

CUADERNOS DE CRÍTICA 45

DANIEL DENNETT

**Condiciones
de la
cualidad de persona**

Traducción de
LORENA MURILLO

INSTITUTO DE INVESTIGACIONES FILOSÓFICAS

Colección: CUADERNOS DE CRÍTICA

Director: DR. LEÓN OLIVÉ

Secretaria: MTRA. CORINA YTURBE

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

México 1989

BF

698

.D4618

1989

El ensayo de Daniel Dennett, "Conditions of Personhood" fue publicado en *The Identities of Persons*, editado por Amélie Oksenberg Rorty, University of California Press, Berkeley-Los Angeles-London, 1976. Esta editorial cedió a *Crítica* los derechos de la versión castellana.

UE-CEES



9300449320

37965.



x 53 496951

DR © 1989, Universidad Nacional Autónoma de México
Circuito Mario de la Cueva
Ciudad de la Investigación en Humanidades
04510 México, D.F.

Impreso y hecho en México

ISSN 0185-2604

CONDICIONES DE LA CUALIDAD DE PERSONA

Yo soy una persona al igual que usted. Hasta aquí no hay ninguna duda. Soy un ser humano, y *probablemente* usted también. Si se siente ofendido por el "probablemente" es porque se supone víctima de cierto tipo de racismo, pues lo que importa respecto de nosotros no es que pertenezcamos a la misma especie biológica, sino que ambos seamos personas, y de ello no tengo la menor duda. La dignidad propia no depende de nuestros orígenes, sea que hayamos nacido de mujer o que simplemente hayamos nacido. Generalmente ignoramos esto, y consideramos la humanidad como la marca decisiva de la cualidad de persona, sin duda debido a que los términos son localmente coextensivos o casi coextensivos. Hoy por hoy, los seres humanos son las únicas personas que reconocemos, y reconocemos a casi todos los seres humanos como personas; pero, por una parte, podemos fácilmente considerar la existencia de personas biológicamente muy diferentes —quizá habitantes de otros planetas— y, por la otra, reconocemos condiciones que niegan la cualidad de persona a los seres humanos, o al menos algunos elementos muy importantes de ella. Por ejemplo, a los seres humanos de poca edad, a los seres humanos con deficiencias mentales y a los seres humanos declarados dementes por psiquiatras facultados, se les niega la cualidad de persona, o cuando menos, elementos fundamentales de ésta.

Uno podría esperar que un concepto tan importante, aplicado y denegado con tal seguridad, tuviera condicio-

nes de atribución suficientes y necesarias, y claramente formulables; pero, de ser así, aún no las hemos descubierto. En última instancia, podría ser que no hubiera nada que descubrir. En último caso, podríamos llegar a darnos cuenta de que el concepto de persona es incoherente y obsoleto. Skinner es uno de los que, entre otros, ha sugerido esto; pero la doctrina no prosperó, sin duda debido, en parte, a que sería difícil, incluso imposible, concebir lo que sucedería si abandonáramos el concepto de persona. Puede argumentarse que la idea de que podamos dejar de ver a los otros y a *nosotros mismos* como personas (si esto no significa, simplemente, que nos aniquilemos y que, por ende, dejemos de ver nada como nada) es contradictoria en sí misma.¹ Por lo que, dejando a un lado lo correcto o erróneo de los fundamentos de la afirmación de Skinner, resulta difícil ver cómo podría ganarle la partida a una noción tan intuitivamente invulnerable. Si, entonces, el concepto de persona es, en cierta forma, una parte imprescindible de nuestro esquema conceptual, podría hallarse aún en peores condiciones de lo que quisiéramos. Podría resultar, por ejemplo, que el concepto de persona no sea sino un nebuloso término honorífico que gustosamente aplicamos a nosotros mismos y a los demás, como mejor nos parece, guiados por nuestras emociones, nuestra sensibilidad estética, cuestiones de costumbre y similares —así como los *chic* son sólo aquellos que pueden lograr que los consideren *chic* otros que a sí mismos se tienen por *chic*. El ser una persona es, sin duda, algo similar a esto, y si no fuera más que esto, tendríamos que reconsiderar, de ser posible, la importancia que hoy atribuimos al concepto.

¹ Véase mi "Mechanism and Responsibility" en T. Honderich (ed.), *Essays on Freedom of Action* (Londres: Routledge & Kegan Paul, 1973). [Versión castellana: *Mecanicismo y responsabilidad*, No. 42 de esta misma colección, 1985. Traducción de Myriam Rudoy.]

Suponiendo que el ser una persona *sí es* algo más, el investigador de condiciones necesarias y suficientes encontrará aún dificultades en el caso de que haya más de un concepto de persona, y existen motivos para sospecharlo. Parece, a grandes rasgos, que se encuentran aquí entreteljadas dos nociones, que llamaremos la noción moral y la noción metafísica. Locke dice que "persona"

es un término causídico, referente a las acciones y sus méritos, y se aplica, por tanto, sólo a agentes inteligentes, capaces de tener leyes, de ser felices y desdichados. Esta personalidad se proyecta a sí misma más allá de la existencia presente, hacia lo pasado, sólo mediante la conciencia —por la cual adquiere compromiso y responsabilidad (*Ensayo*, Libro II, Cap. XXVII).

¿Acaso la noción metafísica —*grosso modo*, la noción de un agente inteligente, consciente y sensible— coincide con la noción moral —*grosso modo*, la noción de un agente que es responsable, que tiene tanto derechos como obligaciones? ¿O es tan solo que el ser una persona en el sentido metafísico es una condición necesaria pero no suficiente para ser una persona en el sentido moral? ¿Acaso el ser una entidad a la que se le atribuyen estados de conciencia o autoconciencia es *lo mismo* que ser un fin-en-sí-mismo, o es solamente una condición previa? En la teoría de la justicia de Rawls, ¿debe contemplarse la derivación a partir de la posición original como una demostración de cómo las personas metafísicas pueden volverse personas morales, o debe contemplarse como una demostración de por qué las personas metafísicas tienen que ser personas morales?² En terrenos menos técnicos,

² En "Justice as Reciprocity", revisión de "Justice as Fairness", impreso en S. Gorovitz (ed.), *Utilitarianism* (Indianapolis: Bobbs Merrill, 1971) [Versión castellana: *La justicia como imparcialidad*, No. 32 de esta misma colección, 1984. Traducción de Roberto J. Vernengo], Rawls admite que las personas en la posición original pueden incluir "naciones, provincias, firmas comerciales, iglesias, equipos y otros. Los principios

esta distinción se desprende claramente; cuando declaramos demente a un hombre, dejamos de tratarlo como a un ser responsable y le negamos la mayor parte de los derechos, aun cuando nuestras interacciones con él no puedan virtualmente distinguirse de las interacciones personales normales, a menos que, en efecto, su demencia sea demasiado acentuada. Al parecer, tomando la palabra "persona" en un sentido particular, seguimos tratándolo y considerándolo como persona. Al inicio afirmé que es indudable que usted y yo somos personas. No me sería posible pensar —y mucho menos afirmar— que todos los lectores de este artículo están legalmente sanos y son moralmente responsables. Lo único que, si acaso, puede haber quedado fuera de dudas es que aquello a lo que se aludía correctamente con los pronombres personales "yo" y "usted" de la frase inicial, era una persona en el sentido metafísico. De ser esto lo único que quedara fuera de dudas, sería entonces posible que la noción metafísica y la noción moral fuesen diferentes. Sin embargo, aun suponiendo la diferencia entre estas nociones, parece haber motivos suficientes para creer que la cualidad metafísica de persona es condición necesaria de la cualidad moral de persona.³

Lo que deseo hacer ahora es examinar seis tesis co-

de justicia se aplican a reclamos conflictivos presentados por personas de cada una de estas diversas clases. Hay, quizá, cierta prioridad lógica en el caso de los individuos humanos." (p. 245). En *A Theory of Justice* (Cambridge, Mass.: Harvard University Press, 1971), reconoce que las partes en la posición original pueden incluir asociaciones y demás entidades que no sean individuos humanos (por ejemplo, p. 146), y la aparente intercambiabilidad de "partes en la posición original" y "personas en la posición original", sugiere que Rawls afirma que para cierto concepto moral de persona, la persona moral está compuesta por personas metafísicas que pueden o no ser, ellas mismas, personas morales.

³ Haciendo a un lado las posibles personas morales compuestas de Rawls. Para más datos sobre las personas compuestas, ver Amélie Rorty, "Persons, Policies, and Bodies", en *International Philosophical Quarterly*, XIII, 1 (marzo, 1973).

nocidas, cada una de las cuales pretende identificar una condición necesaria de la cualidad de persona, y cada una de ellas, en mi opinión, correcta bajo cierta interpretación. Lo que aquí se verá es, primero, cómo (de acuerdo con mi interpretación) éstas son dependientes las unas de las otras; segundo, por qué son condiciones necesarias de la cualidad moral de persona; y tercero, por qué es tan difícil afirmar que, en conjunto, constituyen condiciones suficientes de la cualidad moral de persona. La primera tesis, y la más obvia, afirma que las personas son *seres racionales*. Ésta aparece, por ejemplo, en las teorías éticas de Kant y Rawls y en las teorías "metafísicas" de Aristóteles y Hintikka.⁴ La segunda tesis sostiene que las personas son seres a los que se les atribuyen estados de conciencia o a los que se les atribuyen *enunciados de intencionalidad*, o psicológicos o mentales. Así, Strawson identifica el concepto de persona como "el concepto de un tipo de entidad tal, que tanto los enunciados que le atribuyen estados de conciencia, como los enunciados que le atribuyen características corpóreas" le son aplicables.⁵ La tercera tesis sostiene que el que una cosa sea considerada persona, depende en cierta forma de la *actitud que se adopta* hacia ella, de una *postura adoptada* con respecto a ella. Esta tesis sugiere que no es una vez establecido el hecho objetivo de que algo es una persona cuando

⁴ J. Hintikka, *Knowledge and Belief* (Ithaca: Cornell University Press, 1962).

⁵ P. F. Strawson, *Individuals* (Londres: Methuen, 1959), pp. 101-102. Con frecuencia se ha señalado que la definición de Strawson es, obviamente, demasiado amplia y abarca a todas las criaturas concientes y activas. Ver, por ejemplo, H. Frankfurt, "Freedom of the Will and the Concept of a Person", *Journal of Philosophy* (14 de enero de 1971). Se puede también discutir (y yo lo discutiría) que los estados de conciencia son sólo un subconjunto adecuado de estados psicológicos o estados caracterizados intencionalmente, pero me parece que es claro que Strawson pretende aquí lanzar sus redes lo suficiente como para incluir los estados psicológicos en general.



empezamos a tratarlo o tratarla de un modo particular, sino que este modo particular de tratarlo o tratarla forma parte, de alguna manera y hasta cierto punto, del hecho de ser persona. Algunas variaciones a esta tesis han sido expresadas por MacKay, Strawson, Amélie Rorty, Putnam, Sellars, Flew, Thomas Nagel, Dwight Van de Vate y por mí mismo.⁶ La cuarta tesis dice que el objeto con respecto al cual se adopta esta postura personal debe ser capaz de *reciprocarse* de alguna manera. Algunas versiones muy variadas de esto han sido expresadas o insinuadas por Rawls, MacKay, Strawson, Grice y otros. En ocasiones, dicha reciprocidad ha sido expresada, de manera más bien deficiente, con el lema: ser una persona es tratar a los otros como personas, y junto con esta expresión, se ha afirmado con frecuencia que tratar a alguien como persona es tratarlo moralmente —quizá obedeciendo a la Regla de Oro, pero esto combina varios tipos de reciprocidad. Como dice Nagel, “una conducta extremadamente hostil hacia alguien es compatible con el hecho de tratarlo como persona” (p. 134), y como observa Van de Vate, una de las diferencias entre algunas formas de homicidio no

⁶ D. M. MacKay, “The Use of Behavioral Language to Refer to Mechanical Processes”, *British Journal of Philosophy of Science* (1962), pp. 89-103; P. F. Strawson, “Freedom and Resentment”, *Proceedings of the British Academy* (1962), reimpresso en Strawson (ed.), *Studies in the Philosophy of Thought and Action* (Oxford, 1968); A. Rorty, “Slaves and Machines”, *Analysis* (1962); H. Putnam, “Robots: Machines or Artificially Created Life?”, *Journal of Philosophy* (12 de noviembre de 1964); W. Sellars, “Fatalism and Determinism” en K. Lehrer (ed.), *Freedom and Determinism* (Nueva York: Random House, 1966); A. Flew, “A Rational Animal” en J. R. Smythies (ed.), *Brain and Mind* (Londres: Routledge & Kegan Paul, 1968); T. Nagel, “War and Massacre”, *Philosophy and Public Affairs* (enero de 1972); D. Van de Vate, “The Problem of Robot Consciousness”, *Philosophy and Phenomenological Research* (diciembre de 1971); mi “Intentional Systems”, *Journal of Philosophy* (25 de febrero de 1971) [Versión castellana: *Sistemas intencionales*, No. 40 de esta misma colección, 1985. Traducción de Rafael Vidal].

premeditado y el asesinato, es que el asesino trata a su víctima como persona.

La quinta tesis afirma que las personas deben ser capaces de *comunicación verbal*. Esta condición deniega fácilmente a los animales no humanos la plena cualidad de persona y la responsabilidad moral concomitante, y parece al menos implícita en todas las teorías de contrato social de la ética. Es también una tesis que ha sido sostenida enfáticamente o presupuesta por muchos escritores de la filosofía de la mente, yo entre ellos, y en la cual no se ha sometido a discusión la dimensión moral de la cualidad de persona. La sexta tesis sostiene que las personas se diferencian de otras entidades por ser *conscientes* de una manera particular; *nosotros* somos conscientes de una forma en que ninguna otra especie lo es. En ocasiones, ésta se conoce como *autoconciencia*, de uno u otro tipo. Tres son los filósofos que afirman —de muy diversas maneras— que un tipo especial de conciencia es una condición previa para ser un agente moral; éstos son: Anscombe en *Intention*, Sartre en *La trascendencia del ego*, y Harry Frankfurt en su reciente artículo “Freedom of the Will and the Concept of a Person”.⁷

Debo aclarar que el orden en el que he presentado estas seis tesis es —con una salvedad— el orden de su dependencia. La salvedad es que las tres primeras son mutuamente interdependientes; ser racional es ser intencional y es ser objeto de una postura específica. Estas tres, en conjunto, son condición necesaria pero no suficiente para presentar la forma de reciprocidad que es, a su vez, condición necesaria pero no suficiente para tener la capacidad de comunicación verbal, necesaria⁸ para poseer un

⁷ H. Frankfurt, “Freedom of the Will and the Concept of a Person”, *op. cit.*

⁸ Y suficiente, pero no lo discutiré aquí. Hablo de ello en *Content and Consciousness* (Londres: Routledge & Kegan Paul, 1969) y más

tipo especial de conciencia que, como afirman Anscombe y Frankfurt en diferente forma,⁹ es condición necesaria de la cualidad moral de persona.

Con anterioridad hice referencia a las tres primeras tesis: racionalidad, intencionalidad y postura, para definir no a las personas, sino a una más amplia clase de lo que llamo *sistemas intencionales*, y dado que pretendo partir de esa noción, es preciso hacer aquí un breve resumen. Un sistema intencional es un sistema cuyo comportamiento puede ser (al menos en ocasiones) explicado y predicho recurriendo a atribuciones de *creencias y deseos* (y otros rasgos caracterizados por la intencionalidad —los que aquí llamaré *intenciones*, queriendo referirme a esperanzas, temores, intenciones, percepciones, expectativas, etc.) al sistema. Puede haber *en cada caso* otras formas de predecir y explicar el comportamiento de un sistema intencional —por ejemplo, formas mecanicistas o físicas— pero la postura intencional puede ser la más sencilla o la más efectiva o, en cualquier caso, *una* postura cuya adopción resulta exitosa y que basta para el objetivo que es ser un sistema intencional. Así definidos, es obvio que los sistemas intencionales no son todos personas. Atribuimos creencias y deseos a los perros y a los peces y, por lo tanto, predecimos su comportamiento, e incluso podemos utilizar este procedimiento para predecir el comportamiento de algunas máquinas. Es, por ejemplo, la estrategia que conviene adoptar, de hecho la única conveniente, ante una computadora que juega bien al ajedrez. *Suponiendo* que la computadora tiene ciertas creencias (o información) y deseos (o funciones preferenciales) con respecto al juego de ajedrez que se está llevando a cabo, puedo calcular —bajo circunstancias favorables—

reciente y explícitamente en mi "Reply to Arbib and Gunderson", APA Eastern Division Meetings, 29 de diciembre de 1972.

⁹ No discutiré aquí la tesis de Sartre.

la jugada siguiente más probable de la computadora, a *condición de que yo suponga que la computadora lidia con estos deseos y creencias en forma racional*. La computadora es un sistema intencional en estas circunstancias, no porque tenga algún rasgo intrínseco particular, como tampoco porque realmente tenga creencias y deseos (cualesquiera que éstos sean), sino tan solo porque ella sucumbe ante cierta *postura* adoptada con respecto a ella, a saber, la postura intencional, la postura que procede atribuyendo a la computadora enunciados intencionales bajo las limitantes usuales, la postura que procede considerando a la computadora como un razonador racional práctico.

Es importante observar lo blanda que es esta definición de *sistema intencional* y cuán grande puede ser la correspondiente clase de los sistemas intencionales. Si, por ejemplo, predigo que una planta particular —digamos, una hiedra en maceta— va a crecer dando la vuelta a una esquina, hacia arriba y en dirección de la luz, porque "busca" la luz, "quiere" salir de la sombra en que ahora se encuentra, y "cree" y "espera" que habrá luz al doblar la esquina, habré adoptado una postura intencional con respecto a la planta y, ¡oh, sorpresa!, dentro de límites muy estrechos ésta funciona. Dado que funciona, algunas plantas son sistemas intencionales de muy bajo nivel.

Me di cuenta de la utilidad real de adoptar una postura intencional con respecto a las plantas al platicar con unos leñadores de los bosques de Maine. Estos hombres se refieren invariablemente a los árboles no como "eso", sino como "él", y de un joven retoño dirán: "Él quiere desplegar sus ramas, pero no hay que dejarlo; porque así tendrá que estirarse para alcanzar la luz" o bien "A los pinos no les gusta mojarse los pies como a los cedros". Se puede "engañar" a un manzano y "hacerle creer que es primavera", prendiendo una pequeña fogata bajo sus

ramas al final del otoño; florecerá. Esta manera de hablar no es sólo pintoresca y no es en absoluto supersticiosa; es simplemente una forma eficaz de dar sentido, controlar, predecir y explicar el comportamiento de esas plantas de un modo que sutilmente elude nuestra ignorancia de los mecanismos de control. Algunos biólogos más sofisticados pueden preferir hablar de transmisión de la información desde la periferia del árbol hacia otros puntos del mismo. Esto es menos pintoresco, pero sigue siendo intencional. La abstención total del uso del lenguaje intencional respecto de los árboles puede volverse casi tan heroica, engorrosa y carente de sentido como el tabú paralelo de los conductistas estrictos al hablar de ratas y pichones. Y aunque las expresiones intencionales referentes, por ejemplo, a las actividades del árbol tengan un valor heurístico casi imperceptible, me parece más inteligente dar por sentado que dicho árbol es un sistema intencional muy degenerado, soso e insignificante, que tratar de trazar una línea por arriba de la cual las interpretaciones intencionales sean "objetivamente verdaderas".

Es obvio entonces que ser un sistema intencional no es condición suficiente para ser persona, pero sin duda es condición necesaria. Nada hacia lo cual no podamos adoptar satisfactoriamente una postura intencional, con sus presupuestos de racionalidad, podrá considerarse persona. ¿Podemos entonces definir a las personas como una subclase de los sistemas intencionales? A primera vista podría parecer provechoso suponer que las personas son sólo la subclase de los sistemas intencionales que *realmente* tienen creencias, deseos y demás, y de los que no sólo *suponemos* que los tienen por mor de una predicción expedita. Pero parece que todos los intentos por definir aquello de lo que se puede decir que realmente tiene una creencia (de manera que ni un perro ni un árbol ni

una computadora pudieran resultar idóneos) terminan poniendo condiciones a la creencia genuina, condiciones que (1) son demasiado estrictas para nuestras intuiciones, y (2) aluden a distintas condiciones de la cualidad de persona que se encuentran en los últimos lugares de mi lista. Por ejemplo, se podría afirmar que las creencias genuinas son por necesidad *verbalmente expresables* por el creyente,¹⁰ o que el creyente debe estar *consciente* de que las tiene; pero la gente parece tener muchas creencias que no puede expresar en palabras, y muchas que no sabe que tiene —y, en cualquier caso, espero probar que la capacidad de expresión verbal y la capacidad de conciencia encuentran diferentes *loci* en el conjunto de condiciones necesarias de la cualidad de persona.

Considero que se puede lograr un mayor avance si retomamos nuestra cuarta tesis, la de reciprocidad, para ver qué tipo de definición podría recibir en términos de sistemas intencionales. La tesis sugiere que una persona debe ser capaz de reciprocitar la postura, lo que sugiere que un sistema intencional que a su vez adoptara una postura intencional hacia otros objetos pasaría la prueba. Definamos un *sistema intencional de segundo orden* como aquel al cual atribuimos no sólo creencias, deseos y demás intenciones simples, sino creencias, deseos y demás intenciones *respecto de* creencias, deseos y demás intenciones. Un sistema intencional *S* será un sistema intencional de segundo orden si, entre las atribuciones que le hacemos, se encuentran las del tipo *S cree que T desea que p*, *S espera que T tema que q*, y casos reflexivos como *S cree que S desea que p*. (La importancia de los casos reflexivos se hará evidente, sin que esto nos sorprenda, cuando veamos a quienes interpretan nuestra sexta con-

¹⁰ Cf. Bernard Williams, "Deciding to Believe" en H. E. Kiefer y M. K. Munitz (eds.), *Language, Belief, and Metaphysics* (Nueva York: New York University Press, 1970).

dición como *autoconciencia*. Algunos pensarán que los casos reflexivos convierten automáticamente a todos los sistemas intencionales en sistemas de segundo orden e incluso en sistemas de orden n , sobre la base de que creer que p implica creer que se cree que p , y así sucesivamente; pero éste es un error fundamental. La repetición de creencias y demás intenciones no es nunca redundante, y así, aun cuando algunas repeticiones son normales —se espera que lo sean— nunca son triviales o automáticas.)

Ahora bien, ¿son acaso los seres humanos hasta donde sabemos, los únicos sistemas intencionales de segundo orden? Considero ésta una pregunta empírica. Por ejemplo, atribuimos creencias a los perros, los gatos, los leones, los pájaros y los delfines y a partir de ahí predecimos con frecuencia su comportamiento —cuando todo sale bien—, pero es difícil imaginar un caso en el que la conducta del animal fuese tan sofisticada que necesitásemos atribuirle intenciones de segundo orden para poder predecir o explicar su comportamiento. Por supuesto que si alguna versión del fisicalismo mecanicista es verdadera (como creo), no *necesitaremos* nunca atribuir intención alguna a nada, pero, suponiendo que por razones heurísticas y pragmáticas tuviésemos que atribuir intenciones a los animales, ¿sentiríamos acaso el impulso pragmático de atribuirles intenciones de segundo orden? Con frecuencia los psicólogos han recurrido a un principio llamado el canon de la parsimonia de Lloyd Morgan, que puede considerarse como un caso especial de la navaja de Occam; el principio es que se debe atribuir a un organismo tan poca inteligencia o conciencia o racionalidad o pensamiento, como baste para dar cuenta de su comportamiento. Este principio puede interpretarse, y ha sido interpretado, nada menos que como un conduc-

tismo radical;¹¹ pero esto me parece un error, y podemos interpretarlo como el principio que requerimos al adoptar la postura intencional con respecto a un objeto, para atribuirle las creencias, deseos y demás más simples, menos sofisticados y de menor orden, que darán cuenta de su comportamiento. Así pues, admitiremos, por ejemplo, que Fido *quiere* su cena y *cree* que su amo le dará la cena si gime frente a él, pero no necesitamos atribuirle a Fido la *creencia* posterior de que su gemido produce en su amo la *creencia* de que él, Fido, *quiere* su cena. De manera similar, mi *expectativa* al depositar una moneda en la máquina de dulces no depende de la *creencia* posterior de que, al insertar la moneda, se induce en la máquina la *creencia* de que yo *quiero* un dulce. Esto es, que aun si el gemido de Fido se parece mucho a una verdadera interacción de segundo orden (en la que Fido trata a su amo como a un sistema intencional), si suponemos que para Fido su amo es tan solo una máquina proveedora de alimento que se activa mediante el gemido, obtendremos una atribución predictiva igualmente buena, más modesta, pero que sigue siendo, por supuesto, intencional.

¿Son entonces los perros, los chimpancés y demás animales “superiores” incapaces de elevarse al nivel de los sistemas intencionales de segundo orden?; y de ser así, ¿por qué? Yo solía pensar que la respuesta era afirmativa, pensando que la razón era la falta de lenguaje en los animales no humanos, y que el lenguaje era necesario para representar las intenciones de segundo orden. En otras palabras, pensaba que la condición cuatro se apoyaba en la condición cinco. Me atraía la hipótesis de que los animales no pueden, por ejemplo, tener creencias

¹¹ Por ejemplo, B. F. Skinner, “Behaviorism at Fifty”, en T. W. Wann (ed.), *Behaviorism and Phenomenology* (Chicago: University of Chicago Press, 1964).

de segundo orden, creencias acerca de creencias, por la misma razón por la que no pueden tener creencias sobre el día viernes o sobre poesía. Algunas creencias sólo pueden adquirirse, y por lo tanto representarse, por vía del lenguaje.¹² Pero, si bien es cierto que algunas creencias no pueden adquirirse sin el lenguaje, es falso que todas las creencias de segundo orden se encuentren entre ellas, y es falso también que los no humanos no puedan ser sistemas intencionales de segundo orden. Al pedir a la gente que me diera ejemplos de sistemas intencionales de segundo orden no humanos, encontré algunos casos muy plausibles. Considérese por ejemplo éste de Peter Ashley (en una carta):

Una noche estaba yo sentado en una silla en mi casa, en la *única* silla en la que a mi perra se le permite dormir. Ella estaba echada frente a mí gimiendo. Sus intentos por “convencerme” de que le cediera la silla no estaban logrando gran cosa. Su siguiente paso es el más interesante, más aún, la *única* parte interesante de esta historia. Se levantó y se dirigió hacia la puerta de entrada, donde yo podía seguir viéndola con facilidad. Arañó la puerta, dándome la impresión de que había desistido en sus intentos por obtener la silla y que había decidido salir. Sin embargo, tan pronto llegué a la puerta para dejarla salir, se regresó corriendo a través del cuarto y se trepó en la silla, la silla que me había “obligado” a dejar.

En este caso, parece que tenemos que atribuir a la perra la *intención* de que su amo *crea* que ella *quiere* salir —una intención no sólo de segundo, sino de tercer orden. La clave de este ejemplo, aquello que lo hace un ejemplo de la operación de un sistema intencional de orden superior, es que la creencia que la perra pretende causar en su amo es falsa. Para descubrir más ejemplos de

¹² Se hallarán sugerencias ilustrativas sobre la relación del lenguaje con respecto a la creencia y la racionalidad en Ronald de Sousa, “How to Give a Piece of Your Mind; or, A Logic of Belief and Assent”, *Review of Metaphysics* (septiembre de 1971).

animales que se comportan como sistemas intencionales de segundo orden, nos puede ayudar el pensar en algunos casos de engaño, en los que el animal, creyendo *p*, trata de hacer que otro sistema intencional crea *no-p*. En el caso de un animal que trata de provocar en otro una conducta que las creencias *verdaderas* acerca del entorno del otro no provocarían, no podemos hacer una “división interna” y obtener una explicación que sólo recurra a intenciones de primer nivel. Podemos generalizar este punto antes de pasar a la explicación de por qué es así: cuando *x* trata de provocar una conducta en *y*, conducta que resulta inapropiada para el entorno y las necesidades *reales* de *y*, pero apropiada para el entorno y las necesidades que *y* *percibe* o *crea*, nos vemos obligados a atribuir intenciones de segundo orden a *x*. Una vez establecido de esta manera, este aspecto parece familiar, y frecuentemente es utilizado por los críticos del conductismo: se puede ser conductista en la explicación y control de la conducta de animales de laboratorio sólo en tanto se pueda confiar en que no hay una grave dislocación entre el entorno real del experimento y el entorno percibido por los animales. Una táctica para poner en aprietos a los conductistas en el laboratorio consiste en efectuar experimentos que engañen a los sujetos; si el engaño tiene éxito su conducta será predecible a partir de sus *creencias* falsas respecto del entorno, no a partir del entorno real. Ahora bien, un sistema intencional de primer orden es conductista; no atribuye intención alguna a nada. Así pues, si deseamos tener evidencia efectiva de que un sistema dado *S* *no* es conductista —sino que es un sistema intencional de segundo orden—, esto sólo ocurrirá en aquellos casos en que las teorías conductistas no se adecuen a los datos, sólo en aquellos casos en que el conductismo no explicaría el éxito del sistema *S* cuando manipula la conducta de otro sistema.

Esto sugiere que el ejemplo de Ashley no es, después de todo, tan convincente; que puede ser rechazado suponiendo que su perra es una conductista menor. No es preciso que ella crea que el arañar la puerta llevará a Ashley a creer que ella quiere salir; basta con que ella crea, como buena conductista, que ha condicionado a Ashley para que vaya a la puerta cuando ella la araña. Así, ella aplica el estímulo usual, obtiene la respuesta usual, y eso es todo. El caso de Ashley se derrumba si ésta es la forma normal como su perra logra que se le abra la puerta, como probablemente lo es, pues entonces una hipótesis más modesta será que la perra cree que su amo está condicionado para ir a la puerta cuando ella la araña. Si la perra hubiera hecho algo *novedoso* para engañar a su amo (como correr hacia la ventana, mirar hacia afuera y gruñir sospechosamente), entonces tendríamos que reconocer que el levantarse de la silla no fue simplemente una respuesta condicionada de Ashley, y no podría ser "considerada" como tal por su perra, pero entonces, tal virtuosidad en un perro resultaría altamente improbable.

Pero, ¿cuál es la diferencia entre el caso improbable y los muy conocidos casos en los que un pájaro que tiene su nido al ras del suelo finge que se le ha roto un ala, para alejar al depredador del nido? El efecto logrado es una novedad, en el sentido de que el pájaro, con toda seguridad, no ha condicionado repetidamente a los depredadores de los alrededores mediante este estímulo; por ello, nos vemos obligados a explicar la artimaña como un pequeño y genuino engaño, mediante el cual el pájaro *busca* provocar una *creencia* falsa en el depredador. Llevados a aceptar esta interpretación de la conducta, nos podría impresionar enormemente la ingeniosidad del pájaro, de no ser por el hecho de que sabemos que dicho comportamiento es "meramente instintivo". Pero, ¿por qué se minimiza el truco al llamarlo "meramente instintivo"? Afir-

mar que es instintivo es afirmar que todas las especies de pájaros lo hacen; lo hacen aun cuando las circunstancias no son totalmente apropiadas; lo hacen cuando existen mejores motivos para permanecer en el nido; el patrón de comportamiento es rígido, un tropismo de menor relevancia, y presumiblemente los controles fueron transmitidos genéticamente, no aprendidos ni inventados.

Debemos tener cuidado de no llevar muy lejos esta actitud minimizadora; no es que el pájaro realice el truco "irreflexivamente", pues si bien es claro que en ningún momento cruzó por su mente un argumento o esquema ("Veamos, si yo agito el ala como si estuviera rota, la zorra pensará que..."), un hombre puede hacer algo igualmente astuto y con genuina inteligencia, novedad y adecuación, sin por ello valerse tampoco de "pensamientos conscientes". El *pensar los pensamientos*, como quiera que esto se caracterice, no es lo que hace realmente inteligente a una conducta inteligente. Anscombe dice en cierto momento: "Si supusiéramos que [semejante expresión del razonamiento] describe los procesos mentales reales, en general sería bastante absurda. El interés de la explicación es que describía un orden que está presente siempre que las acciones se realizan con intenciones."¹³ Pero, el "orden está presente" tanto en el caso del pájaro como en el del hombre. Es decir, cuando nos preguntamos por qué los pájaros evolucionaron con este tropismo, lo explicamos subrayando la utilidad de poseer un medio para *engañar* a los depredadores, o de provocar en éstos creencias erróneas; lo que debe explicarse es el origen de las intenciones de segundo orden del pájaro. Yo sería el último en negar o ignorar la enorme diferencia que hay entre una conducta instintiva o tropista, y la conducta más versátil e inteligente de los humanos y de

¹³ G. E. M. Anscombe, *Intention* (Oxford: Blackwell, 1957), p. 80.

otros; pero en lo que quiero insistir aquí es en que, si uno está preparado para adoptar sin escrúpulos una postura intencional como herramienta para predecir y explicar la conducta, el pájaro es entonces un sistema intencional de segundo orden, como cualquier hombre. De ser este el caso, debemos tener mucha cautela al utilizar el argumento que me pareció antes tan atractivo, a saber, que las *representaciones* de intenciones de segundo orden dependerán en alguna forma del lenguaje.¹⁴ Pues no es en absoluto evidente que todas —o siquiera algunas— las creencias y demás intenciones de un sistema intencional necesiten *representarse* “dentro” del sistema de alguna manera particular para que podamos obtener un punto de apoyo para predecir su conducta, *atribuyéndole* dichas intenciones.¹⁵ La situación que expusimos al referirnos al deseo del pájaro de provocar en el depredador una creencia falsa, parece no tener cabida en una representación de esta sofisticada intención; ni necesidad de ella, en los “pensamientos” o “mente” de entidad alguna, dado que ni el pájaro, ni la historia de la evolución, ni la madre naturaleza, necesitan pensar estos pensamientos para que nuestra explicación quede asegurada.

Así pues, la reciprocidad, siempre que se entienda por ésta simplemente la capacidad de los sistemas intencionales para exhibir intenciones de orden superior, si bien depende de las tres primeras condiciones, es independiente de la quinta y la sexta. La justicia que esta noción haga a la reciprocidad de la que hablan otros escritores, sólo aparecerá con claridad cuando se vea cómo ésta se en-

¹⁴ Cf. Ronald de Sousa, “Self-Deception”, *Inquiry*, 13 (1970), esp. p. 317.

¹⁵ Discuto lo anterior con más detalle en “Brain Writing and Mind Reading” en K. Gunderson (ed.), *Language, Mind, and Knowledge* (Minneapolis: University of Minnesota Press, 1975), y en mi “Reply to Arbib and Gunderson”.

treteje con las dos últimas condiciones. Para la quinta condición, la capacidad de comunicación verbal, recurriremos a la teoría del significado de Grice. Grice trata de definir lo que él llama significado no natural, lo que pretende quien se expresa al expresar algo, en términos de las *intenciones* del que se expresa. Su definición inicial es la siguiente:¹⁶

“U pretendió decir algo expresando x” es verdad si, para algún oyente A, U expresó x con la intención de que:

- (1) A produjera una respuesta particular r.
- (2) A pensara (reconociera) que U tenía la intención de (1).
- (3) A cumpliera (1) sobre la base de su cumplimiento de (2).

Obsérvese que la intención (2) atribuye a U no sólo una intención de segundo, sino de tercer orden: U debe *tratar de* que A *reconozca* que U *trata* que A produzca r. No importa en absoluto que Grice se haya visto obligado por una serie de contraejemplos a ir de esta definición inicial hacia versiones mucho más complejas, ya que todas ellas reproducen la intención de tercer orden de (2). Del análisis del significado no natural de Grice se desprenden dos puntos de gran importancia para nosotros. Primero, dado que el significado no natural, pretender significar algo al decir algo, debe ser un rasgo de toda comunicación verbal verdadera, y dado que depende de las intenciones de tercer orden por parte del que se expresa, sostenemos que la condición cinco se apoya en la condición cuatro, y no a la inversa. Segundo, Grice nos muestra que no bastan las simples intenciones de *segundo* orden para producir reciprocidad genuina; para ello se requieren intenciones de *tercer* orden. Grice introduce

¹⁶ Los escritos clave son “Meaning”, *Philosophical Review* (julio de 1957), y “Utterer’s Meaning and Intentions”, *Philosophical Review* (abril de 1969). Su formulación inicial, desarrollada en el primer artículo, está sujeta a una serie de revisiones en el segundo artículo, del cual se ha extraído esta formulación (p. 151).



la condición (2) para excluir casos como éste: "Dejo tirada la porcelana china que mi hija rompió para que mi esposa la vea." Éste no es un caso en el que pretendo significar algo haciendo lo que hago, con la intención de lograr lo que intento, pues aun cuando yo estoy tratando mediante esto de lograr que mi esposa piense algo sobre nuestra hija (una intención de segundo orden de mi parte), el éxito no depende de que ella reconozca mi intención o de que en algo reconozca mi intervención o existencia. No ha habido, para usar el atinado término de Erving Goffman, un *encuentro* real entre nosotros, un *reconocimiento mutuo*. Debe haber un encuentro entre el que se expresa y el oyente para que el que se expresa logre significar algo, pero los encuentros pueden darse en ausencia de significado no natural (la perra de Ashley), y las artimañas que dependen de intenciones de tercer orden no necesitan involucrar encuentros (por ejemplo, *A* puede tratar de que *B* crea que *C* desea que *p*). Así, las intenciones de tercer orden son una condición necesaria pero no suficiente para que ocurran encuentros que a su vez son condición necesaria pero no suficiente para la ocurrencia de casos de significado no natural, esto es, casos de comunicación verbal.

No es por accidente que los casos de Grice de significado no natural pertenecen a una clase en la cual los otros miembros son sujetos de engaño o manipulación. Consideremos el ingenioso contraejemplo de Searle a una de las formulaciones de Grice: el norteamericano capturado en Italia tras las líneas enemigas durante la Segunda Guerra Mundial, que trata de engañar a sus captores italianos haciéndoles creer que es un oficial alemán, repitiendo la única frase en alemán que conoce: "*Kennst du das Land, wo die Zitronen blühen?*".¹⁷ Como apunta Grice, estos

¹⁷ John Searle, "What is a Speech Act?" en Max Black (ed.), *Philo-*

casos comparten con los casos de significado no natural el hecho de confiar en la racionalidad de la víctima o de explotarla. En estos casos, el éxito depende de inducir a la víctima a que se lance en una cadena de razonamientos a la cual uno contribuye directa o indirectamente con premisas. En el engaño, las premisas no son creídas por quien las dice; en la comunicación normal sí lo son. La comunicación, a la manera de Grice, aparece como una especie de manipulación colaborativa del oyente, por parte del que se expresa; y depende, no sólo de la racionalidad del oyente, quien debe descubrir las intenciones del que se expresa, sino también de la *confianza* del oyente en quien se expresa. La comunicación como especie de manipulación no podría funcionar dada la racionalidad que se requiere del oyente, a menos que la confianza del oyente en el que se expresa esté *bien fundada* o sea razonable. Así, la *regla* para la expresión es la sinceridad; si la expresión no es naturalmente confiable, no alcanzará su propósito.¹⁸

La mentira, como forma de engaño, sólo puede funcionar gracias a un antecedente de verdad, pero otras formas de engaño no dependen de la confianza de la víctima. En estos casos, el éxito depende de que la víctima sea *algo* lista, pero no lo suficientemente lista. Los jugadores de póker tontos son la ruina de los jugadores de póker astutos, ya que no logran ver los *bluffs* y ardidés que se les lanzan. Tales engaños sofisticados no necesitan depender de encuentros directos. Existe un libro acerca de cómo descubrir antigüedades falsas (que, inevitablemente, es también un libro acerca de cómo *hacer* antigüedades falsas) que brinda este astuto consejo a quienes quieren engañar al comprador "experto": una vez terminada su

sophy in America (Londres: Allen & Unwin, 1965), discutido por Grice en "Utterer's Meaning and Intentions", p. 160.

¹⁸ Cf. "Intentional Systems", pp. 102-103.

mesa, o cualquier otro objeto (habiendo utilizado todos los medios usuales para dar el efecto de envejecimiento y uso) tome un moderno taladro eléctrico y haga un agujero que atravesase el objeto en un lugar extraño pero visible. El probable comprador pensará: nadie habría hecho una perforación tan desfigurante sin un motivo (de ninguna manera podría suponerse que es "auténtico"), así que debe de haber servido para algún propósito, lo que significa que esta mesa debió utilizarse en algún hogar; dado que fue utilizada en algún hogar, no fue expresamente fabricada para ser vendida en esta tienda de antigüedades... así es que debe de ser auténtica. Aun cuando esta "conclusión" deje lugar a dudas persistentes, el comprador estará tan ocupado imaginando usos para esa perforación, que pasarán meses antes de que las dudas afloren.

Lo importante en estos casos de engaño es el hecho de que, al igual que en el caso del pájaro que fingía, el éxito no depende de que la víctima *realice conscientemente* estas cadenas de razonamientos. No importa que el comprador note apenas la perforación y "tenga el presentimiento" de que la pieza es genuina. Posteriormente, él *podría* aceptar el razonamiento que se hizo como el "motivo" que lo llevó a considerar genuina la pieza, pero podría negarlo, y al negarlo podría estarse engañando a sí mismo, aun cuando los *pensamientos* nunca pasaron por su mente. La cadena de razonamientos explica por qué la perforación funciona como lo hace (si es que funciona), pero, como dice Anscombe, no necesariamente "describe los procesos mentales reales", si suponemos que los procesos mentales reales son procesos o acontecimientos conscientes. Por supuesto que lo mismo se aplica en el caso de las comunicaciones de Grice; no es necesario que el que se expresa o el oyente consideren conscientemente las complejas intenciones que el primero esboza, y lo que

sorprende un poco es que nunca nadie haya utilizado este hecho como una objeción a Grice. Las condiciones de Grice para el significado han sido frecuentemente criticadas por no ser suficientes, pero al parecer existe un argumento aún no utilizado para demostrar que ni siquiera son necesarias. Con toda seguridad, antes de que Grice las señalara poca gente concibió conscientemente esas ingeniosas intenciones, y aun así, la gente se ha estado comunicando durante años. Antes de Grice, si uno hubiera preguntado: "¿Pretendía usted que su oyente reconociera su intención para así provocar esa respuesta en él?", lo más probable es que uno recibiera esta respuesta: "Nunca pretendí algo tan tortuoso. Lo único que pretendía era informar que no llegaría a cenar esta noche" (o lo que fuera). Por lo que, al parecer, si estas complicadas intenciones subyacen en nuestra comunicación, deben haber sido intenciones inconscientes. En efecto, una manera perfectamente natural de responder a los artículos de Grice es hacer la observación de que *no estamos conscientes* de hacer esas cosas cuando nos comunicamos. Ahora bien, Anscombe ha sostenido con mucha fuerza que semejante respuesta establece que la acción que se describe no era intencional.¹⁹ Dado que al hablar uno *no es consciente* de estas *intenciones*, uno no puede estar hablando *con* estas *intenciones*.

¿Por qué nadie ha utilizado este argumento en contra de la teoría de Grice? Porque, propongo, es muy claro que Grice se trae algo entre manos, que nos está dando condiciones necesarias para el significado no natural. Sus análisis arrojan luz sobre muchas cuestiones. ¿Nos comunicamos con las computadoras en Fortran? Fortran parece ser un lenguaje; tiene una gramática, un vocabulario, una semántica. Las transacciones en Fortran entre

¹⁹ G. E. M. Anscombe, *Intention*, p. 11.

el hombre y la máquina se consideran, en ocasiones, como *casos de comunicación entre un hombre y una máquina*, pero tales transacciones son copias pálidas de la comunicación verbal humana precisamente porque las condiciones de Grice para el significado no natural han sido ignoradas. Su aplicación no tiene cabida. El que cumplamos con nuestro objetivo de transmitir un poco de Fortran a la máquina no depende de que logremos que la máquina reconozca nuestras intenciones. Esto no significa que, en el futuro, toda comunicación con las computadoras vaya a tener estas deficiencias (o fuerza, dependiendo de sus propósitos), sino tan solo que en la actualidad nos comunicamos con las computadoras en el sentido fuerte (de Grice).²⁰

Si no estamos por abandonar el modelo griceano, y sin embargo no estamos conscientes de dichas intenciones en nuestra conversación usual, quizá simplemente tengamos entonces que llevar bajo tierra estas intenciones y llamarlas intenciones inconscientes o preconscious. Son intenciones que manifiestan "un orden que está presente" cuando la gente se comunica, intenciones de las que normalmente no nos percatamos, e intenciones que son una condición previa de la comunicación verbal.²¹

²⁰ Howard Friedman me ha hecho notar que muchas compiladoras Fortran actuales, que "corrigen" las entradas del operador mediante la inserción de signos "más" y paréntesis con el fin de producir expresiones bien formadas supuestamente coinciden con los criterios de Grice, dado que, dentro una esfera muy limitada, diagnostican las intenciones "del que se expresa" y proceden sobre la base del diagnóstico. Pero primero tendría que destacarse que las máquinas actuales sólo pueden diagnosticar lo que podría llamarse las intenciones sintácticas del operador, y segundo, que estas máquinas no parecen coincidir con las definiciones subsecuentes y más elaboradas de Grice, sin que yo quiera afirmar que ninguna computadora podría hacerlo.

²¹ De hecho Grice describe tan solo una pequeña porción del orden que está presente como condición previa de la interacción personal normal. Puede hallarse un análisis de las intenciones de nivel superior con una actitud más abierta en las obras de Erwing Goffman, especialmente

Hemos llegado hasta este punto sin tener que recurrir a ninguna clase de conciencia, por lo que, si acaso existe una dependencia entre la conciencia o la autoconciencia y nuestras otras condiciones, tendrá que ser la conciencia la que dependa de las otras. Pero, para demostrar esto, primero debo mostrar cómo las cinco primeras condiciones podrían desempeñar por sí solas un papel en la ética, como lo sugiere la teoría de Rawls. Central en la teoría de Rawls es el establecimiento de una situación idealizada, de la "posición original" habitada por personas idealizadas, idealización de la que se derivan los primeros principios de justicia que generan y arrojan luz sobre el resto de su teoría. Lo que ahora me interesa no es el contenido de estos principios ni la validez de su derivación, sino la naturaleza de la táctica de Rawls. Él supone que un grupo de personas idealizadas, que define como entidades racionales e interesadas en sí mismas, realizan cálculos bajo ciertas restricciones acerca de los probables y posibles efectos interactivos de sus intereses individuales y antagónicos (lo que quizá exija de ellos diseñar intenciones de un orden superior, por ejemplo, creencias respecto de los deseos de los demás, creencias respecto de las creencias de los demás acerca de los propios deseos, y así sucesivamente). Rawls afirma que estos cálculos tienen una "solución" óptima que sería razonable que toda persona interesada en sí misma adoptara como una alternativa al estado de naturaleza de Hobbes. La solución consiste en ponerse de acuerdo con sus congéneres para guiarse por los principios de justicia que Rawls esboza. ¿Qué clase de prueba de los principios de justicia sería ésta? La adopción de estos principios de justicia puede verse, afirma Rawls, como la solución al

en *The Presentation of Self in Everyday Life* (Garden City: Doubleday, 1959).

“juego del orden superior” o al “problema de la negociación”. Es análoga a las derivaciones de la teoría del juego, a las pruebas en la lógica epistémica de Hintikka,²² y a la “demostración” de que la computadora que juega ajedrez hará determinado movimiento, porque es el movimiento más racional dada la información que posee sobre el juego. Todas ellas se basan en la suposición de que existen calculadoras idealmente racionales, y por ello sus respuestas son intrínsecamente normativas. Por lo tanto, considero las derivaciones a partir de la posición original de Rawls como un continuo de las deducciones y extrapolaciones que se hallan en los usos más simples de la postura intencional para entender y controlar la conducta de entidades más simples. De la misma forma en que la verdad y la consistencia son normas para la creencia,²³ y la sinceridad es la norma para la expresión, así, si Rawls está en lo correcto, la justicia, como él la define, es la norma para las interacciones interpersonales. Pero entonces, así como la capacidad que tengamos para explicar la entidad como *racional* es parte de nuestra justificación para considerar que una entidad tiene creencias u otras intenciones, en esa medida nuestros fundamentos para considerar a la entidad como persona incluyen nuestra capacidad para verla regida por los principios de justicia. Una forma de captar el estatuto particular del concepto de persona como pienso que se emplea aquí, sería diciendo que, si bien Rawls no intenta en absoluto afirmar que la justicia es el resultado inevitable de la interacción *humana*, sí sostiene, en efecto, que es el resultado inevitable de la interacción *personal*. Es decir, que el concepto de persona es, en sí mismo, inevitablemente normativo o idealizado; en la medida en que la

²² Véase Hintikka, *Knowledge and Belief*, p. 38.

²³ Véase Dennett, “Intentional Systems”, pp. 102-103.

justicia no se revela en las transacciones e interacciones de la creaturas, en esa misma medida no son personas. Y de nuevo podemos ver que existe “un orden presente” en una sociedad justa independientemente de cualquier episodio real de pensamiento consciente. La existencia de prácticas justas y el “reconocimiento” implícito en ellas no depende de que alguien, alguna vez, de manera consciente o deliberada, haya realizado los cálculos de la posición original idealizada, llegando de manera consciente a acuerdos recíprocos y adoptando conscientemente una postura con respecto a los demás.

Para reconocer a otro como persona, uno debe responderle y actuar con respecto a él de cierta manera; y esta manera está íntimamente relacionada, *prima facie*, con los diversos deberes. El reconocimiento, hasta cierto grado, de estos deberes, y el tener por lo tanto los elementos de moralidad, no es cuestión de elección o de intuición de cualidades morales, o cuestión de expresión de sentimientos o actitudes... es simplemente el seguimiento de una de las formas de conducta en las que se manifiesta el reconocimiento de los demás como personas.²⁴

Por supuesto que la importancia del intento de Rawls por derivar principios de justicia a partir de la “posición original” es que, aunque el resultado pueda ser reconocido como norma *moral*, no se *deriva como* norma moral. No se presupone la moralidad de las partes en la posición original. Esto significa que la derivación de la norma no nos da, en sí misma, ninguna respuesta a las preguntas de cuándo y por qué tenemos el derecho de hacer *moralmente* responsables a las personas por la desviación de esta norma. Aquí, Anscombe viene en nuestra ayuda y, al mismo tiempo, introduce nuestra sexta condición. *Si ha de hacerse responsable de una acción* (una fracción de mi conducta descrita de una manera particular), debí estar *consciente* de esa acción descrita de esa manera par

²⁴ J. Rawls, “Justice as Reciprocity”, p. 259.



ticular.²⁵ ¿Por qué? Porque sólo estando consciente de la acción puedo *decir* lo que intentaba hacer, y participar desde una posición privilegiada en el juego de preguntas y respuestas para dar los motivos de mis actos. (Si no me encuentro en una posición privilegiada para responder a las preguntas acerca de los motivos de mis actos, entonces no hay una razón especial por la que deba preguntárseme *a mí*.) Y lo importante de poder participar en este juego es que sólo a los que son capaces de participar en el juego de dar explicaciones se les puede o no hacer responsables del curso de sus actos o actitudes, y si somos incapaces de “oír razones” en algún asunto, entonces no se nos puede hacer responsables de dicho asunto. Las capacidades de comunicación verbal y de conciencia de nuestros propios actos son, pues, esenciales para quien esté dispuesto a oír argumentos y abierto a la persuasión, y semejante persuasión, semejante ajuste recíproco de intereses logrado mediante la utilización mutua de la racionalidad, es un rasgo de la modalidad óptima de la interacción personal.

Esta capacidad de participación en la persuasión mutua proporciona las bases para una nueva condición de la cualidad de persona, recientemente expuesta por Harry Frankfurt.²⁶ Frankfurt afirma que las personas son la subclase de los sistemas intencionales, capaces de lo que él llama *voliciones de segundo orden*. Ahora bien, a primera vista esto se parece a la clase de los sistemas intencionales de segundo orden, pero, como veremos, no es así.

²⁵ Se me puede hacer responsable de acontecimientos y estados de cosas de los que no fui consciente y debí serlo, pero éstos no son actos intencionales. En estos casos, soy responsable de estos otros asuntos, en virtud de ser responsable de las consecuencias predecibles de los actos —incluyendo actos de omisión— de los que sí era consciente.

²⁶ H. Frankfurt, “Freedom of the Will and the Concept of a Person”. Frankfurt no dice si concibe su condición como meramente necesaria o si la considera también condición suficiente de la cualidad moral de persona.

Además de querer y escoger, y ser llevados *a hacer* esto o aquello, los hombres pueden también querer tener (o no tener) determinados deseos y motivos. Son capaces de querer ser diferentes, en sus preferencias y propósitos, de lo que son... Sin embargo, no hay al parecer otro animal salvo el hombre que tenga la capacidad de autoevaluación reflexiva que se manifiesta en la formación de deseos de segundo orden (p. 7).

Frankfurt señala que hay casos en los que podría decirse que una persona quiere tener un deseo particular, aun cuando no quisiera que ése fuera un deseo real para él, que fuera “su voluntad”. (Por ejemplo, uno puede querer sentir el deseo de la heroína, tan solo para saber qué se sintió desear la heroína, sin querer en absoluto que este deseo se convierta en un deseo real.) En casos más graves, queremos tener un deseo que actualmente no tenemos, y queremos que este deseo se convierta en nuestra voluntad. Frankfurt llama a estos casos voliciones de segundo orden y, afirma, el tenerlos es “esencial para ser una persona” (p. 10). Su argumento en favor de esta afirmación, que no intentaré justificar aquí, procede de un análisis de la distinción entre tener libertad de acción y tener libertad de volición. Según su análisis, se tiene libertad de volición sólo cuando se puede tener la voluntad que se desea, cuando se pueden satisfacer las voliciones de segundo orden. Las personas no siempre gozan de libre albedrío y, bajo ciertas circunstancias, pueden ser responsables de acciones cometidas en ausencia de libertad de volición, pero una persona debe ser siempre una “entidad para la cual la libertad de su volición pueda resultar un problema” (p. 14) —es decir, una entidad capaz de estructurar voliciones de segundo orden, puedan éstas ser o no satisfechas. Frankfurt introduce el maravilloso término “*wanton*” para referirse a quienes “tienen deseos de primer orden pero... no voliciones de segundo orden”. (Para Frankfurt, las voliciones de segundo orden son to-

das, por supuesto, deseos de segundo orden *reflexivos*.) Él afirma que nuestras intuiciones apoyan la opinión de que todos los animales no humanos, así como los niños pequeños y algunos deficientes mentales son *wantons*, y yo soy de los que, entre otros, no puede pensar en posibles contraejemplos. En efecto, algo que parece reforzar su teoría es que, como él afirma, los seres humanos —las únicas personas que reconocemos como tales— se distinguen por esto de los animales. Pero ¿qué hay de especial en las voliciones de segundo orden? ¿Por qué, de entre las intenciones de orden superior, son éstas propiedad exclusiva de las personas? Porque, creo yo, la “autoevaluación reflexiva” de la que habla Frankfurt es, y debe ser, autoconciencia genuina, la cual sólo se obtiene adoptando hacia *uno mismo* la postura anscombiana del que hace preguntas y persuade, y no sólo la de comunicador. Como apunta Frankfurt, los deseos de segundo orden son una noción carente de sentido, a menos que se *actúe* sobre ellos, y actuar sobre un deseo de segundo orden debe ser lógicamente diferente de actuar sobre su componente de primer orden. Actuar sobre un deseo de segundo orden, hacer algo para obtener la adquisición de un deseo de primer orden, es actuar sobre uno mismo de la misma forma como se actuaría sobre otra persona: uno se *instruye* a sí mismo, uno se ofrece persuasiones, argumentos, amenazas, sobornos, con la esperanza de inducirse a uno mismo a adquirir el deseo de primer orden.²⁷ La postura que uno adopta ante sí mismo y el *acceso a uno*

²⁷ Se me ha hecho notar que los perros sementales se masturban con frecuencia aparentemente con el fin de *aumentar su deseo* de copular. Lo que resta importancia a estos casos es que, aun suponiendo que se pueda decir que el perro actúa sobre un deseo para reforzar un deseo, el efecto se logra en una forma no intencional (“puramente fisiológica”); el perro no recurre o no explota su propia racionalidad para alcanzar su fin. (Como si la única forma en que una persona pudiera actuar sobre una volición de segundo orden fuera tomando una pastilla o poniéndose de cabeza, etc.)

mismo son, en estos casos, esencialmente iguales a la postura adoptada ante otra persona y el acceso a ella. Uno debe *preguntarse a sí mismo* cuáles son sus deseos, motivos y razones reales y, sólo si uno puede responder a esta pregunta podrá darse cuenta de los propios deseos y podrá uno estar en posición de provocar el cambio en uno mismo.²⁸ Me parece que sólo aquí tenemos el caso de que el “orden presente” puede no estar ahí a menos que se presente en episodios de pensamiento consciente, en un diálogo con uno mismo.²⁹

Finalmente, ¿por qué no nos es posible afirmar que estas condiciones necesarias para la cualidad moral de persona son también suficientes? Simplemente porque el concepto de persona es, como he tratado de mostrar, inevitablemente normativo. Los seres humanos u otras entidades sólo pueden aspirar a ser aproximaciones al ideal, y no hay forma de establecer una “calificación aprobatoria” que no sea arbitraria. Aun considerando suficientes las seis condiciones (estrictamente interpretadas), no podrían garantizar que una entidad real cualquiera fuese

²⁸ Margaret Gilbert, en “Vices and Self-knowledge”, *Journal of Philosophy* (5 de agosto de 1971), p. 452, examina las implicaciones del hecho de que “cuando, y sólo cuando, uno cree tener una característica dada, puede uno decidir cambiar en función de ella”.

²⁹ Marx, en *La ideología alemana*, dice: “El lenguaje, como la conciencia, sólo surge de la necesidad, la urgencia, de tener relaciones con otros hombres. . . El lenguaje es tan antiguo como la conciencia, el lenguaje es una conciencia práctica.” Y Nietzsche, en *La gaya ciencia*, dice: “Pues podríamos, de hecho, pensar, sentir, desear y recolectar, podríamos igualmente ‘actuar’ en todos los sentidos del término, y sin embargo nada de ello requiere necesariamente ‘venir a la conciencia’ (como uno dice metafóricamente). . . ¿Cuál es entonces, generalmente, el propósito de la conciencia si es principalmente *superflua*? —Ahora bien, me parece, si quieren oír mi respuesta y su supuesto acaso extravagante, que la sutileza y la fuerza de la conciencia son siempre proporcionales a la *capacidad de comunicación* de un hombre (o de un animal), siendo la capacidad de comunicación, a su vez, proporcional a la *necesidad de comunicación*. . . En pocas palabras, el desarrollo del habla y el desarrollo de la conciencia (no de la razón, sino de la razón que se vuelve autoconsciente), van de la mano.”

una persona, pues nunca nada podría satisfacerlas. La noción moral de persona y la noción metafísica de persona no son conceptos separados y distintos, sino sólo dos puntos de apoyo diferentes e inestables de un mismo continuo. Esta relatividad contamina la satisfacción de las condiciones de la cualidad de persona en todos los niveles. No existe condición suficiente objetivamente satisfactoria que sostenga que una entidad *realmente* tiene creencias y, a medida que descubrimos la irracionalidad que se manifiesta bajo una interpretación intencional de una entidad, se debilitan nuestros fundamentos para atribuir creencias, particularmente cuando tenemos (lo que, en principio, siempre *podemos* tener) una descripción mecanicista, no intencional de la entidad. Exactamente en la misma forma, nuestra suposición de que una entidad es una persona se tambalea justo en los casos en que importa: cuando se ha hecho algo mal y cuando surge la pregunta respecto de la responsabilidad. Pues, en estos casos, las bases para decir que una persona es culpable (la evidencia de que obró mal, de que estaba consciente de que obraba mal y de que obró mal por su propia y libre voluntad) son, en sí mismas, bases para dudar de que estemos de hecho tratando con una persona. Y si se preguntara qué podría *resolver* nuestras dudas, la respuesta sería: nada. Cuando surgen semejantes problemas, ni siquiera podemos decir en nuestro propio caso si somos personas.

Condiciones de la cualidad de persona, No. 45 de la colección Cuadernos de *Crítica* del Instituto de Investigaciones Filosóficas, se terminó de imprimir el 30 de abril de 1989 en los talleres de Olmecca Impresiones Finas, S. A. de C. V. Para su composición y formación, realizadas en computadora en el mismo Instituto, siendo Jefe de Publicaciones Antonio Zirión Q., se utilizó el programa T_EX. La edición consta de 2,000 ejemplares.