos explotando la información del medio no siempre es tan clara como parece.

redactado en inglés no se decide a firmarlo. Bilgrami, 1992, señala que el externalista ortodoxo se enfrenta a un dilema: o considera que Pedro es inconsistente, en cuyo caso va a violar cualquier Principio de Caridad, por mínima que sea su fuerza normativa, o, peor aun, tendrá que escindir el contenido y aceptar que ambas acciones son distintas acciones-tipo. lo que tiene pésimas consecuencias para la capacidad predictiva de cualquier intérprete.⁵ De forma que el primer paso que hemos dado nos induce a tomar alguna forma de partido en la gran división filosófica del siglo: la que enfrenta a externalistas contra internalistas.

Una de las formas de tratar el carácter particular del contenido, una vez que hemos aceptado su naturaleza situada, y, correlativamente, el carácter singular de su adscripción en tercera persona, ha sido el de introducir el "modo de presentación" o su correlato, el "modo de acción", para el caso de la conducta intencional. Se trata de un contenido de carácter singular que es individualizado por el objeto intencional. Se podría pensar que han quedado solucionados todos nuestros problemas permitiendo que existan contenidos particulares, que aparecen como modos de presentación, y suponiendo que, mediante alguna construcción contextualista, podríamos determinar una identificación tipo del contenido. Tal vez mediante una teoría similar a la de Kaplan, añadiendo al contenido de la acción algo así como su carácter, que sería una función de contextos a contenidos. Esta forma de ver las cosas nos permitiría la posibilidad de construir acciones-tipo sobre la base de modos de acción particulares. Esta aproximación ha sido defendida muy convincentemente a mi parecer por Crimmins y Perry, 1989, pero me temo que la discusión ha alcanzado ya aguas más profundas. Por una parte están las insuficiencias del modo de presentación que provienen del campo de la semántica, tal como las ha señalado repetidamente S. Schiffer. Así, al aceptar un modo de presentación, con el objeto de adscribir un estado mental particular a un sujeto, estamos comprometiéndonos con una entidad que tiene un difícil encaje en el papel causal que se exige al contenido para que éste adquiera su función explicativa. Para ello necesitamos incluir el contenido en una red inferencial, que es el espejo de una red causal subyacente, de acuerdo a las teorías racionalistas y funcionalistas más extendidas. Pero hasta el momento esta posibilidad exigía una semántica fuertemente composicional y no está claro cómo hacer compatible la teoría deíctica con la

⁵ De hecho el argumento de Bilgrami es que el externalismo se enfrenta a un trilema: o escinde el contenido, o viola el principio de caridad o implica la falta de autoconocimiento por parte del agente (que traería en el terreno de la acción otra forma de violación del principio de caridad: la admisión de la posibilidad de autoengaño).

composicionalidad clásica.⁶ O dicho de manera inversa: si uno acepta las intuiciones de la teoría del modo de presentación habrá que empezar a poner en cuestión la teoría racionalista del contenido, según la cual el contenido está formado composicionalmente por conceptos. Podemos rechazar la composicionalidad, como hace Schiffer, dado que no resulta consistente sostener simultáneamente la composicionalidad del lenguaje en los contextos de actitud y la teoría del modo de presentación singular.⁷ Pero no podemos renunciar a la productividad y sistematicidad del lenguaje y del sistema conceptual producido por el lenguaje.

Pero tal vez lo que está ocurriendo es que nos hemos tomado demasiado en serio la idea de que hay que aceptar una teoría robusta del contenido como única forma de explicar su carácter causal: los contenidos tienen que ser, en esta concepción, patrones sintácticos individualizables con los instrumentos de la teoría *folk*. Pero también nos cabe la opción de cambiar de tema y contemplar

⁶ No es visible a primera vista esta conexión, sin embargo está bastante clara. Las dos teorías principales sobre el contenido de carácter funcionalista son, por una parte, la teoría del papel conceptual, en la que el contenido está determinado unívocamente por las relaciones internas causales-inferenciales de un estado mental/cerebral y, por otra, la teoría informacional más una estructura interna del tipo del Lenguaje del Pensamiento, en la que tal estructura pone el mecanismo causal y la teoría informacional pone la capacidad representacional. Ambas teorías pretenden salvar al tiempo la relación con el medio y la causalidad sistemática que es necesaria para explicar, por ejemplo, la conducta lingüística. Pero los estados mentales que se individualizan por su relación directa con el medio no pueden mantener la composicionalidad porque tienen encima el viejo problema fregeano de los contextos de actitud. El argumento de Schiffer es el siguiente (Schiffer, 1992). Supongamos la oración

"todo el mundo que ha conocido a Madonna cree que es una gran cantante"

tendríamos un modo de presentación *m* para *Madonna* y un modo de presentación *m'* para *ser una gran cantante*. La oración se representaría cuantificando sobre individuos que creen $\langle m, m' \rangle$, donde el duplo es un modo de presentación. Pero esto es claramente imposible puesto que por su propia naturaleza no hay forma de generalizar sobre los modos de presentación: algunos de los individuos sobre los que generalizamos habrán oído de hecho a Madonna, pero otros simplemente la habrán visto agitarse. No hay forma de construir composicionalmente una oración como la anterior sobre la base de modos de presentación, a menos que construyamos $\langle m, m' \rangle$ como una variable. Pero entonces el problema es de instanciación: ¿cuál sería el referente de la variable $\langle \text{Madonna, gran cantante} \rangle$ bajo modos de presentación que son idiosincráticos?

⁷ La noción misma de composicionalidad no está muy clara: van Gelder, 1990,

el contenido desde una perspectiva menos comprometida con la condición del proyecto cartesiano de una correlación isomórfica entre las relaciones establecidas en el nivel representacional y las operaciones de los mecanismos básicos causales. El proyecto cartesiano comienza a ser socavado en dos direcciones: por un lado, el conexionismo se presenta como una concepción funcionalista no simbólica de la representación; por otro lado, algunos recientes análisis del contenido comienzan a postular posibles componentes no conceptuales del contenido (Peacocke, 1992; Cussins, 1992, 1993; Toribio, 1994; Clark, 1993) como subsistemas cognitivos que permiten al organismo aplicar correctamente un concepto en el sentido habitual del término.

Recordemos que nuestros problemas nacen del hecho de la coincidencia de dos factores: por una parte el carácter situado y circunstancial de la acción, lo que parece que nos lleva a una entidad como el modo de acción singular para analizar el contenido de la acción, por otra parte el hecho de que comprender una acción exige necesariamente una adscripción en tercera persona pues el intérprete no accede al modo de acción.

De la metáfora del ordenador a la metáfora del robot

Antes de acercarnos a nuestra respuesta vamos a dar un rodeo a través de un contexto bien diferente de éste en el que estamos planteando los problemas del contenido. Buena parte de la teoría de la mente y de la semántica ha estado lastrada por la idea de que lo único relevante en la explicación de la conducta debería ser las estructuras de procesamiento de la información (Broncano, 1991; Dennett, 1994) y que el diseño físico del sistema cognitivo no tiene ninguna importancia. Creo sin embargo que del mismo modo que la IA clásica ha servido durante varias décadas como modelo para la teoría de la mente, la construcción de robots está modificando sustancialmente las viejas aproximaciones, al menos tanto como en su día lo hizo la computación simbólica (Dennett, 1994). Mi interés por la relevancia de la robótica no proviene de ninguna fascinación por los juguetes móviles sino por una cuestión teórica del funcionalismo: ¿cuál es la influencia que suponen las constricciones de diseño

distingue la composicionalidad suficiente para explicar la sistematicidad y productividad del lenguaje de la composicionalidad concebida como división de un todo en partes definidas y siempre recuperables (véanse Clark, 1993, cap. 2, y Shastri *et al.*, 1993). Pese a todo, como reconoce A. Clark, 1993, p. 224, la sistematicidad y la conciencia siguen siendo los dos problemas más profundos de la teoría de la mente.

físico de los sistemas motores y sensoriomotores para la arquitectura representacional? Mi respuesta es que la influencia es determinante: el sistema se representará el modo de acción en virtud de cuáles sean sus capacidades físicas para la acción. Cito aquí un párrafo de Dennett, 1994, sobre algunos problemas de diseño del proyecto de robot COG, un proyecto del MIT para construir un robot que interaccione realmente en un ambiente humano:

La mayoría de los problemas a los que se dirige este estadio del proyecto nunca surgirán en el campo de la IA pura descorporeizada. ¿Cuántos motores separados podrán usarse para controlar cada mano? Tendrán que montarse sobre alguna forma de antebrazos, ¿quedará sitio entonces para montar los motores directamente sobre los brazos, junto a las articulaciones que controlan, o estarán situados en el espacio intermedio? ¿Cuánto cableado podrán soportar los brazos antes de ser superados por la fatiga y la falta de destreza? Los brazos han sido construidos para obedecer elásticamente, como nuestras propias articulaciones, esto significa que si COG quiere hacer alguna manipulación muy fina con los dedos tendrá que "quemar" algunos grados de libertad de su movimiento de los brazos apoyando temporalmente el codo o la muñeca en una mesa o en algún sitio conveniente, como nosotros haríamos...

Estos problemas de diseño son correlativos al problema teórico de los modos de acción que acabamos de plantear. Cuando los teóricos tradicionales piensan en la acción, se les olvida sin embargo que la ingeniería de la acción fue el motor selectivo de las capacidades representacionales, y no a la inversa.

Supongamos ahora que estamos diseñando un insecto artificial cuya finalidad sea moverse en un entorno abierto, un entorno que debe descubrir sin poseer mapas previos, de modo que el insecto no tenga la posibilidad de programar su movimiento exclusivamente sobre la base de su memoria; por añadido, el entorno es impredecible hasta un cierto grado, aunque presenta elementos prominentes que el sistema cognitivo aprovechará para organizar el movimiento. Aunque más tarde volveré sobre el diseño interior de este cacharro, me interesa ahora traerlo a colación simplemente como un interesante problema de contenido perceptual y de coordinación con el medio para desarrollar este contenido. Supongamos que nuestro sistema es capaz de alguna forma de aprendizaje que no se realiza en un marco simbólico clásico de Inteligencia Artificial. Por ejemplo, dispone de una red neuronal interna y de algún algoritmo genético que genera variaciones en los pesos de conexión entre las unidades. La dinámica del sistema es tal que, a medida que el insecto aprende a moverse manteniendo el equilibrio y a sortear obstáculos, comienzan a estabilizarse ciertos patrones de conducta que, por supuesto, corresponden a patrones internos de la red neuronal, pero que sin embargo no permiten, desde

el punto de vista externo, asegurar cuáles son los elementos del medio a los que la red se ha sensibilizado. Pueden ser los mismos que los que atraen nuestra atención pero, mucho más verosíblemente, dependen de los elementos a los que reaccionan sus sistemas sensoriomotores, que no hay que imaginar que sean los mismos que los nuestros. El resultado es algo muy parecido a un movimiento explorador del medio inteligente. Obsérvese que no hay programas innatos que ligen la conducta del sistema a estímulos fijos: el sistema aprende cuáles son los estímulos relevantes. Para no convertir este ejemplo en un caso de ciencia ficción, quiero recordar que estos insectos ya circulan por los suelos de los laboratorios desde hace algunos años (Brooks, 1991; Maes, 1990). Como diseñadores, nos importa prioritariamente que registre obstáculos, genere conductas para sortearlos, mantenga el equilibrio, etc., y, si es suficientemente plástico, desarrolle conductas estratégicas como las de aproximarse a un objetivo y encontrar el camino de vuelta a su punto de origen (Gallistel, 1990, caps. 3 y 4). Nuestro insecto no tiene por qué ser demasiado inteligente, pero al menos debe ser tan inteligente como una hormiga que explora el árido suelo de Castilla en agosto y es capaz de encontrar de nuevo el refugio húmedo de su hormiguero. Lo reseñable es que la conducta del insecto muestra patrones de conducta, patrones que podemos determinar asignándole un contenido funcional a sus estados internos, aunque seamos incapaces de encontrar una estructura causal interna como lo sería por ejemplo un mapa. No es imposible, sin embargo, que el insecto haya generado un mapa interno aprovechando algunas propiedades del medio que no conocemos todavía, o por lo menos que no conoceremos sin una investigación empírica sistemática (Clark, 1993).

Para complicar nuestro ejemplo, supongamos que construimos otro insecto tal que su objetivo es coordinar su acción con el primer insecto de forma que persigan algún fin conjunto, como pudiera ser encontrar y transportar algún objeto o, al contrario, algún fin contrapuesto, como es que uno de ellos persiga al otro que, a su vez, tiene que escapar. Nuestro escenario se compone ahora de un entorno abierto y de dos pequeños robots que se mueven a través de él. No es imposible describir la conducta de uno y otro mediante nuestro vocabulario sobre percepciones espaciales y de movimiento, quizá les asignemos mapas internos (sabiendo que un mapa es un isomorfismo que preserva las relaciones geométricas) y algún plan a cada uno como pudiera ser "*mantén la distancia > N sobre el perseguidor*" o "*alcánzale y captúrale*". Sin embargo me parece innecesario atribuir a cada insecto una especie de "principio de caridad" idiosincrático fuerte, como sería el que cada uno de ellos asignase al otro los mismos mapas que él mismo es capaz de generar. Bastaría, por el contrario, con que cada uno de ellos fuera capaz de reaccionar adecuadamente a los patrones de conducta del otro, independientemente de cuáles hayan sido

sus causas internas. Pero, esto es lo importante, para conseguir el éxito de la coordinación de conductas es necesario que cada uno sea sensible a, o que sintonice con, los patrones de la conducta ajena.

Si nos situamos ahora como jueces externos al juego de la presa y el depredador, suponiendo que sabemos que no había mapas previos, tenemos que inferir que los insectos reaccionan sistemáticamente a las configuraciones que van presentando sus órganos sensoriales. Podemos, pues, atribuirles la habilidad de explotar las propiedades situacionales del medio sin necesidad de ningún principio *a priori*, aunque como observadores de otra especie no sepamos exactamente cuáles sean las propiedades a las que está reaccionando. Pero lo importante es que no hay ninguna necesidad de suponer que sus estados internos están constituidos por un sistema de símbolos en el sentido en el que la perspectiva clásica representacional lo considera. Basta con que la conducta sea sistemática: no es necesaria la hipótesis fuerte de que sus contenidos internos explotan la información simbólicamente porque, entre otras razones, hemos de suponer que las redes neuronales internas obran simplemente como sistemas plásticos que adaptan su dinámica a las configuraciones estables del medio. Reparemos también en que ambos sistemas muestran una cierta capacidad de predecirse mutuamente, lo que no implica, por lo que sabemos de cómo funcionan las redes neuronales, que la estructura de las unidades ocultas, y por consiguiente de la red, sea la misma. Cada ser podría estar reaccionando a propiedades diferentes, con tal que su cognición fuera la suficiente para soportar la capacidad predictiva que su tarea les exige.

De cada uno de estos insectos podemos decir que tienen alguna forma primitiva de representación sensoria.⁸ La cuestión es si, observando a nuestros insectos, podríamos atribuirles alguna forma de contenidos-tipo, es decir, contenidos perceptuales que impliquen un reconocimiento de los casos bajo la identidad de tipo. Mi hipótesis es que, dadas estas condiciones, si tuviésemos que asignar una forma de contenido perceptual a ambos sistemas, tendríamos que incluir algo así como un equivalente del modo de presentación mas una condición normativa de corrección que permite que ese modo de presentación se inserte adecuadamente en el escenario. Seguramente necesitaríamos además algún aparato interno generalizador para construir categorías de eventos. Desde

⁸ Repárese en que podemos afirmar tal cosa de los robots que hemos construido, pero no de insectos reales. La diferencia ha sido bien establecida por Dretske: nuestros insectos artificiales han adquirido por aprendizaje la capacidad de indicar alguna cosa en el medio, mientras que ésta es una capacidad que no sabemos todavía hasta dónde llega en los insectos.

luego una semántica composicional, funcional de verdad nos lo permitiría, pero me parece milagroso que emergiese un sistema tan estructurado como es un aparato cuantificacional en programas tan simples como los que controlan a los robots. Sería suficiente, sin embargo, con que pudiesen asociar los varios modos de presentación en cuasicategorías conceptuales que obedeciesen a patrones estables. Por ejemplo, <zanja, atravesar, otro lado>. Estos patrones pueden nacer de procesos de estabilización de conductas que refieren a propiedades del entorno y que se muestran en acciones estables como saltar, rodear... etc. (Cussins, 1992).

El contenido de la acción y la dicotomía externo-interno

El debate y el ejemplo anterior nos llevan a poner entre paréntesis las aproximaciones a la explicación de la conducta que siguen estos tres dogmas:

(1) Solamente pueden atribuirse contenidos a mentes (sistemas) de naturaleza simbólica, en la que las propiedades semánticas estén representadas por propiedades sintácticas de los símbolos (*dogma simbólico*).

(2) El contenido tiene un aspecto interno (*narrow*) responsable de la producción causal de la acción y un aspecto externo (*wide*) responsable de las propiedades semánticas (funcionales de verdad del contenido) (*dogma dual*).

(3) Atribuir contenidos es atribuir relaciones (actitudes) a entidades lingüísticas, sean sentencias o proposiciones (*dogma conceptualista*).

Quienes, como Fodor, tienen una teoría hiperrealista del contenido y aceptan a pie juntillas los tres dogmas, no necesitan principio de caridad, porque de hecho la confluencia de los tres dogmas es lo que el teórico atribucionalista quiere obtener con tal principio y, simétricamente, el escéptico o instrumentalista respecto de los contenidos puede conseguir, mediante el PC, todo lo que el realista ya tiene para explicar la conducta inteligente; lo que ocurre es que, de hecho, ha terminado comprometiéndose en la práctica con los mismos dogmas que el hiperrealista, porque se ha obligado a aceptar, como una consecuencia de su posición, que solamente los seres dotados de lenguaje tienen creencias y deseos, y viceversa.

No tengo ahora ni el espacio, ni el tiempo, ni quizá la osadía de echar abajo los

tres dogmas. Estoy convencido sin embargo de que otros autores y, sobre todo, la investigación científica futura se va a encargar de hacerlo, pero en tanto se cumple esta esperanza, me atrevo a proponer tres versiones muy restringidas de los dogmas:

(1) Sólo algunos contenidos son simbólicos, a saber, aquellos que incluyen relaciones metarrepresentacionales y semánticas dentro del propio contenido.

Dos palabras de aclaración. El lenguaje natural forma parte de una familia de sistemas representacionales que permiten autorreferencia y por consiguiente la capacidad de representar la naturaleza simbólica. Pero éste es un hecho contingente acerca de cómo se han configurado nuestro lenguaje y nuestras capacidades cognitivas, no una condición definitoria de qué sea un sistema representacional que sea capaz de actuar en función del contenido representacional.

(2) El contenido es una entidad caracterizada por varias dimensiones, la primera está constituida por las estructuras dinámicas internas responsables de la causalidad, la segunda por las relaciones semánticas referenciales con el medio, para el caso de los contenidos conceptuales, y la tercera por las relaciones sistemáticas de orden práctico que se expresan en forma de destrezas y habilidades.

(3) El contenido conceptual, tal como es caracterizado por las actitudes proposicionales, es el contenido que tienen algunos sistemas capaces de conducta lingüística, pero no agota la naturaleza del contenido.

Pensemos en las personas con trastornos lingüísticos como los que aparecen en las afasias, en las personas autistas o en los niños. A menos que cometamos el provincianismo científico de caracterizar estos sistemas como personas defectuosas (éste es el caso de Davidson, 1982), debemos encontrar una teoría restringida del contenido, si realmente queremos que este término denote un concepto realmente universalizable.

La idea que estamos insinuando es la siguiente: *es necesario un espacio cognitivo de patrones de conocimiento práctico para asignar contenido a las acciones*. Y la existencia de este espacio plantea la posibilidad de una transición continua entre el contenido de carácter simbólico, que

individualizamos lingüística y conceptualmente, y los patrones puramente físicos que forman parte del medio externo o interno. Los patrones de conocimiento práctico se asientan en la dinámica de los estados del cerebro, pero solamente en la medida en que ese cerebro se sitúa en un cuerpo determinado.

Claro que, una vez que comenzamos a deslizarnos por esta pendiente, la dicotomía entre lo externo y lo interno comienza a ser puesta inmediatamente en cuestión. Si externo/interno es la frontera que divide lo simbólico de lo que no lo es, como probablemente establece la concepción funcionalista, no escaparemos a cierto grado de circularidad, porque simbólicos serían aquellos estados internos que pueden causar la conducta del sistema en virtud de la capacidad representacional que portan, y los estados internos serían causales en virtud de su naturaleza simbólica. Pero si la dicotomía adopta una posición menos comprometida, como podría ser un criterio físico, de piel para afuera o para adentro, digamos, entonces tendríamos razones para pensar que se trata de una frontera muy mal trazada, al menos en lo que respecta a las acciones, dado que el conjunto de procesos que exigen una dinámica de colaboración continua entre el medio y el sistema ponen en cuestión gravemente la dicotomía. Pensemos, por citar un caso, el problema que plantean los instrumentos de acción, algo que compartimos con numerosas especies animales ¿son externos o internos para la producción de la acción?, ¿no es más bien un criterio previo, ajeno a la teoría de la acción, el que nos hace trazar la frontera a uno u otro lado?

Sospecho que una vez que abandonamos el proyecto fregeano de una noción trascendental de los conceptos, la dicotomía entre lo conceptual y lo no conceptual y la no menos consagrada dicotomía entre lo externo y lo interno, las dos grandes herencias del cartesianismo mental, aparecen atadas en su destino: con certeza caerán o se sostendrán juntas. Lo que nos deja abierta una vía a partir de lo que podríamos llamar contenidos corporeizados (que corresponden más o menos a lo que Toribio, 1994, llama contenidos ecológicos). Comprendo que hablar de corporeización de los contenidos no es algo que atraiga demasiadas simpatías en los medios de la filosofía analítica pues es conocido que ha sido un insistente tema de la tradición fenomenológica que se remonta a Merleau-Ponty y al propio Heidegger. Peor aun, en los años recientes se ha convertido en un punto central de extraños puntos de vista que mezclan la teoría de la mente con discusiones contraculturales y visiones místicas (Varela *et al.*, 1992), lo que no obsta que varios desarrollos de la más reciente Inteligencia Artificial, de la ingeniería robótica y de la propia neurofisiología sintonicen con algunas de estas reivindicaciones (Winograd y Flores, 1986).

Volvamos de nuevo a nuestro anterior ejemplo del insecto artificial que se mueve en un entorno abierto. Un entorno abierto de esta clase es una fuente continua de información en la que, al no existir mapas previos, no existen lo que

podríamos llamar categorías dadas *a priori* (en este caso podrían ser, por ejemplo, el conjunto de caminos y direcciones posibles en el espacio del entorno). Pensemos, para hacer una analogía, en un entorno parecido al que encontraría una persona de visión normal a quien le impidiésemos la visión y le situásemos en un lugar desconocido y no familiar (Blanco *et al.*, 1992).

Un mapa puede funcionar aquí como una analogía de un concepto, y viceversa: contiene todas las trayectorias posibles, que se activan en el momento en el que el sujeto es capaz de situarse en el mapa del mismo modo que el concepto contiene objetivamente un poder inferencial que se activa en el momento en el que lo situamos en una proposición. Explotando la analogía, la capacidad de situarse en el entorno sería el análogo del reconocimiento para el concepto. Pero hemos convenido en que ni el insecto ni la persona en la oscuridad están provistos de ningún mapa. Desde este punto de partida se puede escribir una historia acerca de cómo los conceptos son construidos sobre una base de interacciones con el medio, y cómo su aplicación posiblemente exija la activación de las actividades de esquemas no conceptuales resultantes de esta activación. No pretendo que la historia sea verdadera, pero sí que sea una aproximación verosímil al problema de la naturaleza del contenido.

Nuestro insecto comienza a moverse, a chocar y a desarrollar correlativamente en su red neuronal ciertos puntos estables de anclaje que funcionan como representaciones de referencias prominentes del medio. Al menos ésta es nuestra descripción externa, pero pensemos en que el insecto internamente dispone de una red neuronal y no de un programa de IA tradicional. La red neuronal es en esencia un sistema dinámico en el que se establecen múltiples activaciones de grupos de unidades que, una a una, no tienen papel funcional describible como símbolos puesto que las capacidades representacionales nacen de la red entera. Si observásemos los disparos de unidades en la red con algún tipo de sensores, nuestra observación sería el de un sistema caótico parecido, por ejemplo, al que muestra el movimiento de la gente en un gran centro comercial a una hora pico. No encontramos patrones estables, nada que pudiera ser confundido con un símbolo, pero si, por ejemplo, anuncian nuevas rebajas en varias plantas, el movimiento caótico se modifica instantáneamente (Freeman, 1991; Skarda y Freeman, 1988; Foss, 1992). Estos cambios cualitativos sensibles a las conductas de estímulo son los que identificamos funcionalmente como patrones conductuales de respuestas.⁹

⁹ La idea central es que necesitamos una descripción "macroscópica" para encontrar un orden subyacente en la actividad cerebral. En el paradigma simbólico podíamos descender hasta el nivel de los símbolos almacenados, pero no es eso lo que

Pero aun si no encontramos patrones observables, la reacción continua al medio nos indica un grado de orden interno que emerge de la dinámica del sistema. Lo importante es que la dinámica no puede ser entendida solamente como una función del estímulo, pero tampoco puede ser entendida sin la colaboración del estímulo externo. Supongamos, para ejemplificar esta observación, que la dinámica de la red está dirigida por alguno de los llamados algoritmos genéticos, que son sistemas de variaciones (de pesos entre conexiones, por ejemplo) seleccionadas por algo externo hasta que emerge un patrón estable que ya no varía más que lentamente (Dennett, 1991); en nuestro caso el mecanismo selector podría ser la percepción del medio a través del movimiento. Los algoritmos genéticos crean en cierto modo orden a partir del caos inicial. Su método consiste en dejar que la naturaleza sea un componente intrínseco del cálculo. Las constricciones en la dinámica del sistema las pone la realidad. En nuestro insecto, por ejemplo, el estímulo proximal podrían ser constricciones sobre el control del movimiento.¹⁰ Si el insecto genera un mapa, en el sentido de que cualquiera de nosotros le pudiera adscribir una entidad de esa categoría, lo habrá conseguido como un resultado secundario y quizás epifenoménico de los patrones de conexión, de los atractores que rigen la dinámica de su red interna (Smolensky, 1996), de los módulos de control sensoriomotor (Brooks, 1991), y de otros posibles componentes que sostienen las relaciones estables con objetos del medio que al insecto le resultan prominentes pero que no tienen por qué ser los que nos resultan a nosotros. Más arriba propusimos el caso de que un insecto tuviese que predecir la conducta de otro para conseguir alcanzarlo. ¿Hay alguna restricción que obligue a que los mapas que tienen ambos sean idénticos? Sabemos que en ambos insectos es necesaria la capacidad de moverse en el espacio y que ambos deben tener un motor predictivo de la conducta del otro, pero no es necesario que ambos mapas sean el mismo, al menos no lo es más que en un sentido retorcido, v. g., en el sentido de que podríamos construir un isomorfismo de uno en otro. Pues bien, supongamos que es posible construir estos insectos: habríamos alcanzado un nivel de contenido que se asienta sobre

ocurre en un sistema como el cerebro: aquí sólo encontramos atractores caóticos que establecen patrones cualitativos de reacción en respuesta a los estímulos.

¹⁰ Uno de los algoritmos genéticos que ha tenido una cierta importancia filosófica fue el desarrollado por Robert Axelrod y Stephanie Forrest para encontrar una salida adaptativa al dilema del prisionero; el algoritmo desarrollaba variaciones de estrategias, hasta que emergió la cooperativa como solución estable. Holland, J. H., 1992; Brooks, R. A., 1991, se ha convertido en uno de los difusores de la importancia de la robótica para la nueva IA. Véase Broncano, 1991.

habilidades circunstanciales, pero que es suficiente para una mutua coordinación de conductas.

Insisto en que esta ficción no tiene por qué tomarse como una historia realista de la ontogenia de los conceptos, aunque varias pistas empíricas hacen algo más que plausible la moraleja para los seres de nuestra especie.

Los navegantes de la isla Truck y la destreza situacional

Con el objeto de hacer posible este puente hacia las habilidades de orden superior, traeré otro ejemplo de naturaleza muy distinta. Proviene de observaciones etnográficas sobre los métodos de navegación de los aborígenes de las islas de la Micronesia, en especial de la isla Truck. Estas observaciones han recibido cierta atención no solamente por parte de los antropólogos, como cabría esperar, sino también de los psicólogos e incluso de los teóricos de la IA (Hutchins, 1983; Suchman, 1986).

Los habitantes de estas islas han navegado tradicionalmente entre las numerosas islas del archipiélago. Los nativos no solamente no disponen de mapas sino que ni siquiera tienen una noción aproximada de qué pueda ser tal cosa y para qué pueda servir. Con sus pequeñas canoas provistas de una vela recorren a veces distancias en mar abierto de 150 millas sin desorientarse. Para conseguirlo, tienen que identificar y mantener una dirección constante en mar abierto a través de las constantes bordadas a las que obliga la navegación a vela y, en las proximidades del objetivo, tienen que ser capaces de identificarlo con precisión para no pasar de largo. Los navegantes micronesios se guían por las estrellas, por el color de las aguas del mar, por la presencia de aves, la dirección del viento y cualesquiera otras fuentes de información disponibles. El problema mayor es decidir cuándo y cómo cambiar de bordada en el zigzaguo contra el viento, ya que un error puede irse multiplicando e impedir que se alcance el objetivo.

Son muy interesantes las descripciones orales de los métodos de navegación hechas por los pilotos a causa de las dificultades que entraña la interpretación de las heurísticas seguidas. En primer término está la dificultad general para verbalizar cualquier destreza, una dificultad que tiene toda persona y que ha recibido suficiente atención en las filosofías de Wittgenstein y Polanyi. Saber hacer no significa saber decir lo que se sabe hacer, e interpretar las órdenes verbales entraña una dificultad mucho más ardua que lo que parece, como cualquiera puede comprobar tomando en sus manos el primer manual de instrucciones que encuentre. Sólo los expertos, que no necesitan manual, son capaces de interpretar el manual. Pero, en segundo término, está el problema

añadido de la traducción de los términos del lenguaje. Varios estudiosos han discutido cómo interpretar los informes de acción de los pilotos en primera persona. Por ejemplo, los navegantes no entienden los mapas que los expertos usan para preguntarles por sus métodos de navegación. Usan el término intraducible ETAK para referirse a la orientación por estimación. Al parecer conciben que su canoa permanece inmóvil y es el paisaje el que se mueve a través de ella. ETAK es el término para orientar adecuadamente la proa en cada momento del escenario del viaje. No tienen reglas, en el sentido de normas permanentes de navegación, sino numerosos recuerdos que permiten la orientación. Por supuesto, su gasto computacional es enorme para el problema de la navegación; se trata de una habilidad de un dominio muy específico, pero lo interesante es que nos permite estudiar una habilidad no formalizable que recuerda la forma en que los robots anteriores estabilizaban sus rutas para guiarse en un entorno abierto.

Las dificultades de traducción nos recuerdan de nuevo la distancia entre interpretar y explicar, sin que todavía hayamos entrado siquiera en el peligroso territorio de la discusión entre razones y causas. El obstáculo que encontramos para explicar la conducta de los navegantes micronesios es, no obstante, de un carácter diferente a esta cuestión. No es una dificultad lingüística, al menos no en el sentido de Quine. El término ETAK refiere al parecer al rumbo seguido en relación con una tercera isla intermedia entre la salida y el objetivo que puede ser un punto imaginario en el horizonte que permanece fijo en el horizonte y que se localiza con respecto al movimiento de las configuraciones estelares. Las dificultades de traducción del término no son quineanas en el sentido de la opacidad de la referencia en un sentido habitual. Lo prueba el que los hablantes bilingües sean incapaces de explicar el concepto que refiere a habilidades que aprendieron en el mar a medida que fueron haciéndose navegantes diestros sin tener la noción de mapa o dirección en el espacio, que permitiría la traducción.

Observemos que enfrentamos aquí una cierta dificultad para el uso del principio de caridad como restricción en la explicación de grano fino de la conducta: el conceder a los nativos creencias y objetivos similares a los nuestros no nos garantiza que exista un trasfondo común de habilidades que permita dar contenido común a los conceptos, y mucho menos, y por la misma razón, situarnos en el mismo escenario desde una perspectiva compartida o compartible. La dificultad surge al menos en la medida en que el Principio de Caridad funcione exclusivamente como una constricción entre conceptos fregeanos o neofregeanos. Nuestros ejemplos plantean además ineludibles preguntas colaterales sobre la autocomprensión y el autoconocimiento, ya que, desde el punto de vista que estamos adoptando, las dificultades de comprensión de la conducta de los nativos son dificultades del mismo orden que las que ellos mismos tienen

para explicar sus métodos.

No está de más insistir otra vez sobre la inexpresabilidad del conocimiento práctico, una característica que pusieron de moda hace décadas varios autores que derivaron la filosofía wittgensteiniana hacia la imposibilidad de fundamentación empírica del conocimiento. Me refiero a Polanyi, Kuhn y Hanson, entre otros muchos que todos conocemos. Sin embargo, no se trata de una inaccesibilidad como la que se postulaba en la ciencia cognitiva clásica, expresada a través de la diferencia entre competencia y actuación, en tanto que niveles diferentes de computación.¹¹ No se trata de que el conocimiento práctico consista en subrutinas simbólicas encapsuladas y no conscientes que sirven de puente entre los procesos conceptuales y los mecanismos motores. Nuestra posición proviene de un origen doble: por un lado la presión que ejerce el carácter situado de las acciones y por otro lado el posible trasfondo no conceptual del contenido de la acción. Hemos afirmado que el carácter circunstancial conlleva una continua colaboración del medio para determinar la identidad del contenido (donde el acento se pone en la continuidad): puede que el sistema sea sensible a ciertas oportunidades informativas previamente a la acción, pero es necesaria una colaboración con la situación para elaborar el contenido intencional de la acción que se va a producir inmediatamente. Por otra parte nuestra discusión de los mapas espaciales del insecto artificial nos llevaba a la conclusión de que esta colaboración no puede entenderse simplemente como una especie de ocasión disparadora de mecanismos preexistentes, sino como un medio en el que pueden generarse esos mecanismos, explotando las identidades y diferencias que el organismo encuentra en su relación con la situación.

Sin embargo no hemos llegado aún a la región en la que el Principio de Caridad se convierte en una precondition de comprensión y de atribución de contenido a las acciones ajenas, puesto que hasta el momento solamente hemos hablado de acciones y comprensión de acciones que no implican la detección de intenciones como parte, bien del contenido productor de la acción, bien de la comprensión de la acción.

La detección de intenciones en el medio, o el caso del Dr. Spock

La clase de habilidades que nos ocupa es la habilidad de atribuir contenido

¹¹ Vera y Simon, 1993, responden con este argumento a los defensores de las implicaciones de la concepción situada de la acción (Greeno y Moore, 1993; Agre, 1993; Lave, 1988).

y comprender las acciones de los otros seres como seres dotados de intenciones. Es un tipo de habilidad cognitiva que, como toda capacidad cognitiva, consiste esencialmente en la explotación de la información disponible en el medio, pero es un tipo especial de habilidad determinada por un tipo de habilidades básicas más básicas que denominaremos *habilidades sociales*. Algunos piensan que la única o más importante de estas capacidades es el lenguaje, pero no es correcto pensar así. La actuación lingüística, cuando está situada y se convierte en proferencia, explota muchísima más información que la puramente lingüística o la conceptual. Me parece que es algo más que una metáfora considerar una conversación como un caso similar al de la navegación de los nativos micronesios. Y a la inversa, seguramente obtendríamos mucha información acerca de cómo actúan las personas ante contextos de incertidumbre observando cómo lo hacen en contextos en los que son especialmente hábiles como es el desarrollo de una conversación. Una conversación casual exige unas capacidades de explotar la situación que no pueden reducirse a la mera explotación de implicaturas conceptuales. Una conversación se parece mucho más al territorio de caza de nuestros insectos artificiales. En ella aparecen obstáculos naturales, malentendidos y sobreentendidos, paisajes llanos y abiertos en los que es fácil moverse y territorios sinuosos en los que hay que moverse con cuidado; forman parte de la conversación los ademanes, gestos y silencios, tanto o más que las proferencias; hay estrategias y tácticas, trampas y corrección parsimoniosa de errores. Romper una conversación es fácil, pero no lo es llevarla a buen puerto. Y llevar a cabo una conversación no es lo mismo que realizar un plan rígido, por el contrario, el éxito resulta de un número indefinido de acciones que exigen la cooperación de los contertulios para explotar los numerosos recursos que la conversación les ofrece en el lapso de tiempo y en la ubicación en que se desarrolla. Una conversación es probablemente uno de los contextos que ponen más a prueba nuestras mejores capacidades cognitivas. Y, como ocurre en la navegación, hay buenos y malos conversadores, lo que nada tiene que ver con que sean malos o buenos hablantes. Definitivamente, una conversación es un buen ejemplo de lo que hemos denominado *habilidad social*.

Las habilidades son capacidades o disposiciones internas para realizar tareas con un cierto grado de éxito. Se diferencian de las funciones en que son adquiridas y en que son contingentes: su realización no implica ninguna operación vital para el organismo. No se dice de alguien que tiene la habilidad de respirar, aunque sí podríamos decir de una cantante de ópera que tiene la habilidad de "respirar con el estómago". Las habilidades sociales son habilidades en este sentido: dependen de la contingencia de la cultura en la que el organismo realiza su aprendizaje, y por lo mismo son sensibles a variaciones culturales, aunque existan regularidades transculturales en el espacio y en el

tiempo. Así, mientras la capacidad lingüística no es seguramente una habilidad social, sino más bien una capacidad de especie, sí lo es la capacidad comunicativa o el uso del lenguaje con propósitos comunicativos. Decimos que algo es una habilidad social en dos sentidos: en el primero, si su adquisición y posesión exige la existencia de una comunidad; en el segundo sentido, si es una capacidad para modificar las relaciones sociales, como son, por ejemplo, las relaciones de jerarquía en un grupo, las relaciones de agresividad, etc.

Son muchas las especies animales que tienen estructura social. La existencia de jerarquías sociales impone restricciones muy importantes sobre la conducta de los individuos, que está obligada al mantenimiento de estas relaciones sociales. Aparecen de este modo las conductas cuya función es enviar una señal fiable al grupo, o las conductas dirigidas a mostrar información engañosa a otras especies; por ejemplo, algunos pájaros se hacen los heridos para despistar al depredador de su nido. En los primates la estructura social parece tener una importancia cognitiva muy superior a todas las demás especies: la mayor parte de los problemas que debe resolver el sistema cognitivo de un chimpancé a lo largo del día provienen del hecho de que el chimpancé vive en grupos en los que su supervivencia depende de relaciones sociales que establecen las jerarquías en el grupo (Humphrey, 1976; Chance y Mead, 1988; Cheney y Seyfarth, 1990). En todas estas especies puede haber acciones cuyo objetivo o sentido sea el modificar la conducta del otro. Esto es lo que ocurre con las muestras de amenaza, de sumisión etc. Las pautas de acción, en todos estos casos, pueden ser o bien resultado de factores innatos o resultado de aprendizaje condicionado, como lo prueban las habilidades de los animales en régimen de entrenamiento artificial. Son pautas de conducta que tienden a manipular la conducta de otros sin detectar intenciones. Basta con alguna forma específica de aprendizaje que reaccione condicionadamente a las formas de la conducta externa por la pura forma, no por el contenido que manifiesta. Es lo que hace el perro cuando nos ve tomar el collar del lugar donde se cuelga habitualmente. El cerebro del perro explota la forma perceptual de la conducta en la situación, no nuestra intención, aunque ésta se manifieste en una forma específica de conducta en esa misma situación. El tipo de habilidad que nos interesa es, sin embargo, específicamente humana, se trata de la manipulación de las conductas ajenas manipulando el contenido de sus mentes. Sólo en este momento podemos decir específicamente que estamos ante acciones sociales genuinas. Es lo que ocurre, por citar dos casos, en el engaño o en la conducta auténticamente cooperativa. Ambos tipos de acción son imposibles si el objetivo de la acción no es la transformación del contenido mental del otro (Cheney y Seyfarth, 1990).

Cuando un ser es capaz de una conducta de engaño genuina ya podemos afirmar sin discusión que tiene habilidad social. Pero, ¿qué es engañar?

Consiste en realizar una acción que conduzca a que otro crea algo que es falso. No obstante esta definición no capta un elemento fundamental: la existencia de una positiva voluntad de engaño, es decir, el engañador actúa con la intención de que el otro genere un contenido mental que es falso. Necesitamos, para utilizar la terminología de Dennett, 1987, una intencionalidad de varios órdenes de profundidad.

(i) Quiero que [X crea que [las galletas están en la caja]], [que es falso]

(ii) Mirando ostentosamente a la caja conseguiré que X crea que [las galletas están en la caja]

(iii) Miro ostentosamente a la caja asegurándome previamente de que X me está mirando

(i) Muestra al menos una intencionalidad de segundo orden, es decir, muestra creencias acerca de creencias, pero obsérvese que añadimos la cláusula “que es falso”, para indicar un nuevo elemento metarrepresentacional: el hecho de que representamos las relaciones entre las creencias y la realidad y asignamos un valor de verdad. Es discutible si este segundo factor es necesario o no para asignar intencionalidad estricta de primer orden. Para los simpatizantes de un concepto lingüístico de la mente el tener contenidos exige tener ya valores de verdad, luego el elemento metarrepresentacional estaría implícito, a pesar de ello creo que en la conducta de engaño, se exige una formulación explícita del elemento metarrepresentacional. En cualquier caso, una conducta genuina de engaño exige la representación intencional explícita de cambiar el contenido mental, y no solamente la forma externa de la conducta del otro. Si la conducta de engaño muestra claramente la naturaleza social de la acción, en el sentido que estamos indicando, parecería menos claro que la capacidad de inteligir las acciones ajenas *qua* acciones intencionales exija un tipo de habilidad similar a la que se muestra en el engaño. Sin embargo me parece que así es: para comprender la conducta ajena como conducta intencional, se necesita al menos la misma capacidad que para engañar al otro, es decir, se necesita captar las intenciones del otro a través de su acción.

Hay una discusión reciente acerca de cómo interpretar esta habilidad; para unos, como Stich y Nichols, 1992; Gopnik, 1993, y muchos otros, los sujetos intencionales desarrollarían una teoría de la mente para predecir la conducta de los otros igual que usamos otro tipo de teorías; para otros como Gordon, 1986,

y Goldman, 1989, los sujetos usarían su propia mente como simuladores de la conducta ajena. En el primer caso, el sujeto actúa como científico, en nuestro caso como psicólogo natural, interpretando eventos observables y explicándolos mediante conceptos teóricos, cuales son las actitudes proposicionales; en el segundo actúa como una especie de artesano que utiliza su propia mente como instrumento de cálculo predictivo de la conducta ajena, aun sin saber qué elementos entran en juego. La discusión acerca de la naturaleza de esta habilidad social es importante para el estatuto de la psicología popular, y afecta de forma derivada a cuál sea el estatuto del Principio de Caridad, pero es una discusión que debe ser resuelta en el marco de la investigación empírica, por lo que afortunadamente podemos suspender temporalmente el juicio sobre una u otra posición, al menos en lo que respecta al tema que discutimos. Lo interesante son los numerosos estudios empíricos que se han realizado sobre el desarrollo de esta capacidad tanto en el hombre como en primates (Whiten, 1991). Estos estudios nos permiten apoyar la idea de que el dominio del juego de atribuciones es una capacidad cognitiva paralela al lenguaje, a pesar de que recibe un refuerzo determinante con el desarrollo del lenguaje. Los niños usan términos que denotan estados internos antes de ser capaces de explicar las conductas ajenas mediante el contenido de sus deseos y creencias; inversamente, son capaces de reaccionar e identificar conductas que expresan estados mentales en estadios prelingüísticos. También sabemos que los niños con síndrome de autismo no llegan jamás a dominar la capacidad atribucional, a pesar de que en algunos casos lleguen a dominar el lenguaje (Butterworth *et al.*, 1991). Puede que esta facultad sea producto de sistemas innatos que se desarrollan secuencialmente en estadios, puede que no, que sea producto del aprendizaje social en un medio inteligente; puede, por último, que haya elementos innatos inespecíficos y que el medio social determine el desarrollo de la facultad específica de atribuir estados mentales. Lo que a nosotros nos importa es que la capacidad atribucional parece mostrar una cierta arquitectura evolutiva en estadios de progresivo dominio en la explicación de las conductas ajenas. Así, la reacción a los estados emotivos de otras personas es previa a cualquier distinción de estados mentales con contenido conceptual. El dominio del lenguaje de deseos y protointenciones también parece previo al dominio de la atribución de deseos y creencias, el dominio de la capacidad metarrepresentacional (Perner, 1990); por último, parece ser necesaria antes de que el niño pueda explicar la conducta ajena atribuyéndole creencias que quizá sean falsas, desde el punto de vista del intérprete. Quizás estos estadios están determinados por la filogénesis de nuestra especie o quizá solamente influyan aspectos de la ontogénesis de las capacidades cognitivas individuales. Sea cual sea la posición verdadera, el Principio de Racionalidad en la Interpretación es un resultado

fáctico de constricciones sobre las conductas de los seres que pueblan el medio ambiente, en un sentido similar a cómo el espacio tridimensional es un subproducto de cómo está constituido nuestro sistema perceptivo (Marr, 1982). Y del mismo modo que nuestro sistema perceptivo aprovecha relaciones topológicas estables del medio ambiente, las constricciones que denominamos PRI son posibles porque existe, de hecho, una cierta comunidad constitutiva en un numeroso grupo de conductas que pueblan el medio ambiente. De manera que podríamos avanzar este principio fiabilista para considerar el PRI en un contexto naturalista:

Dados los sujetos A y B, el hecho de que A y B compartan R patrones de interacción mutua explica la atribución del estado M que hace A a B como explicación de la conducta C incluida en R.

El desarrollo de especie o de individuo en una cultura va fijando o estableciendo un conjunto indeterminado de patrones de interacción entre individuos, que se mantienen como patrones porque han adquirido una función específica, la de contribuir al mantenimiento y preservación de la estructura social de interacciones. El hecho de que esos mecanismos sean adecuados para este objetivo explica su reforzamiento a lo largo del tiempo y explica, posteriormente, que los individuos sintonicen con ellos y desarrollen mecanismos predictivos basados en ellos (Dennett, 1991).

Repárese en que estos mecanismos predictivos, una vez que se desarrolla el lenguaje, tienen un doble aspecto: por una parte el contenido que captura e identifica el lenguaje en contextos abiertos, pero, en segundo lugar, tienen una compleja arquitectura como es la que se refleja en las actitudes encajadas, como quiero que creas que quiero que... La posibilidad lógica del juego del encaje de actitudes está fundamentada en la realidad psicológica o etológica de una gradación anterior de habilidades sociales que nacieron de la explotación de ciertas conductas significativas, previas al lenguaje, antes que en el producto de principios lógicos internos. El hecho de que nuestros ancestros primates vivieran en un medio ambiente pleno de regularidades sobre las que se asentaban estas habilidades posibilitó el desarrollo de una arquitectura estable que permite la atribución múltiple.

Si esta historia tiene alguna verosimilitud podríamos avanzar una hipótesis sobre la fuerza nómica del PRI:

Dada una comunidad S de sujetos, el hecho de que todos los individuos compartan un conjunto H de habilidades de interacción social hace fiable un principio P predictivo de las conductas de los individuos.

La clase H de habilidades que resultan del desarrollo evolutivo es algo que

debe descubrirse empíricamente. No sabemos muy bien cuáles pueden haber sido los estadios en los que se han ido estabilizando ciertos patrones de conducta, pero sí podemos afirmar que el PRI se construye fragmentariamente sobre estos estadios. Resulta así que no existe una barrera nítida entre lo que sería la explicación y manipulación de la conducta ajena en virtud de sus contenidos y lo que podemos considerar meramente como un aprendizaje condicionado de alguna complejidad.

A pesar de que no sabemos muy bien cuáles pueden haber sido estas fases de estabilización, propondré algunas a modo de ejemplos de lo que sería una construcción fragmentaria:

(1) La expresión de emociones es, sabemos, desde Darwin, una regularidad muy estable entre los mamíferos. Las emociones no se reducen, sin embargo, a su expresión conductual, sino que tienen una rica fenomenología y, lo que es más importante, alguna funcionalidad de carácter cognitivo, como es la de detectar rápidamente, sin usar el sistema conceptual o la memoria a largo plazo, posibles objetivos o dificultades en el desarrollo de la acción, y disparar consecuentemente un abanico de posibles planes de acción. Un estado emotivo no determina un contenido en el mismo sentido en que lo hace el lenguaje. Ahora bien, dada la expresión de tal estado, es posible que se haya generado un campo social de sintonía afectiva, y que probablemente la existencia de este campo explique evolutivamente la misma existencia de expresión emotiva, como un mecanismo de cohesión de grupo. Si ocurrió que en las manadas de primates se generó este mecanismo de cohesión social, es plausible que surgiera un espacio de identificaciones y atribuciones, previo al contenido, en el sentido conceptual y lingüístico, sobre el que se construyeron prácticas atribucionales exitosas. De hecho sabemos que en los primates superiores el control emocional del otro es un mecanismo básico de mantenimiento del grupo frente a presiones del individuo (Goodall, 1989). Sabemos también que los niños de muy pocos meses adecuan su conducta a la expresión afectiva de la madre (Harris, 1990).

(2) Algunos autores (Gómez, 1991) han resaltado la importancia de algunas habilidades prelingüísticas, entre las que destacan los mecanismos de atención visual conjunta. Por ejemplo, sabemos que el ser observado es una restricción que influye de manera determinante en los mamíferos: todos los mamíferos son sensibles a las simetrías verticales del medio, lo que es explicable si suponemos una especial sensibilidad a la regularidad tipo, "ser observado por unos ojos" (si los depredadores de los mamíferos hubieran tenido otro diseño físico, esta sensibilidad informacional no existiría), pues bien, la atención visual conjunta se convirtió en un instrumento fundamental de evolución entre los primates.

(3) La presencia de mecanismos estables de *contrato social*, como los establecidos por la relación *pagar el costo/obtener el beneficio de la acción de grupo*, puede considerarse también un elemento primigenio y desencadenante de otras habilidades sociales. Este mecanismo pudo haber surgido como una forma de protección del grupo. Según algunas observaciones empíricas (Cosmides, 1990; Cosmides y Tooby, 1991) los sujetos muestran sesgos cuando se los enfrenta a la tarea de resolver la validez de una inferencia de *modus ponens*, si en esta inferencia está involucrada la relación de pagar el costo/obtener el beneficio del grupo. El sesgo existiría en todas las inferencias en las que esté implícito el objetivo de detectar el que engaña al grupo. Según Leda Cosmides se trata de un mecanismo innato, aunque bien pudiera ser simplemente una regla implícita aprendida. La cuestión central es si han estabilizado las regularidades grupales que hacen que los individuos sintonicen con esta regla.

No importa que esta lista sea completa, ni siquiera que sea adecuada, lo importante es que sea plausible la existencia de una historia previa, contingente, de regularidades protosociales que fueron conformando la arquitectura especial de nuestros hábitos de atribución. Si realmente se dio esta construcción fragmentaria de habilidades de atribución, resultarían dos consecuencias difícilmente evitables. La primera sería la posibilidad de que existieran diversos Principios de Racionalidad en la Interpretación, en la medida en que el PRI afirma que "en condiciones ideales, el sujeto reacciona racionalmente" relativamente a nuestras capacidades de predicción de su conducta. Pensemos en el caso del Dr. Spock de *Star Trek*, quien está constituido de forma aparentemente similar a nosotros, salvo por el hecho de que no tiene emociones. Ante la conducta de un sistema como el Dr. Spock, el PRI no puede funcionar del mismo modo que ante alguien que es como nosotros. Podemos racionalizar su conducta, pero tenemos que hacer un rodeo para sustituir su falta de emociones por algún otro mecanismo cognitivo de similares funciones, con lo que resultará en una historia causal incluida en nuestra explicación muy diferente a la nuestra.

Lo grave no es solamente que hayamos perdido universalidad y bordeemos el relativismo, sino que la segunda consecuencia es la de poner en cuestión la fuerza normativa del principio que establece que un sistema tiene contenidos si y sólo si es racional, dado que la atribución de racionalidad estaría determinada por hechos constitutivos de la estructura cognitiva de los seres que interpretamos. ¿Qué nos queda entonces? ¿Acaso no hemos tirado al bebé con el agua de la bañera?. Desde mi punto de vista quedaría todavía un margen sustancioso para la aplicación normativa del PRI por cuanto afirma (1) que la comunidad de hecho dispone de ciertos sistemas básicos de habilidades, sin los cuales sería imposible la atribución y (2) que la existencia de estos patrones hace fiables

nuestras prácticas predictivas.

Sin embargo, un principio relativizado de este tipo no es una patente de corso para hacer psicología sin perder el tiempo en realizar observaciones empíricas. Por el contrario, nos obliga a analizar los estadios de desarrollo de la psicología *folk*, porque su fuerza es sensible a su origen histórico. El PRI tiene tanta fuerza como la fiabilidad de nuestras predicciones, lo que no nos produce mucho consuelo, pero al menos nos permite saber dónde estamos.

BIBLIOGRAFIA

- Agre, P. (1993), "The Symbolic Worldview", *Cognitive Science*, 17, pp. 61-69.
- Bennett, J. (1991), "How is the Cognitive Ethology Possible?", en C. Ristau (comp.), *Cognitive Ethology. The Minds of Other Animals*, Hillsdale, NJ, Lawrence Erlbaum.
- Bilgrami, A. (1992), *Belief and Meaning. The Unity and Locality of Content*, Oxford, Blackwell.
- Blanco, F., R. Colom, R. López y E. Leo, "Ceguera y desplazamiento mental en una isla ficticia: del ojo de la mente a la mano de la mente", *Actas del IV Seminario Interuniversitario de Ciencia Cognitiva*, Sagunto, 1991.
- Bratman, M. (1992), "Shared Cooperative Activity", *The Philosophical Review*, 101, pp. 327-341.
- Broncano, F. (1991), "I Robot", *Actas del II International Colloquium on Cognitive Science*, San Sebastián, 1991.
- Brooks, R. (1991), "Intelligence Without Reason", *Twelfth International Conference on Artificial Intelligence*, Morgan Kaufmann Publishers.
- Butterworth, G., P. Harris, A. Leslie y H. Wellman, comps. (1991), *Perspectives on the Child Theory of Mind*, Oxford, Oxford University Press.
- Chance, R. M. y A. P. Mead (1988), "Social Behavior and Primate Evolution", en R. Byrne y A. Whiter, comps., *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes and Humans*, Oxford, Clarendon (original de 1953).
- Cheney, D. y R. Seyfarth (1990), *How Monkeys See the World. Inside the Mind of Another Species*, Chicago, Chicago University Press.
- Cheney, D. y R. M. Seyfarth (1991), "Truth and Deception in Animal Communication", en C. A. Ristau, comp., *Cognitive Ethology. The Minds of Other Animals*, Hillsdale, NJ, Lawrence Erlbaum Publishers.
- Clark, A. (1993), *Associative Engines. Connectionism, Concepts and Representational Change*, Cambridge, MA, MIT Press.

- Cosmides, L. (1989), "The Logic of Social Exchange: Has Natural Selection Shaped How Human Reason?", *Cognition* 31, pp. 187-276.
- Cosmides, L. y J. Tooby (1991), "From Evolution to Behavior", en Dupre, E. (comp.), *The Latest on the Best*, MIT Press, Cambridge, Mass.
- Crimmins, M. y J. Perry (1989), "The Prince and the Phone Booth: Reporting Puzzling Beliefs", *The Journal of Philosophy*, LXXXVI, pp. 685-711.
- Cussins, A. (1993), "Non Conceptual Content and the Elimination of Misconceived Composites", *Mind and Language*, 8, pp. 234-252.
- Cussins, A. (1992), "Content, Embodiment and Objectivity: The Theory of Cognitive Trails", *Mind*, 101, pp. 651-688.
- Davidson, D. (1982), "Rational Animals", *Dialéctica*, 36, pp. 318-327.
- Dennett, D. (1987), "Los sistemas intencionales en la etología cognitiva: defensa del 'Paradigma Panglossiano', en *La actitud intencional*, Barcelona, Gedisa.
- Dennett, D. (1994) "The Practical Requirements for Making a Conscious Robot", inf. manus. del Center for Cognitive Studies of Tufts University.
- Dennett, D. (1991), "Real Patterns", *The Journal of Philosophy*, 89, pp. 27-51.
- Evans, G. (1982), *The Varieties of Reference*, Oxford, Oxford University Press.
- Foss, J. (1992), "Introduction to the Epistemology of the Brain: Indeterminacy, Micro-specificity Chaos, and Openness", *Topoi*, 11, pp. 45-57.
- Freeman, W. (1991), "Fisiología de la percepción", *Investigación y Ciencia*, abril, pp. 30-38.
- Gallistel, C. (1990), *The Organization of Learning*, Cambridge, MA, MIT Press.
- Goldman, A. (1989), "Interpretation Psicologized", *Mind and Language*, 4, pp. 161-185.
- Gómez, Juan C. (1991), "Visual Behavior as a Window for Reading the Mind of Other in Primates", en A. Whiten, comp. (1991).
- Goodall, J. (1989), *En la senda del hombre*, Barcelona, Salvat (original de 1967).
- Gopnik, A. (1993), "How We Know our Minds: the Illusion of First-Person Knowledge of Intentionality", *Behavioral and Brain Sciences*, 16, pp. 1-15.
- Gordon, R. M., "Folk Psychology as Simulation", *Mind and Language*, 1, pp. 158-171.
- Greeno, J. G. y J. L. Moore (1993), "Situativity and Symbols", *Cognitive Science*, 17, pp.49-59.
- Harris, P. (1989), *Los niños y las emociones*, Madrid, Alianza.
- Holland, J. H., "Algoritmos genéticos", *Investigación y Ciencia*, sep. 1992.

- Holland, J. H. (1992), *Adaptation in Natural and Artificial Systems*, MIT Press.
- Humphrey, N. K. (1976), "The Social Function of Intellect", en P. G. Bateson y R. A. Hinde, comps., *Growing Points in Ethology*, Cambridge, Cambridge University Press.
- Hutchins, E. (1983), "Understanding Micronesian Navigation", en D. Gentner y A. L. Stevens, *Mental Models*, Hillsdale, Lawrence Erlbaum Publishers.
- Israel, D., J. Perry y S. Tutiya (1993), "Executions, Motivations and Accomplishments", *The Philosophical Review*, 102, pp. 545-589.
- Jolly, A. (1991), "Conscious Chimpanzees?", en A. Ristau (comp.).
- Karmiloff-Smith, A. (1992), *Beyond Modularity: a Developmental Perspective on Cognitive Science*, Cambridge, MA, MIT Press.
- Lave, J. (1988), *Cognition in Practice. Mind, Mathematics and Culture in Everyday Life*, Cambridge UK, Cambridge University Press.
- Maes, P. (comp.) (1990), *Designing Autonomous Agents: Theory and Practice from Biology to Engineering and Back*, Cambridge, Mass., MIT/Elsevier.
- Marr, D., *La visión*, Madrid, Alianza.
- Perner, J., *Understanding the Representational Mind*, Cambridge, MA, MIT Press.
- Perry, J y D. Israel (1991), "Fodor and Psychological Explanations", en B. Loewer y G. Rey (comp.), *Meaning in Mind. Fodor and its Critics*, Oxford, Blackwell.
- Perry, J. (1986), "Circumstantial Attitudes and Benevolent Cognition", en J. Butterfield (comp.) *Language, Mind and Logic*, Cambridge, Cambridge University Press.
- Ristau, C. (comp.), *Cognitive Ethology. The Minds of Other Animals*, Hillsdale, NJ, Lawrence Erlbaum.
- Schiffer, S. (1992), "Belief Ascription", *The Journal of Philosophy*, 89, pp. 490-521.
- Shastri, L. y V. Aijjanagadde (1993), "From Simple Associations to Systematic Reasoning: a Connectionist Representation of Rules, Variables and Dynamic Bindings Using Temporal Synchrony", *Behavioral and Brain Sciences*, 16, pp. 417-494.
- Skarda, C. y W. Freeman (1987), "How Brains Make Chaos in Order to Make Sense of the World", *Behavioral and Brain Sciences*, 19, pp. 161-195.
- Smolensky, P. (1986), "Information Processing in Dynamical Systems", en D. Rumelhart et al., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. I, Cambridge Mass, MIT Press.
- Stich, S. y S. Nichols (1992), "Folk Psychology: Simulation or Tacit Theory",

- Mind and Language*, 7, pp. 35-71.
- Suchman, L. (1986), *Plans and Situated Actions*, Cambridge, UK, Cambridge University Press.
- Toribio, J. (1994), "From Narrow Content to Ecological Content" (próxima publicación).
- Vaina, L. (1987), "Visual Texture for Representation", en Vaina, L. (comp.), *Matters of Intelligence*, Dordrecht, Reidel.
- van Gelder, T. (1990), "Compositionality: a Connectionist Variation of a Classical Theme", *Cognitive Science*, 14, pp. 335-384.
- Varela, F., E. Thomson y E. Rosch (1992), *De cuerpo presente. Las ciencias cognitivas y la experiencia humana*, Barcelona, Gedisa.
- Vera, A. H. y H. A. Simon (1993), "Situated Action: a Symbolic Interpretation" *Cognitive Science*, 17, pp. 7-48.
- Whiten, A. (comp.) (1991), *Natural Theories of Mind. Evolution, Development and Simulation of Everyday Mindreading*, Oxford, Blackwell.
- Winograd, T. y F. Flores (1986), *Understanding Computers and Cognition*, Reading, MA, Addison Wesley.

DEPARTAMENTO DE FILOSOFÍA Y LÓGICA Y FILOSOFÍA DE LA CIENCIA - UNIVERSIDAD DE SALAMANCA

ABSTRACT

In this paper, a naturalistic version of the Charity Principle as an *a priori* constraint in the interpretation of actions is proposed. Given its normative and *a priori* character, the classical CP from Quine, Davidson and Dennett, meet some difficulties to be considered as an empirical constraint. We argue that some troubles could be removed if we change the underlying notion of content ascribed to intentional actions. We present, in first time, the situated character of actions, that involves an analog to the mode-of-presentation problem of indexicals. The necessary relation to a context drives us to the controversy about externalism and internalism. Our proposal is that a non-conceptual dimension of content grounded on the skills of agent is a necessary condition to ascribe content to the intentional action. From this view, we propose to consider the Charity Principle as an historical constraint on the evolution of some classes of cognitive systems, i. e., systems with a rich social environment that become a main source of cognitive problems and hence constrain the possible trajectories of cognitive development.