

# *La estructura consecuencialista del utilitarismo*

GILBERTO GUTIÉRREZ  
(Universidad Complutense)

## 1. **Ética, política y económica**

El Utilitarismo es una de las tradiciones más sólidas, coherentes y vivas del pensamiento moral y político de los dos últimos siglos. Durante el prolongado eclipse que sufrió la filosofía normativa en la primera mitad de la presente centuria la teoría utilitarista subsistió como única defensora de un modelo ético de dudosa reputación filosófica<sup>1</sup>. Pero la presión de la crítica la obligó a refinar y reelaborar muchas de las crudezas e ingenuidades teóricas presentes en las primeras formulaciones de Bentham, Mill o Sidgwick. Así, por ejemplo, rompió con el psicologismo que aquejaba a la noción benthamiana de felicidad entendida como sensación placentera sustituyéndola por la de satisfacción de preferencias mensurables sobre una escala de utilidad. Al eliminar presupuestos epistemológicos y psicológicos caducos que no le eran consustanciales se manifiesta con mayor nitidez el núcleo consecuencialista de la teoría utilitaria. Como este a su vez representa una de las formas básicas del propio razonamiento moral, resulta inevitable referirse a él cuando se analizan los fundamentos racionales de la filosofía moral y política<sup>2</sup>.

La pervivencia del Utilitarismo no se debe únicamente a razones sistemáticas sino también históricas, ya que desde sus orígenes no se limitó a ofrecer una teoría de la moralidad individual sino, en medida mayor aún, de las decisiones sobre política económica adoptadas en el marco inter-

---

1. J. J. C. SMART: *An outline of a system of utilitarian ethics*. En J. J. C. SMART y B. WILLIAMS: *Utilitarianism for and against*. (Cambridge. University Press. 1973) p. 3. Hay traducción española: *Utilitarismo, pro y contra* (Madrid, Tecnos. 1981).

2. La bibliografía sobre la teoría utilitarista en lengua inglesa es sumamente extensa. Entre las escasas publicaciones en español destacan las buenas introducciones generales de J. M. COLOMER: *El Utilitarismo: una teoría de la elección racional* (Barcelona. Montesinos. 1987) y D. M. FARRELL: *Utilitarismo: ética y política* (Buenos Aires. Abeledo-Perrot. 1983); véase también O. HOEFFE: *Estrategias de lo humano* (Buenos Aires. Alfa. 1979); G. PATZIG: *Ética sin metafísica* (Buenos Aires. Alfa. 1975); y la o.c. en nota 1.

vencionista y redistributivista del Estado del Bienestar. Por eso ha conservado y acrecentado su vigencia intelectual como filosofía «pública» de la socialdemocracia, tanto o más que como filosofía «privada» o moral en sentido estricto, asumiendo sistemáticamente las cuestiones éticas vinculadas a la (Nueva) Economía del Bienestar y a la gestión política de las democracias pluralistas<sup>3</sup>.

La Economía Clásica siempre aspiró, en principio al menos, a formular las leyes objetivas que regulan los fenómenos económicos, de manera análoga a lo que lograra Newton con los fenómenos físicos, aunque el modelo ideal del mercado suponía unas condiciones sumamente restrictivas y abstractas. En este marco mantenía su vigencia la prohibición humeana de extraer de las formulaciones de la teoría juicios de valor que sirviesen de fundamento a prescripciones o recomendaciones de política económica. Sin embargo dos factores contribuyeron al nacimiento de una rama de la Economía que, sin abandonar el rigor analítico de la teoría ni la referencia empírica que le proporcionaba la información estadística, permitía formular juicios éticos sobre el ordenamiento económico. Uno, interno, la propia evolución de la Teoría Económica que desarrolla instrumentos analíticos que permiten incorporar al modelo los mecanismos de competencia imperfecta y monopolística o las virtualidades del concepto

---

3. De entre la extensa bibliografía sobre Economía del Bienestar, lo que más interesa a las cuestiones aquí expuestas es: K. J. ARROW y T. SCITOWSKI (eds): *Ensayos sobre la economía del bienestar* (México, Fondo de Cultura Económica, 1974); A. BERGSON: *A reformulation of certain aspects of welfare economics*. «Quarterly Journal of Economics» 52 (1938); *Essays in normative economics* (Cambridge MA, Harvard U.P. 1966); A. BOHNEN: *Die utilitaristische Ethik als Grundlage der Wohlfahrtsökonomie* (Gotinga, 1964); A. CASAHUGA: *Fundamentos normativos de la acción y organización social* (Barcelona, Ariel, 1985); M. H. DOBB: *Economía del bienestar y economía del socialismo* (Madrid, Siglo XXI, 1971); I. GOUGH: *Economía política del estado del bienestar* (Madrid, Blume-Hermann 1982); J. de V. GRAAF: *Teoría de la economía del bienestar* (Buenos Aires, Amorrortu, 1967); J. P. GRIFFIN: *Well-being, its meaning, measurement and moral importance* (Oxford, Clarendon Press, 1986); I. LITTLE: *A critique of Welfare Economics* (Oxford, Clarendon Press, 1957); E. J. MISHAN: *Economía del bienestar* (Madrid, Rialp, 1969); M. MOIX: *De la justicia social al bienestar social* (Madrid, Partenón, 1976); *El bienestar social* (Madrid, Partenón, 1978); S. K. NATH: *Una perspectiva de la economía del bienestar* (Barcelona, Vicens Vives, 1976); C. OFFE: *Contradictions of the Welfare State* (Londres, Hutchinson, 1984); E. S. PHELPS (ed.): *Economic justice* (Harmondsworth, Penguin Books, 1973); A. C. PIGOU: *La economía del bienestar* (Madrid, Aguilar); J. C. PITT (ed.): *Philosophy in economics* (Dordrecht, D. Reidel, 1981); M. RESCHER: *Welfare. The social issues in philosophical perspective* (Pittsburgh PA, University Press, 1972); L. C. ROBBINS: *Ensayo sobre la naturaleza y el significado de la ciencia económica* (México, Fondo de Cultura, 1944); A. K. SEN: *Elección colectiva y bienestar social* (Madrid, Alianza, 1976); *Choice, welfare and measurement* (Oxford, Blackwell, 1982); H. VAN DEN DOEL: *Democracy and Welfare Economics* (Cambridge, University Press, 1979); C. WELLMAN: *Welfare rights* (Totowa NJ, Rowman Littlefield, 1982). Lo expuesto son lugares comunes a la teoría.

de utilidad marginal. Otro, externo, la creciente voluntad y capacidad de los estados modernos para intervenir en la economía por medio de la planificación con el propósito de corregir los llamados fallos del mercado y las aberraciones del puro «laissez faire». Y alegando para ello razones de justicia, equidad o eficacia; es decir, precisamente razones éticas. Este fue el origen de la Economía del Bienestar, que adoptó como criterio básico para decidir entre las políticas alternativas la valoración de sus respectivos efectos sobre el bienestar de la comunidad.

En sus inicios la Economía del Bienestar dio por supuesto el carácter evidente de los principios éticos del Utilitarismo al formular sus proposiciones de bienestar y asignó al ordenamiento económico el objetivo de maximizar el bienestar de la sociedad concebido como la suma de los bienestares individuales. Para hacer esta suma matemáticamente posible y asignar un sentido preciso a la noción de bienestar social, se aceptó la hipótesis de que las utilidades marginales que obtienen los distintos individuos —por ejemplo, de su renta— son medibles cardinalmente y, por lo tanto, estrictamente comparables. O, lo que es lo mismo, que los diversos individuos poseen gustos similares y, por lo tanto, relevantemente iguales en el sentido que inspirara la recomendación benthamiana de que, en la suma de utilidades, cada individuo contase como uno y ninguno como más de uno.

Pero conceder objetividad científica al supuesto de la similitud de gustos entre los diversos individuos enmascara su verdadera naturaleza normativa. Esta confusión acerca del verdadero status epistemológico de las proposiciones de bienestar hacía naufragar la teoría en los escollos del intuicionismo ético. Aunque la crítica, más que negar la viabilidad de las recomendaciones en la economía, exigía el reconocimiento explícito de su carácter ético, se la interpretó como una exhortación a la búsqueda de proposiciones de bienestar, bien «libres de ética»<sup>4</sup> o bien, al menos, poseedoras de un contenido ético «mínimo», que las dejara prácticamente al margen de toda polémica.

En consecuencia, así como la Economía del Bienestar se ocupaba de formular recomendaciones éticas utilitaristas sobre política económica, la «Nueva» Economía del Bienestar se preocupó más bien por encontrar la forma de derivarlas de premisas cuasi factuales por el procedimiento de buscar un acuerdo en torno a criterios valorativos débiles, que garanticen la ausencia de conflicto interpersonal. Aunque el discutible supuesto implícito parece ser que si todos están de acuerdo sobre un juicio de valor entonces éste deja de ser un juicio de valor para convertirse en algo perfectamente «objetivo». En el contexto teórico de la Nueva Economía del Bienestar, junto a divergencias radicales en la interpretación de la naturaleza y la función del agente decisor, se coincide al menos en el plantea-

---

4. S. NATH: *o.c.* p. 20.

miento de un mismo tipo de cuestión ética fundamental: cómo derivar una preferencia colectiva a partir de un conjunto de preferencias individuales y alcanzar un acuerdo sobre una función de bienestar social. Este nuevo enfoque —el llamado «giro paretiano» de la Economía del Bienestar— ha condicionado a su vez una nueva y profunda revisión de muchos de los propios supuestos del Utilitarismo clásico. En sus versiones actuales la teoría utilitaria intenta formular una definición precisa y operativa de la justicia en términos de bienestar social, concibiendo este a su vez como una función de la satisfacción de las utilidades de los individuos reflejadas en sus preferencias.

Una constante en la historia del Utilitarismo ha sido la búsqueda de algún tipo de síntesis con la teoría kantiana. Pero subsiste la extendida convicción de que ambas representan, si no *las* dos, si al menos dos alternativas radicales de la teoría ética y que, de una u otra forma, muchas de las restantes teorías históricas podrían reducirse en lo esencial a una u otra. Ambas, a su vez, se excluirían rigurosamente entre sí, hasta tal punto que, desde la perspectiva de cada una de ellas, la otra aparece como la negación misma del concepto de moralidad.

## 2. Lo bueno y lo correcto en la teoría Ética

Aristóteles, y con él una gran parte de la tradición filosófica, incluyó la Ética dentro del género único de la Filosofía Práctica, basándose para ello en la unidad de su objeto, que es una forma de movimiento voluntario. Como todo movimiento, se dirige a un fin en el que se actualiza la potencialidad del ser móvil que alcanza así cierta perfección (*teleíosis*)<sup>5</sup>. En cuanto voluntario tiene en el agente mismo el principio del que parte, que es, según se traduzca el término *proairesis*, la intención, el propósito o la elección<sup>6</sup>. Puesto que la actualización de su potencia es el bien propio del compuesto, todo agente capaz de moverse a actuar intencionadamente lo hace en razón del fin que es su propio bien. A diferencia del obrar que se consume en la perfección de un artefacto externo al agente (*tò poiétón*) el actuar moral (*tò praktón*) perfecciona el agente mismo y le procura su bien<sup>7</sup>.

La distinción aristotélica entre los bienes que se buscan y aman por sí

---

5. MET 1023a34; 1021b20-21. Todo lo que es bueno «es en sí y por su propia naturaleza un fin y, por tanto, una causa, puesto que en vista de ello se hacen y son las demás cosas; y el fin y aquello en vista de lo cual se hace algo son fin de alguna acción, y todas las acciones se ejecutan con movimiento» (MET 996a24-27).

6. MET 1025b24; EN 1139b30-32.

7. Así pues, es prudente quien «discurre bien sobre lo que es bueno y conveniente para él mismo, no en sentido parcial... sino para vivir bien en general» (EN 1140a24-27).

mismos y los que se aman por razón de aquellos<sup>8</sup>, o entre los verdaderos y los aparentes<sup>9</sup> no parece prejuzgar la fundamental homogeneidad del bien y, en consecuencia, la propia unidad y continuidad de la acción racional. Todo agente, por el hecho mismo de deliberar prácticamente y de resolverse a obrar ha de concebir el objeto de su acción como productivo de un bien.

Aunque con reservas, Kant acepta la que llama «vieja fórmula de las escuelas» según la cual *nihil appetimus nisi sub ratione boni, nihil aversamus nisi sub ratione mali*<sup>10</sup>. Sus reservas se deben a la ambigüedad que cree advertir en el uso unívoco de las nociones de bien y de mal. Esta enmascara no sólo la distinción entre lo agradable o bueno sensible (*angenehm, wohl*) y lo bueno racional (*gut*) sino, dentro de este, entre los tipos de bondad determinados por los modos de necesidad objetiva expresada en los imperativos prácticos; el imperativo moral debe su carácter categórico a la necesidad apodíctica del juicio que expresa la obligación moral<sup>11</sup>. La radical heterogeneidad de la noción de bien como virtud y como felicidad fundamentan la exigencia kantiana de autonomía para la moralidad. Aunque ello la convierte en un universo literalmente sui generis escindido del resto de la racionalidad práctica, Kant intenta, con su concepto de *bonum consummatum*, integrar en una totalidad plena la supremacía cualitativa del bien moral con el justificado anhelo de felicidad, mostrando al menos que es racional la aspiración a la unidad de la razón práctica.

La estructura de toda teoría ética depende por tanto en gran medida del modo como define y conecta ambos polos nocionales: uno primariamente valorativo en torno al concepto de «bueno» y otro predominantemente deóntico en torno a las nociones de «justo», «correcto», «obligatorio». El Utilitarismo se caracteriza en esencia por definir lo bueno con in-

8. *EN* 1094a18-22; 1096b8-34.

9. *MET* 1013b27-28.

10. *CRP*, Analítica, c.II; L. W. BECK reconoce no haber sido capaz de identificar literalmente dicha fórmula, aunque «el sentido y algunos de sus términos» están en Wolff y Baumgarten: *A commentary on Kant's Critique of Practical Reason* (Chicago, University Press, 1960) p. 131n.

11. La distinción entre *bonitas problematica, pragmatica et moralis* se corresponde con el carácter problemático, asertórico o apodíctico del principio que relaciona la acción con un propósito ulterior, en los dos primeros casos, o que expresa simplemente su necesidad objetiva sin referencia a propósito alguno (*FMC* c.2: *Lecciones de Ética* (Barcelona, Crítica, 1988). Como contraste, la división tomista del *bonum* en *honestum, utile et delectabile* se funda en la distinción sustantiva entre lo que es término del movimiento apetitivo —ya sea como algo deseado por sí mismo (*honestum*) o como algo en lo que descansa el apetito cuando lo alcanza (*delectabile*)— o lo que es medio para tender hacia otra cosa (*utile*). Lo bueno así dividido no es unívoco, sino análogo, pues *praedicatur secundum prius et posterius; per prius praedicatur de honesto, et secundario de delectabili, tertio de utili* (*STh* 1 q.5 a.6; 2-2 q.145 a.3).

dependencia de lo justo<sup>12</sup>. Desde una perspectiva kantiana el Utilitarismo no parece, pues, ofrecer tanto una forma alternativa de justificación moral de la conducta, cuanto una guía heterónoma de conducta que se presenta, pura y simplemente, como alternativa a la moralidad misma<sup>13</sup>. Pero desde la perspectiva propia del Utilitarismo lo que se intenta es más bien integrar la moralidad en una teoría general de la racionalidad práctica<sup>14</sup>.

### 3. La teoría consecuencialista

Para fundamentar de forma sistemática la respuesta a la pregunta por lo que se ha de hacer, toda teoría moral debe proponer algún criterio de lo que es objetiva o intrínsecamente bueno y de la acción moralmente correcta. Formalmente hablando, una acción es objetivamente correcta si es sistemáticamente implicada por el principio que define lo que es objetiva o intrínsecamente bueno, y la obligación objetiva del agente es ejecutar la acción objetiva correcta. Todo cuanto contribuya a que el agente actúe moralmente deriva su valor moral del mismo principio. Si la teoría presupone, además, la posibilidad de establecer comparaciones entre los esta-

---

12. La formulación es de J. RAWLS: *A theory of justice* (Oxford. University Press. 1971) pp. 24-25. La constante recurrencia de esa dualidad conceptual en la historia de la ética parece confirmar que responde a una propiedad intrínseca de la moralidad misma: cfr. G. GUTIÉRREZ: *La congruencia de lo bueno y lo justo* «Revista de Filosofía», 2.ª serie 2(1979)33-54; G. H. v. WRIGHT en *The varieties of goodness* (Londres. Routledge. 1963) discute la naturaleza de la unidad entre las variedades del bien y, en particular, del bien moral, especialmente en los cc. 1 y 8.

13. B. GERT: *Moral rules* (Nueva York. Harper. 1973) pp. 8-9.

14. Así lo propone J. C. HARSANYI en *Avances in understanding rational behavior*, reproducido en sus *Essays on ethics, social behavior and scientific explanation* (Dordrecht. D. Reidel. 1976) pp. 96-97; también en R. BUTTS y J. HINTIKKA (eds): *Foundational problems in the special sciences* (D. Reidel. Dordrecht. 1977), pp. 323-4; de forma más directa en *Morality and the theory of rational behavior*, reproducido en A. SEN y B. WILLIAMS (eds): *Utilitarianism and beyond* (Cambridge University Press. 1982), pp. 39-62. Las teorías contractualistas comparten en general una misma aspiración a deducir la moralidad de la racionalidad no-moral; el intento más sistemático de generarla como «una restricción racional de las premisas no-morales de la elección racional» entendiendo la teoría moral esencialmente como «una teoría de las restricciones optimizantes sobre la maximización de la utilidad» es el de D. GAUTHIER: *Morals by agreement* (Clarendon Press. Oxford. 1986); cfr. pp. 4, 16-7, 78. En la medida en que las morales deontológicas se resisten a esta deducción —o, si se prefiere, reducción— aparecen como la negación misma de la racionalidad; cfr. J. MOSTERIN: *Racionalidad y acción humana* (Madrid. Alianza. 1977) pp. 34-35. En términos kantianos, por el contrario, el intento mismo es tan vano como esperar peras de un olmo; cfr. G. GUTIÉRREZ: *Ex pumice aquam?* en J. MUGUERZA y R. R. ARAMAYO: *Kant después de Kant* (Madrid. Tecnos. 1989) pp. 142-7.

dos de cosas o entre las acciones de los agentes por referencia a su bondad intrínseca o a su corrección u obligatoriedad moral, ha de definir asimismo el punto de vista y la escala de valoraciones que permite jerarquizarlos de mejor a peor.

La teoría consecuencialista sostiene que, en una situación dada, la acción moralmente correcta es la que produce el mejor resultado global posible desde una perspectiva impersonal que concede igual peso a los intereses de todos los afectados<sup>15</sup>; es decir, que la corrección moral de una acción depende de sus consecuencias buenas o malas<sup>16</sup>. Pero como no se puede afirmar, so pena de incurrir en una regresión infinita, que todo lo moralmente valioso tiene valor en virtud de sus consecuencias, es lógicamente necesario afirmar que hay cosas que poseen valor intrínseco<sup>17</sup>.

Lo distintivo del Consecuencialismo es afirmar que lo único intrínsecamente valioso son precisamente los estados de cosas que las acciones contribuyen a producir<sup>18</sup>. Entre las diversas interpretaciones materiales — no todas necesariamente utilitaristas, ni necesariamente monistas— de este principio formal se cuenta, por ejemplo, la tesis utilitarista clásica que sostiene que «la felicidad o placer universal es lo único absolutamente bueno o deseable; el fin al que la acción de un agente razonable como tal debe dirigirse»<sup>19</sup>, o la fórmula mooreana según la cual «ciertos estados de

---

15. A diferencia de lo que ocurre con otras teorías morales vinculadas a un autor, no existen «autoridades» análogas en la teoría consecuencialista. Si se insiste en hablar de una *teoría* lo cierto es que, a pesar de ser un lugar de referencia común en la filosofía moral contemporánea, no ha sido sistematizada por ningún autor en particular. Más bien se ha tratado de hacer explícitos los supuestos consecuencialistas implícitos en otras teorías, sobre todo, aunque no exclusivamente, en el Utilitarismo. Una muestra representativa de lo que se sostiene en nombre de la teoría, precedida de una buena introducción la ofrece S. SCHEFFLER (ed): *Consequentialism and its critics* (Oxford. University Press. 1988), a quien se debe la caracterización citada (p.1); del mismo autor: *The rejection of consequentialism* (Oxford. Clarendon Press. 1982). La primera parte de la obra de D. PARFIT: *Reasons and persons* (Oxford. University Press. 1985) es el mejor análisis de la estructura formal de la teoría. También B. WILLIAMS: *A critique of utilitarianism*, en: J. J. C. SMART y B. WILLIAMS: *o.c.* en nota 1, p. 9. Escribir con mayúscula los nombres de ambas teorías no es más que un recurso tipográfico que no implica hipostasiarlas.

16. J. J. C. SMART: *art.cit.*, p. 9.

17. El razonamiento es análogo al aristotélico sobre el fin último: cfr: W. F. R. HARDIE: *Aristotle's ethical theory* (Oxford. Clarendon Press. 1980) p. 16-7; G. E. M. ANSCOMBE: *Intention* (Oxford. Blackwell. 1957) p. 33-4; o al humeano en *An enquiry concerning the principles of morals*. Ed. L. A. SELBY-BIGGE (Oxford. Clarendon Press. 1966) Appendix I, p. 293.

18. B. WILLIAMS: *art. cit.* en nota 15, p. 83.

19. H. SIDGWICK: *The methods of ethics* (Indianapolis/Cambridge. Hackett. 1981) p. 421.

consciencia... como los placeres de la comunicación humana y el disfrute de objetos bellos (son) algo bueno en sí mismo»<sup>20</sup>.

La noción de estados de cosas pretende designar configuraciones del mundo específicamente definidas por (o por sus efectos en) los estados —en el sentido más amplio del término— de las personas que los integran y que, como tales, son tanto objeto de las preferencias y las elecciones de los agentes como producto de sus acciones individuales y colectivas. Dado que estos estados de cosas son esencialmente configuraciones complejas de vectores individuales o personales, resulta más preciso, dentro de la inevitable vaguedad del concepto, hablar de «estados sociales»<sup>21</sup>.

Sostener que los estados sociales poseen valor intrínseco sólo significa que, en tanto que metas finales y productos terminales de la actividad racional, no se ordenan a la producción ulterior de ningún otro estado de cosas. Constituye esta una de las interpretaciones posibles del axioma antropocéntrico de la mayor parte de las teorías éticas que afirma que sólo el hombre es un fin en sí. El cometido que la teoría consecuencialista atribuye a la Ética concebida como parte de una teoría general de la elección (o de la acción) racional depende a su vez de las peculiaridades que presenta la producción de los estados sociales.

Por otra parte, el concepto de estados sociales responde formalmente a la exigencia utilitarista tradicional de considerar la conducta objetivamente correcta como aquella «que producirá la mayor cantidad de felicidad *en conjunto*, esto es, teniendo en cuenta a todos aquellos cuya felicidad sea afectada por la conducta»<sup>22</sup>. Lo específicamente consecuencialista no es el contenido de lo que ha de maximizarse sino la exigencia misma de hacer máximo un valor total para un conjunto definido de individuos. Ya se ha visto cómo esta misma indefinición formal ha permitido a la teoría consecuencialista conservar su identidad teórica a lo largo de su historia y desprenderse de sus propias interpretaciones coyunturales desde el primitivo concepto hedonista de felicidad presente en las formulaciones de Sidgwick hasta las elaboraciones contemporáneas del concepto de bienes-

20. G. E. MOORE: *Principia ethica* (Cambridge. University Press. 1976) p. 188-9. Traducción española en UNAM. México. 1959.

21. Es el término corriente empleado para referirse a algo tan complejo como «una descripción completa de la cantidad de cada tipo de bienes en manos de cada individuo, la cantidad de trabajo que cada individuo ha de aportar, la cantidad de cada tipo de factor productivo invertido en cada tipo de actividad productiva y las cantidades de los diversos tipos de actividades colectivas»: K. J. ARROW: *Elección individual y valores sociales* (Madrid. Instituto de Estudios Fiscales. 1974) p. 77). Véase también A. SEN: *Elección...*, citado en nota 2, p. 17-21; P. J. HAMMOND: *Utilitarianism, uncertainty and information*, en SEN y WILLIAMS: *o.c.* en la nota 14, p. 86; I. LEVI: *Liberty and welfare*, *ib.* p. 239-40).

22. H. SIDGWICK, *o.c.* p. 411.

tar social como satisfacción de las preferencias, necesidades o intereses, etc.<sup>23</sup>.

El concepto mismo de maximización global supone que tiene sentido considerar un estado social como la resultante de algún tipo de integración o totalización de los estados individuales, lo que implica que es posible compararlos entre sí<sup>24</sup>. La comparabilidad de los estados es un supuesto esencial del Consecuencialismo<sup>25</sup> y una implicación necesaria de la primacía que concede a lo bueno sobre lo moralmente correcto<sup>26</sup>.

En todo caso, el estado social o de cosas así definido permite en principio a la teoría consecuencialista identificar determinadas situaciones cuya complejidad no les impide ser globalmente valoradas y objetivamente descritas con independencia de toda noción *moral* previa. La definición responde a su proyecto metodológico de naturalizar la ética, esto es, de integrar la moralidad en el ámbito común de la racionalidad práctica tal como ésta se ejerce en el resto de las actividades humanas, sin apelar a la concepción, tenida por oscura o incluso irracional, de la autonomía de la moral. Lo que implica el propósito humeano de «introducir el método experimental de razonamiento en los asuntos morales» es, a fin de cuentas,

23. Sobre la dimensión moral del concepto de bienestar, cfr. las *oo.cc.* de J. GRIFFIN y N. RESCHER en nota 3; sobre el de preferencias, cfr. M. HOLLIS: *The cunning of reason* (Cambridge. University Press. 1987) p. 64s; J. ELSTER: *Sour grapes* (Cambridge. University Press. 1983) *passim*; hay traducción española: *Uvas amargas* (Barcelona. Península. 1987).

24. La comparación puede ser estrictamente *cardinal* dentro de una escala numérica, discontinua o interválica, o bien, de forma más compleja, *ordinal* según la ordenación definida por una función lineal y continua; cfr. M. FLEMING: *A cardinal concept of welfare* (1952) reproducido en J. PHELPS (ed.): *o.c.* en nota 3, p. 245-65; K. J. ARROW: *Some ordinalist-utilitarian notes on Rawls' theory of justice* «Journal of Philosophy» 70 (1973) p. 245-63.

25. B. WILLIAMS *art. cit.* en nota 14, p. 88.

26. Aparte de ser objeto de críticas de principio, esta exigencia plantea considerables dificultades a la versión bienestarista de la teoría utilitaria cuya operatividad práctica depende de la posibilidad de llevar a cabo el tipo de comparaciones interpersonales de utilidad que permiten definir una «función de bienestar social» en el sentido de A. BERGSON: *art. cit.* en nota 3. Gracias a ella podrían ordenarse diversos estados sociales alternativos según una escala de bienestar social entendido como una función matemática de diversas variables que se pretende dependientes, en última instancia, de las preferencias valorativas individuales. Lo que se convierte en problemático es entonces la posibilidad misma de integrar estas preferencias de forma que satisfagan simultáneamente cierto conjunto de axiomas formales y de condiciones materiales intuitivamente asociadas a una sociedad democrática. El locus clásico de esta cuestión es la obra de K. J. ARROW: *Social choice and individual values* (New Haven CT. Yale U. P. 1951), Véase J. KELLY: *Arrow's impossibility theorems* (Nueva York. Academic Press. 1978); A. MACKAY: *Arrow's theorem: the paradox of social choice* (New Haven CT. Yale U.P. 1980).

que nuestros juicios de valor sobre las cosas que son buenas pertenecen al dominio de la experiencia no-moral y que la obligación moral básica no es más que una aplicación del principio práctico general de maximización que nos ordena elegir la mejor alternativa. En definitiva, pues, la experiencia nos dice qué estados de cosas valoramos como buenos, y la ética nos dice que la acción moralmente correcta, y por tanto obligatoria, es la que produce el mejor estado de cosas.

Se ha hecho notar que la versión monista de la teoría consecuencialista —como la de Sidgwick antes citada— no es la única posible, pues una teoría consecuencialista puede apelar a múltiples principios y valores<sup>27</sup>, incluidos aquellos que a primera vista aparecen habitualmente asociados a teorías que no son consecuencialistas. Podría entonces creerse que, si se lo define de forma suficientemente amplia, el Consecuencialismo pierde todo carácter específico para convertirse en una condición trivialmente necesaria de cualquier teoría moral, incluida la kantiana, y tan formal y vacía de contenido como la expresada en el axioma *bonum est faciendum, malum est vitandum*<sup>28</sup>. De ser así se borraría la diferencia, para muchos esencial, entre las llamadas éticas deontológicas —típicamente, la kantiana— y las teleológicas —las consecuencialistas en general y la utilitarista en particular.

El contraste radical punto entre ambas se manifiesta, por ejemplo, en el valor que cada una concede a la integridad y a la autonomía de los agentes personales, que para el deontologismo implica la condición inviolable de la persona y prohíbe considerarla como mero medio y no como fin. Esta condición fundamenta su derecho inalienable a no ser tratada de cierta forma por otros e impone a los demás la prohibición estricta de tratarla de esa forma. Pero si en términos consecuencialistas nada impide considerar como intrínsecamente valiosos aquellos estados terminales de cosas en los que se logra que desaparezcan o se minimicen las violaciones de este derecho la diferencia entre ambas concepciones parece desvanecerse. Como veremos, esta apariencia es ilusoria.

### 3. La interpretación utilitarista del consecuencialismo

El Utilitarismo no es sino uno de los desarrollos posibles de los supuestos consecuencialistas básicos según los cuales: 1) existe un último objetivo moral y es que los resultados sean tan buenos como sea posible; 2) cada uno de nosotros debe hacer todo lo que produzca un resultado mejor; 3) si alguien hace lo que cree que producirá un resultado peor, ac-

27. Cfr. p. ej. C. D. BROAD: *On the function of false hypothesis in ethics* (1916), reproducido en D. R. CHENEY (ed): *Broad's critical essays in moral philosophy* (Londres, Allen & Unwin, 1971) p. 53-54; D. PARFIT: *o.c.*, p. 48-9.

28. J. SMART: *art. cit.*, p. 13; B. WILLIAMS, *id.*, p. 83.

túa inmoralmente<sup>29</sup>. Por esta razón, quien hace aquello cuyos resultados sabe que serán peores obra mal en términos objetivos, pues produce una cantidad de bien menor que la que está a su alcance y es axiomáticamente verdadero que el bien es moralmente preferible al mal. Negar esto parece comprometernos a afirmar que la moralidad nos prohíbe hacer todo el bien que podríamos o nos manda hacer menos del que seríamos capaces<sup>30</sup>.

Como teoría moral el Consecuencialismo propone como objetivo moral definitivo lograr los mejores resultados posibles y para ello extiende sus exigencias absolutamente a todo cuanto contribuye a maximizar el bien total, aplicando con todo rigor el axioma según el cual *bonum ex integra causa, malum ex quocumque defectu*. Por esta razón, en última instancia, carece de sentido «subdividir la conducta humana en conducta que plantea cuestiones morales y conducta que no las plantea»<sup>31</sup> ya que toda cuestión acerca de lo que es racional hacer, evitar o incluso impedir que ocurra versa en definitiva sobre la producción de las mejores consecuencias y se convierte, por definición, en una cuestión moral. Esta característica amplía ilimitadamente la noción de responsabilidad moral y hace del Consecuencialismo, en especial en su formulación como «Consecuencialismo del acto», una doctrina exigente hasta extremos poco realistas, lo que afecta a su propia viabilidad como teoría libre de paradojas<sup>32</sup>.

El atractivo intelectual del Consecuencialismo radica en la simplicidad de su principio básico que afirma que la corrección moral de los actos es un valor esencialmente instrumental, función de su capacidad de producir consecuencias buenas. La bondad de las consecuencias deriva en última instancia de los fines últimos o metas terminales, que son los únicos estados de cosas intrínsecamente buenos. Para la mayor parte de las teorías éticas es axiomático que sólo el hombre o la naturaleza huma-

29. D. PARFIT, *o.c.*, p. 24.

30. S. SCHEFFLER, *o.c.*, p. 1.

31. D. REGAN: *Utilitarianism and cooperation* (Oxford. Clarendon Press. 1980) p. 237.

32. Véase la formulación de W. GODWIN en *Equity concerning political justice* (Harmondsworth. Penguin. 1976) p. 174; cfr. J. L. MACKIE: (Harmondsworth. Penguin. 1977) p. 129-134; al hacer al agente responsable *positivamente* de lo que hace y *negativamente* de lo que no impide que ocurra, el Consecuencialismo convierte a este en un mero canalizador de los proyectos de todos, incluido el suyo, hacia una decisión optimílica impersonal; con ello invade su autonomía e integridad, alienándole de *sus propios* proyectos como persona individual: B. WILLIAMS, *art. cit.* en nota 15, p. 93s., 116-7; véanse las propuestas de S. SCHEFFLER: *The rejection...* (citada en nota 15) y de M. SLOTE: *Common-sense morality and consequentialism* (Londres. Routledge. 1985) para exonerar al Consecuencialismo de la carga de implausibilidad que arrastra su concepto de responsabilidad negativa; sobre las paradojas en que incurre, cfr. D. PARFIT, *o.c.* p. 29s.

na es un fin en sí. Pero a pesar de la habitual asociación del Utilitarismo con el individualismo y hasta con el liberalismo, lo cierto es que, como teoría práctica, interpreta la naturaleza humana en sentido colectivo, de forma que ser fin en sí no es un atributo esencialmente inherente a las personas individuales sino a los estados terminales, que en aspectos muy decisivos desconocen la singularidad y el carácter único de aquellas.

Puesto que la lógica misma del valorar implica necesariamente la referencia a un punto de vista y los estados sociales se componen de una pluralidad de agentes individuales con intereses y valores dispares, es patente la necesidad del recurso consecuencialista a un punto de vista que sea imparcial por definición y asigne impersonalmente un peso igual a los intereses individuales. Como forma suprema de la racionalidad colectiva, la moralidad es por definición imparcialidad<sup>33</sup> y consiste esencialmente en seguir la regla de oro que exhorta a tratar a los demás como uno desea ser tratado. Porque sólo si «cada uno cuenta como uno y nadie como más de uno», según la fórmula benthamiana, podrá satisfacerse la condición necesaria para agregar las utilidades individuales en el conjunto social. Estas mismas premisas ya llevaron a Adam Smith a introducir la figura del espectador imparcial e indiferente que alienta en el pecho de todo hombre razonable y juzga de los sentimientos y motivos de cada uno con la suficiente distancia, y con cuya aprobación nos identificamos por medio de la simpatía<sup>34</sup>. Gran parte de la teoría utilitarista contemporánea acepta la definición de las reglas morales como aquellas reglas de conducta hipotético-imperativas «que recomendaría para su observancia general un observador bien informado, inteligente y dotado de imparcial simpatía»<sup>35</sup>.

En su versión «bienestarista» el Consecuencialismo ha desarrollado un modelo de la toma de decisiones públicas o colectivas que opera con el supuesto de un único agente que «lleva a cabo la requerida organización de los deseos de todas las personas en un sistema coherente de deseos... fundiendo así las muchas personas en una»<sup>36</sup>. Este elige entre los diferentes estados sociales con el propósito de maximizar la utilidad general —que no tiene por qué ser necesariamente la *suma* neta de las utilidades individuales, sino cualquier otra fórmula matemática que satisfaga determinados principios normativos, como los de la diferencia, la utilidad promedio, etc. —de forma análoga a como el agente individual maximiza la suya propia. Ya vimos que ha merecido por ello el fundado reproche de

33. B. WILLIAMS, *art. cit.*, p. 95-96.

34. A. SMITH: *The theory of moral sentiments*. En D. D. RAPHAEL (ed): *British moralists 1650-1800*. Vol. 2. (Oxford. Clarendon Press. 1969) p. 226.

35. Por ejemplo, J. HARSANYI: *Ethics in terms of hypothetical imperatives* (1958), reproducido en sus *Essays on ethics, social behavior and scientific explanation* (Dordrecht. D. Reidel. 1976) p. 28s.

36. J. RAWLS. *o.c.*, p. 27.

extrapolar a la escala de la racionalidad estratégica los criterios que sólo son aplicables en la de la racionalidad paramétrica, en la cual es un único agente individual el que toma las decisiones, y hacer así plausible en principio la violación de los derechos de un individuo para optimizar el estado social resultante. En efecto, si intrínsecamente valiosos son sólo los estados terminales y todo lo demás —incluidos los agentes, su autonomía, su integridad y sus acciones— es bueno en la estricta medida de su contribución causal a la producción de aquellos, parece no sólo permisible sino incluso obligatorio «victimizar» al agente individual.

Esta conclusión se sigue de la exigencia consecuencialista de considerar la bondad de todas las cosas «desde el punto de vista del universo», lo que implica la esencial irrelevancia de la perspectiva tanto individual como temporal de los agentes, tomadas en sí mismas. Ya Sidgwick advertía que la felicidad de una persona, precisamente en cuanto *suya* «no puede ser una parte más importante del Bien, tomado universalmente, que la felicidad igual de cualquier otra persona»<sup>37</sup>. Por la misma razón, el agente que actúa racionalmente en beneficio de su propio futuro no debe anteponer sus preferencias puramente temporales a la preocupación imparcial respecto a los diversos períodos de una misma vida, ni debe privilegiar por tanto ninguno de ellos por el mero hecho de ocupar una determinada localización temporal. Este postulado de la racionalidad consecuencialista lleva a Sidgwick a afirmar que la obligación que cada uno tiene de tender a su propio bien en conjunto supone un principio de «preocupación imparcial por todas las partes de la vida consciente» del agente. Es decir, que «el Después, como tal, no debe ser considerado ni más ni menos que el Ahora», lo que no quiere decir «que el bien del presente no haya de ser preferido al del futuro por razón de su mayor certeza», sino simplemente que «la mera diferencia de prioridad y posterioridad en el tiempo no es un fundamento razonable para prestarle mayor atención a la consciencia de un momento que a la de otro»<sup>38</sup>. El Consecuencialismo impone por su propia naturaleza una perspectiva cósmica, neutral respecto del agente —un «punto de vista desde ninguna parte»<sup>39</sup>.

El Utilitarismo, por tanto, justifica en principio la posibilidad de efectuar transferencias redistributivas interpersonales de todo tipo de bienes, con compensaciones o incluso sin ellas, si sirven al propósito de maximizar el bien *conjunto* de los integrantes de un estado social. Las éticas deontológicas, por el contrario, circundan a los individuos con una barrera infranqueable de derechos «naturales» —en el sentido de no ser socialmen-

---

37. H. SIDGWICK, o.c., p. 420-1.

38. *Ibid.*, p. 381.

39. T. NAGEL: *The view from nowhere* (New York. Oxford University Press. 1986), c.9; cfr. J. FINNIS: *Natural law and natural rights* (Oxford. Clarendon Press. 1980) p. 111-33; *Fundamentals of ethics* (Oxford, Clarendon Press. 1983). p. 86s.

te conferidos— e imponen restricciones morales absolutas respecto a lo que puede hacer o hacérsele a un agente aunque sea con la finalidad de procurar a *otro* agente o al conjunto de ellos el mismo tipo de bien intrínseco (o evitarle el mal intrínseco) que define la propia teoría<sup>40</sup>.

#### 4. Los objetivos de una teoría moral

Dos distinciones ayudan a entender hasta qué punto entre ambos tipos de teorías subsiste una diferencia esencial. La primera, entre los objetivos *formales* y los objetivos *sustantivos* o materiales de una teoría moral. Toda teoría moral distingue entre acciones correctas e incorrectas, buenas y malas, morales e inmorales. El objetivo formal común a todas las teorías es actuar moralmente. El objetivo sustantivo es propio de cada una y lo constituye su definición de aquello en lo que consiste actuar moralmente. No es necesario, pero tampoco imposible, entender este objetivo como un fin para el que la actuación correcta es un mero medio en el sentido de algunas éticas teleológicas, ya que incluso una ética deontológica que prohibiese cinco tipos de actos «nos propondría a cada uno de nosotros como objetivo (sustantivo) no actuar nunca de esas cinco maneras»<sup>41</sup>.

La segunda distinción diferencia las teorías *neutrales respecto del agente*, que proponen a todos los agentes objetivos comunes, de las teorías *relativas al agente*, que proponen a diferentes agentes objetivos diferentes<sup>42</sup>. Una teoría que considerase intrínsecamente bueno un estado terminal de cosas en el que no se violasen derechos fundamentales sería consecuencialista si definiese como objetivamente correctas las acciones que contribuyesen a producirlo y propusiese su realización como objetivo sustantivo común para todo agente. Sería «relativa al agente» y, en esa

---

40. R. NOZICK sostiene que extrapolar a la colectividad el mismo tipo de balance que una persona individual hace, a la hora de imponerse sacrificios en áreas de *su* propio bien superior, entre *sus* propios costes y beneficios implica suponer falazmente que existe «una entidad social con un bien que se impone algún sacrificio por su propio bien. (Pero) sólo existen personas individuales, diferentes personas individuales con sus propias vidas individuales, diferentes personas individuales con sus propias vidas individuales. Usar a una de estas personas en beneficio de otras la usa a ella y beneficia a las otras... (algo que) no respeta suficientemente ni tiene en cuenta el hecho de que ella es una persona separada, que la suya es la única vida que tiene. *Ella* no consigue ningún bien que le compense de su sacrificio y nadie tiene derecho a imponérselo... No hay forma de contrapesar moralmente una de nuestras vidas con otras para alcanzar un bien *social* global. No existe ningún sacrificio justificado de alguno de nosotros en favor de otros»: *Anarchy, state and utopia* (Oxford. Blackwell. 1974). p. 32-33, subrayados suyos; cfr. los ya citados SCHEFFLER y SLOTE, *passim*; también J. RAWLS, *o.c.*, p. 26.

41. D. PARFIT: *o.c.*, p. 3, 9, 37, 45.

42. *Id.*, p. 54-55.

medida, no-consecuencialista, si propusiese a todo agente como objetivos de sus acciones no que *no haya* al final violaciones de derechos, sino que *él no los viole*. Juzgaría moralmente incorrecto que *yo* violase los derechos de los otros, incluso aunque violándolos se lograra hacer desaparecer al final toda violación de aquellos<sup>43</sup>. En el primer caso la teoría no incluiría entre sus objetivos *sustantivos* el obrar moralmente y podría considerarlo como un mero medio para lograr el único objetivo que verdaderamente lo es<sup>44</sup>.

Para el Consecuencialismo, pues, las acciones no poseen necesariamente valor intrínseco, sino que valen según su capacidad causal de producir estados de cosas cuya bondad es en principio independiente de su forma de causación<sup>45</sup>. Mas incluso en el caso de que se considere la propia acción intrínsecamente buena «su corrección moral se deriva de la bondad de cierto estado de cosas y es correcta *porque* el estado de cosas que consiste en que se ejecute es mejor que cualquier otro estado de cosas accesible al agente, mientras que para los no-consecuencialistas ocurre, al menos a veces, lo contrario, y un estado de cosas es mejor que sus alternativas porque consiste en que se ejecute la acción correcta»<sup>46</sup>.

## 5. Las paradojas de la teoría

La experiencia enseña que el hecho de que los agentes posean ciertos tipos de motivos y disposiciones contribuye eficazmente a producir estados de cosas intrínsecamente valiosos<sup>47</sup>. Como el Consecuencialismo «cubre todo cuanto puede hacer el resultado mejor o peor, incluidos motivos y disposiciones», una extensión natural de sus principios lleva a concluir que «los motivos mejores posibles son aquellos de los que puede afirmarse que, si los tenemos, el resultado será mejor»<sup>48</sup>. Si somos consecuencialistas estaremos pues obligados a introducir en el mundo y en nosotros mismos cuantos cambios contribuyan a mejorar el resultado —en este caso, nuestros motivos para actuar.

---

43. *Id.*, p. 27; R. NOZICK: *o.c.*, p. 28-33; B. WILLIAMS: *o.c.*, p. 108-111. La preocupación por la propia pureza moral sería, en términos consecuencialistas, un residuo de la posición privilegiada que irracionalmente se atribuye a sí mismo el sujeto en el cosmos o en la historia. El terrorismo revolucionario es en este sentido el paradigma del más consecuente Consecuencialismo. Véase T. HONDERICH: *Violence for equality* (Harmondsworth. Penguin Books, 1980).

44. *Id.*, p. 37, 45-9; M. SLOTE: *o.c.*, p. 5.

45. *Id.*, p. 26-27; B. WILLIAMS: *art. cit.*, p. 84.

46. B. WILLIAMS: *art. cit.*, p. 87; subrayado suyo.

47. Véase, p. ej., R. M. ADAMS: *Motive utilitarianism*. «Journal of Philosophy» 73 (1976).

48. D. PARFIT: *o.c.*, p. 26.

Siguiendo con el ejemplo anterior, sería no sólo recomendable sino incluso obligatorio, en términos estrictamente consecuencialistas, fomentar una aversión a engañar, coaccionar o traicionar a otros. Una disposición negativa, en suma, a violar sus derechos morales. Consecuencia de esta disposición, y condición de su efectividad, es que el agente considere inmoral obrar de esa forma, y experimente remordimientos y sentimiento de culpa al hacerlo. En general, el Consecuencialismo no rompe el vínculo que se establece entre la creencia de que un acto es inmoral y la experiencia del remordimiento<sup>49</sup>. Al disuadir de la comisión de actos inmorales esta disposición produce *en general* resultados mejores. Pero ese vínculo habitual puede romperse en aquellos casos especiales en los cuales lo que produce un resultado mejor es precisamente actuar de manera inmoral. En esta situación la teoría exige al agente que venza su repugnancia a obrar inmoralmemente, algo que la propia teoría define como moralmente incorrecto. Su aversión a hacerlo es irracional en términos estrictamente utilitarios, pues nada se gana con tenerla<sup>50</sup>.

Pero al romper la conexión entre nuestra conciencia de *obrar* inmoralmemente y nuestra conciencia de *ser* inmorales, dejamos de considerar la moral con suficiente seriedad. Si no creemos que nuestra aversión a obrar inmoralmemente ha de prevalecer sobre cualquier otra consideración y no sentimos remordimientos cuando no es así, la propia disposición se vería socavada y dejaría de ser eficaz en términos consecuenciales. Pero como la teoría misma nos manda hacernos con los mejores motivos y disposiciones, con sus correspondientes creencias y emociones, se lograría un resultado mejor si revisásemos nuestras creencias morales, dejásemos de creer en la teoría consecuencial y adoptásemos en su lugar alguna versión refinada de la moralidad del sentido común. En términos coloquiales diríase que el mejor favor que se hace la teoría a sí misma es recomendar que no se la crea. Afirmaría así su propia verdad, a la inversa del principio de Berkeley, desvaneciéndose de la percepción.

En sus diferentes versiones, pues, las teorías consecuencialistas nos proponen como objetivo común que las cosas resulten lo mejor posible, pero no necesariamente que nunca actuemos de modo inmoral<sup>51</sup>. Una teoría moral que concibiese el actuar conforme a los principios de la propia teoría como un mero medio para alcanzar los objetivos de ésta, podría incluso recomendar que no se le hiciera caso, que se la olvidara o que se la sustituyera por otra. Poseería la paradójica propiedad de ser, en la terminología de Parfit, *autoevanescente* (*self-effacing*).

49. *Id.*, p. 35.

50. B. WILLIAMS: *art.cit.*, p. 101.

51. Para «la forma más conocida de Consecuencialismo, el Utilitarismo... evitar obrar incorrectamente en un mero medio para el logro del único objetivo moral sustantivo, y no es en sí mismo un objetivo sustantivo»: D. PARFIT: *o.c.*, p. 37.

Ya advertía Ryle del «aroma de absurdidad» que exhala el propósito expreso de «olvidar la diferencia entre el bien y el mal», sobre todo cuando la única razón para hacerlo no se basa en dudas intelectuales sobre nuestras viejas creencias, sino en la convicción de que se lograrían mejores resultados si tuviésemos creencias distintas<sup>52</sup>. Entre los posibles mecanismos para borrar el Consecuencialismo de nuestra mente, Parfit menciona uno de raigambre orwelliana: hipnotizar una generación y educar a la siguiente de forma diferente, haciéndola olvidar cómo y por qué adquirió las nuevas creencias y ocultando el proceso a los historiadores futuros. Pero como las consecuencias de borrar la teoría consecuencialista de la conciencia moral colectiva pueden ser desastrosas, debe adoptarse la precaución de mantener a unos pocos en el conocimiento de la verdad y en posesión de las pruebas del engaño, por si fuere necesario reintroducir la creencia en la verdadera teoría moral.

## 6. El Utilitarismo como moralidad pública

Hemos visto que el Utilitarismo —en sus diferentes formulaciones y como versión más conocida, aunque no única, del Consecuencialismo— define el mejor estado de cosas como aquél que procura a los individuos la mayor suma neta de beneficios, felicidad, bienestar, satisfacción de sus preferencias, etc. Si lo distintivo del Consecuencialismo es sostener que el valor de las acciones no es nunca intrínseco sino siempre consecuencial o derivativo y referido siempre a sus consecuencias sobre los estados finales de cosas, la aportación distintiva del Utilitarismo es su específica interpretación de las variadas características que hacen «buenas» a las consecuencias, entre las que figuran las versiones contemporáneas que requieren que las decisiones sociales satisfagan criterios tales como el de maximizar la utilidad promedio<sup>53</sup> o la de los menos favorecidos<sup>54</sup>.

Para el Consecuencialismo, el *valor* moral de las *acciones* depende de las «buenas» consecuencias que produzcan, mientras que la *obligación* moral objetiva que los *agentes* tienen de producirlas se funda en el carác-

---

52. G. RYLE: *On forgetting the difference between right and wrong*. En A. MELDEN: *Essays in moral philosophy* (Seattle, Washington University Press, 1958) p. 155-59. Acechan aquí las paradojas que suscitan el autoengaño o la decisión de creer; cfr. D. PARFIT: *o.c.*, p. 40-43; J. ELSTER: *Sour grapes*, p. 47-54 sobre la apuesta de Pascal; p. 172-9 sobre el autoengaño; p. 148s. sobre las creencias interesadas y el creer a voluntad; también la recopilación de ELSTER: *The multiple self* (Cambridge, University Press, 1985); B. WILLIAMS: *Deciding to believe* (1970) incluido en *Problems of the self* (Cambridge, University Press, 1973).

53. p. ej., J. HARSANYI en *oo.cc.* en nota 14.

54. p. ej., J. RAWLS en *o.c.*, aunque no pueda considerársele un utilitarista en sentido estricto.

ter axiomático de la preferibilidad del bien sobre el mal. La distinción entre lo que es correcto objetiva y subjetivamente afecta a la diferencia entre aquello que constituye la razón objetiva para hacer algo y aquello que resultaría racional hacer dada la información que poseemos. El Utilitarismo es primariamente una teoría sobre la norma objetiva de la moralidad, es decir, sobre lo que hace buena o mala una acción. Y sólo derivadamente ofrece un procedimiento de decisión que incluye, por ejemplo, recomendaciones acerca de los motivos más eficaces para la promoción de los fines señalados por la teoría<sup>55</sup>. Por ello el propio Sidgwick advertía contra el error de interpretar que «la doctrina según la cual la Felicidad Universal es el *criterio* último implica que la Benevolencia Universal es siempre el mejor, o el único correcto, de los *motivos* de actuación». No sólo no lo implica, sino que incluso podría implicar lo contrario, pues «si la experiencia muestra que la felicidad universal se logrará de forma más satisfactoria si los hombres obran con frecuencia por motivos diferentes de la pura filantropía universal, es obvio que esos otros motivos deben preferirse razonablemente con arreglo a principios utilitaristas»<sup>56</sup>.

Este resultado resulta paradójico sólo en apariencia. Nada impide en principio a una teoría coherente con el principio consecuencialista que atribuye valor intrínseco sólo a los estados terminales de cosas desaconsejar, en nombre de ese mismo principio, su propia adopción como moralidad pública. Sidgwick reconoce que, con criterios utilitaristas, podría ser moralmente correcto hacer y recomendar en privado lo que no sería correcto defender en público. Como la conciencia moral del hombre ordinario repudia la noción de una moral esotérica y existen razones utilitarias para mantener esta opinión común con carácter general, «la conclusión utilitarista, cuidadosamente formulada, podría ser la siguiente: que la opinión que afirma que el secreto podría hacer correcta una acción que en caso contrario no lo sería debería ser mantenida comparativamente en secreto; y, de igual modo, la doctrina que defiende la conveniencia de la moralidad esotérica debería seguir siendo esotérica. O, si es difícil mantenerla oculta, podría ser deseable que el sentido común repudie las doctrinas que es conveniente reservar para una minoría ilustrada»<sup>57</sup>.

La mayor parte de las críticas que se han dirigido a este tipo de conclusiones del principio consecuencialista han interpretado su carácter paradójico de manera sustantiva, resaltando su carácter contraintuitivo o la repugnancia que producen a nuestras convicciones morales más arraigadas o a nuestro «sentido común moral»<sup>58</sup>. Así, por ejemplo, se ha hablado de

---

55. R. E. BALES: *Act-utilitarianism: account of right-making characteristics or decision-making procedure?* «American Philosophical Quarterly» 8 (1971) 257-258.

56. H. SIDGWICK: *o.c.*, p. 413; subrayados suyos.

57. *Id.*, p. 489-490.

58. D. PARFIT: *o.c.*, p. 388.

una especie de Utilitarismo colonial, en cuyo nombre una minoría utilitarista ilustrada controla una sociedad que mayoritariamente no lo es<sup>59</sup>, incumpliendo así una de las condiciones asociadas a la noción misma de regla moral, esto es, que sea públicamente conocida<sup>60</sup>.

Se han señalado asimismo el escaso peso que el Utilitarismo concede a las consideraciones de justicia y equidad en la suma total de utilidades<sup>61</sup> y las restricciones que impone a la información que acepta de los individuos para la valoración de los estados sociales<sup>62</sup>; su ya mencionada incapacidad para tener en cuenta el carácter único y separado de las personas, debida a sus mecanismos de agregación de las utilidades individuales; el atentado a la integridad personal que supone la obligación de producir las mejores consecuencias desde un punto de vista impersonal e imparcial<sup>63</sup>; y, por no alargar la lista, los efectos negativos sobre la fiabilidad de las personas o las instituciones legales que se seguirán del conocimiento general de que todos los miembros de una sociedad son utilitaristas<sup>64</sup>.

## 7. El Consecuencialismo y la «moral del sentido común»

En términos generales, todas estas objeciones coinciden en argumentar que nuestras convicciones morales encuentran inaceptables muchas de las conclusiones lógicas de los principios consecuencialistas. Y en esa misma medida no escapan a la sospecha de dar por supuesto lo mismo que tratan de demostrar. En efecto, de forma implícita o explícita la tradición mayoritaria en la historia de la filosofía moral comparte la convicción de que la tarea propia de la teoría ética —o al menos una parte muy importante de ella— es la de fundamentar sistemáticamente los datos de la conciencia moral común<sup>65</sup>. Los propios consecuencialistas suelen argu-

59. A. SEN y B. WILLIAMS, Introducción a su *o.c.*, p. 16; B. WILLIAMS: *art. cit.*, p. 138.

60. Cfr. R. HARROD: *Utilitarianism revised* «Mind» 45 (1936) 152-154; J. RAWLS: *o.c.*, p. 55, 133.

61. J. A. MIRRLEES: *The economic uses of utilitarianism*. En SEN y WILLIAMS: *o.c.*, p. 74-78; J. RAWLS: *o.c.*, p. 26; B. WILLIAMS: *art. cit.*, p. 143; G. DWORKIN: *Taking rights seriously* (Londres. Duckworth. 1977) p. 232s; traducción española: *Los derechos en serio* (Barcelona. Ariel. 1984).

62. P. J. HAMMOND: *Utilitarianism, uncertainty and information*. En SEN y WILLIAMS: *o.c.*, p. 98-102; P. DASGUPTA: *Utilitarianism, information and rights*, *ib.* p. 199-200.

63. Véase la nota 32.

64. D. H. HODGSON: *Consequences of utilitarianism* (Oxford. Clarendon Press. 1967) p. 84s; y la réplica de D. REGAN en *o.c.* en nota 27, p. 67s.

65. Así, por ejemplo, para Aristóteles, las *éndoxa* u opiniones generalmente admitidas son los *phainómena* de los que parte la indagación ética (en 1145b2-7); se parte del *hóti* (que) y si está suficientemente claro, no habrá necesidad del *dióti* (por qué) (EN 1095b6-7); «le fait... c'est, par exemple, que dans les circonstances concrètes où je me

mentar que «a pesar de las apariencias en contra, es posible reconciliar de una u otra forma nuestros valores morales de sentido común con los principios consecuencialistas, al menos hasta cierto punto». Pero su teoría posee además una dimensión constructiva y revisionista que se manifiesta en su tratamiento de los casos en los que la reconciliación es imposible: la conclusión apropiada es que «la moralidad del sentido común es irracional y hay que revisarla»<sup>66</sup>.

La presunción en favor de las obligaciones que nos impone la moral común se justifica asimismo por razones consecuencialistas, pero ello no las convierte en obligaciones en sentido estricto, sino todo lo más en obligaciones *prima facie* o condicionales<sup>67</sup>. Al basarse en la propiedad que posee un acto en virtud de su pertenencia a un determinado género, estas úl-

---

trouve actuellement, il est beau pour moi de tenir ferme à mon poste de combat; percevoir ce *fait*, c'est le propre du vertueux...; le moraliste recueille ce *fait*, c'est-à-dire ces jugements de valeur singuliers, et en dégage par induction la notion générale de la vertu; il n'a pas pour ce faire nul besoin *pourquoi*... il était alors beau pour moi de tenir bon»: R. A. GAUTHIER: *L'éthique à Nicomaque*. T.II: Commentaire (Louvain. Publications Universitaires. 1959) p. 19-20. También Hume argumenta a favor de la obediencia civil apelando al «consentimiento universal de la humanidad» y opone los sentimientos del vulgo a cualquier razonamiento filosófico pues, en este caso, las opiniones de los hombres «ostentan una peculiar autoridad y son en gran medida infalibles»: *A treatise of human nature*. Ed. L. A. SELBY-Bigge (Oxford. Clarendon Press. 1978) p. 546. Aunque para Kant la tarea de la filosofía moral es establecer las condiciones trascendentales de posibilidad de los hechos que se dan en la conciencia moral y no la de inducir por generalización la noción de bien moral, afirma no obstante que la razón vulgar, sin ser capaz de formular abstracta y universalmente el principio de la moralidad, se sirve de él constantemente como de un compás para distinguir el bien del mal, pues no hace falta ninguna filosofía para saber qué hacer; de todo lo cual se deduce la superioridad del juicio práctico sobre el teórico en el ámbito del sentido común: *Grundlegung zur Metaphysik der Sitten* (Ak 403-4). De manera más explícita W. D. ROSS señala la diferencia de sentido común en el razonamiento científico y en el juicio moral: en estos últimos no hay ningún acceso más directo a los hechos morales que el pensar sobre ellos, de forma que «las convicciones morales de las personas reflexivas y bien educadas son los datos de la ética como las percepciones sensoriales lo son de la ciencia natural; (...) el cuerpo de convicciones morales de los mejores es el producto acumulado de la reflexión moral de muchas generaciones, (... algo que) el teórico no puede sino tratar con el máximo respeto (...pues son) el cimiento sobre el que hemos de construir: *The right and the good* (Oxford. Clarendon Press. 1930) p. 39-41. Por último, para J. RAWLS existe una clase definida y limitada de hechos con los que cabe contrastar los principios de la justicia, y estos son nuestros juicios ponderados en equilibrio reflexivo; concibe así la teoría moral como «un intento de describir nuestra capacidad moral» para formular tales juicios: o.c. p. 46-51. Sobre este particular cfr. G. GUTIÉRREZ: *Sobre el sentido y el sentimiento morales* «Agora» 2 (1982) y *El agente fantasma* «Agora» 5 (1985).

66. S. SCHEFFLER. Introducción a *Consequentialism...* citado en nota 15, p. 2.

67. W. D. ROSS: o.c., p. 19-20.

timas se subordinan lógicamente a aquellas otras que se fundan en las propiedades intrínsecas del acto mismo. La cláusula «prima facie» expresa el carácter revisionista del Consecuencialismo respecto de la llamada «moralidad del sentido común» y que consiste en su propuesta de, o bien reformar o abandonar los principios de ésta incompatibles con el Consecuencialismo, o bien, y por razones consecuencialistas, convertir la doctrina consecuencialista en patrimonio esotérico de una élite paternal<sup>68</sup>. Todo lo cual demuestra una vez más que la contraposición entre el Consecuencialismo y sus adversarios no es la que existe entre dos teorías morales sino que expresa una concepción alternativa de lo que es —o, más propiamente, lo que debe ser— la moralidad misma.

Por esa razón ninguna de las anteriores objeciones establece el carácter paradójico, en términos técnicos, de la teoría consecuencialista. Sobre todo en un sentido análogo al que ponen de manifiesto las llamadas paradojas pragmáticas —las que sobreviven «cuando las condiciones de las acciones expresadas en el enunciado se definen de tal modo que su anuncio implica que las propias acciones no podrán ejecutarse»<sup>69</sup>. Pero lo verdaderamente significativo de esta conclusión de las premisas consecuencialistas es que suscitan sospechas sobre el carácter pragmáticamente paradójico de una teoría cuya puesta en práctica implicase su *autoevanescencia* (*self-effacingness*) o incluso su *autoderogación* (*self-defeatingness*)<sup>70</sup>.

## 8. Actos y reglas

Al señalar como objetivo último común a todo agente el producir los mejores resultados posibles, el principio consecuencialista tomado estrictamente impone a cada agente la ineludible obligación de ejecutar *el acto* objetivamente correcto, es decir, el que de hecho conducirá a lograr el objetivo óptimo. Quien hace lo que sabe o cree que producirá un resultado peor, actúa incorrectamente, moralmente mal. Pero aun dejando aparte los efectos paradójicos o contraintuitivos de algunas de sus aplicaciones, es un hecho de experiencia que, incluso en términos consecuencialistas, fiar al exclusivo juicio del agente individual la ponderación del conjunto de factores que en cada caso concreto aseguran el resultado óptimo puede acarrear peores consecuencias para los estados sociales.

Para evitar los efectos no deseados de la aplicación de un principio exclusivamente orientado a la realización de *actos* se hizo necesario prestar una atención mayor a lo que implica el hecho mismo de aplicar un *princi-*

68. D. PARFIT: *o.c.*, p. 100-110; cfr. más arriba la nota 57.

69. G. PARETI: *I paradossi pragmatici*. «Rivista di Filosofia» 69 (1978) p. 27. Véase más arriba la nota 52.

70. D. PARFIT: *o.c.*, p. 24-28; 40-43.

pio o una regla. O, lo que es lo mismo, a la dimensión estructural, coordinativa o cooperativa de las acciones propias de los contextos estratégicos de la acción<sup>71</sup>. De aquí nace la distinción entre dos tipos básicos de Utilitarismo, habitualmente conocidos como *Utilitarismo del Acto*, extremo o directo, y *Utilitarismo de la Regla*, restringido o indirecto<sup>72</sup>.

Hume ya había aducido como fundamento de su distinción entre las virtudes socialmente útiles de la benevolencia y la justicia que la utilidad de aquella se produce de forma directa, sin la mediación de «ningún esquema o sistema» y sin que las consecuencias de su observancia resulten «de la concurrencia, imitación o ejemplo de los demás». Mientras que los beneficios de la segunda, en cambio, «no son consecuencia de cada acto individual, sino que surgen del esquema o sistema conjunto en el que concurre la totalidad o la mayor parte de la sociedad»; aun a pesar de que «el resultado de los actos individuales sea, en muchos casos, directamente contrario al del esquema conjunto de acciones». El beneficio de los actos propios de la justicia «procede únicamente de la observancia de la regla general». Ambas especies de virtud contribuyen a la «felicidad y prosperidad de la humanidad», pero la benevolencia se asemeja «a un muro construido por muchas manos, que crece con cada piedra que se le amontona y aumenta en proporción a la diligencia y cuidado de cada obrero»; la justicia, a su vez, puede compararse «a la construcción de una bóveda,

---

71. Utilizando la terminología de J. ELSTER en *Ulysses...*, p. 18, 117; en *Sour grapes*, p. 12-13; o en su introducción a *Rational choice* (Oxford. Blackwell. 1986), p. 7, podemos caracterizar el entorno *paramétrico* de las decisiones como aquel en el cual el agente considera su propia decisión como la única variable y los restantes factores —sean estos agentes propiamente naturales o bien agentes personales tenidos por invariantes— como constantes o parámetros de su problema de decisión. El contraste es con el entorno *estratégico*, en el cual el agente tiene en cuenta que existen otros agentes con sus propias funciones de utilidad que forman parte de su entorno y que saben que él forma parte del suyo. Todos los cuales saben a su vez que el efecto de sus decisiones individuales es función de la agregación de estas. O sea, que es por definición conjunto o colectivo. Cuando los fines alternativos de las acciones son los estados sociales se produce una transformación estructural del concepto de racionalidad al trasladarlo de la escala individual a la colectiva. Este es el caso de las elecciones y actuaciones que determinan los estados sociales, las cuales, por afectar a la producción de bienes públicos suscitan cuestiones fundamentales de índole ética y política. Sobre el concepto de bienes públicos, véase M. OLSON: *The logic of collective action* (Cambridge MA. Harvard University Press. 1965) p. 14; J. RAWLS: *o.c.*, p. 266; sobre las complicaciones derivadas de la interacción de expectativas en los entornos estratégicos, véase el c.7 de M. HOLLIS: *o.c.* en la nota 23, el conjunto de D. REGAN, *o.c.* en nota 31 y, para los contextos económicos, S. M. SHEFFRIN: *Expectativas racionales* (Madrid. Alianza. 1985), c.1.

72. La denominación parece haberla introducido en 1959 R. Brandt *Teoría ética* (Madrid. Alianza. 1982) p. 439, 456; también en su *Toward a credible form of utilitarianism*, en: H. N. CASTAÑEDA (ed): *Morality and the language of conduct*. (Detroit MI. Wayne State University Press. 1963) p. 109.

cada una de cuyas piedras individuales caería de suyo a tierra» y sólo se mantiene erecta «gracias a la asistencia mutua y a la combinación de sus correspondientes partes»<sup>73</sup>.

La concurrencia de los actos individuales en la producción del beneficio común es tarea relativamente sencilla cuando el número de los agentes es pequeño, pero el problema se plantea cuando el número de agentes cuya concurrencia total o mayoritaria se requiere para producirlo es tal que se hace imposible o harto difícil detectar las consecuencias de los actos o las omisiones individuales. Dos vecinos pueden ponerse de acuerdo para desecar una pradera que poseen en común «porque es fácil para cada uno saber lo que piensa el otro, y cada uno puede percatarse de que la consecuencia inmediata de su incumplimiento del acuerdo es el abandono completo del proyecto; pero es muy difícil, de hecho imposible, que mil individuos se pongan de acuerdo para una actuación de este tipo»<sup>74</sup>.

En términos consecuencialistas la obligación que tiene el agente en el primer caso afecta directa e inmediatamente al acto particular. Para prevenir los del segundo caso es necesario recoger en un principio, que por su propia naturaleza universal no menciona actos particulares, la referencia a aquel *tipo* de actos que, *si* son ejecutados por todos los agentes (o por un número suficiente de ellos) que se encuentren en circunstancias relevantemente similares, etc. producirán las mejores consecuencias. En estos casos la obligación afecta primariamente al principio general y derivadamente al acto particular. Lo que es preciso entonces justificar con criterios consecuencialistas son las instituciones —en el sentido de principios y reglas que tipifican formas de actuación— y no las acciones individuales que las ponen en práctica y, de paso, la convicción, sostenida por la conciencia moral común, de que ciertos actos son obligatorios en virtud de ser de un determinado tipo. Por ejemplo, por ser un caso de cumplimiento de promesa, de veracidad, de cooperación, etc.

En la tradición utilitaria pronto se cobró conciencia de la necesidad de complementar el cálculo utilitarista centrado en los efectos de los actos particulares mediante alguna versión del principio kantiano de universalizabilidad de la máxima. En la interpretación de Sidgwick este principio, «al menos en tanto que evidente por sí mismo, significa tan sólo que un acto, si es correcto para un individuo cualquiera, debe ser correcto por razones generales y, consecuentemente, para alguna *clase* de personas». Cree no obstante que debe admitirse una cualificación esencial «que modifica de manera importante su fuerza práctica»: su aplicación al caso en el que «la creencia en que el acto en cuestión no será ampliamente imitado». Pero esta redefinición le sirve precisamente para justificar el abandono del principio general en aquellos casos en los que, a la superior utili-

73. D. HUME: *An enquiry concerning the principles of morals*, ed. cit., p. 303-305.

74. D. HUME: *A treatise...*, ed. cit., p. 538.

dad del acto particular que lo infringe, se une la posibilidad de «definir esta clase mediante la característica de la creencia en que el acto seguirá siendo excepcional»<sup>75</sup> y, en definitiva, para hacer verosímil la dimensión esotérica y elitista de la moralidad utilitarista ya mencionada.

### 9. La producción cooperativa de bienes

Una forma indirecta de averiguar qué tipo de actos permitirían que su máxima fuese universalizada y convertida en un principio de obligación incondicionada la proporciona la pregunta acerca de lo que pasaría *si todo el mundo hiciese lo mismo*. Al analizar «la función de las hipótesis falsas en la ética» Broad llama a este argumento «el principio de falsa universalización», usado negativamente para condenar una determinada acción. En su opinión, «algo muy parecido a él fue considerado por Kant como *la ley fundamental de la ética*»<sup>76</sup>, a pesar de la paradoja que resulta de pedirnos que creamos «que la corrección o incorrección de muchas de nuestras acciones depende de las consecuencias probables, no de lo que juzgamos ser verdadero, sino de lo que sabemos que es falso», esto es, que no todo el mundo hará lo mismo. Ciertamente, como reconoce Broad, la paradoja sólo se suscita cuando se considera que la bondad o maldad moral de las acciones depende de sus consecuencias efectivas o probables, y desaparece cuando se la cree totalmente dependiente de sus cualidades intrínsecas<sup>77</sup>.

Una de las funciones más paradójicas que desempeñan las hipótesis falsas o contrafácticas es la de demostrar a los agentes que deben cooperar en la producción de ciertos bienes que sólo cooperativamente pueden obtenerse; pero por regla general este argumento sólo se emplea «cuando la parte con la que cada miembro del grupo contribuye al bien conjunto es muy pequeña»<sup>78</sup>. Broad considera el caso de un individuo invitado a unirse a un grupo que coopera en la producción de un bien. Es razonable suponer que, a medida que aumenta el número de miembros del grupo, disminuyen, aunque en proporción decreciente, los sacrificios que se impone cada individuo, y que la cantidad de bien producido por el grupo asimismo aumenta, aunque en proporción menguante. En estas condiciones puede alcanzarse un punto en el que el individuo encuentre que el excedente del bien total producido por el grupo sobre el total de sacrificios sufridos por el grupo incluyéndole a él es menor que ese mismo excedente excluyéndole a él<sup>79</sup>. Es decir, que los resultados globales son mejores sin él

75. H. SIDGWICK: *o.c.*, p. 485-486; subrayado suyo.

76. C. D. BROAD: *art. cit.*, p. 43; subrayado suyo.

77. *Id.*, p. 44.

78. *Id.*, p. 50.

79. *Id.*, p. 52.

que con él, por lo que sería *obligación* del individuo —de todos los individuos— rehusar unirse al grupo, con la consecuencia de que, si todos hiciesen lo mismo, dejaría de producirse el bien. Pero si al individuo le consta que la hipótesis es falsa, esto es, que no todos harán lo mismo, lo cierto es que el bien se producirá, con su cooperación o sin ella. Y mejor aún sin ella, tanto para él *como para el grupo*, pues la suma de beneficios es superior si él no coopera «aunque» —la adversativa es meramente retórica— se beneficie.

Broad, que ciertamente no es utilitarista, aunque sí consecuencialista, señala que la cuestión de fondo es la equidad en la distribución de beneficios y sacrificios. Aunque «la existencia de ciertos estados mentales es intrínsecamente buena y es el deber de cada uno tender al que cree que será el mejor estado de cosas en conjunto», un conjunto de bienes distribuidos de cierta forma entre un conjunto de personas «puede constituir un estado de cosas intrínsecamente mejor que el mismo conjunto distribuido de manera diferente», y la apelación a la equidad parece basarse en el principio de que «el estado de cosas mejor posible se alcanza cuando el grupo de productores y el de quienes disfrutan es todo lo idéntico que se pueda»<sup>80</sup>.

Este es en esencia el problema que plantea al Utilitarismo y al Consecuencialismo en general del fenómeno del *free rider*, que se suscita fundamentalmente a propósito de la producción de bienes públicos<sup>81</sup>. Es decir, aquellos que sólo se pueden obtener mediante la cooperación de un cierto número de personas, la cual implica ciertos sacrificios para todas ellas, pero cuya característica esencial consiste en ser bienes «que por su propia naturaleza tienen que ser disfrutados por todos los miembros de una cierta clase, se identifique o no esta clase con el grupo de quienes lo producen»<sup>82</sup>. Parfit los considera variantes de un idéntico esquema del dilema del prisionero multipersonal en su versión «del contribuyente»: para cada uno «puede ser cierto que, si contribuye, incrementará la suma de beneficios. Pero su parte en los beneficios del incremento podría ser muy pequeña. Su contribución podría no compensarle. Tal vez fuera mejor para cada uno no colaborar. Y esto independientemente de lo que hagan los demás.

80. *Id.*, p. 53-54.

81. Para D. GAUTHIER el *free rider* es aquel que «obtiene un beneficio sin pagar la totalidad o parte de su coste», mientras que el *parásito* es aquél que, «al obtener un beneficio, desplaza la totalidad o parte de su coste a otra persona»: *o.c.*, p. 96. Sobre el fenómeno del *free rider* en relación con los bienes públicos existe abundante bibliografía; además de la citada en la nota 67, véase G. GUTIERREZ: *La congruencia entre lo bueno y lo justo* «Revista de Filosofía» 2.<sup>a</sup> Serie, 2 (1979) 51-4; D. PARFIT: *Prudence, morality and the Prisoner's Dilemma* «Proceedings of the British Academy» 65 (1979), que incluye numerosas referencias bibliográficas; está reproducido en S. SCHEFFLER: *o.c.* y traducido, sin las referencias bibliográficas, como *Prudencia, moralidad y el dilema del prisionero* en «Diálogo Filosófico» 13 (1989) 4-30.

82. C. D. BROAD: *art. cit.*, p. 53.

Pero sería peor para cada uno que fuesen menos los que contribuyen. Y si ninguno contribuyese, eso sería peor para cada uno que si todos lo hicieran»<sup>83</sup>. El que puedan ser lógicamente generados por una teoría que ofrece razones para actuar parece implicar que esa teoría se deroga a sí misma —es *self-defeating*<sup>84</sup>.

Ya en 1936, para evitar que la aplicación de la «cruda máxima utilitarista» empleada en la casuística utilitaria «tendiese a desarrollar un código contradictorio», Roy Harrod intentó incorporar a la filosofía utilitarista el que llamó «principio kantiano»<sup>85</sup>. Por tal entiende «la proposición según la cual la moralidad requiere que se actúe conforme a una máxima que pueda convertirse en una ley general»<sup>86</sup>. Pero la máxima utilitaria que manda «elegir siempre la acción que contribuirá a la mayor felicidad» ¿no es una ley general?. En el caso de la mentira, por ejemplo, parecería aconsejar que se mintiese siempre que las pérdidas derivadas del acto particular de mentir —la pérdida de crédito del mentiroso, la disminución de la confianza en la función comunicativa del lenguaje, entre otras— fuesen compensadas por la mayor felicidad producida al hacerlo. En efecto, «si todo el mundo mintiese en esas circunstancias y sólo en ellas, aparentemente todo iría bien». Si así fuera, «las consecuencias indicadas por el crudo principio utilitarista serían siempre idénticas a las que se deducirían de la aplicación del principio kantiano: las consecuencias del acto, considerado en sí y por sí no diferirían de las consecuencias de tal acto siempre que fuese realizado en circunstancias relevantes precisamente semejantes». Pero no ocurre así, pues «hay actos que, cuando se realizan en *n* ocasiones similares tienen consecuencias que son más de *n* veces mayores que las que resultan de su ejecución»<sup>87</sup>; son estos casos los que dan pie a la creencia —correcta, en opinión de Harrod— de la conciencia moral común de que ciertos tipos de actos son obligatorios<sup>88</sup>.

Para Harrod, un Utilitarismo refinado preferirá el principio kantiano al utilitarista crudo, debido a la pérdida mayor de beneficios cuando la mentira se generaliza. Y juzgará moralmente incorrecto aquel acto que, siendo conveniente en una circunstancia particular, se volvería inconveniente cuando fuera hecho por todos en las circunstancias relevantes precisamente semejantes. El principio kantiano no proporciona el fundamento de toda moralidad —aunque el propio Kant erróneamente así lo creyera— pero sí el de aquellos actos morales tenidos por obligatorios —por ejemplo, la veracidad, a diferencia de la prestación de ayuda al próxi-

83. D. PARFIT: *Prudence...*, p. 541.

84. *Id.*, p. 539.

85. R. HARROD: *art. cit.*, p. 148-149.

86. *Id.*, p. 147.

87. *Id.*, p. 148.

88. *Id.*, p. 147.

mo<sup>89</sup>. La máxima utilitaria vulgar peca de excesiva generalidad y no presta suficiente atención a las complejidades de la condición humana, entre ellas el hecho histórico de que «en el proceso de evolución de la sociedad estable han sobrevivido aquellos sistemas que establecieron prácticas e instituciones reconocidas que ponían en práctica el principio kantiano»<sup>90</sup>.

Esto permite explicar tanto el rigor con el que consideramos las obligaciones como la profunda aversión que suscita su incumplimiento e incluso «la opinión *prima facie* de que hay algo en la naturaleza reconocida de una obligación que entra en conflicto con cualquier filosofía de fines». Aunque el conflicto es aparente, los sentimientos de aversión tienen su razón de ser precisamente como mecanismo de resistencia a la tentación a la que puede sucumbir el individuo que no entiende el principio kantiano. Una vez más es preciso recurrir a un argumento esotérico que evoca el de Sidgwick: «son la sutileza y la dificultad del principio, que no puede ser explicado al hombre medio, las que han convertido el elemento arbitrario y autoritario en la esfera moral en algo necesario para la evolución de la sociedad estable». Como esta era ilustrada tiene sus peligros, tal vez los filósofos «de la obligación indefinible» —los deontologistas, intuicionistas, etc.— aun hayan de seguir representando su papel, y «podría ser inconveniente que los partidarios de la conveniencia los expusieran a la vergüenza pública»<sup>91</sup>.

Harrod discute expresamente las dificultades que plantea, incluso a su versión refinada de Utilitarismo, el Principio de Publicidad<sup>92</sup>. En el caso de las mentiras útiles que pueden mantenerse en secreto, si se sabe que los individuos virtuosos profesan el crudo principio utilitario cuando es posible guardar el secreto, se perderá la confianza. Lo deseable será entonces «que todos abjuren totalmente del principio utilitarista y, al mismo tiempo, obren de acuerdo con él cuando se pueda mantener en secreto». Pero ¿qué doctrina habría que predicar? Tanto el crudo principio utilitario, pues se desea que todos actúen conforme a él, como algún principio anti-utilitario, porque es deseable que todos crean que nadie actúa conforme a él. Este sería un sistema contradictorio, pero nada impide que sirviera mejor al interés común. Sin embargo, no es el tipo de sistemas al que suele llamarse «moralidad». Y no se trata únicamente de una cuestión semántica, pues el término «obligación moral» designa convenientemente «un sistema de conducta comúnmente reconocido por sus participantes» y que, para ser efectivo, ha de estar «estrechamente conectado con la emoción y la expresión de aprobación y desaprobación»<sup>93</sup> —condiciones que

---

89. *Id.*, p. 149-50.

90. *Id.*, p. 151.

91. *Id.*, p. 150.

92. *Id.*, p. 152.

93. *Id.*, p. 153.

quedan excluidas por la propia naturaleza de aquel. Harrod concluye que «el utilitarista que desee obtener las ventajas que produce la incorporación del principio kantiano a las prácticas públicamente reconocidas tiene que desear que éstas se observen tanto si es posible mantener en secreto su transgresión como si no lo es»<sup>94</sup>. Pero esta «obligación de desear» que se impone al utilitarista, o bien se justifica en términos «crudamente» consecuencialistas, o bien, como se ha tenido ocasión de observar anteriormente, no escapa a la objeción de auto-evanescencia o de auto-derogación formulada a la teoría consecuencialista.

### 10. La justificación de los principios prácticos

Son numerosas las manifestaciones del inestable dualismo que caracteriza el razonamiento consecuencialista cuando se aplica a la justificación de los principios de la moralidad *práctica* en cuanto tales. Por limitarnos a sólo dos ejemplos puede citarse el argumento de Rawls<sup>95</sup> según el cual la existencia de las instituciones penales como tales se justifica por las consideraciones de utilidad pública que deben inspirar a los legisladores; mientras que la aplicación de la sanción a un delincuente particular por parte de un juez se justifica en términos de coherencia con el sistema legal aun cuando en ese caso concreto no existan razones utilitaristas para hacerlo. Una institución como el «telishment» —un «punishment» o castigo justificado en términos estrictamente teleológicos, que permitiese condenar a un inocente si fuera conveniente para los intereses de la sociedad— sería indefendible en términos utilitaristas. Algo semejante ocurre con la institución de cumplir las promesas. Su existencia se defiende por consideraciones utilitaristas, pero una vez que existe, el individuo particular no puede aplicarlas para eximirse de la obligación de cumplir la que él hizo, porque «el quid de la práctica es abdicar el título propio para actuar de acuerdo con las consideraciones utilitaristas y prudenciales... El prometiende está obligado porque prometió: no depende de él juzgar el caso según lo merezca»<sup>96</sup>.

Otro ejemplo más reciente es la distinción que hace Hare entre dos tipos de principios correspondientes a sendos niveles del razonamiento moral. El primer nivel comprende el entramado de principios generales e intuitivos característicos de la moral del sentido común cuya validez *ut in pluribus* permite aplicarlos al razonamiento moral de la vida cotidiana en circunstancias normales. Se adquieren por medio de la educación, incluso la auto-educación y van acompañados de emociones y sentimientos más o

94. *Id.*, p. 154.

95. J. RAWLS: *Dos conceptos de reglas* (1955), reproducido en P. FOOT (ed): *Teorías sobre la ética*. México. Fondo de Cultura. 1974.

96. *Id.*, p. 227.

menos específicos y profundamente arraigados. Por su propia generalidad dan origen a conflictos en apariencia insolubles, como en aquellos casos en los que dos deberes *prima facie* igualmente estrictos exigen del agente actuaciones incompatibles entre sí. Es el nivel propio del Utilitarismo de la regla y en él los conflictos entre principios han de resolverse de forma intuitiva<sup>97</sup>.

Hare reserva, en cambio, para su segundo nivel la función de dirimir críticamente los conflictos entre los principios generales del nivel intuitivo por medio de un juicio «restringido únicamente por las propiedades lógicas de los conceptos morales y por los hechos no-morales, y por ninguna otra cosa»<sup>98</sup>. Los principios de ambos niveles son igualmente *universales* por «estar gobernados por un cuantificador universal y no contener constantes individuales», pero difieren en la dimensión generalidad-especificidad: los principios propios del nivel intuitivo han de ser relativamente simples y generales, esto es inespecíficos, para poder desempeñar su función, mientras que los del nivel crítico pueden ser tan *específicos* como haga falta<sup>99</sup> para adaptarlos a una situación particular sin que ello obste a su universalidad, pues «en teoría no es necesario mencionar individuos para describir una situación plenamente, o al menos tan plenamente como precisemos» de forma que «pueda aplicarse igualmente a cualquier situación precisamente similar que incluya personas, lugares, etc. precisamente similares»<sup>100</sup>. Ello basta para garantizar la universalizabilidad del principio moral, que exige —¿lógicamente? ¿moralmente?— al agente que esté dispuesto a formular en cualquier situación precisamente similar el mismo juicio moral que formula en una situación particular. Como en cualquiera de ellas él podría ocupar cualquiera de las posiciones que en esta situación particular ocupan otras partes «no le parecerá aceptable ningún juicio que no resulte en lo mejor, a fin de cuentas, para todas las partes», con lo que su juicio vendrá a coincidir con el que emitiría «un cuidadoso utilitarista del acto»<sup>101</sup>.

Como ejemplo hipotético de un ser que en el límite emplease única-

---

97. R. M. HARE: *Ethical theory and utilitarianism* (1976), reproducido en A. SEN y B. WILLIAMS (eds), *o.c.*, p. 23-38 y *Moral thinking. its levels, methods and point* (Oxford. Clarendon Press. 1981). «Intuición» no ha de entenderse aquí en el sentido estrictamente epistemológico del término, sino más bien en aquel con el que Rawls designa la opinión de quienes sostienen que, alcanzado cierto nivel de generalidad, existe «una familia de primeros principios irreductibles que es preciso ponderar entre sí preguntándonos cuál es el balance más justo en nuestro juicio meditado», sin que por definición quepa remitirse a principios de un nivel superior: *o.c.*, p. 34. Véase B. BARRY: *Political argument* (Henley-on-Thames. Routledge. 1965) p. 2-15.

98. R. HARE: *Moral thinking*, p. 40.

99. *Id.*, p. 41.

100. *Id.*, p. 42.

101. *Id.*, p. 42-3.

mente el razonamiento crítico Hare imagina uno dotado de «poderes sobrehumanos de pensamiento, conocimiento sobrehumano y sin ninguna debilidad humana»: el *arcángel*<sup>102</sup>. En el otro extremo, la persona afectada por todo tipo de debilidades y por completo incapaz de pensar críticamente, que adquiriría sus principios mediante la educación o por la imitación de otros: el *prole*<sup>103</sup>. En este caso, por fin, la reminiscencia orwelliana se convierte en referencia explícita.

La última y suprema instancia de la moralidad es, pues, el principio consecuencialista directo, inmediato o «crudo». Por estarle lógicamente y epistemológicamente subordinadas, las normas, reglas y principios de la moralidad común son estrictamente contingentes y provisionales. Los principios *prima facie* del primer nivel son seleccionados en el segundo nivel atendiendo a las consecuencias de inculcarlos en otros o en nosotros mismos. Son consideraciones consecuencialistas —del acto— las que desaconsejan recomendar el Consecuencialismo del acto o la ética de situación como doctrina moral apta para todos los públicos. Por esta razón se ha hecho notar que los Utilitarismos del acto y de la regla son fundamentalmente lo mismo por ser extensionalmente equivalentes. Es decir, que desde un punto de vista utilitario «si consideramos todos los hechos relevantes en un caso determinado no importa que preguntemos '¿qué pasaría si todos hiciesen lo mismo?' en vez de '¿qué pasaría si se realizase este acto?'<sup>104</sup>.

## 11. Apéndice: agentes personales y racionalidad colectiva

Las reflexiones más recientes sobre la teoría consecuencialista se han centrado en las posibilidades que brinda al desarrollo del concepto de racionalidad colectiva o pública, sin necesidad de recurrir al falso supuesto —consustancial a las diversas formas de estatalismo socialista— de que hay o debe haber un centro soberano de decisión que determina lo que es correcto para la sociedad en conjunto. Aunque la noción de racionalidad pública aplicada a una compleja sociedad moderna exige disponer y aplicar reglas generales, sólo teorías muy primitivas de la soberanía suponen la existencia de un centro único de decisión que desempeñe una función análoga a la del yo en el agente<sup>105</sup>. De ahí que el interés se haya centrado

---

102. *Id.*, p. 44.

103. *Id.*, p. 45.

104. D. LYONS: *Forms and limits of utilitarianism* (Oxford. Clarendon Press. 1965) p. 118.

105. A. SEN y B. WILLIAMS: *Introducción a su o.c.*, p. 2-3; cfr. P. DASGUPTA, *art. cit.* en nota 58.

en el análisis de aquellas formas de acción colectiva que hacen posible la cooperación racional espontánea entre los individuos<sup>106</sup>.

1) Donald Regan considera que las teorías consecuencialistas pueden poseer tres propiedades que afectan a su capacidad de producir consecuencias óptimas. La *primera* (UA) es la propiedad fundamental que según el Utilitarismo del acto debe poseer toda teoría consecuencialista. La idea central es que el agente dispone de una lista de actos que puede realizar, fuera de la cual todos los demás factores causalmente relevantes acerca del mundo están dados. A cada acto de la lista corresponde un estado futuro de cosas con un valor especificado que se realizará si se ejecuta el acto. Lo que la teoría requiere del agente es que ejecute el acto que produzca un estado de cosas cuyo valor sea al menos tan grande como el de cualquier otro estado del mundo correspondiente a otro acto de la lista del agente. Poseerá esta propiedad aquella teoría tal que *cualquier agente* que la satisfaga en cualquier situación producirá las mejores consecuencias que le es posible producir en esa situación. Incluyendo en la situación todas las características causales relevantes del resto del mundo: tanto la conducta de los demás agentes en la que no puede influir como las disposiciones para la conducta de aquellos otros que pueden ser influidas por los actos del agente<sup>107</sup>.

La *segunda* (OC) es la propiedad de «optimización coordinada» que poseerá aquella teoría tal que, si *todos los agentes* la satisfacen en todas las situaciones, entonces *la clase de todos los agentes* produce por sus actos tomados *en conjunto* las mejores consecuencias que les es posible producir por cualquier pauta de conducta.

La *tercera* (AD) es la propiedad de «adaptabilidad», que poseerán sólo aquellas teorías en las cuales, en cualquier situación que implique elecciones por cualquier número de agentes, los agentes que *satisfacen* la teoría en esa situación producen por medio de sus actos tomados en conjunto las mejores consecuencias posibles para cualquier pauta de conducta *dada la conducta de los agentes que no la satisfacen*.

El supuesto consecuencialista básico según el cual los agentes morales deben maximizar las buenas consecuencias permite dos interpretaciones distintas y enfrentadas. Según una de ellas —el Utilitarismo del acto— es el *individuo* quien debe maximizar las buenas consecuencias de sus ac-

---

106. Sobre la posibilidad de la cooperación entre agentes racionales puede consultarse, además de las obras ya citadas de D. GAUTHIER, M. OLSON, D. PARFIT y D. REGAN, las de R. HARDIN: *Collective action* (Baltimore MD. Johns Hopkins U.P. 1982); R. AXELROD: *La evolución de la cooperación* (Madrid. Alianza. 1986); H. MARGOLIS: *Selfishness, altruism and rationality* (Cambridge. University Press. 1982); M. TAYLOR: *Anarchy and liberty* (Cambridge. University Press. 1982) y *The possibility of cooperation* (Cambridge. University Press. 1987).

107. D. REGAN: *o.c.* p. 3-4.

ciones como tal individuo. Como el Utilitarismo del acto prescribe para cada individuo lo que sería mejor que él hiciese, no lo que sería mejor que hiciese algún grupo, la satisfacción universal de esta teoría es condición necesaria pero no suficiente para la producción de las mejores consecuencias totales, pues nada dice acerca de aquellos bienes que sólo pueden producirse de forma colectiva. La segunda interpretación busca subsanar esta deficiencia encomendando al *grupo* de agentes que actúe de tal manera que las consecuencias de su conducta *colectiva* sean las mejores en las circunstancias en las que se encuentra su grupo como tal grupo. Comparten esta interpretación tanto el Utilitarismo de la regla —según el cual cada agente debe seguir el conjunto de reglas que sería mejor que *todos* siguiesen— como el principio de generalización utilitaria —que manda hacer a cada agente lo que sería mejor que hiciesen él y todos aquellos que se encuentran en una situación similar<sup>108</sup>.

El Utilitarismo del acto no posee la segunda ni menos aún la tercera de las propiedades ante numeradas: un universo de agentes todos los cuales satisfacen la primera de ellas no necesariamente producen las mejores consecuencias posibles en cuanto grupo. Los dos tipos de Utilitarismos colectivos tampoco poseen la primera ni la tercera porque, entre otras complejas razones, carecen de una teoría eficaz para la producción de las mejores consecuencias a pesar de la existencia en el grupo de agentes que no siguen las reglas. La adaptabilidad es una propiedad más fuerte que la mera conjunción de las otras dos. Esquemáticamente:

SI POSEE	ENTONCES	PERO NADA SE SIGUE
UA	todo agente individual que la satisface produce las mejores consecuencias posibles en sus circunstancias	de las colecciones de agentes
OC	si todos la satisfacen, se producen las mejores consecuencias posibles	para un individuo o grupo si uno o más agentes no satisfacen la teoría
AD	con independencia de cuantos la satisfagan los que la satisfacen producen colectivamente las mejores consecuencias posibles	

108. *Ib.*, p. 21-3.

Una teoría adaptable establecerá para cada agente el deber de cooperar con cualquier agente que a su vez coopere en la producción de las mejores consecuencias posibles dada la conducta de los no cooperadores. La idea básica del «Utilitarismo cooperativo» propuesto por Regan es indicar a todo agente que proceda en dos pasos. Primero ha de identificar otros agentes deseosos y capaces de cooperar en la producción de mejores consecuencias, y a continuación debe contribuir con su parte en el mejor plan de conducta para el grupo que consta de él mismo y de los demás cooperadores identificados, teniendo en cuenta la conducta de los no-miembros de ese grupo<sup>109</sup>. Entre las teorías consecuencialistas «sólo el Utilitarismo cooperativo es una teoría para la comunidad de agentes morales»<sup>110</sup>.

2) La obra de Derek Parfit es suficientemente compleja como para que un resumen esquemático de ella resulte inteligible en el espacio aquí disponible. Es un fino análisis de las condiciones que permiten producir mejores resultados mediante la coordinación de las acciones de los agentes, incluyendo los efectos paradójicos que nacen de ignorar las diferencias de escala entre los pequeños grupos y las grandes colectividades y de creer erróneamente que, porque los efectos de los actos individuales son a menudo imperceptibles y triviales, carecen de consecuencias dañosas cuando lo cierto es que la concurrencia de muchos agentes hace que *nos* causemos —a nosotros mismos y a otros— grandes daños. Es preciso avanzar hacia un grado mayor de altruísmo racional, comenzando por erradicar las falsas creencias acerca de los efectos de nuestros propios actos. El movimiento hacia el altruísmo racional lo es al mismo tiempo hacia el grado mayor de impersonalidad, incluyendo la revisión de muchas de las creencias que respalda la moralidad del sentido común. En gran parte ello implica descartar la teoría «egoísta» de la moralidad, que se caracteriza precisamente por conceder la máxima importancia al carácter distinto y separado de las personas —«la teoría que *me* dice a mí que haga lo mejor para *mí*»<sup>111</sup>— y, en consecuencia, revisar el concepto mismo de identidad personal. No por ello es necesario aceptar la imparcialidad extrema del «punto de vista del universo» sidgwickiano, sino una teoría según la cual «la unidad fundamental no es el agente a lo largo de toda su vida, sino el agente en el momento de actuar»<sup>112</sup>. Una deseable Teoría Unificada deberá rechazar tanto el egoísmo cuanto el Consecuencialismo y la moralidad del sentido común pues todas ellas conducen a uno u otro tipo de auto-derogación<sup>113</sup>. Aunque mucha gente cree que sólo hay dos

---

109. *Ib.*, p. xi.

110. *Ib.*, p. 211.

111. D. PARFIT: *o.c.*, p. 444.

112. *Ib.*, p. 445.

113. *Ib.*, p. 449.

formas de hacer Ética, una por el camino llano que apela simplemente a nuestras intuiciones y la otra por avenida de la metaética, lo cierto es que la resolución de cuestiones no sólo metaéticas sino, más aún, metafísicas implica conclusiones sobre la ética misma, cuya historia, así como la de la humanidad tal vez no habría hecho sino comenzar<sup>114</sup>.

---

114. *Ib.*, p. 447.