

Responsabilidad semántica y naturalismo

JOSEFA TORIBIO MATEAS

Universidad de Sussex

INTRODUCCIÓN

La principal tesis de este trabajo es que los estados mentales con contenido (las creencias, en particular) se presentan en dos variedades diferentes. La primera variedad —a la que me refiero con el término ‘no-responsable’ (*mindless*)— requiere poco más que la codificación de información encaminada hacia una acción o respuesta específica. La segunda variedad de creencias—presentada aquí bajo el rótulo de ‘responsable’ (*mindful*)— incluye aquellos estados mentales susceptibles de constituirse en objeto de reflexión crítica. Si un sistema cognitivo tiene la habilidad de reconocer críticamente una de sus creencias como (in)apropiada dada la evidencia disponible y/o las circunstancias relevantes, tal estado mental pertenece a la variedad ‘responsable’ (*mindful*).

La distinción entre creencias susceptibles de entrar en procesos argumentativos o revisiones racionales y creencias que no son susceptibles de entrar en esos procesos es central en ciertas discusiones sobre justificación epistémica. Una creencia está adecuadamente justificada, se argumenta algunas veces (e.g. Chisholm, 1982), sólo si está sujeta a razones y/o evidencia empírica, o (en algunas posiciones más complejas, pero también más cercanas a la tesis que defiendo en este artículo) si el sujeto es capaz de reflexionar de manera crítica sobre tal creencia (cf. Kim, 1994). La tesis principal de este artículo es, sin embargo, que la capacidad de asumir ciertas responsabilidades doxásticas (especialmente responsabilidades que incluyen reflexión crítica) es igualmente central en la determinación de la *naturaleza* (y no sólo la justificación) del contenido de los estados mentales, i.e., que la disponibilidad de un esta-

do mental para entrar en procesos de corrección racional es crucial en la *individuación* del propio estado mental. No prestar atención a este factor normativo en la individuación y caracterización de ciertos estados mentales tiene como consecuencia, concluiré, la imposibilidad de evaluar correctamente algunos de los denominados enfoques naturalistas sobre el contenido. La razón es que los retos que los defensores de tales enfoques tienen que enfrentar son diferentes en los casos de creencia ‘responsable’ y creencia ‘no-responsable’.

I. CREENCIAS RESPONSABLES

En el mundo natural, existe una división razonablemente definida entre sistemas que son capaces de examinar críticamente sus propios estados mentales y sistemas que —aunque poseedores, sin duda, de estados mentales con contenido semántico— los aceptan siempre sin ulterior análisis. Una rata, por ejemplo, puede aprender a ser precavida y (tras algunos refuerzos negativos) responder a los estímulos de su entorno de forma distinta. No parece probable, sin embargo (aunque quizá no imposible), que estos animales sean capaces de exhibir un tipo de conducta que pueda describirse como de reflexión crítica sobre la relación entre sus creencias y la evidencia que las sustenta. La tesis que pretendo defender en este trabajo es que este hecho marca una diferencia esencial si no en la *justificabilidad* de las creencias¹, sí ciertamente en su *naturalidad*. Las creencias susceptibles de evaluación crítica son *conceptualmente* especiales y constituyen una *clase cognitiva* en sí misma. Ignorar esta clase especial de estados mentales crea confusiones y problemas innecesarios, especialmente en lo que respecta a la plausibilidad de ciertos proyectos de naturalización de lo mental.

La observación es, así pues, que no todos los estados mentales que pueden caracterizarse como estados de creencia tienen un contenido con respecto al cual el sistema cognitivo guarda una cierta relación, a saber, una relación en la que el contenido es, en sí mismo, objeto potencial de reflexión crítica. Lo que intentaré defender es que el papel conceptual de los estados mentales que exhiben esta susceptibilidad de auto-crítica es tan esencialmente diferente del papel conceptual de los estados mentales cerrados a ella que resulta intelectualmente fructífero considerar los estados mentales como pertenecientes a dos *clases cognitivas* diferentes.

1 La verdad del punto de vista que se conoce como *responsabilismo epistémico* es independiente de mi tesis. Esta doctrina se utiliza aquí simplemente como objeto de una cierta analogía estructural.

La distinción no es nueva. Haugeland (1998) marca una diferencia similar a través de una parábola sobre familias y permutaciones. Haugeland nos invita a imaginar una familia de seres humanos en la que cada miembro de la familia se define a través de un conjunto de características, como, por ejemplo, apariencia física, olor, tono de voz, etc. y en la que, un buen día, ciertas permutaciones empiezan a ocurrir:

La hermana suena como el padre, se mueve como la abuela y huele como el hermano pequeño. Incluso los diferentes miembros corporales se mezclan: la cabeza de la madre (pero el pelo del padre) sobre el torso del tío con los brazos del bebé —o sólo dos cabezas, y no brazos o torso (con una voz de camionero y un olor a sandía). Y momentos más tarde, cambian de nuevo, con nuevas divisiones y nuevos participantes. ¿Qué diría Ud.?

(Haugeland, 1998, p. 261)

No sé Ud., pero yo diría que alguien le ha echado un ácido a la casera. Y, en esencia, ese es precisamente el mensaje que Haugeland quiere transmitir. Cualquier ser humano, sugiere Haugeland, tenderá muy probablemente a rechazar lo que parece percibir en estas circunstancias porque tales creencias perceptivas están sujetas —y no se adecuan— a criterios mínimos de escrutinio racional. Pero, ¿que pasaría —se sigue preguntando Haugeland— si nos centramos en el perro de la familia al que consideramos presumiblemente no-afectado?:

Creo que el perro ladraría ... aparecería desorientado y confuso, quizá incluso asustado. Pero no puedo imaginar, por más que lo intente, que esta reacción del perro pueda entenderse como un rechazo de la escena o una desestimación de su realidad basándose en que es imposible que eso ocurra.

(Haugeland, 1998, p. 262)

Lo que Haugeland pretende enseñarnos con la parábola es que la ausencia de tales reacciones críticas muestra que el perro «no entiende nada» (*op. cit.*, p. 262). El perro no ve las cosas como *objetos*. Sólomente posee lo que Haugeland denomina «habilidades mundanas», pero carece de «habilidades constitutivas». Las habilidades mundanas son clases básicas de saber-como (*know-how*), tales como correr, pasear en bicicleta o coger una pelota. Las habilidades constitutivas se centran en la ca-

pacidad para juzgar la aceptabilidad o legalidad de ciertas acciones, tales como reconocer un fuera de juego en fútbol (o, en el ejemplo de Haugeland, el orden apropiado de batear en beisbol. *Op. cit.*, pp. 323-324). Son, en esencia, habilidades de reconocimiento de lo legal y lo racional y en tanto tales sustentan clases diferentes de *política cognitiva*, tanto al nivel personal, como al nivel de la comunidad a la que el sujeto pertenece. Sin tal *política cognitiva*, Haugeland afirma, es estrictamente inapropiado hablar de comprensión, de pensamiento, o —si el sistema cognitivo dispone de un lenguaje— del significado de ninguna de sus expresiones.

Tal conclusión, sin embargo, parece demasiado radical. Parece más plausible permitir que un sistema no sometido a tal noción de *política cognitiva* (e.g., el perro de la familia permutante) pueda considerarse, no obstante, sujeto de estados mentales dirigidos a sí mismo y a su entorno, *pero* que esos estados mentales sean simplemente de una *clase* diferente a los nuestros. Esa es mi propuesta. Consideren, por ejemplo, el caso del denominado Síndrome de Capgras. Los sujetos que padecen esta enfermedad creen que su pareja ha sido reemplazada por una copia perceptualmente indistinguible (a menudo, un robot). El paciente aparentemente no adopta (de hecho, no puede adoptar) la estancia crítica que Haugeland favorece. Sin embargo, resulta altamente implausible suponer que tales pacientes no tienen pensamiento alguno sobre sus parejas o que no existe nada que pueda llamarse genuinamente ‘comprensión’².

Los movimientos dialécticos que se podrían introducir en este punto son, sin duda, muchos. Pero en lugar de sumergirnos en esta controversia de carácter local, me parece más fructífero proponer en su lugar una divi-

2 Una posible explicación del síndrome es que el enfermo de Capgras carece (debido a una disfunción neurológica) de una respuesta afectiva normal hacia la otra persona y que racionaliza el desajuste entre su falta de afecto y la idea de que debería sentirlo interponiendo la figura de un impostor (cf. Ellis y Young, 1990). Stone y Young (1997) sostienen, por otra parte, que el enfermo con Síndrome de Capgras exhibe una respuesta *racional* a lo que puede considerarse una experiencia perceptiva anómala. Stone y Young argumentan que el proceso de formación de creencias está suficientemente bien estructurado como para tener como resultado un estado genuino de creencia: «[el enfermo de Capgras] realmente cree las cosas que dice creer» (Stone y Young, 1997, p. 334).

La cuestión sobre si el proceso de fijación de creencias que resulta en una creencia ilusoria es en algún sentido un proceso racional es, así pues, una cuestión compleja y ampliamente debatida, pero el contraste con la posición de Haugeland es, no obstante, significativo. Lo que parece claro, en cualquier caso, es que (por razones desconocidas) estos enfermos son incapaces de llevar a cabo la clase de actividad crítica de revisión de creencias que Haugeland introduce en su parábola de la familia permutante.

sión en el espacio de estados mentales. La división se basa en la siguiente caracterización de lo que podríamos denominar *responsabilidad semántica*: un sujeto S cree responsablemente que P si y sólo si las dos condiciones siguientes se cumplen: i) S es un sistema cognitivo que funciona correctamente; y ii) se podría esperar razonablemente de S que cumpla sus responsabilidades críticas con respecto a P. Cumplir con las responsabilidades críticas con respecto a una creencia, en este contexto, es equivalente a apreciar los diferentes compromisos que un sujeto asume cuando sustenta o rechaza una creencia y estar dispuesto a exhibir una actividad de sopeso de las razones y/o evidencia relativas a esa creencia.

Para perfilar más detalladamente esta noción de responsabilidad semántica, examinemos en primer lugar un caso en que la primera de sus condiciones no se cumple, un caso de 'deficiencia cognitiva'. Tomemos, por ejemplo, a la persona que sufre el síndrome de Capgras y cree, falsamente, que su pareja es un robot. Resulta plausible considerar que tal sujeto no ha de ser juzgado responsable (o, si prefieren una terminología legal, *culpable*) por la falsedad de su creencia porque al menos una de las dos condiciones para creer responsablemente no se cumple: el sujeto no es un sistema cognitivo que funciona de forma adecuada. En un caso como este, podemos hablar de creencias *patológicas*, i.e., creencias que son el resultado de un malfuncionamiento físico y de las que, por tanto, el sujeto no es responsable a nivel semántico.

Por supuesto, el concepto de deficiencia cognitiva no es un concepto preciso. Alston (1988) describe un caso interesante en el que un alumno de filosofía no puede entender correctamente algunas de las tesis de Locke. El caso se introduce con el afán de ilustrar la tesis de que la noción de responsabilidad *epistémica* es insuficiente para caracterizar la noción de justificación *epistémica*³. Tal y como Alston presenta la situación, la interpretación es que el alumno no puede evitar el creer algo falso. En particular, el alumno cree (falsamente) que:

(L) según Locke, todo es una cuestión de opinión.

En lugar de creer (correctamente) que:

(K) según Locke, el conocimiento de una persona se restringe a sus ideas.

Alston insiste en que no hay nada que el alumno pueda hacer para mejorar su comprensión de los textos de Locke. El alumno « ... no tiene

3 Cf. Steup, 1988, pp. 78-81. Allí se encontrará una discusión crítica de este ejemplo desde el punto de vista epistemológico.

lo que fuera que sea necesario tener para seguir una exposición o razonamiento filosófico de carácter abstracto» (*op. cit.*, p. 287). Este pasaje sugiere una interpretación del caso en términos de una (débil) deficiencia cognitiva, una interpretación que lo convierte en un caso similar al del paciente de Capgras. Consideren, sin embargo, una lectura diferente de la situación. Imagínen, por ejemplo (como suele ser el caso), que el alumno tiene la capacidad para identificar su creencia como un error pero que (por razones desconocidas) simplemente nunca pone en práctica tal capacidad. En este caso podemos afirmar que en un sentido muy concreto —que se explicará en la sección siguiente— el alumno *debería saber qué hacer*, i.e., debería ser capaz de reconocer su error. Siendo así, tiene sentido representar al alumno dentro del espacio de responsabilidad del que estamos hablando y esperar que sea capaz de *aprender* a través de un discurso razonado. Las creencias del alumno, en este escenario, pertenecerían a la categoría de creencias *responsables*. Esta categoría abarca todos aquellos tipos de contenido mental capaz de participar en lo que algunas veces se denomina el «espacio de razones» (Sellars, 1963). Tales estados están sujetos a un cierto «ideal constitutivo de racionalidad» (cf. Davidson, 1970, p. 98). Lo que venimos llamando ‘creencias responsables’ son así las creencias que más genuinamente pertenecen al ámbito semántico, puesto que cuandoquiera que sustentamos este tipo de creencias, nos revelamos como participantes en una variedad de prácticas de carácter fuertemente normativo. Las creencias responsables representan, como intentaré mostrar en la última sección de este trabajo, el caso más duro para un tratamiento naturalista del contenido.

Hasta ahora he introducido casos de creencias *patológicas* —i.e., casos en los que los sujetos son sistemas cognitivos que no funcionan adecuadamente— y casos de creencias *responsables* (*mindful beliefs*). A continuación paso a introducir una tercera categoría, la categoría de creencia no-responsable (*mindless belief*). Estas creencias son sostenidas por sujetos que, aunque funcionan de manera adecuada, no cumplen la segunda de las condiciones constitutivas de lo que hemos denominado *responsabilidad semántica*. Muchas de nuestras creencias perceptivas pertenecen a esta categoría. Piénsen, por ejemplo, en la representación mental que forma una persona al salir de su casa, en virtud de la cual cree que hay tres escalones en frente de la puerta. Los movimientos del cuerpo de esa persona se ajustan para adecuarse a las propiedades de la representación. Sin embargo, en el *cuarto* escalón, tropieza y se cae. Una cierta ilusión perceptiva ha tenido lugar. El entorno le ha jugado una pasada (había cuatro escalones, no tres). Esta es la clase de situación en la que no tiene sentido hablar del sujeto como *responsable* de la ac-

ción desencadenada por el contenido de su creencia. El sujeto no es (semánticamente) responsable de tal creencia porque no se podría haber esperado razonablemente de él que examinara críticamente su representación mental, i.e., la creencia que desencadena su caída.

La distinción entre creencia responsable y no-responsable es probablemente cercana a la (no muy apreciada) distinción propuesta por Daniel Dennett entre creencias y opiniones⁴. Dennett sigue a de Sousa en una caracterización de las creencias como fenómenos básicos y «menos intelectuales» que el fenómeno involucrado en el asentimiento a una proposición. Éste último se presenta como una actividad más sofisticada que se da en situaciones como, por ejemplo, decidirse a hacer una cosa o cambiar de opinión. Asentir a una proposición, desde el punto de vista de Dennett, implica tener una cierta actitud hacia un enunciado que se representa mentalmente. El estado más básico de creer tal o cual cosa, por otra parte, está constituido —de nuevo, de acuerdo con Dennett— por el patrón no verbal de acciones al que la criatura está predispuesta (cf. Dennett, 1978, p. 307). Dennett utiliza así de forma técnica los términos ‘opinión’ y ‘creencia’ para referirse a estas dos clases de estados cognitivos. Las opiniones están «infectadas verbalmente ... [son] estados del usuario de un lenguaje», mientras que las creencias son «estados más básicos que compartimos con ranas, perros y demás (cf. Dennett, 1987, p. 233). Las dos clases de estados cognitivos son tan profundamente diferentes que deberíamos —sugiere Dennett— «simplemente examinar algunas de nuestras intuiciones [sobre la noción de creencia] como pertenecientes realmente a una categoría diferente de estado mental: el estado de asentir, i.e., el estado de opinión, no de creencia» (Dennett, 1978, p. 304).

La propuesta de Dennett es así cercana, tanto en motivación, como en ejecución, a la nuestra. Pero existen dos diferencias notables. La primera —y menos importante— es que, a diferencia de Dennett, aquí se caracteriza lo que Dennett denomina ‘estado de opinión’ como un cierto *tipo* de creencia. Una caracterización diferente significaría alejarse demasiado del uso normal del término⁵. La segunda diferencia —más importante— es que la distinción propuesta aquí entre creencia responsable y no-responsable sólo se relaciona *contingentemente* con la capacidad para usar un lenguaje y, ciertamente, con la capacidad para repre-

4 Cf. Dennett (1978). En ese trabajo Dennett reconoce la influencia de R. de Sousa (1971).

5 Esta puede ser, de hecho, la razón por la que la distinción de Dennett nunca ha echado raíces en la literatura sobre estos temas.

sentar mentalmente enunciados en ese lenguaje. Es posible que la capacidad para tratar las creencias propias (y las de los demás) como objetos de evaluación crítica se facilite y consolide a través del uso de un lenguaje —y de ciertas prácticas sociales. Pero no existe una conexión *conceptual* entre estas dos esferas. Es perfectamente posible que un sistema que no disponga de un lenguaje público sepa cuándo y cómo tratar algunos de sus propios estados mentales como estados con un contenido erróneo y, por tanto, perfectamente posible que tales sistemas puedan revisar esos estados de manera apropiada. Resulta, por supuesto, notoria e inevitablemente difícil expresar en palabras exactamente cómo se habría de desarrollar tal proceso. Sin embargo, no es del todo imposible imaginar, por ejemplo, una cierta clase de sistemas dotados de una *inteligencia arquitectónica* que *piensa* en imágenes sobre posibles diseños de construcción de edificios y es capaz, en ciertos momentos cruciales, de re-examinar críticamente aspectos del plan que emerge en el diseño para comprobar que, e.g., ciertos requisitos indispensables se han seguido correctamente. El ejemplo es ciertamente sofisticado, pero la intuición general que lo sustenta no lo es: la intuición es que Dennett tiene toda la razón cuando distingue dos categorías distintas de estados mentales que se conocen normalmente bajo el nombre de ‘creencia’, pero está equivocado en explicar la diferencia entre esas dos categorías en términos de «infección lingüística». La división entre estas dos categorías está determinada por una profunda diferencia en los patrones de compromiso normativo y en la sensibilidad crítica con relación a la cual el estado mental se caracteriza. El uso de un lenguaje puede contribuir a extender el dominio de estos estados *hambrientos de normatividad*, pero de ninguna manera *determina* su naturaleza.

2. SABER CUANDO PREOCUPARSE

Las creencias que vengo denominando *responsables* son susceptibles de revisión a través de procesos distintivos de evaluación crítica. Esta posibilidad de evaluación racional es la que les permite jugar un papel cognitivo especial. Están constituídas de tal manera que pueden participar en cadenas inferenciales y —cuando se expresan de forma lingüística— en episodios de argumentación interpersonal. Lo que conocemos como ‘ciencia’ es probablemente la versión más clara de esta clase especial de economía cognitiva.

Ahora bien, una duda importante emerge casi de inmediato. La marca distintiva de las creencias responsables —tal y como se han descrito aquí— es que, especialmente en el caso de creencias erróneas, hemos de

considerar al sujeto *responsable* de su error y confiar en que re-examine tales creencias de forma crítica. Ante esta expectativa, podría parecer que consideramos que el sujeto puede ser cognitivamente *culpable* con respecto a todas sus creencias (responsables) erróneas. Sin embargo, como Tidman (1996) ha señalado en su análisis de la noción análoga de justificación epistémica, no sería razonable esperar de un sujeto que evalúe críticamente todas sus creencias de forma continua⁶. De la misma manera, sería ilusorio imaginar que los sujetos evalúan críticamente todas sus creencias *responsables*, i.e., todas aquellas creencias que podrían, en principio, ser evaluadas de tal manera. Pero, si esto es así, ¿que es exactamente lo que, de acuerdo con la posición defendida en este trabajo se requiere del sujeto que posee creencias responsables? Intentaré mostrar que lo que se requiere es simplemente la posesión de (en términos de Haugeland) una habilidad mundana más: la habilidad mundana de *saber cuándo hay que preocuparse*, i.e., saber (en el sentido de saber-cómo) cuándo un cierto proceso de reflexión es necesario. Las creencias responsables están todas constituídas, desde mi punto de vista, por la presencia de una capacidad bien definida de saber cuándo hay que reflexionar sobre ellas, una capacidad que (a diferencia de lo que Haugeland piensa) no supone más que el desarrollo de una clase adicional de habilidades, —una habilidad que, ciertamente, se nutre de forma especial del entorno lingüístico y social que rodea a la especie humana, pero que no lo requiere *necesariamente*.

Para empezar la justificación de esta tesis, consideremos una vez más algunos ejemplos. Imaginemos un sujeto que está mirando una vara (recta) a través de un prisma, lo que hace que la vara aparezca ligeramente curvada. Resulta interesante observar que, bajo tales circunstancias, si se toca la vara, también se percibirá de forma curvada al tacto⁷. Imaginemos ahora tres sujetos enfrentados a esta situación, i.e., los tres sujetos perciben la curvatura de la vara tanto cuando miran a través del prisma como cuando la tocan. El primer sujeto es un niño pequeño. El segundo es un adulto, pero no un experto en cuestiones de percepción u óptica. El tercero es un especialista en optometría. Ahora imaginemos, finalmente, que se somete a los sujetos a un cuestionario en el que tienen que contestar de la forma más rápida posible si la vara es recta o curva. Si contestan de forma correcta, se les abonará una cantidad considerable

6 Una discusión muy interesante y reciente de este tema en el ámbito de la justificación epistémica puede encontrarse en Katzoff (2000).

7 Una descripción más detallada de éste y otros ejemplos puede verse en R. Welch y D. Warren (1986).

de dinero. Si la respuesta es incorrecta, no perciben nada. Supongan, como parece probable, que el niño pequeño (al que podemos imaginar correctamente influenciado tanto por la información visual como por la táctil), responde erróneamente. No parece razonable esperar otra cosa y, por tanto, cualquier acción basada en esa creencia perceptiva sería (de acuerdo con nuestra terminología) no-responsable. Otro sería el caso, sin embargo, con el especialista en optometría. Si cometiera un error (quizá debido a la ansiedad que le provoca el experimento), la intuición más fuerte es que *debería haber sabido* contestar de otra manera. En particular, la intuición es que el especialista hubiera debido saber que ésta era una situación en la que hubiera debido —de acuerdo con nuestra terminología— *preocuparse*, es decir, una situación que requería una reflexión ulterior sobre la creencia inmediata fruto de la percepción.

El adulto no-especialista presenta un caso intuitivamente borroso. No está claro si esta persona debería haber considerado esta situación como un caso en el que una evaluación crítica de la creencia perceptiva original era necesaria. La clarificación de este tercer tipo de escenario requerirá posiblemente la introducción de otros factores, tales como su educación, su entorno cultural y social y, en general, toda una variedad de circunstancias *atenuantes*. Pero es muy posible que situaciones como ésta sean intrínsecamente borrosas y que, en estos casos, no haya una línea divisoria clara con respecto a la pregunta de si el sujeto *debería haber sabido qué hacer*. Dada la relativa claridad de los dos extremos del continuo, estas situaciones intermedias han de aceptarse sin mayor dificultad porque tal aceptación nos permite admitir una cierta indeterminación en nuestra taxonomía de estados mentales precisamente en aquellos casos en los que nuestras intuiciones no son claras, i.e., en aquellos casos en los que no está claro cuál sea la *naturaleza* de las creencias del sujeto y, por tanto, cuáles exactamente sean sus responsabilidades semánticas.

La explicación que subyace a estos casos inciertos es instructiva. La incertidumbre refleja la naturaleza *relacional* de la individuación de las representaciones propuesta en este artículo. Lo que convierte a una creencia en una creencia *responsable*, desde nuestro punto de vista, es, en parte, las relaciones que esa creencia establece con el conjunto de habilidades del sujeto, entre las cuales está la habilidad de saber cuándo se requiere una evaluación crítica, la habilidad de saber cuándo hay que preocuparse. Pero tal conjunto de habilidades se presenta de forma escalonada. La cuestión no es si se tienen o no, sino del *grado* en que se tienen y, en consecuencia, las nociones de creencia responsable y no-responsable que resultan de tal gradación reflejan la misma naturaleza escalonada.

Ahora podemos empezar a resolver algunos de los problemas mencionados anteriormente. Un sujeto es capaz de saber cuándo ha de preocuparse (en nuestra acepción del término) cuando resulta ser un especialista en el dominio al que se refiere la creencia. Parte de lo que hace que un sujeto sea un buen cocinero, una buena jugadora de ajedrez o una experta controladora aérea es aprender, a través de tanteos y práctica, a reconocer el tipo de situaciones en las que se necesita parar un momento y reflexionar sobre las representaciones mentales que se han hecho de una parcela del mundo, en lugar de confiar simplemente en una respuesta automática. El proceso de reconocimiento de estas situaciones debe ser, sin embargo, so pena de circularidad, automático. Tales situaciones saltan, sin más, a la vista del especialista. Esta habilidad cognitiva esencial, a la que me gustaría denominar «saltar a la vista crítico» (*critical pop-out*) ha sido injustamente olvidada en las discusiones filosóficas y de ciencia cognitiva⁸. Haugeland (1998), a pesar de su insistencia tanto en las habilidades caracterizadas en términos de *saber-cómo* cuanto en la importancia de una cierta clase de *estar alerta* cognitivo (que permite detectar, por ejemplo, los movimientos ilegales en un juego de ajedrez), nunca conecta estas dos nociones de la manera descrita aquí: la conexión de estas nociones requiere una habilidad más, la de detectar situaciones en las que es necesario tener una cierta *vigilancia cognitiva práctica*. Churchland (1996) desarrolla un modelo de cognición moral que asimila las representaciones de tinte moral a un modelo conexionista en el que se ajustan y coordinan ciertos patrones de conducta basándose en la experiencia adquirida por el sistema sobre el mundo real. Sin embargo, también Churchland ignora la habilidad ulterior y esencialmente moral de saber cuándo el sistema debería cuestionar las respuestas básicas del proceso mismo de coordinación de patrones (cf. Clark, 1996). Nuevos estudios neuro-cognitivos (por ejemplo, estudios de neuro-imágenes que registran las interacciones entre respuestas automáticas y ‘pensadas’, tales como los desarrollados por Drevets y Raichle (1998) y Raichle, 1998) empiezan a aproximarse al estudio de esta clase de fenómenos.

Si este *saltar a la vista crítico* (la habilidad de saber cuándo evaluar críticamente tanto las creencias propias como las de los demás) es, de hecho, sólo una clase más de habilidad que se adquiere a través de entrenamiento y experiencia, la cuestión inmediata es por qué sólo los seres

8 Dos excepciones importantes son el análisis llevado a cabo por Peggy DesAutels sobre cómo se aprende a percibir situaciones de carácter moral (DesAutels, 1996) y las críticas a los modelos conexionistas de cognición moral desarrollado por Andy Clark (1996).

humanos y no, aparentemente, perros, gatos y otras especies, son capaces de desarrollar y exhibir esta forma de representación. Ésta es, sin embargo, una cuestión empírica sobre la que no tengo competencia. Parece probable asumir, no obstante, que los regímenes de entrenamiento específico que proporciona el dominio de un lenguaje público y la pertenencia a una cultura como la humana juegan un papel empíricamente importante. Se podría así contar una historia (como lo hace, por ejemplo, Dennett, 1996) de acuerdo con la cual usos inicialmente no-responsables de gruñidos y señales se hacen progresivamente responsables en la medida en que se convierten en prácticas de uso común gobernadas por procedimientos de aprendizaje y corrección. Una vez que tales usos se han asentado, el lenguaje, progresivamente adquiriría el papel concebido por Jackendoff (1996) —el papel de permitirnos *congelar* nuestros pensamientos e ideas en un formato que los convierte de forma sencilla y generalizada en objetos de futura reflexión y análisis. Sospecho que este rico entorno social y lingüístico es el que proporciona el terreno de entrenamiento esencial dentro del cual se desarrollan y se nutren las habilidades de auto-reflexión y de *saltar a la vista crítico*. He de confesar que no sabría cómo contestar la pregunta sobre si existe una versión no-lingüística de este marco de entrenamiento en el que se puedan desarrollar estas habilidades. Sin embargo, estoy convencida —tal y como he señalado anteriormente— de que el papel de un lenguaje público no es *constitutivo* con respecto al fenómeno de creencia responsable, pero es muy posible que sea empírica y evolutivamente central en la ampliación y desarrollo de lo que podemos denominar *inteligencia crítica*.

Quiero hacer hincapié, finalmente, en que la noción de ‘saltar a la vista crítico’ no ha de ser asimilada a una noción mucho más simple relacionada con la presentación de una situación novedosa (*novelty pop-out*). Las situaciones novedosas, por supuesto, tienden a requerir nuestras facultades críticas y reflexivas. Sin embargo, la clase de habilidad que he sugerido aquí se exhibe también, de forma natural, en situaciones con las que estamos familiarizados. Por ejemplo, cuando una experta controladora aérea reconoce inmediatamente (y justo a tiempo) que su orden de bajar a 300 pies de altura y girar 30 grados hacia el oeste causaría un choque frontal, o cuando la profesora de lógica de pronto descubre un error elemental en la demostración que está llevando a cabo. Por supuesto, no siempre conseguimos identificar ese tipo de errores. Pero si no lo conseguimos, hemos de asumir la responsabilidad por esos fallos, precisamente porque las creencias involucradas en esas acciones son creencias responsables, creencias cuya necesidad de evaluar críti-

camente debería *saltar a la vista*. Aquí es donde descansa, desde mi punto de vista, la noción de *responsabilidad semántica* que he estado manejando, porque en este tipo de situaciones somos capaces de exhibir las habilidades críticas que nos permitirían corregir nuestros errores.

3. LA NATURALIZACIÓN DE LO MENTAL

Me gustaría terminar hablando de una consecuencia importante del análisis que he introducido en este trabajo. Tiene que ver con la forma y la plausibilidad de los proyectos de naturalización en semántica y filosofía de la mente⁹. Si se acepta el punto de vista defendido aquí, tales proyectos intentan abarcar dos tipos de fenómenos mentales muy diferentes: las representaciones que hemos denominado creencias responsables y las no-responsables. Sin embargo, mientras la naturalización del segundo tipo de estado mental es relativamente simple, naturalizar las creencias responsables resulta ser no sólo un proyecto mucho más complejo, sino quizá imposible. La ventaja de delimitar las dos clases de creencias como *clases cognitivas* distintas es que nos permite criticar, por una parte, lo que hay de erróneo en algunas de las críticas anti-naturalistas mientras que, por otra, nos permite determinar mucho más claramente lo que resulta problemático en los proyecto de naturalización del amplio y complicado mundo de lo mental.

Consideren, por ejemplo, la crítica de Haugeland (1998) a los intentos de naturalizar la noción de representación errónea a través de un análisis teleosemántico. Desde tales puntos de vista, las representaciones mentales se caracterizan como ‘objetos teleofuncionales’, i.e., como objetos producidos por mecanismos biológicos que han sido diseñados o seleccionados a través de una cierta historia evolutiva para realizar una ‘función apropiada’. Toda vez que la noción de representación se define en términos de función apropiada —mantiene Haugeland— tales análisis no pueden dar cuenta de la distinción esencial entre ‘representación errónea’ y ‘malfuncionamiento’. Así es como Haugeland presenta su caso:

Imáginese una especie de pájaro insectívoro que ha evolucionado en un entorno en donde la mayoría de las mariposas amarillas son venenosas, y la mayoría de las mariposas de otros colores no lo son. Supóngase además

⁹ Un tratamiento mucho más detallado de este problema puede encontrarse en Toribio (1999).

que esta especie de pájaro ha desarrollado un mecanismo que le permite detectar y evitar las mariposas amarillas. El punto en cuestión puede expresarse de la siguiente manera: si un pájaro que funciona de forma apropiada (y está rodeado de la luz adecuada, etc.) detecta y rechaza una (excepcionalmente rara) mariposa amarilla no venenosa, la sugerencia de que el pájaro representó —erróneamente— la mariposa como venenosa no está justificada. Lo mismo ocurriría si el pájaro detecta y acepta una (de nuevo, excepcionalmente rara) mariposa naranja que resulta ser venenosa ... En tales casos ... la respuesta del pájaro no «significa» nada, nada que no sea lo que de hecho produce esa respuesta en pájaros normales y en condiciones normales.

(Haugeland, 1998, p. 310)

Estamos de acuerdo con Haugeland en la necesidad de mantener una distinción clara entre la idea de cometer un error como resultado de un malfuncionamiento en el sistema y la idea de cometer un error en situaciones en las que el sistema funciona normalmente. Pero no estamos de acuerdo con el veredicto final con respecto a los análisis teleosemánticos. El problema principal es que la ontología de Haugeland es demasiado austera para poder acomodar las virtudes genuinas (aunque limitadas) de esta clase de proyecto de naturalización semántica.

Para Haugeland, «cuando no hay malfuncionamiento, la respuesta es simplemente correcta. En otras palabras, no existen bases biológicas para poder caracterizar un sistema como funcionando normalmente pero procesando erróneamente la información entrante» (*ibid.*). Haugeland rechaza así la idea de que el pájaro de su ejemplo representara erróneamente, en ningún sentido del término, la mariposa naranja como una mariposa comestible. Sin embargo, esta afirmación —me gustaría sugerir— es innecesariamente extremista. La representación puede caracterizarse como un caso claro de *representación no-responsable*. El entorno le ha jugado al pájaro una mala pasada, una mala pasada análoga a la ilusión perceptiva presentada en la Sección 1 relativa al número de escalones que una persona creer ver al salir de su casa. Recuerden que en aquella situación la persona —por mor del ejemplo— ‘funcionaba’ normalmente, no tenía ninguna deficiencia cognitiva, pero ciertamente *representaba erróneamente* el entorno que la rodeaba de forma específica, a saber, como conteniendo tres escalones en lugar de cuatro. Lo interesante del caso es que el sujeto del ejemplo no es responsable de la creencia que desemboca en su caída. El ejemplo de Haugeland nos presenta,

de la misma manera, un caso de representación errónea sin malfuncionamiento y sin error 'culpable': un caso de lo que venimos denominando representación (errónea) no-responsable.

Aquí es precisamente donde podemos localizar el problema principal del análisis llevado a cabo por Haugeland: su análisis no deja sitio para esta importante categoría. Sin embargo, sin esta distinción, tendemos a olvidar el valor real (aunque limitado) de muchos de los proyectos de naturalización en semántica. El problema no es que una explicación teleosemántica de los contenidos mentales no pueda dar cuenta de la distinción entre malfuncionamiento y representación errónea. Los análisis teleológicos pueden reconocer, y de hecho reconocen, esa distinción en el sentido siguiente: los análisis teleológicos del contenido semántico de las representaciones pueden dar cuenta de la distinción entre lo que aquí hemos llamado *creencia patológica* (debida a un malfuncionamiento del sistema) y *creencia no-responsable*. La distinción que no pueden explicar es la que existe entre la idea de malfuncionamiento y la idea —no de representación errónea *simpliciter*, sino— de representación errónea *responsable*. En otras palabras, las explicaciones teleosemánticas pueden dar cuenta, y de hecho la dan, de la distinción entre malfuncionamiento y representación no-responsable, pero fracasan en su explicación de las formas *responsables* de errar (y acertar).

Dado que las representaciones responsables son las más distintivamente humanas, el *espíritu* de la crítica de Haugeland permanece intacto, pero el factor de responsabilidad semántica que nuestro análisis introduce es importante porque nos ayuda a apreciar la lección siguiente: no es que los proyectos de naturalización semántica no puedan dar cuenta de la normatividad constitutiva de los estados mentales con contenido *en general*. Lo que ocurre es que, desde un punto de vista teleosemántico, no se pueden re-utilizar los recursos teóricos que dan cuenta de forma adecuada de la clase de normatividad ejemplificada por los casos de creencia no-responsable para que abarquen también a los casos más complejos de normatividad racional que he descrito en este trabajo y que he denominado creencias responsables.

La subclase de creencias responsables constituye el fenómeno más problemático para una explicación naturalista del contenido. Este fenómeno, sin embargo, es el más distintivamente humano porque alude a nuestra habilidad, no solamente para representar el mundo, sino también, y especialmente, a nuestra habilidad para representarlo de una forma en la que las nociones de verdad y corrección adquieren su significado pleno. La mente humana no es simplemente una máquina que genera res-

puestas adaptativas. Tampoco somos (simplemente) máquinas de razonamiento contrafáctico. Somos máquinas que pueden razonar sobre el proceso de razonamiento mismo y que pueden dar cuenta de ese proceso. El papel cognitivo y conceptual de este tipo de estados mentales responsables es suficientemente específico para que esté justificado su tratamiento como una subclase especial de estados cognitivos. No distinguirlos de sus vecinos no-responsables es asumir un alto riesgo filosófico y científico.

BIBLIOGRAFÍA

- Alston, W. P. 1988: The Deontological Conception of Epistemic Justification. En J. E. Tomberlin (ed.), *Philosophical Perspectives*, Volume 2, pp. 257-299. Atascadero, CA: Ridgeview Press.
- Brandom, R. 1994: *Making It Explicit*. Cambridge, MA: Harvard University Press.
- Chisholm, R. 1982: *The Foundations of Knowing*. Minneapolis: University of Minnesota Press.
- Churchland, P. 1996: The Neural Representation of the Social World. En L. May, M. Friedman, y A. Clark (eds.), *Minds and Morals. Essays on Cognitive Science and Ethics*, pp. 91-108. Cambridge, MA: MIT Press.
- Clark, A. 1996: Connectionism, Moral Cognition, and Collaborative Problem Solving. En L. May, M. Friedman, y A. Clark (eds.), *Minds and Morals. Essays on Cognitive Science and Ethics*, pp. 109-127. Cambridge, MA: MIT Press.
- Davidson, D. 1970: Mental Events. En L. Foster y J. W. Swanson (eds.), *Experience and Theory*, pp. 79-101. London: Duckworth.
- Dennett, D. 1978: How To Change Your Mind. In D. Dennett, *Brainstorms*, pp. 300-309. Hove, Sussex: Harvester Press.
- Dennett, D. 1987: Reflections: The Language of Thought Reconsidered. In D. Dennett, *The Intentional Stance*, pp. 227-235. Cambridge, MA: MIT Press.
- Dennett, D. 1996: *Kinds of Minds. Toward an Understanding of Consciousness*. New York, NY: Basic Books.
- DesAutels, P. 1996: Gestalt Shifts in Moral Perception. En L. May, M. Friedman, y A. Clark (eds.), *Minds and Morals. Essays on Cognitive Science and Ethics*, pp. 129-143. Cambridge, MA: MIT Press.
- de Sousa, R. 1971: How To give a Piece of Your Mind: or, the Logic of Belief and Assent. *Review of Metaphysics*, 25, 52-79
- Drevets, W. C. y Raichle, M. E. 1998: Reciprocal Suppression of Regional Cerebral Blood Flow during Emotional versus Higher Cognitive Processes: Implications for Interactions between Emotion and Cognition. *Cognition and Emotion*, 12 (3), 353-385.

- Ellis, H. D. y Young, A. W. 1990: Accounting for Delusional Misidentifications. *British Journal of Psychiatry*, 157, 239-248.
- Haugeland, J. 1998: Truth and Rule-Following. En J. Haugeland, *Having Thought. Essays in the Metaphysics of Mind*, pp. 305-361. Cambridge, MA: Harvard University Press.
- Jackendoff, 1996: *The Architecture of the Language Facility*. Cambridge, MA: MIT Press.
- Katzoff, Ch. 2000: Counter-evidence and the duty to critically reflect. *Analysis*, 60.1, 89-96.
- Kim, J. 1994: The Deontological Conception of Epistemic Justification and Doxastic Voluntarism. *Analysis*, 54, 282-284.
- Raichle, M. E. 1998: The Neural Correlates of Consciousness: An Analysis of Cognitive Skill Learning. *Phil. Trans. R. Soc. Lond. B*, 353, 1889-1901
- Sellars, W. 1963: *Science, Perception and Reality*. London: Routledge and Kegan Paul.
- Steup, M. 1988: The Deontic Conception of Epistemic Justification. *Philosophical Studies*, 53, 65-84.
- Stone, T. y Young A. W. 1997: Delusions and Brain Injury: The Philosophy and Psychology of Belief. *Mind and Language*, 12 3/4, 327-364.
- Tidman, P. 1996: Critical reflection: an alleged epistemic duty. *Analysis*, 56, 268-276.
- Toribio, J. 1999: «Normativity: The Hard Problem». *Facta Philosophica*. Enviado.
- Welch, R. B. y Warren, D. H. 1986: Intersensory Reactions. En K. Boff, L. Kaufman, y J. Thomas (eds.) *Handbook of Perception and Human Performance*, vol. 1-25. New York: Wiley.