ACCIÓN ABIERTA, UTILIDAD Y UTILITARISMO*

HÉCTOR-NERI CASTAÑEDA
Indiana University

(i) ... when the question is which among several courses still open to a man he *ought* to choose.
G. E. Moore, *Ethics*, capítulo 5

(ii) ... an obligation must be an obligation, not to *do* something, but to perform an activity of a totally different kind, that of setting or excerting ourselves to do something, i.e., to bring something about.
H. A. Prichard, "Duty and Ignorance of Fact"

Este artículo es una nueva contribución al examen continuado de los fundamentos lógicos del utilitarismo.[1] Aquí discutiremos el utilitarismo de la acción.

El utilitarismo es una familia de doctrinas que tratan de articular *in concreto* el ideal de que siempre se debe hacer lo mejor. Por ejemplo, de acuerdo con el utilitarismo simple de la acción, un agente X debe siempre elegir, hacer una acción A y hacerla, si y sólo si, el que X haga A produce al menos un valor tan grande, de un determinado tipo, como el que causaría cualquier otra acción alternativa abierta a X en las mismas circunstancias. Llamamos a esta postura "simple" porque *ella* convierte la condición de producir el máximo valor (esto es, el causar al menos un valor tan grande) en una condición suficiente y necesaria para la obligatoriedad. Las

* Este estudio fue escrito durante el tiempo en que el autor fue Becario de la Fundación Guggenheim.

[1] La primera parte del argumento aparece en "A problem for Utilitarianism", *Analysis*, Vol. XXVIII (1968), y la segunda en "Ought, Value, and Utilitarianism", *American Philosophical Quarterly*, Vol. VI (1969).

críticas morales a semejante tesis son bien conocidas y aquí no nos conciernen. El caso es que éstas no eliminan el utilitarismo de la acción como una fuente de deberes *prima facie* o como un criterio para la solución de algunos conflictos de deberes. Lo que nos interesa aquí es la fundamentación lógica, la inteligibilidad misma, del utilitarismo de la acción.

Resulta obvio que un utilitarismo, y en verdad cualquier teoría de la obligación, es inteligible en la medida en que se posea una idea clara de lo que es en general una acción[2] y, en particular, en la medida en que se posean claros conceptos de *acción alternativa, acción abierta* y *la mejor consecuencia.*[3] Éstos son los tres conceptos que nos interesan aquí y su análisis es crucial para la formulación misma del utilitarismo de la acción. De hecho ya en otro sitio,[4] donde

[2] Últimamente hemos logrado obtener un mejor entendimiento de la naturaleza de la acción. La literatura reciente sobre el tema de la acción es bastante amplia pero entre los importantes estudios que desarrollan tesis en consonancia con la que está detrás de este ensayo debemos mencionar: (*i*) H. A. Prichard, "Duty and Ignorance of Fact", A British Academy Lecture of 1932, reproducido en H. A. Prichard, *Moral Obligation*, Oxford: Clarendon Press (1949); (*ii*) Arthur C. Danto, "What We Can Do", *The Journal of Philosophy,* Vol. LX (1963).

[3] Una excelente discusión, pionera del concepto de *acción alternativa* y *consecuencia,* aparece en *The Alternatives and the Consequences of Actions* de Lars Bergström, Stockholm, Aqvist & Wiksell, 1966. Bergström hace un análisis de alternativas el cual no garantiza que exista un único conjunto de alternativas aunque haya sido escogido un punto de vista. Discute una lista de condiciones meramente necesarias, pero el resultado no es conclusivo. Como una reacción a lo que dice Bergström, Lennart Aqvist ha suministrado un análisis de alternatividad que garantiza la unicidad una vez que se ha escogido ya el punto de vista; esto lo hace en "Improved Formulations of Act-Utilitarianism", *Noûs* Vol. III (1969). Con independencia de ambos trabajos discuto dos conceptos de acción alternativa en "Ought, Value, and Utilitarianism". De manera bastante interesante mi α— alternatividad se relaciona con la de Bergström y mi μ-alternatividad con la de Aqvist. Ninguno de estos cuatro conceptos de alternatividad implican probabilidad. En este respecto, el análisis que se ofrece en la Sección 4 de este estudio es un adelanto. Un análisis de *acción abierta* ha sido previamente ofrecido en "Ought, Value and Utilitarianism".

[4] El artículo de Bergström "Utilitarianism, and Deontic Logic", *Analysis,* Vol. XXIX (1968) es una buena réplica al mío "A Problem for Utilitarianism", Bergström (*i*) concede el punto de que la condición de producir el máximo valor no es una condición necesaria para las obligaciones, y (*ii*) propone una versión recursiva del utilitarismo de la acción. Desafortunadamente, esta ver-

iniciamos el estudio de los fundamentos lógicos del utilitarismo, argumentamos que la condición de producir el máximo valor no puede ser una condición necesaria para la obligatoriedad, esto es, que un utilitarismo de la acción, que sea viable, no puede ser simple.

Sin embargo, nuestro propósito principal en este ensayo no es crítico, excepto en la cuestión fundamental que sigue. El utilitarismo simple presentado anteriormente no puede suministrar una guía o dirección en los casos más típicos de acción y elección, a pesar de que la mayoría de las formas de utilitarismo de la acción son propuestas como teorías realistas de la obligación, éstas no se interesan en lo que uno debería ideal e irrealizablemente hacer, sino en responder a la pregunta de cuál de los cursos de acción *abiertos* a un agente debe éste elegir. El realismo de la elección requiere que frente a esta cuestión el curso de acción elegido no sea meramente una acción que él agente deba realizar, sino una acción que el agente deba, o bien realizar, o simplemente deba *tratar* de realizar. En muchas ocasiones simplemente no está abierto al agente el llevar a cabo una acción que causaría el mayor bien relevante para el caso, sino solamente el *tratar* de realizar tal acción.

El problema aquí consiste en que el criterio para elegir lo que uno va a *tratar* de hacer no puede ser el *causar* el mayor bien. Lo que uno debe tratar de hacer es en buena parte determinado por las probabilidades de los resultados posibles de los cursos de acción abiertos. Sin embargo, como se mostrará más adelante, no es una tarea fácil combinar las probabilidades y los valores.

Nuestro propósito aquí es constructivo. En la Parte I ofrecemos una contribución al análisis de *acción alternativa abierta*; distinguimos entre un conjunto de acciones para realizar y un conjunto de acciones para elegir y estructu-

sión es en el fondo obscura porque se basa en su análisis de alternatividad en su libro, que no incluye una sola condición suficiente para ser conjunto de alternativas relevantes. Véanse los puntos mencionados en la nota 3.

ramos las relaciones entre los deberes de elegir, de hacer y de tratar de hacer. En la Parte II, desde la posición lograda en la Parte I, examinamos algunas maneras de enfrentarse con el serio problema de combinar valor y probabilidad, las cuales han sido ya exploradas en diversas teorías de la utilidad y en la moderna teoría de la decisión bajo riesgo. Dado que dichas maneras no son satisfactorias, concluimos este trabajo con una propuesta sumamente tentativa para combinar probabilidad y valor en una versión del utilitarismo de la acción. Con esta propuesta, montada sobre el análisis previo de la acción alternativa abierta, ofrecemos aquí al menos una formulación coherente del utilitarismo de la acción. Éste, desde luego, no es un utilitarismo simple.

PARTE I. *Acciones Alternativas Abiertas*

1. *Realizaciones, elecciones y medios.* Considérese a un hombre, al que llamaremos *Utilius,* quien se halla encerrado en un cuarto y sabe que se ha puesto en marcha un mecanismo que hará explotar una bomba en el seno de una reunión política. Afortunadamente, se encuentra frente a Utilius el único objeto en el cuarto, a saber, un artefacto electrónico con tres botones:

(*i*) el botón b_1 que si se oprime simplemente detendrá el mecanismo que opera la bomba;

(*ii*) el botón b_2 que si se oprime, *o bien* nada más pondrá en marcha un mecanismo que opera la bomba, *o bien* detendrá la bomba y a la vez hará sonar una alarma contra incendios;

(*iii*) el botón b_3 que produce los mismos efectos que el *botón* b_2.

Así pues, Utilius tiene que hacer una elección entre tres alternativas:

1. D~A detener la bomba sin hacer sonar la alarma contra incendios.

2. $D{\sim}A \vee DA$: o bien detener la bomba sin hacer sonar la alarma contra incendios, o bien detener la bomba y a la vez hacer sonar la alarma;
3. ${\sim}D{\sim}A$ no hacer nada para detener la bomba o para hacer sonar la alarma.

Por supuesto que Utilius también tiene que elegir entre:

1'. oprimir el botón b_1;
2'. oprimir el botón b_2 o b_3;
  *2'*a: oprimir el botón b_2
  *2'*b: oprimir el botón b_3
3'. no oprimir ningún botón.

Sin duda hay otros conjuntos de acciones entre las cuales seguramente Utilius tendrá que hacer futuras elecciones, v.gr., entre oprimir el botón que desea con la mano izquierda o con la mano derecha, entre oprimirlo mientras está parado o mientras está arrodillado. Pero la consideración principal de Utilius es el asunto de la bomba. Con respecto a este asunto las acciones más relevantes son la *1* la *2* y la *3* antes mencionadas. Las acciones *1'*, *2'* y *3'* son relevantes para este asunto sólo como medios para dar lugar a las acciones *1*, *2* y *3*. Todas las demás acciones que puede realizar en ese cuarto son, o bien completamente irrelevantes respecto al asunto que le preocupa, o bien son derivativamente relevantes como medios para los medios *1'* *2'* o *3'*, o como medios para eslabones más remotos en la cadena de medios que pueden dar lugar a *1*, *2* o *3*.

Con respecto al asunto de la bomba, el triple de acciones (1, 2, 3), es decir, ($D{\sim}A$, $D{\sim}A \vee DA$, ${\sim}D{\sim}A$), agota los cursos alternativos de acción abiertos a Utilius en sus circunstancias. Obviamente Utilius no está en la posición de realizar ${\sim}DA$, es decir, hacer sonar la alarma sin detener la bomba, no se diga ya escoger racionalmente aun creyéndose sinceramente capaz de realizar ${\sim}DA$. Más aún, en función de la circunstancia en que se encuentra Utilius los miembros de ($D{\sim}A$, $D{\sim}A \vee DA$, ${\sim}D{\sim}A$) son de elección incompatible por pares. Utilius tiene que elegir entre D~A y (D~A v DA). Por el contrario, los conjuntos (*1'*, *2'*a, *3'*) y (*1'*, *2'*b, *3'*), cada uno de los cuales representa una lista completa de los medios que puede utilizar Utilius para hacer algo con respecto al asunto de la bomba, no son exhaustivos con respecto a oprimir botones; cada conjunto deja otro botón disponible. El conjunto (*1'*, *2'*, *3'*) es exhaustivo con respecto a oprimir botones, pero puede no ser de realización incompatible por pares. Puede suceder, por ejemplo, que Utilius pueda oprimir simultáneamente los botones b_1 y b_2. Naturalmente que si el circuito del artefacto eléctrico está en conformidad con las relaciones lógicas, oprimir los botones b_1 y b_2 simultáneamente tiene exactamente el mismo resultado que oprimir sólo el botón b_1, pues (D~A) & (D~A v DA) es equivalente a (D~A). Sin embargo, que Utilius realice una de las acciones en el conjunto (*1*, *2*, *3*) es idéntico, aunque sólo contingentemente, a que realice la acción correspondiente en el conjunto (*1'*, *2'*, *3'*). Esto es, las acciones en estos dos conjuntos son abstractamente diferentes, pero sólo hay un acontecimiento o cambio en el mundo que Utilius va a introducir, en la estructura causal de sus circunstancias, para que se produzcan otros cambios que lo convertirán en agente de las acciones constituidas por dichos cambios.

En cualesquiera circunstancias el conjunto de acciones (D~A, D~A v DA, ~D~A) es redundante en cuanto *conjunto de ejecuciones o realizaciones*. Si Utilius realiza D~A, independientemente de cómo lo haga, automáticamente realiza D~A v DA. Más aún, si realiza D~A v DA, realiza o bien D~A o bien DA. Si oprime el botón b_2 y detiene la bomba mientras suena la alarma, Utilius realiza DA. En este caso es tan susceptible o digno de encomio como si hubiera oprimido el botón b_1. Claro está, el conjunto no redundante de realizaciones que puede probablemente producir es (D~A, DA, ~D~A). Con todo, en circunstancias tales como las descritas anteriormente, el conjunto (D~A, D~A v DA,

~D~A) no es redundante en cuanto *conjunto de elecciones*. Primero, Utilius tiene a su disposición un mecanismo que le permite elegir D~A v DA, sin elegir ninguno de los miembros de la disyunción. Segundo, en sus circunstancias Utilius no puede en rigor, esto es, con efectividad, elegir realizar DA. Si quiere realizar DA sólo puede elegir D~A v DA y esperar que el segundo miembro de la disyunción sea actualizado por la acción del mecanismo.

Debemos pues distinguir: (*i*) entre conjuntos de acciones como realizaciones alternativas de un agente y conjuntos de acciones como elecciones alternativas para su realización; (*ii*) entre conjuntos de estos tipos y conjuntos de acciones como medios para realizaciones.

2. *La estructura de la acción: máquinas de acción.* En general, haciendo abstracción de los detalles irrelevantes y de los rasgos específicos del caso de Utilius, podemos ver que una situación en la que un agente debe elegir qué hacer, es una compleja estructura causal que incluye los siguientes elementos:

(A) Un propósito o punto de vista dominante W que organiza en una jerarquía de acciones un subconjunto de nexos causales y de cambios posibles que resultan de las más simples acciones del agente;

(B) El conjunto de ejecuciones posibles del agente, determinado tanto por el propósito o punto de vista dominante W como por la red de nexos causales que surgen de las más simples acciones del agente, las cuales se llamarán *las acciones resultantes (output actions) del agente con respecto a W;*

(C) El conjunto de elecciones posibles a disposición del agente, determinado por las acciones resultantes y la red causal: llamémoslo *acciones entrantes (input actions) del agente con respecto a W;*

(D) El conjunto de posibles medios básicos asequibles al agente;

(E) Conjuntos de acciones posibles que son medios para otros medios que permiten al agente llevar al cabo sus realizaciones;

(F) El conjunto de todas las posibles acciones restantes que son completamente irrelevantes para el propósito dominante y para propósitos subsidiarios relativos a la elección de medios.

Nos referiremos a (D) y a (E) como a la mecánica de las acciones entrantes *(input)* y resultantes *(output)*. El principio crucial aquí es:

EP. La ejecución por el agente de una acción resultante A con respecto al punto de vista W es idéntica a su ejecución de una acción entrante B con respecto a W, y ésta última es idéntica a su ejecución de una acción C que sirve de medio para B, y ésta es idéntica a su ejecución de una acción D, inferior en la escala de los medios, que sirve de medio para C, y así sucesivamente hasta llegar a alguna de las acciones básicas más simples, que, en el caso de los terráqueos humanos contemporáneos, son movimientos corporales.

Vamos a ocuparnos aquí de (C), el conjunto de acciones entrantes con respecto a algún punto de vista W. Por lo tanto asumiremos simplemente que cada agente está en el punto focal de una red causal y que esta red incluye la mecánica de las acciones entrantes y resultantes que estamos considerando. No necesitamos aquí distinguir entre diferentes medios alternativos para una misma acción entrante (v.gr., oprimir el botón b_2 y oprimir el botón b_3 en el caso de Utilius antes mencionado). Más aún, la naturaleza de los medios carece aquí de importancia. Nos es indiferente, por ejemplo, que Utilius camine dos millas para poner una carta en el buzón o que la envíe pidiéndole a su hijo que la entregue al cartero en la puerta, u oprimiendo un botón en una máquina que deposite cartas en un buzón. Así, podemos por conveniencia imaginar que nuestros agentes actúan simplemente oprimiendo botones en una *máquina de acción* semejante al artefacto que Utilius operó en la Sección 1 de esta Primera Parte. Lo haremos así pero con un cambio importante.

Recordaremos que Utilius tenía la oportunidad de elegir la disyunción (D~A v DA), esto es detener la bomba sin hacer sonar la alarma, o detener la bomba y hacer sonar la alarma. Deliberadamente nos abstuvimos entonces de introducir consideraciones de probabilidad; pero es perfectamente obvio que el botón b_2 puede muy bien estar conectado a un mecanismo que puede, o bien activar, con mayor probabilidad, un mecanismo que realiza el primer miembro de la disyunción, o bien activar un mecanismo que realiza el segundo miembro de la disyunción. Así, podemos suponer que nuestra máquina de acción tiene, en lugar de botones para oprimir, discos giratorios con unidades de radio divididas en sectores de diferentes áreas. Cada sector representa un miembro de la disyunción asignada a esa parte del disco y el área del sector representa la probabilidad de que el mecanismo conectado con el disco dé lugar al miembro de la disyunción representado por el sector. Así pues, la acción que el agente puede realizar como medio final, es simplemente hacer girar la manecilla del disco que representa la acción entrante que ha elegido.

3. *Acciones abiertas.* Las alternativas abiertas son, las acciones entrantes. Esto es, un curso de acción abierto a un agente en circunstancias C es un curso que el agente no solamente puede ejecutar en C, sino que también puede en rigor elegir ejecutar en C. Obviamente el sentido de "puede" que se usa aquí es un sentido fuertemente contingente que entraña tanto poderes y habilidades como circunstancias causales definidas. Sin embargo, a pesar de la conexión íntima entre este sentido de "puede" y el sentido en que una acción alternativa está abierta, estos sentidos pueden contrastarse de la siguiente manera:

E. 1. "X puede ejecutar A v B en C" implica "X puede ejecutar A en C o X puede ejecutar B en C";

A. 1. "A v B está abierta a X en C" *no implica* "A está abierta a X en C o B está abierta a X en C".

Para dejar constancia, hagamos una lista de otros principios cruciales de la lógica de la apertura frente a la lógica de nuestro "puede" en sentido fuerte. Para mayor conveniencia abreviémosla de la siguiente manera:

Ax = X ejecuta A en C.
$\Diamond(Ax)$ = es lógicamente posible que X ejecute A en C.
$\Diamond\!\!\!\cdot\;(Ax)$ = X puede (en nuestro sentido fuerte) ejecutar A en C.
$o\ (Ax)$ = está abierto para X ejecutar A en C.
$\rightarrow$ = implica.

Entonces nuestros principios cruciales son:

E. 2. $Ax \rightarrow \Diamond\!\!\!\cdot\;(Ax)$
E. 3. $\Diamond\!\!\!\cdot\;(Ax) \rightarrow \Diamond(Ax)$
E. 4. $\Diamond\!\!\!\cdot\;(Ax \vee Bx) \rightarrow \Diamond\!\!\!\cdot\;(Ax) \vee \Diamond\!\!\!\cdot\;(Bx)$
E. 5. Si $(Ax \rightarrow Bx)$, entonces $\Diamond\!\!\!\cdot\;(Ax) \rightarrow \Diamond\!\!\!\cdot\;(Bx)$
A. 2. $o\ (Ax) \rightarrow \Diamond\!\!\!\cdot\;(Ax)$
A. 3. $o\ (Ax) \rightarrow o\ (Ax)$
A. 4. Si $(Ax \rightarrow Bx)$, entonces $o\ (Ax) \rightarrow o\ (Bx)$

4. *Ámbitos de elección: alternativas abiertas.* Determinemos el conjunto de acciones entrantes, esto es, el conjunto de alternativas abiertas de las cuales un agente debe elegir qué hacer. Como dijimos en la Sección 2, este conjunto está determinado tanto por el propósito o punto de vista dominante como por la red causal en torno al agente. Podemos dividir *grosso modo* la contribución de cada uno de la siguiente manera: (*i*) el punto de vista determina el contenido o tipo de acción de la acción entrante; (*ii*) la red causal determina la apertura y, especialmente, las oportunidades o grados de probabilidad de cada acción resultante con respecto a una acción entrante dada.

Aquí no estamos analizando el punto de vista; pero sí nos interesa que el punto de vista determina lo que es importante en la situación del agente, al determinar el conjunto de acciones resultantes y el contenido de las acciones entrantes. Así, podemos tratar aquí el punto de vista simplemente como la asignación de un conjunto de acciones relevantes o importantes para la situación del agente. Por razones que se harán claras más adelante, podemos tomar el conjunto de acciones relevantes que caracterizan un punto de vista dado, como un conjunto de acciones que son unidades de relevancia, es decir, acciones que pueden ser lógica y ontológicamente tan complejas como se quiera, pero tales que, desde el punto de vista considerado, sólo ellas o compuestos de ellas son relevantes, esto es, tienen el valor propio del punto de vista considerado.

Comencemos, pues, con un conjunto α de acciones que son unidades de relevancia y a las cuales llamaremos *acciones atómicamente relevantes* características del punto de vista considerado; abreviémoslas AARas. Digamos que:

$$\alpha = \{A_1, \ldots, A_n\}.$$

Es claro que α determina un conjunto β de acciones que son conjunciones de AARas o sus negaciones, pero no ambas. Hay 2^n conjunciones de la forma:

$(A_1)' \& (A_2)' \& \ldots \& (A_n)'$ donde cada $(A_j)'$ es, o bien A_j o bien $(\sim A_j)$.

Estas conjunciones son todas las alternativas que el agente puede ejecutar. *Qua* realizaciones son éstas los resultados básicos que surgen del elegir y del hacer del agente. Las llamaremos ABRas como abreviación de *acciones básicas resultantes* determinadas por el punto de vista considerado. Así:

$\beta = \{C_1, \ldots, C_{2^n})$, donde cada C_i es una ABRa.

Podemos considerar las ABRas como los *átomos axiológicos* relevantes, esto es, como las acciones a las que el punto de vista dominante les asigna valor de inmediato. Esto no es un requisito pero tiene la ventaja de eludir el problema de si el valor de una conjunción es la suma de los valores de los miembros de la misma conjunción o no lo es. Como es bien sabido en base a la lógica elemental, todo compuesto de AARas que se forme utilizando las conectivas lógicas es, o bien equivalentes a una ABRa, o a una disyunción de ABRas. Así pues, al tratar a las ABRas como axiológicamente básicas reducimos el problema de computar el valor de una acción al problema de calcular los valores de disyunciones cuyos miembros son ABRas cuyo valor está dado.

Puesto que el que X ejecute una ABRa implica el que X ejecute cualquier acción implicada por una ABRa, consideremos las acciones de la forma siguiente:

$(C_1)' \vee (C_2)' \vee \ldots \vee (C_{2^n})'$ en donde cada $(C_j)'$ es, o bien C_j; o bien un miembro nulo de la disyunción:
Llamemos a estas acciones ARas, forma abreviada de *acciones resultantes* características del punto de vista considerado. Es claro que las ABRas son ARas. Cuando toda $(C_j)'$ es nula, tenemos entonces la acción nula que es por lo tanto también una ARa. Hay en total 2^{2^n} ARas. Tenemos pues el conjunto:

$\gamma = \{D_1, D_2, \ldots, D_{2^{2^n}}\}$, donde toda D_i es una ARa.

Introducimos ahora la contribución al conjunto de alternativas provenientes de la red causal en torno al agente. Consideremos ahora la máquina de acción; algunas de las marcas en el disco giratorio corresponderán a ABRas y algunas otras a ARas disyuntivas. Las primeras simplemente asignan una probabilidad 1 a las ABRas, en tanto que las segundas asignan una probabilidad 1 a disyunciones y reparten esta probabilidad entre los miembros de la disyunción en cuestión. Más precisamente, cada marca en el disco giratorio m_j está dividida en 2^{2^n} sectores (algunos de los cuales son nulos) de área P_i^j y se corresponden de la siguiente manera:

$$(C_1)', (C_2)', \ldots, (C_{2^n})'$$
$$p_1^j\ , \quad p_2^j, \ldots ; \quad p_{2^n}^j;$$

en donde $p_1^j + p_2^j + \ldots + p_{2^n}^j = 1$ y cada $p_i^j \geqslant 0$.

Cada una de estas asignaciones de una probabilidad distribuida trae como consecuencia una PARa, esto es, una *acción resultante probabilizada* con respecto a un determinado punto de vista y a la red causal del agente.

Puesto que a cada una de las ABRas se les asigna un valor, podemos considerar una PARa M_j como un conjunto de triples. Tenemos así:

$$M_j = (C_1\ v_1\ p_1^j);\ (C_2\ v_2\ p_2^j), \ldots \quad (C_{2^n};\ v_{2^n},\ p_{2^n}^j),$$

en donde v_i es el valor de la ABRa C_i y p_i^j es la probabilidad asignada por M_j a C_i.

Es claro que hay infinitas maneras en que pueden asignarse a cada ARa probabilidades distribuidas. No es claro, sin embargo, que las redes causales en torno de los agentes tengan mecanismos infinitos que determinen, para un punto de vista dado, PARas infinitas. Con todo, incluso en el caso en que la naturaleza permitiera infinitos mecanismos, parece legítimo asumir que las creaturas actuantes tienen un umbral en su percepción de probabilidades. Parece probable que la mayoría de las veces, un agente considera sólo un conjunto finito de PARas; de hecho, su red causal puede incluso no proporcionar mecanismos para determinar una PARa para cada ARa. Este era precisamente el caso de Utilius en la Sección 1 precedente. En general, si continuamos usando el modelo de la máquina de acción, tenemos que notar que la máquina de acción de un agente, en un determinado momento, puede tener algunos discos marcadores fallidos, esto es, que no funcionen. Así pues, de todo el conjunto de PARas que podemos formar en principio, tenemos que seleccionar las PARas que están a la disposición del agente, esto es, aquellas PARas para las cuales puede realizar una acción que sirva como medio. Éstas serían las APARas o PARas abiertas. Llegamos entonces al siguiente conjunto:

$\omega = \{M_1, M_2, \ldots\}$, donde cada M_j es una APARa.

El conjunto ω, que puede muy bien ser infinito, es, entonces, el conjunto de alternativas abiertas de las cuales un agente debe elegir que hacer. El conjunto ω es el ámbito actual de elecciones en el tiempo *t* para un agente X, dado un punto de vista W y la red causal de X en *t*. Nuestro análisis final de *alternativa abierta a X en el tiempo t en la situación S con respecto al punto de vista W* es simplemente este: una alternativa abierta a X en *t* en S con respecto a W, es un resultado abierto probabilizado para X en *t* en S con respecto a W, esto es, una acción resultante probabilizada M tal que X está en S y X puede efectivamente elegir realizar en S en *t* una acción A que sirva de medio con respecto a W y, el que realice A sería idéntico a que realizara M. Por otra parte, las acciones entrantes de X en *t* con respecto a W son sus APARas en *t* con respecto a W.

5. *Deberes de elegir, de hacer y de tratar.* Ahora podemos dar cuenta de las relaciones entre lo que un hombre debe ejecutar, lo que debe tratar de ejecutar y lo que debe elegir ejecutar o tratar de ejecutar. Antes que nada, *lo que un agente siempre debe elegir ejecutar es una APARa.* Esto es verdad no solamente con respecto al punto de vista dominante considerado, sino también para los puntos de vista derivados, introducidos cuando el agente está eligiendo los medios para sus APARas primarias. En general, dado un conjunto ω de APARas, hay un conjunto de APARas de las que el agente debe elegir una. Llamemos a éstas *APARas debidas* (provenientes de un sentido o tipo de deber) o DAPARas, y al conjunto de ellas $\triangle\omega$ o simplemente $\triangle$, cuando está claro cuál es el conjunto ω.

En la situación S, entonces, un agente tiene un conjunto $\triangle$ de DAPARas de las cuales tiene que elegir lo que ha de producir como una acción entrante para su máquina de acción; también tiene que elegir una acción medio para dar lugar a la DAPARa que haya elegido.

Si $\triangle\omega = \omega$, entonces normalmente no hay una acción que el agente deba ejecutar o tratar de ejecutar con respecto al punto de vista que determina parcialmente a ω. Si $\triangle$ tiene solamente un miembro y éste es una PARa M_j que asigna probabilidad 1 a una ABRa dada C_i; entonces M_j asigna probabilidad 0 a las otras 2^n–1 ABRas. En este caso el agente tiene que realizar C_i. Si para $j = 1, 2, \ldots, k$ y $k > 1$, cada APARa M_j en $\triangle$ asigna una probabilidad de 1 a la ABRa C_{j_i} entonces no es el caso que el agente tenga que realizar ninguna ABRa dada, sino que debe elegir una de las ABRas $C_{j_1}, \ldots, C_{j_k}$. En este caso decimos que el agente debe realizar la disyunción C_{j_i} v...v C_{j_k}. También decimos que el agente debe ejecutar una disyunción C_{j_i} v...v C_{j_k} si el único miembro de $\triangle$ asigna la misma probabilidad a cada uno de $C_{j_i} \ldots C_{j_k}$ y asigna una probabilidad de 0 a todas las otras 2^n–k ABRas. Finalmente, de un modo general, si un agente debe realizar una acción A también tiene que realizar cualquier acción B cuando el hacer A implica el hacer B.[5]

Supongamos que $\triangle$ tiene algunos miembros que no asignan probabilidad 1 a ninguna ABRa. Supongamos que hay una PARa M_h que a la vez que asigna probabilidad 1 a una ABRa C_i no está en $\triangle$ por la razón de que M_h no está abierta al agente en su situación, esto es, M_h no está en $\triangle\omega$ por la simple razón de que M_h no está en ω. Más aún, supongamos que C_i es la única ABRa a la cual cada miembro de $\triangle$ le asigna una probabilidad diferente de 0. Entonces, la situación del agente puede ser más brevemente descrita como aquélla en que, aunque no es el caso que deba ejecutar C_i, sin embargo debe tratar de hacer C_i. De una manera más general, supongamos que D_k es una disyunción C_i v ... v C_h de ABRas tal que: (*i*) cada miembro de $\triangle$ asigna una probabilidad diferente a 0 a cada una de $C_i, \ldots, C_h$; (*ii*) la

[5] Digo *grosso modo* porque el principio en cuestión es más exactamente como sigue: "si la orden 'X, realiza A' implica la orden 'X, realiza B', entonces, 'X debe de realizar A' implica 'X debe realizar B'." Las razones para órdenes o prescripciones implicadas aparecen, por ejemplo, en "Actions, Imperatives and Obligations", *Proceedings of the Aristotelian Society*, Vol. XLVIII (1967-68), y en "Acts, the Logic of Obligation, and Deontic Calculi", *Philosophical Studies*, Vol. XIX (1968), y CRÍTICA, Vol. I (1967).

PARa que asigna probabilidad $1/h$ a cada $C_i, \ldots, C_h$ estaría en $\triangle$ si estuviera en ω (esto es, la PARa en cuestión satisfaría todas las condiciones para estar en $\triangle$ excepto la de ser miembro de ω), y (*iii*) D_k es la única disyunción de ABRas que satisface (*i*) y (*ii*). Entonces, podemos brevemente decir que el agente en cuestión debe; *no* hacer D_k, sino *tratar* de hacer D_k.

Naturalmente que el deber de realizar o de tratar de realizar una acción entrante M_h lleva consigo el deber de hacer una acción A que serviría como medio para la realización de M_h tal que, el hacer A del agente en su situación es idéntico a su hacer M_h y, *a fortiori*, a su hacer la acción resultante (cualquiera que ésta sea) producida por A. Pero, lo que A sea y lo que sean sus alternativas no puede decidirse simplemente examinando los miembros de $\triangle\omega$ o el punto de vista que determina a ω.

6. *La matriz de elección.* Podemos representar la estructura de un APARa de una manera simplificada. Ordenamos las ABRas, o bien arbitrariamente, o, si lo preferimos, en orden de valor decreciente y por un *fiat* en el caso en que tengan el mismo valor. Después ordenamos los miembros de cada APARa de la siguiente manera: el *i*avo triple en una APARa es el que incluye la *i*ava ABRa. Así, no necesitamos referirnos explícitamente a las ABRas cuando describimos una APARa M_j, que es; por lo tanto, un 2^n–tuple ordenado $((v_1; p^j_1); \ldots; (v_{2^n}; p^j_{2^n}))$, de pares, siendo los miembros de estos últimos un valor y una probabilidad, esto es, $0 \leqslant p^j_1 \leqslant 1$ y $p^j_1 + \ldots + p^j_{2^n} = 1$.

Ahora bien, puesto que a cada ABRa se le asigna sólo uno y el mismo valor por el punto de vista considerado, la característica de una APARa M_j es simplemente el vector de probabilidad $(p^j_1; \ldots, p^j_{2^n})$. Así, todo el conjunto ω o ámbito de elección para un agente X en una situación dada S en un cierto tiempo t puede representarse por una matriz de probabilidad a la que llamaremos *la matriz de elección para X en*

S *en* t. Por ejemplo:

Esquema I:

	$C_1\ldots$ 600	$C_h\ldots$ 300	$C_i\ldots$ 100	$C_j\ldots$ 100	$C_k\ldots$ —400	C_{2^n} —800
M_1	.5	0	0	0	.5	0
M_2	0	0	.5	.5	0	0
M_3	0	1	0	0	0	0
M_4	0	.4	.3	.3	0	0

PARTE II. *Utilitarismo*

7. *Utilitarismo y utilidad esperada.* Consideremos de nuevo el caso de Utilius. Éste se enfrenta al problema de cuál de las alternativas que le están aún abiertas debe elegir. De acuerdo con el utilitarismo de la acción la respuesta es muy sencilla: debe elegir una de las alternativas abiertas que tengan el más alto valor. Pero esta respuesta no tiene sentido a menos que haya un método para calcular el valor de las APARas que asignan una probabilidad diferente de cero a más de una ABRa; como es el caso de M_1, M_2 y M_4 en el Esquema I. En tanto que una elección de entre las APARas no es más que una elección de entre vectores de probabilidad, es claro que ahora estamos tratando con un sentido amplio, probabilizado, de valor. Sin embargo, es evidente que no podemos elegir simplemente el vector que tenga el componente con la más alta probabilidad, ya que tal vector puede muy bien ser aquél que asigna la más alta probabilidad a la peor ABRa posible.

Una APARa es, como dijimos, un 2^n–tuple de pares de probabilidades y valores. Esto es exactamente lo que en teoría de la utilidad se llama *lotería,* esto es, una asignación de probabilidades a ciertos premios. En nuestro caso los premios son las ABRas. Ahora bien, la teoría clásica de la utilidad computa la *utilidad esperada,* u(L), de una lotería L con m premios por medio de la fórmula:

(1) $u(L) = v_1p_1 + \ldots + v_mp_m$,
en donde v_i es el valor del *i*avo premio:[6]

Así, conectando la teoría de la utilidad con el utilitarismo podemos usar (1) para determinar el valor de cada una de las APARas entre las que Utilius debe elegir una. Consideremos, a manera de ilustración, el conjunto ω_1 que tiene sólo los dos miembros M_1 y M_2 descritos en el Esquema I. Las computaciones tendrían la forma siguiente:

$$u(M_1) = .5 \times 600 + .5 \times -400 = 100$$
$$u(M_2) = .5 \times 100 + .5 \times -100 = 100$$

Si la utilidad esperada es la extensión relevante del valor de las ABRas, en este caso no hay ninguna alternativa abierta a Utilius que éste deba elegir. Sin embargo, en la gran lotería de la vida, con respecto a la cual la moralidad y el utilitarismo quieren constituirse en guías, no podemos aceptar que la utilidad esperada sea considerada como la extensión propia de las PARas de los valores de las ABRas. Podría ser muy bien el caso de que 400 unidades de desvalor representen una catástrofe terrible y que el agente moral deba abstenerse de correr el riesgo de que tal catástrofe se produzca. Es cierto que una probabilidad de .5 es muy grande. Aunque las APARas de un agente moral sean loterías (en sentido técnico), el agente moral no debe ser un jugador que tome demasiados riesgos. El riesgo de una catástrofe echa por tierra el prospecto de una bonanza.

En la teoría moderna de la utilidad, el valor de una lotería no necesita ser la utilidad esperada de la lotería. Von Neumann y Morgenstern muestran cómo establecer un rango de loterías que comprenden un conjunto de premios en el caso en que tengan: (a) un rango o jerarquía de premios, y (b) una comparación entre cada premio y una lotería que tenga solamente el más alto o el más bajo de los premios. El supuesto crucial es este: supongamos que C_1 y C_{2^n} sean

[6] Véase, por ejemplo, R. Duncan Luce y Howard Raiffa, *Games and Decisions,* New York, John Wiley and Sons Inc., 1967 cap. 3.

el más alto y el más bajo premio respectivamente; entonces, para cada premio C_j hay una probabilidad p_j ($\neq 0$) tal que resulta indiferente elegir C_j o elegir una lotería que asigne una probabilidad p_j a C_1 y una probabilidad $1-p_j$ a C_{2^n}. Por medio de la substitución de loterías indiferentes y de otros principios, para el caso de las tres primeras loterías en el Esquema 1, habría unas probabilidades p y q que permitirían sostener las siguientes ecuaciones, donde "$\approx$" expresaría igualdad de valor, esto es, de utilidad esperada:

$$M_1 = ((.5, 600), (5, -800))$$
$$M_2 = ((.5, 100), (.5, \; 100)) \approx ((p, 600), (1-p, -800))$$
$$M_3 = ((1, \; 300)) \qquad\qquad\qquad \approx ((q, 600), (1,q, \; -800))$$

En tanto que p y q son más grandes que .5, M_2 y M_3 son preferibles a M_1. Esto parece bien: Sin embargo, no podemos adoptar este método. Mientras debemos abstenernos de arriesgarnos a producir la catástrofe (C_{2^n}, —800), es indiferente que consideremos M_2 o M_3 como alternativas. Lo único que importa es la cuantía de la probabilidad comprendida. Así pues, dado que se descarta la catástrofe (C_{2^n}, —800), las probabilidades p y q deberían ser ambas despreciables e idénticas o, en último término, deberían tener un idéntico umbral: Pero si $p = q$, entonces, $M_2 \approx M_3$, lo cual es obviamente falso.

Es una cuestión discutible si hay una pequeña probabilidad que anule el descartar una catástrofe. En términos puramente matemáticos podemos considerar una probabilidad de la magnitud .00001e(1000e(1000e(1000))), en donde 'e' significa que el número mencionado a la derecha es un exponente del mencionado a la izquierda. Ya que una probabilidad de este orden está indudablemente más allá de nuestra percepción de probabilidades, podemos asumir que en esta discusión la probabilidad 0 y la probabilidad 1 son umbrales 0 y 1. Esto es, para los propósitos de elegir lo que uno debe realizar, el descartar una catástrofe se anula solamente por la probabilidad (umbral) 0.

Nos apresuramos a advertir que el hecho de que la utilidad esperada clásica, o la utilidad esperada según la teoría de Von Neumann-Morgenstern, no sea el valor que determine lo que uno deba elegir moralmente, no es una objeción a la teoría de la utilidad o a las importantes aplicaciones de esta teoría en economía y en otras ciencias sociales, o a su uso en el estudio psicológico de la toma de decisiones. En estas disciplinas a lo sumo se da por sentado que cada persona actúa de acuerdo con una función propia de utilidad, de manera que la teoría de las decisiones, no solamente la teoría de la utilidad, suministra a lo sumo una teoría normativa *condicional*, esto es, una teoría que nos dice lo que el agente debería elegir supuesto que tuviese ciertos valores y cierta función de utilidad. Aquí nos las habemos con un punto de vista absoluto y no tenemos nada que decir acerca de los estudios fácticos sobre la toma de decisiones o las teorías matemáticas usadas y desarrolladas en el proceso de aquéllos estudios.

8. *Utilitarismo, minimax y máximo.* Nuestro problema de encontrar el mejor curso de acción es, esencialmente, un caso especial del problema de encontrar la mejor estrategia en un juego. Se trata de un juego en el cual el otro jugador es la naturaleza, esto es, el que elegirá la ABRa que debe realizarse. Sin embargo, nuestro problema es más simple. En un juego de dos personas se da una matriz tal que en la intersección de la columna *j* y el renglón *i*, se encuentra una descripción o nombre del par de premios $(a_{ij}; b_{ij})$ que ganarán respectivamente cada uno de los jugadores. Cuando se incluyen estrategias de probabilidad, el problema es encontrar un vector de probabilidad que maximizará el ganar o, alternativamente, minimizará el perder. En nuestro caso hay solamente un renglón de premios y se nos dan *todas* las probabilidades permitidas. Parece, entonces, que todo lo que tenemos que hacer es encontrar la estrategia que maximice el valor (o las oportunidades del valor más grande), o, al-

ternativamente, encontrar la estrategia que minimice el riesgo de desvalor.

De acuerdo con el principio *minimax*, las mejores estrategias son aquéllas que incluyen el mínimo de los premios máximos de todas las estrategias disponibles. Este principio y su contraparte, el principio *maximin*, tiene la ventaja de no requerir un rango completo de las estrategias; van directamente a la médula de la cuestión y determinan el conjunto de las mejores estrategias, sin importarles el rango de las que no son las mejores.

Debemos advertir que los criterios para las mejores estrategias propuestos por los partidarios de la teoría de la decisión, se aplican a los valores de los premios. Cuando el problema consiste en encontrar un vector de probabilidad, este vector será usado para computar la utilidad esperada de la estrategia.[7] Pero nuestro caso es diferente; hemos visto que la utilidad esperada no es satisfactoria y, además, debemos considerar valores y probabilidades. De cualquier modo, debemos remover todos los obstáculos para encontrar un utilitarismo de la acción que sea viable. Aplicaremos, entonces, los criterios existentes propuestos a los riesgos de valor representados por los productos de valores y sus probabilidades correspondientes.

Supongamos que Utilius tiene como conjunto de APARas con respecto a un determinado punto de vista, el conjunto ω descrito en el Esquema I. Entonces, encontramos los siguientes riesgos de valor:

Esquema II:
Riesgos de valor

	$C_1 \dots$	$C_h \dots$	$C_i \dots$	$C_j \dots$	$C_k \dots$	$C_{2^n} \dots$	máximos	mínimos
M_1	+ 300	0	0	0	—200	0	+ 300	—200
M_2	0	0	+ 50	+ 50	0	0	+ 50	0
M_3	0	+ 300	0	0	0	0	+ 300	0
M_4	0	+ 120	+ 30	+ 30	0	0	+ 120	0

[7] Véase, por ejemplo, R. Duncan Luce y Howard Raiffa, *op. cit.*, cap. 13.

El principio *minimax* ofrece una respuesta a partir de la penúltima columna. Como podemos ver, el mínimo máximo es 50; así, de acuerdo con el principio *minimax* la alternativa M_2 debe ser elegida. Esto es, sin embargo, incorrecto, ya que la certidumbre de 300 convierte a M_3 en la alternativa que realmente debería elegirse. En este ejemplo el principio *maximin,* que se aplica a la última columna, resulta más obviamente inadecuado: solamente excluye a M_1 y asigna el mismo riesgo de valor a M_2 y a M_3.

Restrinjámonos a los riesgos de valor de las ABRas a las que se asigna una *probabilidad diferente de 0.* Entonces, los máximos y los mínimos relevantes son:

Esquema III:

	Máximos	Mínimos
M_1	+ 300	— 200
M_2	+ 50	+ 50
M_3	+ 300	+ 300
M_4	+ 120	+ 30

Aquí de nuevo el principio *minimax* no logra convertir a M_3 en la alternativa que tenga el valor más alto y, por lo tanto, en aquélla que Utilius deba preferir. Por otro lado, el máximo mínimo es 300 y el principio *maximin* convierte, correctamente; a la acción M_3 en aquélla que Utilius debe elegir.

Sin embargo, el principio *maximin* no es siempre satisfactorio. Si Utilius estuviera forzado a elegir nada más entre M_2 y M_4, el principio *maximin* ofrecería la respuesta equivocada; convertiría a M_2 en la acción mejor, aunque M_4, con su 40% de probabilidades de producir 300 unidades de valor, es una acción muy superior.

El economista L. J. Savage ha propuesto un mejoramiento del principio *minimax.* Savage aplica este principio, no a la matriz original de un juego, sino a la *matriz de lamento (regret matrix)* que se deriva de la primera de la siguiente manera: cada entrada a_{ij} de la matriz original se reemplaza por la diferencia a_{ij}–A_j; en donde A_j es el máximo de la columna *j*. Aprovechando la idea de Savage podemos distinguir entre una *matriz vertical de lamento* (tal como la acabamos de describir) y una *matriz horizontal de lamento* (que se obtendría reemplazando a_{ij} por a_{ij}–A_i; en donde A_i es el máximo valor en el *i*avo renglón). Así pues, tomando los principios *minimax* y *maximin* obtenemos cuatro criterios. En el caso de Utilius y su conjunto de APARas descrito en el Esquema I, encontramos las siguientes matrices:

Esquema IV

1. *Matriz Vertical de Lamento*

ω	C_1	C_h	C_i	C_j	C_k	Máximos	Mínimos
M_1	0	—300	—50	—50	—200	0	—300
M_2	—300	—300	0	0	0	0	—300
M_3	—300	0	—50	—50	0	0	—300
M_4	—300	—180	—20	—20	0	0	—300

2. *Matriz Horizontal de Lamento*

ω	C_1	C_h	C_i	C_j	C_k	Máximos	Mínimos
M_1	0	—300	—300	—300	—500	0	—500
M_2	—50	—50	0	0	—50	0	—50
M_3	—300	0	—300	—300	—300	0	—300
M_4	—120	0	—90	—90	—120	0	—120

Patentemente, el único criterio descriminador es el principio *maximin* aplicado a la matriz horizontal de lamento. Pero incluso esta produce un resultado incorrecto, ya que convierte a M_2, y no a M_3, en la APARa que Utilius debe elegir.

9. *Utilidad esperada calibrada.* El utilitarismo de la acción no parece, entonces, tan prometedor como pudiera pensarse a primera vista, ya que no hay un criterio viable de la

'mejor acción'. De hecho, la expresión 'la mejor acción' puede ser utilizada como un mero sinónimo de 'acción obligatoria' o 'acción que es miembro de un conjunto del cual una acción debe ser elegida (o ejecutada)'. Entonces, para que el utilitarismo no sea una tautología vacía, debe considerarse que es esencialmente la teoría acerca del criterio para determinar la mejor acción u obligatoriedad. Pero, como hemos visto, la formulación de tal criterio no es una cuestión

Concluiremos esta investigación con una propuesta *sumamente* tentativa de un criterio formal utilitarista de obligatoriedad. Para simplificar asumamos que las ABRas determinadas por el punto de vista relevante están ordenadas en valor decreciente y que las primeras r ABRas tienen valor positivo, entonces:

$$C_1, \ldots C_r, C_{r+1}, \ldots, C_{2^n}.$$

En donde los valores $v_i > 0$, para $i = 1, 2, \ldots, r$,
y $v_i \leqslant 0$; para $i = r + 1, r + 2, \ldots, 2^n$,

Necesitamos las siguientes definiciones:

D1. L es la cantidad mínima de desvalor considerada como catástrofe en la situación en cuestión con respecto a un método de medición axiológica.

D2. Ω es el subconjunto de ω que resulta de ω al sustraer todas las APARas en ω que asignan una probabilidad diferente a cero a una ABRa C_j tal que $v_i \leqslant L$.

D3. Ω' es el subconjunto de ω que resulta de ω al sustraer todas las APARas en ω que asignan una probabilidad diferente a cero a C_{2^n} y $v_{2^n} \leqslant L$.

D4. La *utilidad esperada negativamente L-calibrada* de una APARa M_j es:

$$E^*_L(M_j) = p^j_1 v_1 + \ldots + p^j_r v_r + p^j_{r+1} \frac{(Lv_i)}{(L-v_i)} + \ldots + p^j_{2^n} \frac{(Lv_{2^n})}{(L-v_{2^n})}$$

D5. $E^*_{L_{(m)}}(M_j)$ = La suma de los primeros *m* términos del polinomio anterior.

Ahora podemos formular lo siguiente:

10. *Criterio del utilitarismo de la acción.* Dado un conjunto $\triangle\omega$ de APARas de las cuales el agente debe elegir un curso de acción en la situación S en el tiempo *t*,

(A) X debe elegir cualesquiera de los miembros del conjunto $\triangle\omega$ que se describe a continuación y

(B) X no debe de elegir ninguna de las APARas que no estén en $\triangle\omega$.

El conjunto $\triangle\omega$ se determina como sigue:

(1) Si no hay ningún valor L, tal como se le caracteriza en D1, entonces $\triangle\omega$ es el conjunto de todas las APARas en ω que tienen una utilidad esperada clásica tan alta como la de cualquier otro miembro de ω;

(2) Si hay un valor catastrófico L y $L < v_{2^n}$, entonces $\triangle\omega$ es el conjunto de todas las APARas en ω que tienen al menos una utilidad esperada negativamente L-calibrada tan grande como la de cualquier otro miembro de ω.

(3) Si hay un valor catastrófico L y $L \geqslant v_{2^n}$, y el conjunto Ω es idéntico a ω, o Ω no está vacío, entonces $\triangle\omega$ es el conjunto de todas las APARas en Ω que tienen al menos una utilidad esperada negativamente v_{2^n}-calibrada tan grande como la de cualquier otro miembro de Ω;

(4) Si hay un valor catastrófico $L \geqslant v_{2^n}$ y Ω está vacío, pero Ω' no está vacío, entonces $\triangle\omega$ es el conjunto de todas las APARas en Ω' que tienen al menos una utilidad esperada negativamente v_{2^n}-calibrada tan grande como la de cualquier otro miembro de Ω';

(5) Si tenemos el caso (4) pero Ω' está vacío, entonces $\triangle\omega$ es el conjunto de APARas M_j en ω para las cuales $E^*_{v_{2^n}\,(2^n-1)}(M_j) \geqslant E^*_{v_{2^n}\,(2^n-1)}(M_h)$; para cualquier M_h en ω.

SUMMARY

This essay is another installement in a continuing examination of the logical fundations of utilitarianism. Here we discuss act-utilitarianism. Utilitarianism is intelligible to the extent that there is an understanding of the concepts of *alternative action*, *open action*, and *best consequence*. We have already argued elsewhere, that the value-maximizing condition (i.e., causing at leats as great a value) cannot be necessary for obligation, i.e., that simple act-utilitarianism cannot be a viable act-utilitarianism.

Our main purpose here is not critical but constructive. In Part I, we offer a contribution to the analysis of *open alternative action*, distinguish between a set of actions for performance and a set of actions for choice, and structure the relationships among duties to choose, to do, and to try to do. In Part II, we examine some ways of meeting the serious problem of combining value and probability that have already been explored in utility theory and the modern theory of decision under risk. Since the ways in question are not satisfactory, we conclude the essay with a very tentative proposal for combining probabilities and values in a version of act-utilitarianism. But it is not a simple utilitarianism.

I. OPEN ALTERNATIVE ACTIONS

1. *Performances, choices and means.* From the example given in this section, we come to the conclusion that we must distinguish (*i*) sets of actions as alternative performances of an agent from sets of actions as alternative choices for performance; and (*ii*) each of these types of sets from sets of actions as means for performances.

2. *The structure of action: action machines.* A situation in which an agent ought to choose what to do is a complex causal structure that includes:

(A) A dominant point of view W that organizes a subset of the causal links and possible changes issuing from the agent's simplest actions into a hierarchy of actions;

(B) The set of possible performances by the agent, determined both by the dominant purpose W and the network of causal links that start from the agent simplest actions: to be called *the agent's output actions with respect to W*;

(C) The set of possible choices available to the agent, determined

by the output actions and the causal network: to be called *the agent's input actions with respect to W*;

(D) The set of possible ultimate means available to the agent;

(E) Sets of possible actions which are means to means to the agent's bringing about his performances;

(F) The set of all remaining possible actions, which are wholly irrelevant to the dominant purpose.

The crucial principle here is:

EP. The agent's performing an output action with respect to W is identical with his performing an input action B with respect to W, and the latter is identical with his performing a means action C for B, and so forth down to some simplest basic action.

We are concerned here with (C). We need not differentiate between alternative means for an input action. The nature of the means is of no importance here. For convenience we may imagine that our agents act by simply pushing buttons in an *action machine*. Thus, the ultimate means action the agent can perform is simply to operate the action machine in a way to produce the input action he has chosen.

3. *Open actions*. The open alternatives are in fact the input actions. That is, a course of action open to an agent in some circumstances C is a course which the agent not only can perform in C, but can also choose effectively to perform in C. The sense of "can" in use here is a very contingent one involving both powers and abilities as well as definite causal circumstances. This sense of "can" and the sense in which an alternative action is open, contrast as follows:

Perf. "X can perform A v B in C" entails "X can perform A in C or X can perform B in C";

Op. "A v B is open to X in C" does *not* entail "A is open to X in C or B is open to X in C"

4. *Choice spaces: open alternatives*. We must tackle the problem of determine the set of input actions, i.e., the set of open alternatives from which an agent ought to choose what to do. As said in § 2, this set is determined both by the dominant purpose or viewpoint and the causal network around the agent. We can roughly divide the contribution each makes as follows: (*i*) the viewpoint determines the content or type of action of the input action; (*ii*) the causal network determines the openness, and particularly the chances or

amounts of probability of each output action with respect to a given input action.

We may treat the viewpoint simply as the assignment of a set of important or relevant actions to the agent's situation. We may take these actions as a set of actions that are units of relevance, and we shall call this set α. Actions in α are to be called *atomically relevant actions*, and for short *aras*.

The set α determines a set β of actions which are conjunctions of aras or their negations but not both. These conjunctions are the full alternatives that the agent can bring out. *Qua* performances they are the basic outcomes that will issue from the agent choosing and doing. We shall call them *boas*, as short for *basic output actions* determined by the viewpoint under consideration.

The boas may be regarded as the relevant *axiological atoms.* This has the convenience of reducing the problem of computing values to the problem of reckoning the values of disjunctions given the values of the boas which are their disjuncts.

Since X's performing a boa entails his performing any action entailed by the boa, we consider the action of the form,

$$(C_1)' \vee (C_2)' \vee \ldots \vee (C_{2^n})',$$

where each $(C_j)'$ is either C_j or a null disjunct. We shall call these actions *oas*, as short for *output actions.* There are altogether $_{2^{2^n}}$ oas. Thus we have the set

$$\gamma = \{D_1, D_2, \ldots, D_{2^{2^n}}\}, \text{ where each } D_i \text{ is an oa.}$$

Now we bring in the contribution to the set of open alternatives from the causal network surrounding the agent. In our action machine, some of the buttons correspond to boas and some to disjunctive oas. The former simply assign probability 1 to boas; the latter assign probability 1 to disjunctions, and partition this probability into the disjuncts of the disjunctions in question. Each of the assignments of a probability distribution to an oa yields a *poa*, i.e., a *probabilitized output action* with respect to the given viewpoint and the agent's causal network.

Since each of the boas is assigned a value, we may consider a poa M_j as a set of triples, thus:

$$M_j = (C_1, v_1, p_1^j),\ (C_2, v_2, p_2^j),\ \ldots\ (C_{2^n},\ v_{2^n}, p_{2^n}^{j_n}),$$

where v_i is the value of boa C_i and p_i^j is the probability assigned by M_j to C_i.

From the whole set of poas that we can form in principle, we must select the poas which are available to the agent. These are the *opoas*, or *open poas*. Hence we arrive to the set

$$\omega = \{M_1, M_2 \ldots\}, \text{ where each } M_j \text{ is an opoa.}$$

The set ω is the actual space of choices at time t for an agent X, given a viewpoint and X's causal network at t. An *open alternative for X at time t in situation S with respect to viewpoint W* is simply an open probabilitized output for X at t in S with respect to W, i.e., a probabilitized output action M such that X is in S and X can effectively choose to perform in S at t a means action A, and his doing A would be identical with his doing M.

5. *Duties to choose, to do, and to try*. What an agent ought to choose to perform is always an opoa. Let us call *dutiful opoas*, or *dopoas*, to those opoas from which the agent ought to choose one, and the set of dopoas $\triangle\omega$ or simply $\triangle$, when it is clear what the set ω is.

If $\triangle\omega = \omega$, then there is normally no action that the agent ought to do or try to do with respect to the viewpoint that partially determines ω. If $\triangle$ has just one member and this is a poa M_j that assigns probability 1 to a given boa C_i, then the agent ought to *do* C_i. If for $j = 1,2,\ldots, k$ and $k > 1$, each opoa M_j in $\triangle$ assigns probability 1 to boa C_{j_i}, then the agent ought to *choose* one of the boas $C_{j_i}, \ldots, C_{j_k}$.

Let $\triangle$ have some members that assign probability 1 to no boa. Let there be a poa M_h that both assigns probability 1 to a boa C_i and is not in $\triangle$ just because M_h is not open to the agent in his situation. Further, let C_i be the only boa which is assigned by each member of $\triangle$ a probability different from O. Then the agent ought to *try* to do C_i.

6. *The choice matrix*. We can represent the structure of an opoa in the following way: We order the boas wholly arbitrarily, or, as we prefer, by decreasing value. Then we order the members of each opoa as follows: the *i*th triple in an opoa is the one which includes the *i*th boa. Thus we need not refer explicitely to the boas in describing an opoa M_j, which is therefore, an ordered 2^n-tuple $((v_1, p^j_1, \ldots, (v_{2^n}, p^j_{2^n}))$ of pairs, the members of the latter being a value and a probability. Since each boa is assigned just one value by the viewpoint under consideration, the characteristic of an opoa M_j is simply the probability vector $(p^j_1; \ldots ; p^j_{2^n})$: Thus; the choice space (the set ω) for an angent X in situation S at time t can be represented by a probability matrix, to be called the choice matrix for X in S at t.

II. UTILITARIANISM

7. *Utilitarianism and expected utility.* According to act-utilitarianism one ought to choose one of the open alternatives that have the highest value. But this is nonsensical unless there is a method for computing a value for opoas that assign nonzero probabilities to more than one boa. It is clear that we are now dealing with an extended, probabilitized sense of value.

An opoa is a 2^n-tuple of pairs of probabilities and values. This is what in utility theory is called a *lottery,* i.e., an assignment of probabilities to certain prizes. Classical utility theory computes the *expected utility* u(L) of a lottery L with *m* prizes by means of the formula:

(1) $$u(L) = v_1p_1 \ldots v_mp_m$$
where v_i is the value of the *i*th prize.

Thus, connecting utility theory with utilitarianism we can use (1) to determine the value of each of the opoas from which an agent ought to choose. Yet, we cannot accept expected utility as the proper extension for poas of the values of the boas. Even though a moral agent's opoas are lotteries, the moral agent ought not to take the risk of bringing about a catastrophe.

In modern utility theory the value of a lottery need not be the lottery's expected utility. Von Neumann and Morgenstern showed how to establish a ranking of the lotteries involving a set of prizes if there is: (a) a ranking of the prizes, and (b) a comparison between each prize and a lottery involving just the highest and the lowest prizes. However, as long as we ought to refrain from risking bringing about a catastrophe, we cannot accept this method niether. We hasten to note that the fact that classical expected utility or the Von Neumann-Morgenstern expected utility is not the value that determines what one ought morally to choose is not an objection to utility theory, or to the applications of the theory in other sciences.

8. *Utilitarianism, minimax and maximum.* Our problem of finding the best courses of action is essentially a special case of the problem of finding the best strategy in a game. In a two person game there is a given matrix such that at the intersection of column *j* and row *i* lies a description of the pair of prizes to be gained by each of the players, respectively. When probability strategies are included, the problem is to find a probability vector that will maximize gain or will minimize loose. Our case is simpler. We have just one row of prizes and we are given *all* the probabili-

ties that are allowed. We have to find the strategy that maximizes the chances of greater value or the strategies that minimizes the risk of disvalue.

According to the *minimax principle* the best strategies are the ones that include the minimum of the maximal prizes of all available strategies. This principle, and its dual, the *maximin principle,* go directly to the heart of the matter and determine the set of best strategies, without concern for the ranking of the non-best ones.

The criteria for best strategies proposed by practitioners of decision theory apply to the values of the prizes. We must consider values and probabilities. We shall, therefore, apply the existing proposed criteria to the risks of value represented by the product of values and their corresponding probabilities. However, if we apply the minimax, or alternatively, the maximum principle, we see that the answer they yield to our problem is not always satisfactory.

The economist L. J. Savage has proposed an embellishment of the minimax principle. He applies this principle to the *regret matrix* of the original matrix of a game. But even this yields incorrect results.

9. *Calibrated expected utility.* In order for utilitarianism not to be an empty tautology, utilitarianism must be taken to be essentially the theory about the criterion of best action or obligatoriness. But the formulation of such a criterion is not an easy matter.

We conclude this discussion with a very tentative proposal of a formal utilitarian criterion of obligatoriness. For convenience we assume that the boas are ordered in decreasing value and that the first r boas have possitive value. We need the following definitions:

D1. L is the minimal amount of disvalue regarded as a catastrophe.

D2. Ω is the subset of ω that results from ω by substracting all opoas in ω that assign a nonzero probability to boa C_j such that $v_j \leqslant L$.

D3. Ω' is the subset of ω that results from ω by substracting all opoas in ω that assign a nonzero probability to $C_2{}^n$ and $v_2{}^n \leqslant L$.

D4. The *negatively L-calibrated expected utility* of an opoa M_j is:

$$E^*_L(M_j) = p^j_1 v_1 + \dots + p^j_{r+1} \frac{(Lv_i)}{(L-v_i)} + \dots + p^j_{2^n} \frac{(Lv_{2^n})}{(L-v_2{}^n)}$$

D5. $E^*_{L_{(m)}}(M_j)$ = The sum of the first m terms of the above polynomial.

10. *Act-utilitarian criterion.* Given a set ω of opoas from which an agent X ought to choose a course of action in a situation S at time t,

(A) X ought to choose any one of the members of the set $\triangle\omega$ to be described below, and

(B) X ought to choose none of the opoas not in $\triangle\omega$.

The set $\triangle\omega$ is determined as follows:

(1) If there is no value L as characterized in D1, then $\triangle\omega$ is the set of all opoas in ω that have as great classical expected utility as any other member of ω;

(2) If there is a catastrophic value L and $L < v_2{}^n$, then $\triangle\omega$ is the set of all opoas in ω that have at least as great negatively L-calibrated expected utility as any other member of ω.

(3) If there is a catastrophic value L and $L \geqslant v_2{}^n$, and the set Ω is the same as ω or Ω is not empty, then $\triangle\omega$ is the set of all opoas in Ω that have at least as great negatively $v_2{}^n$-calibrated expected utility as any other member of Ω;

(4) If there is a catastrophic value $L \geqslant v_2{}^n$ and Ω is empty, but Ω' is not empty, then $\triangle\omega$ is the set of all opoas in Ω' that have at least as great negatively $v_2{}^n$-calibrated expected utility as any other member of Ω';

(5) If as in (4) but Ω' is empty, then $\triangle\omega$ is the set of opoas M_j in ω for which
$E^*_{v_2{}^n\,(2^{n-1})}(M_j) \geqslant E^*_{v_2{}^n\,(2^{n-1})}(M_h)$, for any M_h in ω.