

Ocho *desiderata* metodológicos de las teorías sociales normativas *

ANTONI DOMÈNECH

Universidad de Barcelona

Que la filosofía y la ciencia social normativas han experimentado en los últimos treinta años una explosión de productiva creatividad es cosa de la que apenas cabe dudar, si más no, a tenor del éxito conseguido en la lucha por arbitrar recursos y generar publicaciones. Desde finales de los sesenta, áreas de investigación enteras, como la filosofía política, la filosofía del derecho, la economía normativa y la ética social, durante décadas escasa o rutinariamente cultivadas, han conseguido atraer la atención de un número creciente de investigadores entusiastas y excelentes, que han removido los otrora apáticos cimientos de sus disciplinas.

Como suele ocurrir cuando un área de trabajo altera subitánea y radicalmente su faz—como ocurre ahora mismo, por ejemplo, en la revolución intelectual que están provocando las ciencias cognitivas—, el estallido innovador en filosofía y ciencia social normativas, a la vez que un campo imantado de fascinación, ha generado un poco de desconcierto y desorientación: proliferan los conceptos y los argumentos a tal punto, que, de un lado, se hace por momentos difícil, si no imposible, mantener una perspectiva global e informada; del otro, se hacen perdidosos los criterios

evaluativos generales que permiten juzgar de esos conceptos y de esos argumentos.

Las páginas que siguen son un modesto ejercicio metodológico de orientación. Presentan y discuten brevemente ocho *desiderata* metodológicos, ocho criterios de evaluación de la bondad de las teorías sociales normativas. Algunos de los criterios presentados tienen su correlato en criterios metodológicos amplios, que no aproblemáticamente, aceptados para las teorías (sociales) positivas o descriptivas; otros, responden a inquietudes y problemas específicamente suscitados por la reflexión social normativa.

Los ocho criterios aquí presentados se examinan y discuten de manera informal, a título de exploración heurística; huelga, pues, decir que no se ofrece prueba concluyente alguna en favor de la independencia conceptual de unos respecto de los otros (ni menos, claro está, de su mutua compatibilidad). No obstante, cuando el espacio lo permite, se trata explícitamente la interrelación de varios de ellos, apuntando tal vez a vías por las que un tratamiento más formal y técnico de esos *desiderata* podría llegar a reducir algunos a la condición de meros teoremas, derivables de otros (o a mostrar incompatibilidades mutuas).

* Una versión anterior de este texto fue escrita como material didáctico en 1993 para el Seminario permanente de Ética, Economía y Sociedad de la Facultad de Ciencias Económicas de la Universidad de Barcelona. Algunos amigos y colegas, como Andrés de Francisco, Félix Ovejero, Fernando Aguiar, Angel Puyol, Daniel Raventós y Jordi Mundó han tenido la generosidad de citar repetidamente en sus escritos o en sus conferencias aquella versión inédita, y, con Paco Álvarez y Salvador Giner, han acabado por convencerme de que podía tener alguna utilidad que no fuera puramente didáctica. La nueva versión que ahora se publica está escrita en buena medida respondiendo a su solicitud. Espero al menos que no les defraude.

1. Consistencia

1.1. Consistencia lógica

La consistencia lógica es el primero y más natural de los requisitos cuyo cumplimiento cabe exigir a una teoría social normativa. Tiene un obvio correlato en el *desideratum* de consistencia exigible a las teorías positivas: la teoría no puede sostener cosas lógicamente contradictorias; de los axiomas de una teoría no pueden derivarse A y no-A a la vez. Como es harto sabido, de una teoría lógicamente inconsistente puede inferirse deductivamente cualquier enunciado, en particular en el caso de las teorías sociales normativas, puede inferirse que una sociedad esclavista es justa, y su contrario, que es injusta. En el mejor de los casos, pues, una teoría social normativa lógicamente inconsistente sería inútil; en el peor, un arma al servicio de las causas más deleznable que imaginar quepa¹.

Pero las teorías sociales normativas pueden traer consigo formas específicas de inconsistencia, desconocidas o insólitas en el ámbito de la ciencia positiva. A evitarlas van dirigidos los dos subcriterios que siguen.

1.2. No-autonulificatoriedad

Parece exigible que las teorías sociales normativas no aboguen por una conducta de los agentes sociales que viole sistemáticamente los valores fundamentales defendidos por la teoría en cuestión. Tal vez el ejemplo más llamativo de inconsistencia, en este sentido de autonulificación, lo proporcione la teoría normativa paretiana de la «utilidad» y la «ofelinidad» (Pareto, 1913). De acuerdo con ella, los agentes sociales tienen dos tipos de preferencias: las personales (representables por una función de «ofelinidad») y las sociales (representables por una función de «utilidad»); las primeras tocan a lo que promueve el bienestar del propio agente; las segundas, tienen que ver con lo que el agente considera lo mejor para la sociedad, globalmente considerada. Según la doctrina ético-social paretiana —que puede entenderse como una versión técnicamente refinada del viejo ideologema de Mandeville sobre los vicios privados que traen consigo el beneficio público—, la mejor manera de promover las preferencias sociales altruistas de los agentes económicos (de maximizar la función de «utilidad») sería que los agentes económicos tuvieran preferencias personales egoís-

¹ He aquí un ejemplo casi insuperable de apología de la inconsistencia lógica en la teorización normativa: «Nosotros los fascistas hemos tenido el valor de hacer a un lado todas las teorías políticas tradicionales, y somos aristócratas y demócratas, revolucionarios y reaccionarios, proletarios y antiproletarios, pacifistas y antipacifistas.» Esta cita de Benito Mussolini procede de Franz Neumann (1983: 511).

tas y actuaran racionalmente de acuerdo con ellas (maximizando una función de «ofelinidad» egoísta).

1.3. *Consistencia performativa*

Cuando alguien dice cosas por el estilo de: «Llueve, pero yo no creo que llueva», incurre en una inconsistencia performativa. El contenido de lo que afirma no es propiamente inconsistente; lo que es inconsistente es el acto de afirmarlo. Una ilustración clara de esa inconsistencia performativa son las doctrinas, de estirpe sofista, que afirman el «egoísmo moral». De acuerdo con ellas, no hay ningún bien —moral, o de otro tipo— que no sea servicial, es decir, que no sea sostenido o abrazado o afirmado *pro domo mea*, puesto, esto es, al servicio del interés propio, y por eso mismo, embozado hipócritamente en el cendal de las causas comunes. Ahora bien; sostener eso, o es autorrefutatorio (si suponemos que quien lo dice, lo dice desinteresadamente), o es performativamente inconsistente, en la medida en que no se ve qué interés propio puede promover la paladina declaración de que nada se hace o se dice desinteresadamente (Domènech, 1997b). En substancia, es una inconsistencia performativa decir: «Todo el mundo encubre su voluntad de poder tras una máscara de bienes pretendidamente comunes, y yo también»².

2. *Informatividad*

También el criterio de informatividad tiene su correlato en la ciencia positiva. Una teoría —positiva o normativa— es *ceteris paribus* tanto mejor cuanto más informativa, es decir, cuantos más mundos posibles excluya. Una teoría —positiva o normativa— es tanto más informativa cuantos más mundos posibles sean incompatibles con ella. La informatividad de una teoría positiva suele asociarse con su capacidad predictiva: la teoría positiva más informativa de que disponemos hoy es la electrodinámica cuántica, capaz de hacer predicciones con una exactitud de hasta once decimales [que es, más o menos, como predecir la distancia entre Los Ángeles y Nueva York con un margen de error del diámetro de un pelo humano (Feynman, 1985: 7)]. Pero lo que aquí nos interesa es la informatividad de las teorías sociales normativas.

Una teoría social normativa es tanto más informativa cuantos más mundos sociales posibles excluye como ético-socialmente indeseables. En el límite, cumpliría óptimamente con este *desideratum* una teoría que, de un conjunto infinito

² La posición de los defensores de la llamada «ética comunicativa» o discursiva (Habermas, 1983) se puede entender como un intento —fallido (Tugendhat, 1993: 166 ss.)— de usar la prohibición de incurrir en contradicciones performativas como la pieza angular de una ética social firmemente anclada en la racionalidad, es decir: como un intento de encargar a un solo *desideratum* metodológico todo el trabajo de determinación del espacio normativo.

de ordenamientos sociales reputados posibles por ella, seleccionara como ético-socialmente deseable sólo uno de esos ordenamientos, y excluyera a todos los demás. En el extremo opuesto estarían las «teorías» ético-sociales compatibles con todo o con casi todo: lo mismo con tiranías abyectas, que con los más regalados libertinajes; al par con sociedades procusteanamente igualitarias y con órdenes sociales en la cúspide de la desigualdad y la polarización. Estas últimas, las «teorías» poco o nada informativas, serían incapaces de seleccionar mundos sociales posibles; serían «teorías» sociales normativas que se conformarían si no con cualquier cosa, con demasiadas.

La teoría utilitarista clásica (cardinalista, con utilidades cardinales interpersonalmente comparables), por ejemplo, cumple excelentemente el criterio de informatividad: de un conjunto infinito de mundos sociales reputados posibles por ella, selecciona como justo sólo uno de ellos, el que maximiza la función social de utilidad agregada, excluyendo a todos los demás. En efecto: el dominio de la función de utilidad social —que representa los mundos socialmente posibles para la teoría— es infinito, pero sólo hay un valor máximo en el contradominio de la función, el cual valor se corresponde funcionalmente con el mundo social elegido como justo por el utilitarismo clásico.

En cambio, el utilitarismo ordinalista de la llamada «nueva economía de bienestar» (dominante desde finales de los años treinta del siglo xx) no cumple bien el criterio de informatividad. Pues, al computar sólo la información procedente de las utilidades ordinales de los individuos, no puede agregar esas utilidades con operaciones algebraicas como las de la suma o el producto. El utilitarismo ordinalista tiene técnicamente vedado el criterio clásico de maximizar la suma (o el producto) de las utilidades individuales. El criterio ordinalista de justicia tiene que limitarse a la eficiencia u optimalidad paretiana. De acuerdo con el cual, es justa aquella sociedad en la que no puede incrementarse la utilidad de nadie, sin generar desutilidad al menos a uno. Como es bien sabido, en principio, hay muchas distribuciones del producto social compatibles con el criterio de optimalidad o eficiencia, desde sociedades insuperablemente desiguales, hasta sociedades muy igualitarias. Sin duda, este criterio utilitarista-ordinalista no es totalmente ininformativo, porque excluye muchos mundos sociales caracterizados por la ineficiencia económica; pero no es lo bastante informativo, porque en la llamada frontera de la optimalidad paretiana pueden coexistir las más dispares distribuciones del producto social³.

³ En obsequio a la brevedad, supongamos una sociedad compuesta por sólo dos individuos, A y B. Supongamos que el producto social que hay que distribuir entre A y B asciende a un millón de euros. La frontera paretiana de eficiencia son todas aquellas distribuciones del producto social entre A y B que suman un millón de euros. Por ejemplo, si A recibe medio millón, y B, el otro medio millón, estamos en un óptimo de Pareto; si A recibe un millón, y B, cero, también estamos en un óptimo; si B recibe un millón, y A, cero, también estamos en un óptimo; si A recibe veinte mil euros y B recibe novecientos ochenta mil euros, también estamos en un óptimo. Sólo quedan excluidas por el criterio ordinalista de optimalidad aquellas situaciones sociales en las que lo que reciban A y B juntos sea inferior a un millón de euros: por ejemplo, si A recibe

La informatividad es un *desideratum* muy importante. Y el que el utilitarismo clásico cumpliera a satisfacción este criterio metodológico puede explicar en parte la larga pervivencia del cardinalismo, a pesar de las devastadoras críticas a que fue sometido casi desde sus orígenes, incluidas las censuras procedentes de la «nueva» economía ordinalista de bienestar. Tal vez explique también, parcialmente, el éxito fulminante de la teoría de la justicia como equidad, de John Rawls, a partir de los años setenta. Rawls no ha sido muy original en su pliego de críticas a las teorías utilitaristas; en cierta medida, se limitó a compendiar y a sistematizar objeciones que se callaban por sabidas. Pero, por vez primera, consiguió articular conceptualmente una teoría global alternativa, casi tan informativa como la teoría utilitarista clásica, y en cualquier caso, mucho más informativa que la de la nueva economía de bienestar.

Sorprendente y admirable en la teoría de la justicia como equidad es que sea tan selectiva —tan informativa—, sirviéndose de instrumentos formales de agregación muchísimo más sencillos y pedestres que los de la «vieja» y la «nueva» economía de bienestar. Desde el punto de vista de las técnicas formales empleadas, Rawls procede del modo siguiente.

En primer lugar, parte de un conjunto de bienes —los bienes primarios— que no podrían ser más heterogéneos, porque no admiten comparaciones entre ellos, ni tienen en común sino el ser bienes aptos, en principio, para todos los propósitos, bienes instrumentales, de uso universal, necesarios para promover los más variados planes de vida concebibles (al menos, en nuestra cultura). Rawls agrega esos bienes heterogéneos mediante el sencillo expediente de ordenarlos lexicográficamente: libertades públicas básicas, igualdad equitativa de oportunidades, ingresos, etcétera ⁴.

A continuación, Rawls formula un criterio de distribución de esos bienes primarios. Y ese criterio de distribución tiene también una estructura formal lexicográfica. Primero, las máximas libertades públicas iguales para todos, luego la igualdad equitativa de oportunidades, y finalmente, la distribución de ingresos de acuerdo con un principio maximin: entre dos sociedades que respeten por igual los dos primeros principios (libertades e igualdad equitativa de oportunidades), preferiremos aquella en la que los más pobres vivan mejor; entre dos sociedades que respeten por igual los dos primeros principios y en las que los más pobres tengan el mismo nivel de ingresos, preferiremos aquella en la que los segundos más pobres vivan mejor; entre dos sociedades que respeten los dos primeros principios y en las que los más pobres y los segundos

cuatrocientos noventa y nueve mil novecientos noventa y nueve euros, y B, otro tanto, estamos fuera de la frontera de eficiencia, porque hay un despilfarro de dos euros.

⁴ Un orden lexicográfico —el de las palabras en el diccionario— se define formalmente así: el par (X_1, Y_1) es lexicográficamente preferido al par (X_2, Y_2) , si y sólo si:

1. $X_1 > X_2$; o bien:
2. si $X_1 = X_2$, entonces $Y_1 > Y_2$.

Para una demostración formal de que el índice de bienes primarios rawlsiano tiene que ser un orden lexicográfico véase Blair (1988).

más pobres tengan el mismo nivel de ingresos, preferiremos aquella en la que los terceros más pobres tengan mejores ingresos. Y así sucesivamente.

El tipo de agregación de los *distribuida* y el criterio de distribución rawlsianos (que, juntos, componen lo que se conoce formalmente como un procedimiento *bilexicográfico maximin*) funcionan como un eficiente mecanismo de desempate, merced al cual, entre dos sociedades cualesquiera, siempre podemos seleccionar la más justa (Domènech, 1996). Por motivos técnicos que no hacen aquí al caso, los índices lexicográficos no pueden ser maximizados, de manera que, en el esquema conceptual rawlsiano —a diferencia de lo que ocurre con el utilitarismo cardinalista clásico—, no tendría siquiera sentido plantearse el problema de qué tipo de sociedad, de entre todas las posibles, satisface máximamente todos los requisitos de la teoría. Pero el poder decidir siempre, entre dos sociedades cualesquiera, cuál sea la más justa de acuerdo con la teoría de la justicia como equidad parece cumplir muy razonablemente el *desideratum* de la informatividad.

Menos prometedora, en punto a informatividad, es la teoría nozickiana de la justicia (Nozick, 1974). Justa es, en substancia, para esta teoría, toda sociedad que satisfaga las siguientes restricciones: 1) respetar el derecho de autopropiedad (en el que se incluye el derecho a los frutos del propio trabajo); 2) respetar unas determinadas condiciones de apropiación originaria de recursos externos (la famosa «cláusula de Locke»), y 3) respetar los resultados de los intercambios libremente consentidos de bienes y servicios (ya se trate de intercambios egoístas, mediados por el mercado; ya de intercambios altruistas, como las donaciones, las contradonaciones y los legados). En el supuesto de que una sociedad no haya respetado alguna de estas restricciones, puede aún caminar por la senda de la justicia nozickiana procediendo 4) a reparaciones que corrijan y rectifiquen las violaciones históricamente registradas de los derechos postulados por la teoría. La teoría es, pues, puramente procedimental, o según lo expresa Nozick, no-configurada: no prefigura, ni perfila, el tipo de sociedad que habrá de resultar del respeto de esas restricciones. Siendo esto así, cabe imaginar como nozickianamente justa no sólo una sociedad en la que todo intercambio de bienes y servicios se produce en la esfera egoísta del mercado —resultado que tan frecuente como injustificadamente se considera dilecto de la teoría nozickiana—, sino también una sociedad de reciprocidad altruista general (Kolm, 1985). Ni siquiera es descartable, en este esquema teórico, una sociedad con una fuerte presencia invasora del Estado, hecha necesaria para proceder a reparaciones históricas a gran escala y acaso apremiantemente solicitada por la cuarta restricción de la teoría. Así pues, la concepción nozickiana de la justicia no satisface adecuadamente el *desideratum* de informatividad o selectividad: es compatible con mundos sociales muy diversos. La conocida cautela con la que Robert Nozick ha defendido el procedimentalismo deontológico de su concepción de la justicia, y a tenor de la cual las restricciones bastan a determinar la justicia de una sociedad, «siempre que no se produzcan

catástrofes morales»⁵, es seguramente el precio que este inteligente filósofo se ha creído en el caso de pagar por la debilidad informativa de su teoría social normativa.

3. *Equilibrio reflexivo*

Una antigua y venerable tradición de filosofía de la ciencia y del conocimiento sostiene que las teorías positivas se contrastan con los hechos y quedan determinadas por ellos. Esta tesis ha sido puesta bajo sospecha en las últimas décadas. Es bien sabido que, a barlovento de esa sospecha, se ha llegado incluso a negar que haya hechos brutos, independientes de todo esfuerzo teorizador. Pero, extremismos folclóricos aparte⁶, la crítica más solvente de la tesis clásica procede de Quine (1962), quien en substancia sostuvo convincentemente hace ya cuatro décadas que todos los hechos del mundo no bastan a determinar concluyentemente la aptitud o inaptitud de una teoría: las teorías están *subdeterminadas* por los hechos. Esto equivale a la tesis moderada, según la cual, a la hora de elegir entre dos teorías positivas cualesquiera, la prueba de los hechos no es la prueba definitiva, sino *una* prueba, todo lo importante que se quiera, pero no más que *una* prueba, *un* criterio de selección, entre otros. Si nuestra capacidad para imaginar hipótesis teóricas es potencialmente ilimitada, entonces necesitamos cribas, filtros, que nos permitan seleccionar entre esas hipótesis; criterios o dispositivos criteriosales que no hagan totalmente *indeterminada* nuestra selección, que determinen el espacio de las hipótesis teóricas. Es verdad que los hechos no determinan completamente ese espacio, que no nos llevan derechamente a la elección de una teoría o de una hipótesis teórica. Es verdad que las teorías están siempre *subdeterminadas* por los hechos. Pero se puede decir que, en las teorías positivas, los hechos *contribuyen* a determinar el espacio teórico.

Aparentemente, no hay equivalente alguno de ese factor determinante en las teorías sociales normativas. En principio, las teorías normativas no pueden contrastarse con los hechos. (Sostener lo contrario, sin mayores cualificaciones, sería cometer la falacia naturalista.) Ahora bien; parece que el espacio conceptual de las teorías sociales normativas, como el de las teorías positivas, necesita determinación: ¿cómo, si no, elegir entre teorías normativas en conflicto? La negativa a determinar el espacio teórico-normativo equivale a abandonar por completo y sin restos toda posibilidad de discusión racional en asuntos de ética social, y aun de ética *tout court*: sería reconocer como un *fatum* trágico la existencia de un «politeísmo ético» irreductible a la deliberación racional,

⁵ Amartya Sen (1981) exploró una de esas «catástrofes morales», resultado de procesos exquisitamente respetuosos del derecho de propiedad privada: la miseria y la hambruna en la Bengala de 1943, y en Etiopía, Bangladesh y los países del Sahel en los setenta.

⁶ Para una crítica, tan mesurada como devastadora, de esos extremismos, que implican un desatentado relativismo ontológico, epistemológico y, naturalmente, ético, véase Searle (1997).

por decirlo con la celebrada expresión de Max Weber. (Lo mismo que la negativa a determinar el espacio teórico-positivo, sosteniendo, por ejemplo, que toda realidad es realidad socialmente construida —y toda verdad, verdad servicial y banderiza—, equivale a abandonar por completo toda posibilidad de discusión gnoseológica racional.)

Si la contrastación con los hechos empíricos es, cuando menos, una fuente importante de determinación de las teorías positivas, y si aparentemente las teorías normativas no pueden contrastarse con los hechos, ¿en contraste con qué habríamos de empezar a determinar y a restringir el espacio de posibilidades de estas últimas?

El criterio del equilibrio reflexivo, popularizado por John Rawls⁷, sostiene que las teorías normativas, y en particular, las teorías sociales normativas, son susceptibles de contrastación con una peculiarísima clase de hechos: la de nuestras intuiciones morales. Ahora bien; esa singular clase de hechos que son nuestras intuiciones morales presenta también una serie de problemas específicos, que hacen del ejercicio de contrastación de las teorías normativas una empresa particularmente delicada. He aquí un breve repaso de algunos de esos problemas:

3.1. *La posible naturaleza tornadiza e incoherente de nuestras intuiciones y emociones morales*

Nuestras intuiciones morales son el precipitado de complicadas —y aun no cabalmente entendidas— pautas de evolución, de evolución biológica y de evolución cultural. Es verdad que los procesos evolucionarios son procesos algorítmicos de diseño (Dennett, 1995), y en este sentido, más que ser un precipitado, cabría decir de nuestras intuiciones y de nuestras emociones morales que son un producto o un resultado con bastante articulación funcional. Sin embargo, todo lo que sabemos en general sobre los procesos algorítmicos de diseño biológico (sobre los procesos de diseño propios de la evolución cultural tenemos hipótesis aún menos firmes) apunta a que sus productos no son cabalmente fiables en punto a coherencia, sistematicidad y estructura globales. La selección natural —una de las principales; si no la principal fuerza diseñadora en la evolución biológica— procede oportunista y miope: ni retrocede a la hora de aprovechar ventajas coyunturales, ni ve más allá de esas ventajas; nunca construye *ex amorphos hylés*, sino aprovechando y reaprovechando material ya estructurado. Muy bien pudiera ser que nuestras intuiciones y nuestras emociones morales de raíz biológica (por no hablar de las culturalmente formadas) se solaparan, fueran globalmente incoherentes, y hasta, en determinados

⁷ Con el interesante antecedente de Nelson Goodman (1965), que formuló explícitamente este criterio para justificar las reglas básicas de inferencia deductiva e inductiva. Véase Domènech (1997a: 250 ss.)

contextos, contradictorias. La universal experiencia humana de los dilemas éticos, tan explotada como recurso literario y de reflexión filosófica, bien pudiera echar sus raíces en el carácter tornadizo y poco coherente de nuestras intuiciones y emociones morales. Por eso el equilibrio que se busca en el presente *desideratum* entre nuestras teorías normativas y nuestras intuiciones morales tiene que ser un equilibrio *reflexivo*, esto es, un equilibrio con ida y vuelta. Nuestras teorías normativas deben ser a satisfacción de nuestras intuiciones (es más: en buena medida, lo que tratan de hacer las teorías normativas es captar adecuadamente el núcleo de nuestras intuiciones)⁸, pero tampoco hay que descartar que, una vez ordenadas y sistematizadas nuestras intuiciones morales por una teoría normativa consistente e informativa, ésta nos ayude a ver las limitaciones o las incoherencias de esas intuiciones, conminándonos a enmendarlas y aun a podarlas⁹. El equilibrio reflexivo se alcanzaría idealmente cuando nuestra

⁸ Gracias a las investigaciones psicológico-cognitivas realizadas en los últimos años sobre lo que se ha venido en llamar *folk-science*, es decir, sobre la sistematización teórica de los complejos de intuiciones con que está dotada la arquitectura cognitiva humana en diversos dominios (geometría, física, biología, psicología, sociología), estamos ahora, según creo, mejor preparados para entender lo que quiere decir «captar adecuadamente el núcleo de nuestras intuiciones». En este sentido, la axiomatización de Euclides captó el núcleo de nuestras intuiciones sobre la estructura geométrica del espacio. Una reivindicación de la física de Aristóteles —entendida esta última como intento de captar el núcleo de nuestras intuiciones físicas— puede verse en McCloskey (1983a; 1983b). Sobre biología *folk*, véase Atran (1990). Sobre sociología *folk*, véase Hirschfeld (1994). Sobre *folk psychology* la bibliografía es ya literalmente inabarcable: el disparo de salida lo dio el polémico Churchland (1981). A propósito de todo lo cual, vale la pena observar lo siguiente. Cuando hacemos ciencia positiva o descriptiva podemos contar con una doble determinación de nuestras teorías: la que viene de las intuiciones ínsitas en nuestra arquitectura cognitiva (que nos proporciona o determina el repertorio de nuestras primeras conjeturas e hipótesis) y la que viene de los hechos a los que sometemos nuestras hipótesis (hechos que pueden dar en la concluyente refutación de nuestras teorías *folk*: como ocurrió con la física aristotélica, o como ocurrió con la geometría euclídea en tanto que teoría empírica sobre la estructura del espacio). En cambio, cuando hacemos ciencia o filosofía social normativas, sólo podemos contar con la primera de las determinaciones, lo que monta tanto como decir que ninguna teoría normativa puede rebasar holgadamente el estadio *folk*.

⁹ El que podamos enmendar y podar nuestras intuiciones no implica que podamos rebasar el carácter *folk* de nuestras teorías sociales normativas (véase la nota anterior). Ocurre simplemente que las intuiciones arraigadas en nuestra arquitectura cognitiva y emocional están verosíblemente regimentadas en dominios específicos (Hirschfeld y Gelman, 1994; Barkow, Cosmides y Tooby, 1992). Al aplicar a esas intuiciones esquemas generales o universales de razonamiento, podemos tratar de trascender esa regimentación y buscar una coherencia interespecífica. Se trata de un ejercicio de distanciamiento intelectual respecto de los dominios específicos en que están estructuradas nuestras intuiciones —morales o de otro tipo—, de un ejercicio en el que se despliega una intencionalidad de segundo orden. Ésta, en cualquier caso, podría ser una de las justificaciones psicológicas más profundas de la idea, según la cual no hay vida moral ni buen conocimiento sin intencionalidad de segundo orden (sin preferencias sobre preferencias, o sin creencias sobre creencias): véase Frankfurt (1971), Domènech (1989, 1997), Lehrer (1990).

teoría acabara casando con nuestras intuiciones, tras un período más o menos dilatado de reflexión y enmienda mutua ¹⁰.

3.2. *La diversidad de nuestras intuiciones y emociones morales*

Sin embargo, no hay por qué presuponer que, aun mudadizas e incoherentes, nuestras intuiciones y emociones morales son idénticas para todos los individuos. En cierto sentido —que habrá de aclararse más adelante—, es de presumir lo contrario. Por lo pronto, un teorema de la genética de poblaciones nos asegura la diversidad biológica dentro de cualquier clan darwiniano. Consecuencia de lo cual es: podemos estar razonablemente seguros de que, en lo que a sus raíces biológicas hace, nuestras intuiciones y nuestras emociones morales son, *hasta cierto punto* (ya tendremos ocasión de notar cuán importante es este subrayado), diversas; no todos los individuos de la especie *homo sapiens* compartimos exactamente las mismas. Más patente aún se hace este problema cuando consideramos las intuiciones y las emociones culturalmente moldeadas. Si hay un sentido en el que se puede decir que la vergüenza es una emoción moral universal, en tanto que biológicamente anclada (Demócrito presentó la *aídesthai autón*, la vergüenza de sí propio, como origen y fundamento de toda ética), de ninguna manera cabe decir lo mismo de la emoción moral de la culpa, que tiene una impronta de todo punto cultural y es, civilizatoriamente hablando, parroquiiana (Singer, 1953; Gibbard, 1990: 145 s.).

Pero ni siquiera dentro de una cultura dada hay que suponer homogeneidad en las intuiciones y en las emociones morales de la gente. Las sociedades están atravesadas por conflictos, escindidas en intereses de clase social y de otros tipos, crecientemente patentes. Es de temer que las intuiciones y las emociones morales estén, si no moldeadas, sí al menos sesgadas por esos intereses encontrados.

A la vista de todo esto, ¿se puede defender todavía, sin avilantez ni ingenuidad, el criterio del equilibrio reflexivo? Vayamos por partes. La diversidad biológica, el hecho de que no todos los humanos tengan exactamente las mismas

¹⁰ Una objeción común al equilibrio reflexivo es la de que es circular. Apela, por un lado, a nuestras intuiciones para fundar las teorías normativas, pero luego usa esas mismas teorías para corregir nuestras intuiciones. Esta objeción valdría también para un buen número de análisis aparentemente circulares de fenómenos lingüísticos, epistemológicos, computacionales y sociales. Entre estos últimos está el fenómeno del «conocimiento común» (yo sé que tú sabes que yo sé que tú sabes que yo sé que tú sabes que yo sé que tú sabes, etc.), que es un supuesto que parece imprescindible para la aplicación de la teoría de los juegos a los problemas sociales. Cuando esta objeción se hace técnicamente refinada sostiene dos tesis: que cualquier cosa que pretenda rigor matemático debe poder reducirse a la teoría de conjuntos, y que la única concepción coherente de la teoría de conjuntos prohíbe este tipo de circularidades. Para una refutación de estas dos tesis, pero especialmente de la segunda, véase Barwise y Moss (1996), en donde, apoyándose en la nueva teoría de los hiperconjuntos (o de los «conjuntos no-bien-fundados»), se construye un interesante esquema formal en el que cobijar a este tipo de análisis de fenómenos circulares.

intuiciones morales de raíz biológica —el hecho, por ejemplo, de que no todos los humanos experimenten la emoción de la vergüenza con la misma intensidad—, no presenta un problema irresoluble. Puede incluso ser una ventaja, si tenemos en cuenta la posibilidad a la que atendimos al comienzo de la presente sección: la posibilidad de que nuestras intuiciones morales individuales sean incoherentes o estén deficientemente articuladas. Pues de las dos tesis juntas (la de los «defectos» de nuestras intuiciones y la de la diversidad de las mismas), se sigue que habrá también cierta diversidad en la distribución de los «defectos», lo que abre la vía para la mutua corrección de los mismos: podemos servirnos de dos herramientas diversamente defectuosas para reparar o mitigar con una los defectos de la otra.

También nuestras intuiciones cognitivas e inferenciales individuales son deficientes: la investigación empírica de nuestras capacidades y habilidades cognitivas, por ejemplo, ha ilustrado abrumadoramente en los últimos años, entre otros, abundantes fallos inferenciales en el razonamiento humano. Por poner un ejemplo extremo, puestos ante una implicación material y ante el consecuente de esa implicación, en muchos contextos la mayoría de la gente (más de dos tercios, incluyendo a académicos intelectualmente entrenados) comete la falacia de la afirmación del antecedente, es decir, viola una regla elemental de inferencia deductiva de la lógica de enunciados (Cosmides, 1992: 181-4). Si, a pesar de que en muchos contextos la mayoría de los humanos cometemos la falacia de afirmación del antecedente, todos aceptamos, sin embargo, la validez normativa de la regla lógica que la prohíbe, es porque, tras la debida reflexión, la minoría que no comete espontáneamente la mencionada falacia de inferencia —pero cuyas intuiciones inferenciales tienen seguramente «defectos» que otros no tienen— ha conseguido reparar o corregir o mitigar el «defecto» inferencial de la mayoría (y sin duda también, obvio es decirlo, porque en otros contextos —en contextos de intercambio social— la mayoría tampoco comete esta falacia).

Eso explica en buena medida el que una condición necesaria —aunque no suficiente— del progreso del conocimiento, lo mismo que del progreso ético, sea el uso público de la razón: *nemo solus satis sapit*. Lo que viene directamente en apoyo de la idea de que el equilibrio reflexivo entre nuestras teorías normativas (tanto morales como epistémicas) requiere la deliberación pública racional¹¹.

Ahora bien, los —imprescindibles— procesos de deliberación pública racional, que ponen en común nuestras diversas y «defectuosas» intuiciones y las corrigen y reparan, podrían verse frustrados en al menos dos casos: *a)* cuando la divergencia de esas intuiciones es demasiado radical, o *b)* cuando esas intuiciones están demasiado fuertemente sesgadas por intereses pugnaces:

¹¹ Eso es casi tanto como decir que el equilibrio reflexivo sólo es posible en condiciones de democracia deliberativa.

a) Como ya se sugirió, las intuiciones y las emociones morales más culturalmente moldeadas —menos biológicamente arraigadas— pueden presentar una diversidad tal, que haga imposible la deliberación racional entre culturas muy distintas. No digo que esto sea así. Pero si lo fuera, entonces habría que restringir a una sola tradición histórico-cultural la aspiración a un equilibrio reflexivo. (Rawls, por ejemplo, y de manera señalada en su última época, parece haber tomado este partido al insistir en que su teoría de la justicia vale sólo para el ámbito de nuestras sociedades industrialmente avanzadas y democráticas.)

b) Pero aun restringiéndonos a un solo tipo de sociedad, con una determinada tradición histórico-cultural, se nos plantea el problema de la división de la sociedad en grupos de intereses encontrados y aun hostiles. Se nos plantea el problema de que las intuiciones y las emociones morales de la gente pueden estar irreparablemente sesgadas por sus intereses¹². No hay deliberación pública racional posible ni en condiciones de guerra de clases abierta ni en el *bellum omnium contra omnes*. Lo único que cabe decir aquí es que estas situaciones extremas —pero ni mucho menos infrecuentes— caen fuera del espacio de la justicia, por lo que sale sobrando en ellas cualquier teoría social normativa¹³.

¹² Se puede plantear el problema, aún más complicado, de que una diferencia de intuiciones morales de origen biológico sea amplificada por la elaboración cultural de esas intuiciones. Así, por ejemplo, determinadas hipótesis biológico-evolucionarias sobre la filogénesis humana y el estudio etológico del comportamiento de nuestra especie coinciden en insistir en el llamado «lado oscuro de la sexualidad masculina» (Wilson y Daly, 1992). De acuerdo con esta tesis, el comportamiento sexual masculino estaría en buena medida guiado por el temor del «cuco»: por el temor a la promiscuidad de su pareja femenina y a la consiguiente inversión de recursos propios en la cría de hijos ajenos. De ahí derivarían unas tendencias «propietaristas» sobre las mujeres, es decir, unas intuiciones morales tendentes a considerar a la mujer como una propiedad. De esas intuiciones digamos «naturales» —por seguir con la hipótesis— pueden hacerse elaboraciones culturales muy distintas: desde la «mujer, la pata quebrada y en casa» de nuestra cultura, hasta la ablación de clítoris, habitual en ciertas culturas norte y centroafricanas, pasando por el *chador* islámico y el vendaje y la molturación de los huesos de los pies de la tradición china. En este tipo de culturas, que amplifican hasta lo execrable disposiciones de raíz presumiblemente biológica, parece difícil hallar soluciones menos radicales que la puesta por obra por el gobierno revolucionario de China en 1949: barrer sin contemplaciones toda una tradición cultural.

¹³ Una dilatada línea intelectual que, arrancando de los sofistas y pasando por Hobbes y Nietzsche, desemboca en cierta posmodernidad pretendidamente radical sostiene que las sociedades humanas viven permanentemente en estas situaciones extremas, y que no hay posibilidad de deliberación racional en modo alguno. Ahora bien, aparte del presumible carácter autorrefutatorio de este tipo de tesis, hay que decir que están además infundadas históricamente. La deliberación pública institucionalizada es un invento de la democracia republicana antigua. En la primera tipología que nos es conocida de los regímenes políticos, la de Heródoto (III, 80) en el siglo V antes de nuestra era, se hace el elogio de la democracia —contra la monarquía y contra la oligarquía— diciendo que en ella «se rinden cuentas al pueblo de la autoridad que se ejerce, y todas las deliberaciones están sometidas al público». Pero la democracia deliberativa republicana fue precisamente inventada en un momento extraordinariamente agudo de lucha de clases: fue inventada por la Atenas de la revolución aeropagita. Basta leer la obra maestra del clasicista Ste. Croix (1981) para convencerse. Y dicho sea de paso: los inauguradores de la línea de pen-

3.3. Las «funciones propias» de nuestras intuiciones morales

Venimos al problema seguramente más interesante que se le plantea al *desideratum* del equilibrio reflexivo. Y es el siguiente: es verdad que hay que suponer que, biológicamente hablando, nuestras intuiciones y nuestras emociones morales son diversas. Ya nos hemos ocupado de este asunto, introduciendo la cautela de que son diversas «hasta cierto punto»; es decir, que no son radicalmente diversas; no al punto de impedir o comprometer seriamente los procesos de puesta en común deliberativos. Parece firme la conclusión de que esto es así, entre otras razones de peso, por la muy evidente de que una especie esencialmente social como la nuestra no hubiera podido prosperar biológicamente de otro modo. Esto equivale a la tesis de que nuestras intuiciones morales innatas (como nuestra arquitectura cognitiva innata) tienen una estructura relativamente homogénea y funcionalmente integrada, a par de regimentada en módulos o dominios específicos. El problema es, empero, que el diseño funcional de nuestras intuiciones y nuestras emociones morales es un diseño adaptado a circunstancias ecológicas y culturales distintas de las actuales. En una palabra: nuestra arquitectura moral y emocional fue diseñada por la selección natural para las condiciones de vida del Pleistoceno (Barkow, Cosmides y Tooby, 1992; Mariansky y Turner, 1992), no para sociedades modernas, industriales y democráticas. Para bandas de entre 70 y 150 cazadores y recolectores ubicados en la sabana; no para competidores en un mercado libre de bienes y servicios, ni para ciudadanos de un estado liberal de Derecho, ni para anónimos habitantes de una megápolis contemporánea, ni para esquimales del Ártico, ni para súbditos de una teocracia islámica, ni para patriotas de las *póleis* mediterráneas antiguas, ni para intrépidos tripulantes de naves espaciales. Es decir: las funciones que en nuestros días y en nuestros contextos ecológicos y culturales pueden cumplir muchas de nuestras intuiciones y emociones morales *no son funciones propias*¹⁴.

Establecer las funciones propias de nuestra arquitectura cognitiva y moral es una empresa fascinante para la psicología y para la ciencia social positiva. Abre sin duda vías muy potentes y fértiles de explicación de la conducta social

samiento que rechaza por imposible la institucionalización de la deliberación pública racional —los sofistas— fueron precisamente enemigos encarnizados del régimen democrático ateniense.

¹⁴ La noción de «función propia» fue acuñada por Ruth Millikan (1984). El último Hayek se mostró muy preocupado por este hecho: en las funciones propias de nuestras intuiciones morales veía una amenaza y una fuente incesante de descontento y oposición al orden capitalista ultraliberal que era de su gusto: «Los instintos innatos del hombre no son a propósito para una sociedad como la que vivimos. Los instintos estaban adaptados a la vida en pequeños grupos (...). Sólo la civilización trajo individualización y diferenciación. El pensamiento primitivo consiste capitalmente en sentimientos comunes de los miembros de los pequeños grupos. El colectivismo moderno es una recaída en ese estado salvaje, un intento de reconstruir esos fuertes vínculos que se dan en los grupos limitados...» Hayek (1983: 164-165).

humana (Boyer, 1995; Fiske, 1993), en particular, de un amplio abanico de conductas maladaptadas a las circunstancias actuales (Barkow, 1992). Pero presenta varios problemas de importancia para la «contrastación» de las teorías sociales normativas. He discutido con cierto detalle en otro lugar los que me parecen más decisivos (Domènech, 1998). Para lo que aquí importa, se pueden decir dos cosas: en primer lugar, es fundamental que las teorías sociales normativas procuren diseños institucionales lo más amigables posibles para con las funciones propias de nuestras intuiciones morales; y en segundo lugar, cuando eso no sea enteramente posible, que procuren diseños institucionales que eviten la posible manipulación de esas intuiciones¹⁵.

Por lo demás, y a los efectos de esta discusión, podemos considerar el hecho de que las funciones cumplidas actualmente por nuestras intuiciones morales sean en parte funciones impropias como un caso especial del problema, más general, antes discutido de la falta de coherencia y sistematicidad de esas intuiciones. Con el resultado siguiente: así como hay que admitir con Quine que los hechos empíricamente brutos subdeterminan el espacio de las teorías positivas, *con mayor motivo aún* hay que admitir que nuestras intuiciones y emociones morales subdeterminan el espacio de nuestras teorías sociales normativas. Así como la prueba de los hechos no es la última palabra, el criterio definitivo, en la evaluación de las teorías positivas, *aún tenemos razones de más peso* para afirmar que el equilibrio reflexivo, con ser muy importante, no es la última palabra, el criterio definitivo, en la evaluación metodológica de nuestras teorías sociales normativas.

Lo que sigue es una breve exploración de cinco criterios complementarios, que tratan de determinar algo más el espacio lógico de las teorías sociales normativas.

4. Estabilidad

Una teoría social normativa es tanto más estable cuantos menos incentivos proporcione para que los agentes sociales violen sistemáticamente los principios básicos de la teoría.

El llamado utilitarismo de las reglas es una teoría social normativa típicamente inestable. Esta teoría se distingue del utilitarismo cardinalista clásico

¹⁵ Un ejemplo ilustrará lo que se quiere decir. La curiosidad insaciable que experimentamos los humanos por las vidas privadas de los demás, sobre todo por las informaciones sobre sus aficiones sexuales y su *status* social, tenía una función propia obvia en las pequeñas bandas de cazadores y recolectores del Pleistoceno (Barkow, 1992). En nuestro mundo, es evidente que no hay forma decente de dar satisfacción a esa curiosidad: no podemos verosímelmente diseñar un ambiente institucional demasiado amigable para ella. Pero podemos al menos buscarle un entorno institucional que impida o dificulte su manipulación grosera e indecorosa por empresarios de la prensa y la televisión amarillas, ávidos de los beneficios que trae consigo el explotarla («impropiamente»).

porque sostiene que, más que las acciones consecuencialistas de cada agente, lo que puede de verdad promover la utilidad global es el respeto o la observancia por parte de todos de determinadas normas sociales, como, por ejemplo, la que manda cumplir las promesas hechas. El utilitarismo de las reglas trató de limar los cantos más ásperos del utilitarismo tradicional o de los actos, cuyo decidido consecuencialismo chocaba frontalmente con el grueso de las intuiciones morales que nos instan a respetar determinadas normas y reglas (como la de cumplir las promesas), con relativa independencia de las consecuencias que ello traiga consigo. Aparentemente, el utilitarismo de las reglas corregía este defecto al invitarnos a considerar que buena parte de las normas sociales deben su existencia al hecho de que su observancia contribuye al bienestar colectivo, a la promoción de la utilidad global. De acuerdo con eso, la mejor manera de maximizar la utilidad agregada sería, en general, observar las normas sociales establecidas. Ahora bien; si el agente X considera que el objetivo moral perseguido es maximizar la utilidad social agregada, entonces, en cualquier momento, T, y respecto de cualquier norma o regla social, Y, X tiene un incentivo para plantearse si en esta ocasión, en T, lo mejor, en punto a maximizar la utilidad social agregada, no sería precisamente violar la norma Y. Es decir, el utilitarismo de las reglas da incentivos sistemáticos a los agentes sociales para violar un principio básico de la teoría (observar las reglas), y en este sentido, es una teoría inestable.

Otra teoría que parece claramente inestable es la teoría deontológica de la justicia de Robert Nozick. Ya tuvimos ocasión de ver las cuatro condiciones básicas en que se resume esta teoría. Esas cuatro condiciones se formulan como derechos, y esos derechos hacen las veces de restricciones laterales de la acción de los individuos: cualesquiera que fueren las consecuencias que acarree el respeto de esos derechos, la situación social resultante será justa en el sentido de Nozick. Nozick considera que sólo un estado mínimo (por el estilo del «guardián nocturno» del liberalismo decimonónico) es compatible con el respeto de esos derechos. No obstante, y por motivos que no hacen aquí al caso, la mejor forma de captar conceptualmente los derechos no es entendiéndolos como restricciones laterales, sino —en términos de teoría de los juegos de estrategia— como estrategias socialmente abiertas a los individuos (Gibbard, 1974; Gaertner, Pattanaik y Suzumura, 1992). X tiene derecho a vallar su jardín como le dé la gana, si y sólo si ésa es una estrategia socialmente abierta para él. Pero, una vez entendidos de este modo, los derechos no conducen a equilibrios de Nash, que son las soluciones naturales de los juegos de estrategia (cualquiera que sean sus dificultades como criterio matemático de solución de los juegos, un equilibrio de Nash goza de particular estabilidad porque, una vez en él, por definición, ningún jugador puede cambiar unilateralmente de estrategia sin salir perjudicado; no estar en un equilibrio de Nash quiere entonces decir que todos los individuos están permanentemente tentados a cambiar de estrategia). Fiar la viabilidad de una sociedad justa

al mero respeto de derechos «libertarianos» implica entonces medidas públicas —o privadas— draconianas, tendentes a hacerlos cumplir (Riley, 1989, 1990; Sen, 1994), y significa dar incentivos a los agentes sociales para construir estados o sociedades civiles policíacos. Lo cual, compadeciéndose mal con un estado «mínimo» y con una sociedad civil «libertariana», amenaza gravemente la estabilidad metodológica de la teoría.

5. Saciabilidad

Toda teoría social normativa —ya tenga un formato consecuencialista, ya deontológico— se propone determinados objetivos —ya el logro de una situación social determinada, ya el respeto de unos determinados procedimientos—. El *desideratum* de saciabilidad significa que las teorías sociales normativas están obligadas a ofrecer criterios que permitan saber de antemano cuándo pueden considerarse cumplidos esos objetivos. Una teoría que no está en condiciones de dar esos criterios es una teoría insaciable. Pongamos el caso de un juez utilitarista clásico que, a la hora de dictar una sentencia, tuviera un amplio margen de discrecionalidad. Dentro de ese margen, sólo estaría guiado por el criterio utilitarista general de la maximización de la utilidad global de la sociedad. Su sentencia tendría que contribuir a esa maximización. Obviamente, el aspecto más importante de la sentencia sería, en ese esquema teórico-normativo, su eficacia *ad deterrendum*, es decir, el impacto disuasorio de la sentencia sobre las intenciones delictivas futuras del condenado y de cualquier otro miembro de la sociedad. Pues bien; mientras la sentencia cumpla con el requisito de aumentar la utilidad social global, no hay límite alguno establecido por la teoría para la dureza de la pena impuesta (dicho sea de paso: independientemente de la gravedad del delito cometido). Puede ser una mera multa, o una semana de arresto, o cinco años de cárcel, o treinta años de trabajos forzados: dependiendo de las circunstancias sociales del momento —de la alarma social causada por el delito, de la notoriedad pública del delincuente, etc.—, la sentencia tiene que imponer aquella pena que contribuya a aumentar —por su eficacia disuasoria— la utilidad global; no está restringida de antemano por ningún criterio que ponga límites a la dureza del castigo¹⁶. En este sentido, la teoría social normativa en que se inspira nuestro juez es una teoría insaciable.

Otro ejemplo de amenaza de insaciabilidad nos lo brinda una teoría social normativa en las antípodas del utilitarismo clásico. Se trata de la interesante teoría dworkiniana de la igualdad de recursos externos e internos (Dworkin, 1981a, 1981b y 1987). Ronald Dworkin parte de la idea de base rawlsiana de que una teoría de la justicia tiene que cribar adecuadamente entre aquello de lo que puede hacerse responsables a los individuos —y en lo que no debe

¹⁶ El ejemplo procede de Braithwaite y Pettit (1990: 45 ss.).

entrar la justicia distributiva— y aquello de lo que *no* puede hacerse responsables —en lo que debe entrar rectificativamente la justicia distributiva.

Pero mientras Rawls se limita a considerar como ámbito primordial de intervención rectificadora de la justicia distributiva el del azar social (nadie es responsable de su dotación social de partida, de la familia o de la clase social en la que ha nacido), Dworkin está interesado también en corregir redistributivamente los yerros, los infortunios y las desigualdades del azar natural (lotería genética, accidentes naturales sobrevenidos, etc.).

Para rectificar los azares sociales, Rawls ofrece un criterio (el maximin) que satisface el requisito de saciabilidad: es justa aquella estructura social básica en la que (atendidos los mundos tecnológica y productivamente posibles) los que menos tienen, tienen más. Alcanzada esa pauta de distribución, la teoría de la justicia como equidad queda saciada. Pero compensar redistributivamente el azar natural es mucho más complicado. Dworkin ha avanzado provisionalmente el siguiente criterio. Supongamos que la gente estuviera tras un velo de ignorancia —más tenue, por cierto, que el de Rawls—, y desconociera sus dotaciones naturales. Supongamos que hubiera un hipotético mercado perfectamente competitivo en el que todo el mundo pudiera contratar seguros para cubrirse de la eventualidad de una discapacidad física. Los hipotéticos agentes sociales dworkinianos tendrían entonces que tomar la decisión de contratar o no determinados seguros contra la posibilidad de que, descorrido el velo de ignorancia, fueran tetrapléjicos, o paráliticos cerebrales, o poco inteligentes, o nada agraciados físicamente, o subdotados para algún deporte, etc. Puesto que ese hipotético mercado sería perfectamente competitivo, los precios de los seguros serían precios de equilibrio, y la decisión de un agente social de contratar a esos precios unos u otros seguros contra unos u otros riesgos de una determinada discapacidad física mediría el coste de oportunidad para el agente en cuestión de padecer esa discapacidad. Pues bien; el criterio esbozado por Dworkin sugiere que la compensación debida en justicia a la gente por las dotaciones naturales que les han tocado en suerte es precisamente el premio del seguro cuyo precio estarían dispuestos a pagar en el hipotético mercado de futuros; la compensación debida en justicia a la gente es la que cubre el coste de oportunidad de su discapacitación física.

El criterio esbozado por Dworkin tiene muchos atractivos. Es verdad que la corrección del azar natural apela directamente a intuiciones morales muy vivas y ampliamente compartidas. El criterio tiene la ventaja adicional de tratar de cribar entre aquello de lo que los individuos son responsables y aquello de lo que no lo son, mediante el mecanismo conceptual de dejarles en libertad de contratar o no un seguro para una amplia gama de discapacidades físicas a precios de equilibrio. Además, los agentes realizan su criba tras un velo de ignorancia más tenue que el de Rawls, lo que podría permitir supuestos

motivacionales menos deflacionarios que los de la teoría de la justicia como equidad¹⁷. Por añadidura, idealmente, el criterio parece saciable.

El criterio dworkiniano hay que entenderlo como una especie de experimento intelectual, apto a hacernos inteligible cómo podría llegar a compensarse con recursos sociales externos la insuficiencia de los recursos internos de los individuos. Pero, en la práctica, no hay mercados de futuros perfectamente competitivos de este tipo. Es más: es teóricamente imposible que los haya¹⁸. Así que no hay manera de saber cuál es exactamente el coste de oportunidad, en recursos externos, para un agente de haber nacido ciego, o de haber contraído el sida en una transfusión de sangre en un hospital público, pongamos por caso. En cierto sentido, ni todo el dinero del mundo —ni todos los recursos sociales externos— pueden compensar a alguien de su ceguera o de una vida expuesta al albur de una enfermedad, sobre crónica, socialmente estigmatizada. Lo que es tanto como decir que, a falta de indicaciones más precisas, o más operativas, el criterio dworkiniano está seriamente amenazado por la insaciabilidad¹⁹.

6. *Fuerza categórica*

Las teorías sociales normativas no se construyen para entrar en las vitrinas de los museos de Historia de las Ideas Buenas, Bonitas y Elegantes, sino que aspiran a transformar la vida social, a orientar las acciones de las mujeres y de los hombres, a convencerles de que el tipo de sociedad propuesto por la teoría es un tipo de sociedad por el que vale la pena luchar, un tipo de sociedad que están moralmente obligados a promover (o, cuando logrado, a conservar). Una teoría social normativa que dimita, explícita o tácitamente, de esa aspiración carece de fuerza categórica: quienes la conozcan, pueden admirarla por su informatividad, por su elegancia conceptual, por su belleza formal, o por lo que sea, pero no se sentirán compelidos a actuar de acuerdo con ella.

En general, las teorías sociales normativas construidas contractualistamente carecen de fuerza categórica. Pues una teoría contractualista razonable no puede suponer, como pretendían ciertos contractualismos tradicionalistas pre-

¹⁷ Pronto tendremos ocasión de ver la posible importancia de que los supuestos de comportamiento de los agentes en las teorías sociales normativas sean deflacionarios e indeterminados, o al revés, proteicos y determinados.

¹⁸ Por definición, los mercados de futuros, cargados de incertidumbres irreductibles y de asimetrías informativas, no pueden ser perfectamente competitivos.

¹⁹ Otro criterio que está amenazado de insaciabilidad de un modo parecido al de Dworkin es el criterio de reparaciones históricas de Nozick. Por ejemplo, el catalán ha sido una lengua perseguida y oprimida políticamente por el Estado español durante décadas. Pocos disputarán ese hecho. Se necesita una reparación histórica, en forma, por ejemplo, de discriminación positiva. Pero, ¿hasta dónde debe llegar esa reparación? Ahí empiezan las disputas, que no pueden zanjarse normativamente a falta de criterios no sólo claros y precisos, sino también saciables.

hobbesianos, que las generaciones presentes estamos o moral o jurídicamente obligadas por un contrato social que cerraron *realmente* nuestros ancestros. Las teorías contractualistas razonables se sirven, antes bien, de un contrato social *hipotético*, a fin de justificar normativamente, de una u otra forma, el tipo de sociedad que consideran justa o buena. Hobbes, por ejemplo, usó la metáfora conceptual del estado de naturaleza *ad deterrendum*, para presentar como razonable el establecimiento de un hipotético contrato social que permitiera salir de ese estado. Y Rawls usa las metáforas conceptuales de la situación original y del velo de ignorancia para presentar como razonable el hipotético contrato social que cerrarían las partes en esa situación imaginaria de incertidumbre.

Ahora bien, un contrato hipotético no tiene fuerza categórica. El agente social X puede realizar el servicio S, que es de indudable utilidad para el agente Y. Sin duda, el agente Y necesita el servicio S, y está dispuesto a pagar un buen dinero a quien se lo realice. Si los agentes X e Y cierran un contrato, el agente Y estará jurídica —y moralmente— obligado a pagar al agente X el dinero que figura en el contrato (digamos, el precio de mercado del servicio S). Pero no tendría sentido decir que, del hecho de que el agente X sabe que el agente Y necesita el servicio S, que está dispuesto a remunerarlo y que el precio de mercado del servicio S es la cantidad C, se infiere que: si el agente X proporciona por su cuenta y riesgo, y sin encomendarse ni a Dios ni al diablo, el servicio S al agente Y, entonces este último está o jurídica o moralmente obligado a pagar a X la cantidad C. Los contratos reales pueden tener fuerza categórica; los contratos hipotéticos, no (Dworkin, 1993: 70 ss.).

La impotencia categórica de los constructos contractualistas se agrava aún más si la teoría social normativa a la que sirven renuncia a elaborar una noción de buena vida individual, es decir, si la teoría social normativa en cuestión, por los motivos que fuere, es éticamente agnóstica. Un contrato social que, sobre hipotético, no deriva más o menos directamente de una determinada noción de buena vida individual, parece la vía más segura a la debilidad categórica de la concepción normativa resultante.

Como es bien sabido, la concepción normativa del último Rawls es paradigma de ambos rasgos (contractualismo hipotético y agnosticismo ético). Los dos son centrales en la versión actual de la teoría de la justicia como equidad; su abandono, entiende Rawls, tendría un precio que él no está en modo alguno dispuesto a pagar. Pero su aceptación, según parece, hace almoneda de la fuerza categórica. Tenemos ahora que considerar si el coste de abandonar los dos rasgos que parecen incompatibles con la fuerza categórica es tan elevado como lo supone Rawls, y, en tal caso, si habría que renunciar al *desideratum* de la fuerza categórica.

En el breve espacio de que aquí se dispone se puede decir lo siguiente: La construcción contractualista hipotética de una teoría social normativa sub-

viene a la necesidad de resolver una dificultad importante, pero no es el único modo de resolverla. La dificultad consiste en lo siguiente:

Cualquier juicio normativo puede entenderse como una exigencia de exclusión de información. (La metáfora icónica que presenta a la justicia con los ojos vendados expresa esto del modo más plástico.) Si el ejercicio intelectual en el que estamos comprometidos es el de ofrecer un juicio normativo de la estructura social básica de nuestras sociedades, entonces estamos obligados a excluir determinada información, o, dicho a la inversa, estamos obligados a determinar con la mayor precisión posible cuál es la información pertinente para juzgar normativamente la estructura social básica. La criba normativa entre la información pertinente y la información excluible tiene que coincidir con la criba entre aquello de lo que puede hacerse responsables a los agentes sociales y aquello de lo que no puede hacerse responsables. Para Rawls, no puede hacerse responsables a los individuos, por ejemplo, de su ubicación en la estructura social, es decir, de su origen de clase, ni de su talento. Esa información (sobre el origen de clase y el talento de los individuos) tiene, pues, que excluirse del juicio normativo sobre la estructura social básica. Y un modo particularmente económico y elegante de hacerlo es presentando el juicio normativo sobre la estructura social como dimanante del hipotético acuerdo al que llegarían unos agentes racionales y razonables —las partes de la situación imaginaria— situados tras un «velo de ignorancia» que les impidiera conocer toda la información excluible (su talento, su origen de clase, etc.), pero no la información pertinente (por ejemplo, su apetencia por los bienes primarios instrumentales).

Pero si esto es así, es perfectamente concebible que la necesidad de excluir determinada información, pero no otra, pueda construirse directamente, en vez de por la vía rodeada de la metáfora contractualista hipotética, apelando a nuestras intuiciones morales y emocionales, y buscando un equilibrio reflexivo con ellas. ¿Por qué, entonces, empeñarse en construcciones contractualistas hipotéticas? He aquí tres posibles respuestas:

1. Para empezar, podría argüirse que los humanos son inevitablemente egoístas, que actúan movidos por una racionalidad estrictamente económica: la situación contractual hipotética rawlsiana serviría, lo mismo que la hobbesiana, para mostrarles el camino que deben seguir para seguir siendo egoístas racionales. Rawls ha rechazado expresamente en muchas ocasiones esta interpretación de su obra, insistiendo, por ejemplo, en que las partes de la situación original son algo más que meros agentes racionales (y, por lo pronto, en que son también «razonables»)²⁰.

2. Una segunda respuesta diría que, aun no estando ni mucho menos el grueso de los individuos estrictamente movidos por el interés propio, a

²⁰ De todos modos, no han faltado reconstrucciones hermenéuticas de la teoría rawlsiana en términos de pura racionalidad económica. Véase, por ejemplo, Binmore (1989), Buchanan (1976), Cooter (1989).

la hora de hacer construcciones normativas es mejor partir *à la Hume* del supuesto de villanía, es mejor ponerse en lo peor y sentar criterios operativos también para el caso de que todos fueran villanos. En esta interpretación, la situación originaria, sin necesidad de suponer que las partes son villanas, estaría, sin embargo, idealmente a prueba de villanía, por privar a los posibles villanos —y a todos los demás— de buena parte de la información que se requiere para serlo. Sin ser éste el lugar adecuado para discutirlo, hay que decir que tampoco ésta parece una interpretación demasiado expedita de la posición de Rawls.

3. Otra posible respuesta sería que, aun no estando los individuos estrictamente movidos por el interés propio, subsistiría el hecho de que, *en situaciones sociales reales*, no serían capaces de excluir y cribar adecuadamente la información, según exige la teoría en cuestión. (La respuesta de los agentes egoístas, antes comentada, sería un caso particular de esta tercera respuesta.) Imaginemos, por ejemplo, que, de repente, todos los empresarios aplicarían el juicio normativo rawlsiano sobre la estructura social básica, no a la globalidad de ésta, sino a su propia actividad empresarial. Imaginemos que, a la hora de contratar o despedir trabajadores actuaran de acuerdo con el criterio maximin, pensado para el conjunto de la estructura básica. Imaginemos que despidieran a los trabajadores más competentes y laboriosos —con el argumento maximin de que no tendrán problemas en el mercado de trabajo para volver a emplearse— y contrataran por regla general a los trabajadores menos competentes y laboriosos —con el argumento maximin de que son los peor situados en el mercado de trabajo—. El resultado agregado de esta conducta empresarial generalizada sería una estructura social rawlsianamente injusta, en la que los que peor estuvieran estarían verosímelmente mucho peor que en una estructura social en la que los empresarios se comportaran como suelen. Y ese resultado habría sido consecuencia, no del egoísmo de los empresarios —huelga decirlo—, sino de una conducta fundada en un juicio normativo descontextualizado, que no habría cribado adecuadamente la información. La tercera respuesta, en suma, justificaría el formato hipotético-contractualista basándose en la hipótesis de que, en situaciones sociales reales, los individuos no pueden cribar adecuadamente por sí propios la información, ni, por lo mismo, mantener juicios normativos ecuanímenes o pertinentes. O en la hipótesis de que la teoría social normativa no puede sino abstenerse de esa discusión. Sea como fuere, la teoría social normativa tendría que tener unos supuestos extremadamente deflacionarios respecto de las motivaciones y las capacidades de criba informativa de los individuos, supuestos deflacionarios que sólo podrían quedar bien servidos dentro de una construcción hipotético-contractualista²¹. Esta tercera respuesta

²¹ Entiendo por supuestos deflacionarios de comportamiento supuestos compatibles con una gran variedad de motivaciones y conductas de los agentes, es decir, supuestos poco comprometidos (y poco comprometedores). Hay varias razones que invitan a construir teorías sociales normativas con supuestos deflacionarios: evitar entrar en disputas laterales complicadas, por ejemplo, o la

tal vez haga más justicia hermenéutica a Rawls que las anteriores, y se compeadece bien con su vieja distinción de 1971 entre la teoría política *ideal* por él desarrollada, que estudia la sociedad justa haciendo abstracción de los problemas de observancia ciudadana de las normas, y la teoría política realista, que no deja de examinar esos problemas.

Así pues, el precio a pagar por abandonar los constructos hipotético-contractualistas no es, como podría temerse, la imposibilidad de excluir información en el juicio normativo de la estructura social básica, sino probablemente el abandono del estilo «ideal» de hacer teoría política y el consiguiente abandono de los supuestos deflacionarios sobre las motivaciones y las capacidades de los agentes sociales.

Venimos ahora a la segunda fuente de amenazas a la fuerza categórica: el agnosticismo ético. El precio que según Rawls hay que pagar por prescindir de él es terrorífico: es el precio del sectarismo político, el precio de la intolerancia. En buena medida, el segundo Rawls se ha consagrado a la construcción de un elaborado argumento en favor del agnosticismo ético de las teorías sociales normativas. La idea se puede resumir del siguiente modo: la tolerancia debe ser un valor fundamental para una teoría social normativa; pero la idea de derivar una teoría social normativa de una determinada concepción ética de la buena vida individual atenta directamente contra ese valor, pues la teoría social normativa resultante sólo será aceptable para quienes compartan esa visión ética fundacional de la buena vida. Consecuencia de lo cual es: una teoría social normativa que quiera respetar el valor de la tolerancia y que pretenda valer para sociedades complejas, y a menudo, multiculturales, no puede sino ser éticamente ecuménica, es decir, agnóstica. Una teoría social normativa tolerante sólo puede fundarse en un consenso entrecruzado del que participen las más variadas concepciones éticas razonables de buena vida individual, y respecto de las cuales, la teoría debe mantenerse escrupulosamente neutral. En el contexto de la presente discusión, esto significaría estar condenados a elegir entre el valor irrenunciable de la tolerancia y el *desideratum*

necesidad de atraerse un sostén ecuménico, necesidad aparentemente bien servida por supuestos motivacionales lo más neutros y desleídos posibles. En la medida en que esos supuestos, por poco comprometidos, son poco informativos, resultarán por eso mismo más difíciles de refutar, y serán más fácilmente aceptables por todos. Pero puede ocurrir que incluso los supuestos más deflacionarios —salvo que sean tautológicos— acaben revelándose falsos, de manera que la teoría que se arme con ellos naufragará en los dos frentes. Así, por ejemplo, por muy deflacionarios que sean los supuestos motivacionales de comportamiento de las partes en la situación originaria rawlsiana, parece claro que implican al menos un modo de procesar la información estrictamente dominio-universal, y la necesidad normativa de restringir con un velo de ignorancia su información respondería precisamente a eso, a que ellos no pueden hacerlo por sí propios. Ahora bien; todo apunta a que los humanos procesamos la información mediante módulos dominio-específicos (Hirschfeld y Gelman, 1994), y a que, en particular, en el contexto de la vida social, estamos cognitivamente dotados con al menos cuatro modos, relativamente independientes, de procesar la información social relacional, cada uno de los cuatro provisto de un peculiar filtro excluyente de información (Fiske, 1993).

de la fuerza categórica. Una elección sin duda trágica. Pero el argumento de Rawls no parece concluyente.

Por de pronto, hay varias formas posibles de construir el valor de la tolerancia dentro de una teoría social normativa. La de Rawls no es sino una de ellas, consistente en elevar el valor de la tolerancia a *axioma metodológico* de la teoría. Pero hay al menos otra forma, no menos legítima: construir la tolerancia como un *teorema* derivado de la determinada concepción ética de la buena vida en la que se pretenda fundar la concepción de la justicia en cuestión. Una teoría social normativa derivada de una ética que implique la tolerancia tiene que ser necesariamente tolerante, y puede aspirar a tener fuerza categórica (Dworkin, 1993).

Está aquí fuera de lugar la exploración de este asunto. Pero al menos hay que observar que la construcción de la tolerancia como un teorema, no como un axioma metodológico, aparte de no bloquear la aspiración a la fuerza categórica de la teoría social normativa, parece tener otras ventajas metodológicas. Pues, por un lado, la estrategia rawlsiana de elaborar una concepción política de la justicia sin solución de continuidad con la ética (la «estrategia de la discontinuidad», como la llama Dworkin) parece necesitar, como ya tuvimos ocasión de ver, unos supuestos motivacionales y de capacitación ética de los agentes extremadamente deflacionarios, unos supuestos que pueden chocar con las exigencias del *desideratum* metodológico de factibilidad, al que nos referiremos en un momento²². Por otro lado, incluso supuestos motivacionales y de capacitación ética extremadamente deflacionarios como los rawlsianos pueden generar inconsistencias en la teoría social normativa, inconsistencias que pasarán más fácilmente inadvertidas en una estrategia teorizadora de la discontinuidad que en una estrategia teórica de la continuidad. El ejemplo, antes mencionado, de unos empresarios que tuvieran un comportamiento éticamente —no políticamente— rawlsiano viene aquí muy a propósito. La conducta de esos empresarios generaría, como vimos, muchos problemas para la teoría de la justicia como equidad. Tantos, que no hay más remedio que pensar que la teoría es incompatible con esta posible noción de buena vida de los empresarios. Lo que suscita al menos dos cuestiones metodológicas interesantes. En primer lugar, una teoría que eleva a axioma metodológico la tolerancia resultaría incompatible con al menos una noción de vida buena (la de los empresarios con una conducta maximin respecto de sus empleados), sin que, aparentemente, pueda decirse de esa noción que sea «irrazonable» en el sentido rawlsiano, es decir, que haya que excluirla del consenso entrecruzado. En segundo lugar, si finalmente, por el motivo que fuera, se la declarara «irrazonable», la teoría estaría amenazada por la autonulificatoriedad, una de las formas de inconsistencia visitadas en la sección 1: la teoría abogaría por

²² Es interesante al respecto la crítica que el lingüista y psicólogo cognitivo Georges Lakoff (1996) ha realizado recientemente de este tipo de teorías deflacionarias de la justicia, y en particular, de la rawlsiana.

una forma de conducta de los agentes sociales que borraría o nulificaría los valores fundamentales de la teoría.

7. Factibilidad

Una teoría normativa factible acepta, por lo pronto, la regla de que *el debe implica el puede*. Pero el *desideratum* de factibilidad exige más que la aceptación expresa de esa regla. Pretende que las teorías sociales normativas estén permanentemente abiertas a los resultados pertinentes de la ciencia empírica positiva; pretende que se preocupen activamente de sus propias condiciones de viabilidad. Este criterio es suficientemente claro, como para que valga la pena explayarse aquí en él. Dos cosas son, sin embargo, de notar.

La primera es que tal vez no sea conceptualmente independiente del *desideratum* de la fuerza categórica: es posible que la fuerza categórica sea un caso particular de la factibilidad (aunque es fácil imaginar teorías con fuerza categórica que no sean técnica, o institucional, o socialmente factibles, se hace más difícil imaginar teorías que, no obstante carecer de fuerza categórica, sean factibles).

La segunda es que, tomado en serio, el *desideratum* de factibilidad pone en graves aprietos a un estilo de teorización social normativa que ha gozado de gran predicamento en las dos últimas décadas. Para decirlo sumariamente, ese estilo va ligado a la idea rawlsiana (Rawls, 1996: 171-2, 186-7) de la «independencia» de la concepción política de la justicia y a la creencia, que va con ella, de que la teorización social normativa avanza mejor, en la medida en que se mantiene neutral respecto de disputas y querellas aledañas: de disputas en torno de cuestiones éticas sobre lo que sea una buena vida, de disputas metafísicas generales sobre el valor de la libertad, de disputas sobre los supuestos de comportamiento de los individuos, etc. Como va dicho, este modo de proceder lleva consigo algunas ventajas importantes: para empezar, permite ceñir la atención a un abanico bien delimitado de problemas y evita la dispersión lateral del esfuerzo. Pero el resultado es la construcción de teorías sociales normativas con supuestos motivacionales superlativamente deflacionarios (lo que puede poner en entredicho, además de su fuerza categórica, su viabilidad psicológica y la factibilidad de su diseño institucional, entre otras cosas), y desde luego, no muy abiertas al careo empírico procedente de los avances de las ciencias positivas pertinentes (ciencia económica, sociología, antropología, psicología cognitiva, biología evolucionaria, etc.). Además, la elaboración de teorías sociales normativas con supuestos deflacionarios, compatibles con un amplio espectro de supuestos más específicos y precisos, seguramente pierde en informatividad mucho más de lo que presumiblemente gana en poder de atracción ecuménico. De la combinación del criterio de factibilidad y del criterio de informatividad parece derivarse la exigencia de que las teorías sociales normativas ganen poder ecuménico, no por la vía trillada de apoyarse en supuestos

motivacionales exageradamente deflacionarios, ni por la de encastillarse en puntos de vista «independientes» o en teorizaciones «ideales» (no siempre ha de ser verdadera la fórmula *via trita, via tuta*), sino por la vía de apoyarse en supuestos motivacionales *correctos*, científica o filosóficamente defendibles, y de penetrarse resueltamente de los resultados pertinentes de la investigación empírica positiva.

8. Parsimonia

Ceteris paribus, entre dos teorías sociales normativas que satisfagan más o menos parigualmente los anteriores *desiderata*, es mejor aquella que los satisfaga con mayor economía de medios. Arrow (1963) escribió hace ya muchos años que la «ventaja de las argumentaciones formales abstractas es que permiten un ahorro considerable de tiempo». Se dice a veces —no sin fundamento— que el impacto que el estilo teorizador de la ciencia económica ha tenido en la ciencia y la filosofía social normativas de los últimos años ha sido responsable en buena medida de la introducción en éstas de supuestos motivacionales inadecuados e irrealistas. Pero en el haber de ese impacto hay que cargar, sin duda, los hábitos de razonamientos claros, precisos y a menudo formalmente refinados que ha adquirido recientemente la filosofía política. Sea el que fuere el futuro de la teorización social normativa, hay que esperar que esos hábitos se mantengan y se robustezcan. Mucho antes que Arrow, Sócrates enseñó que también las «opiniones verdaderas se nos escapan del alma, de suerte que son de poco valor, en tanto no las encadenemos por un razonamiento...» (*Menón*: 98a).

BIBLIOGRAFÍA

- ARROW, J. K. (1963): *Social Choice and Individual Values*, Wiley, Nueva York.
- ATTRAN, S. (1990): *Cognitive Foundations of Natural History*, Cambridge: Cambridge University Press, 1990.
- BARKOW, J. (1992): «Beneath New Culture is Old Psychology: Gossip and Social Stratification», en Barkow, Cosmides y Tooby (1992).
- BARKOW, J.; COSMIDES, L., y TOOBY, J. (1992): *The Adapted Mind*, Oxford, Oxford University Press.
- BARWISE, J., y MOSS, L. (1996): *Vicious Circles*, Stanford, CSLI Publications.
- BINMORE, K. (1989): «Social Contract, I: Harsanyi and Rawls», en *Economic Journal*, 99.
- BLAIR, D. H. (1988): «The Primary-Goods Indexation Problem in Rawl's Theory of Justice», *Theory and Decision*, 24.
- BOYER, P. (1995): *The Naturalness of Religious Ideas: A Cognitive Theory of Religion*, Los Ángeles, California University Press.
- BRAITHWAITE, J., y PETTIT, Ph. (1990): *Not Just Deserts. A Republican Theory of Criminal Justice*, Oxford, Clarendon Press.

- BUCHANAN, J. M. (1976): «A Hobbesian Reinterpretation of the Rawlsian Difference Principle», en *Kyklos*, 29.
- CHURCHLAND, P. (1981): «Eliminative Materialism and the Propositional Attitudes», en *Journal of Philosophy*, febrero, pp. 67-90.
- COOTER, R. (1989): «Rawls's Lexical Orderings are Good Economics», *Economics and Philosophy*, 5.
- COSMIDES, L. (1992): «Cognitive Adaptations for Social Exchange», en Barkow, Cosmides y Tooby (comps.) (1992).
- DENNETT, D. (1995): *Darwin's Dangerous Idea*, Nueva York: Simon and Schuster.
- DOMÈNECH, A. (1989): *De la ética a la política*, Barcelona, Crítica.
- (1996): «Ética y economía de bienestar: una panorámica», en O. Guariglia (comp.), *Cuestiones morales*, vol. 12 de la *Enciclopedia iberoamericana de filosofía*, Madrid, Trotta, pp. 191-222.
- (1997a): «La filosofía del conocimiento como filosofía normativa», en Manuel Cruz (comp.): *Acción humana*, Barcelona, Ariel, pp. 234-262.
- (1997b): «El eterno retorno de Calicles», en F. Álvarez y R. Rodríguez Aramayo, *Disenso e incertidumbre (Javier Muguerza veinte años después de «La razón sin esperanza»)*, volumen en preparación.
- (1998): «El equilibrio reflexivo y el origen de nuestro conocimiento moral», en J. Bertomeu et al. (comps.), *Festschrift para Osvaldo Guariglia*, volumen en preparación.
- DWORKIN, R. (1981a): «What is Equality?, Part I», *Philosophy and Public Affairs*, 10, pp. 185-246.
- (1981b): «What is Equality?, Part II», *Philosophy and Public Affairs*, 10, pp. 283-345.
- (1987): «What is Equality?, Part III», *Iowa Law Review*, 73, pp. 1-73.
- (1993): *Ética privada e igualitarismo político*, trad. A. Domènech, Barcelona, Paidós.
- FEYNMAN, R. (1985): *QED*, Princeton, Princeton University Press.
- FISKE, A. (1993): *Structures of Social Life*, Nueva York, Free Press.
- FRANKFURT, H. (1971): «Freedom of the Will and the Concept of a Person», en *Journal of Philosophy*.
- GAERTNER, W.; PATTANAİK, P. K., y SUZUMURA, K. (1992): «Individual Rights Revisited», en *Economica*, 59.
- GIBBARD, A. (1974): «A Pareto-Consistent Libertarian Claim», en *Journal of Economic Theory*, 7.
- (1990): *Wise Choices, Apt Feelings*, Oxford, Clarendon Press.
- GOODMAN, N. (1965): *Fact, Fiction, and Forecast*, Indianápolis, Bobbs-Merrill.
- HABERMAS, J. (1983): «Diskursethik - Notizen zu einem Begründungsprogram», recogido en *Moralbewusstsein und kommunikatives Handeln*, Francfort, Suhrkamp.
- HAYEK, F. (1983): «Die überschätzte Vernunft», en R. Riedl y F. Freuzer (comps.), *Evolution und Menschenbild*, Hamburgo, Hoffmann und Campe.
- HIRSCHFELD, L. A. (1994): «Is the acquisition of social categories based on domain-specific competence or on knowledge transfer?», en L. A. Hirschfeld y S. A. Gelman (1994).
- HIRSCHFELD, L. A., y GELMAN, S. A. (comps.) (1994): *Mapping the Mind*, Cambridge, Cambridge University Press.
- KOLM, S. Ch. (1985): *Le contrat social libéral*, París, PUF.
- LAKOFF, G. (1996): *Moral Politics. What Conservatives Know that Liberals Don't*, Los Angeles, University of California Press.
- LEHRER, K. (1990): *Metamind*, Oxford, Clarendon Press.

- MARIANSKY, A., y TURNER, J. (1992): *The Social Cage: Human Nature and the Evolution of Society*, Stanford, Stanford University Press.
- MCCLOSKEY, M. (1983a): «Naive Theories of Motion», en D. Gentner y A. Stevens (comps.), *Mental Models*, Hillsdale, N. J., Lawrence Elbaum Assocs., Inc.
- (1983b): «Intuitive Physics», *Scientific American*, abril.
- MILLIKAN, R. (1984): *Language, Thought, and Other Biological Categories*, Cambridge, Mass., MIT Press.
- NEUMANN, F. (1983): *Behemoth. Pensamiento y acción en el nacionalsocialismo*, trad. V. Herrero y J. Márquez, México, FCE.
- NOZICK, R. (1974): *Anarchy, State, and Utopia*, Oxford, Blackwell.
- PARETO, V. (1913): «Il massimo di utilità per una collettività in sociologia», en *Giornale degli economisti*, 3.ª serie, pp. 337-341.
- PLATÓN: *Menon*, en *Oeuvres complètes*, ed. bilingüe, vol. III, 2.ª parte, París, Les Belles Letres.
- QUINE, W. O. (1962): «Dos dogmas del empirismo», cap. II de *Desde un punto de vista lógico*, trad. M. Sacristán, Barcelona, Ariel.
- RAWLS, J. (1971): *A Theory of Justice*, Cambridge, Mass., Harvard University Press.
- (1996): *El liberalismo político*, trad. A. Domènech, Barcelona, Crítica.
- RILEY, J. (1989): «Rights to Liberty in Purely Private Matters, Part I», en *Economics and Philosophy*, 5.
- (1990): «Rights to Liberty in Purely Private Matters, Part II», en *Economics and Philosophy*, 6.
- SEARLE, J. (1997): *La construcción de la realidad social*, trad. A. Domènech, Barcelona, Paidós.
- SEN, A. (1981): *Poverty and Famines. An Essay on Entitlement and Deprivation*, Oxford, Clarendon Press.
- (1994): «Welfare Economics and Two Approaches to Rights»: *Annual Meeting of the European Public Choice Society*, Valencia, 6-9 de abril, mimeo.
- SINGER, M.: «Shame Cultures and Guilt Cultures», en Piers y Singer, *Shame and Guilt*, Springfield, Ill., Charles C. Thomas.
- STE. CROIX, G. E. M. (1981): *The Class Struggle in the Ancient Greek World*, Londres, Duckworth.
- TUGENDHAT, E. (1993): *Vorlesungen über Ethik*, Francfort, Suhrkamp.
- WILSON, M., y DALY, M. (1992): «The Man Who Mistook His Wife for a Chattel», en Barkow, Cosmides y Tooby (comps.) (1992).