Sociedad Colombiana de Matemáticas
XV Congreso Nacional de Matemáticas
2005 **Apuntes**

Lecturas Matemáticas
Volumen Especial (2006), páginas 225–247

Modelos matemáticos del conocimiento con aplicaciones a la economía

Luciano I. de Castro
Universidad Carlos III, Madrid, España

Abstract. This article describes a mathematical model for knowledge, which is used in game theory and economics theory in general. The model is based on logic, and allows the study of important questions in economics. We present examples and applications.

Key words and phrases. Epistemology, Interactive epistemology, Knowledge, Common knowledge, Semantic, Syntactic, Model.

2000 AMS Mathematics Subject Classification. AMS Subject Classification: Primary: 03B80, 91B44 Secondary: 03B42, 91A99.

Resumen. ¿Es posible modelar matemáticamente el conocimiento? ¿Qué herramientas y conceptos matemáticos si puede usar para tanto? En este artículo, respondemos a estas cuestiones al desarrollar los conceptos básicos de epistemología tal cómo es usada en teoría de juegos y finanzas. Así, la respuesta a tales preguntas no es sólo un interesante ejercicio matemático, sino que encuentra mucha utilidad en aplicaciones. Es este artículo también un homenaje a Robert Aumann, el matemático israelí ganador del premio Nobel de Economía de 2005 quien desarrolló gran parte de la teoría aquí expuesta.

1. Introducción

El objetivo de nuestro artículo es construir un modelo matemático para el conocimiento y discutir algunas aplicaciones, principalmente en el campo de la economía matemática y teoría de juegos.

Una advertencia importante es que no tocaremos los problemas filosóficos de la epistemología o de la filosofía de la ciencia ni tampoco los problemas prácticos de la inteligencia artificial.

Nuestra cuestión principal es la siguiente: ¿cómo es posible modelar matemáticamente el conocimiento? Para contestar a tal pregunta, utilizaremos principalmente conceptos de la lógica y de la teoría de sistemas en un nivel elemental. Esperamos que cualquier lector con un conocimiento de las matemáticas a nivel elemental sea capaz de entender el contenido presentado aquí. Solo en la sección **??** utilizaremos conceptos específicos de la teoría de juegos, pero a un nivel muy básico. Además, todo lo que necesitaremos sobre el tema lo definiremos en esa misma sección. Los usos de la teoría presentada aquí los discutiremos brevemente en la sección **??**, incluyendo las finanzas, la teoría de juegos, el equilibrio general, etc.

El artículo se organiza en la forma siguiente: en la sección **??**, presentamos el modelo básico del conocimiento con oraciones, conocido como *modelo sintáctico*. La sección **??** construye entonces el *modelo semántico* usado en las finanzas y la teoría de juegos. La sección **??** introduce el *operador del conocimiento* en el modelo semántico, para establecer la relación entre las de los metodologías. La sección **??** define y discute ejemplos del conocimiento común.

Este artículo está parcialmente basado en los trabajos del ganador del prémio Nobel de Economía, Robert Aumann (1999a, 1999b).

2. Modelo sintáctico

2.1. **Sentencias.** Los objetos básicos que podrían ser conocidos serán aserciones en el mundo. Por ejemplo:

Colombia es un país. (a)

Lloverá mañana. (b)

La bolsa subió 1% ayer. (c)

Tales aserciones (que pueden ser verdades o no) serán llamadas las *oraciones atómicas* y denotadas con letras minúsculas. El sistema simple de las oraciones atómicas (o de hechos) será denotado con S.

De las oraciones atómicas, podemos conseguir *oraciones* o *fórmulas* usando las reglas siguientes:

$$\text{Si } a \in S, \text{ entonces } a \text{ es una oración.} \tag{R1}$$

$$\text{Si } a \text{ es una oración, entonces } \sim a \text{ es una oración.} \tag{R2}$$

$$\text{Si } a \text{ y } b \text{ son oraciones, } (a \wedge b) \text{ es una oración.} \tag{R3}$$

Con estas reglas, tenemos exactamente el concepto de oraciones como en la lógica. Dado que nuestro objetivo es estudiar la noción de conocimiento, introducimos otra regla más para formar oraciones. Sean $I = \{1, ..., n\}$ un conjunto de individuos y para cada $i \in I$, un conjunto de atributos subjetivos, A_i. Por ejemplo, A_i podría ser el conjunto $\{k_i, g_i, c_i, p_i^{\alpha}\}$, donde sus elementos tendrían los siguientes significados: $k_i\,(x)$ significa que el individuo i sabe que la oración

x es verdadera; $g_i(x)$ significa que a i gusta que x sea verdad; $c_i(x)$ significa que i cree que x es verdad; $p_i^\alpha(x)$ significa que i atribuye probabilidad $\alpha \in [0,1]$ a que x sea verdad, etc. Así, introducimos la regla siguiente: si $i \in I$ y $f_i \in A_i$:

$$\text{Si } a \text{ es una oración, entonces } f_i(a) \text{ es una oración.} \tag{R4}$$

Estamos ahora en posición de definir lo que entenderemos por oraciones:

Definición 2.1. Dados S, I y $A = \{A_i\}_{i \in I}$ como arriba, una secuencia finita de símbolos es una *oración* (o *fórmula*) si se obtiene por el uso repetido de las reglas R1, R2, R3 y R4. Una oración se llama *objetiva* si se obtiene de R1, R2 y R3.

Por ejemplo, si $S = \{a, b, c\}$, $I = \{1\}$ y $A_1 = \{k_1\}$ entonces $\sim (a \wedge \sim (b \wedge c))$ y $k_1(a \wedge c)$ son oraciones, mientras que $((a \wedge \sim))\, bc$, $d \wedge b$ y $k_1(\sim aa)$ no lo son.

Para la conveniencia de la notación, escribimos $(a \vee b)$ como abreviatura de $\sim (\sim a \wedge \sim b)$ y $(a \Rightarrow b)$ para $\sim (a \wedge \sim b)$. Semejantemente, definimos $(a \Leftrightarrow b)$. Vamos además a estipular que podemos omitir o agregar paréntesis según reglas usuales, es decir, para $(a \wedge b)$ escribiremos $a \wedge b$.

Notación 2.2. Dados S, I y $A = \{A_i\}_{i \in I}$ como arriba, el conjunto de todas las oraciones se llama la *sintaxis* y se denota con $\mathcal{S}(S, I, A)$, o simplemente con $\mathcal{S}$ si no hay posibilidad de confusión. Al conjunto de las oraciones objetivas lo denotaremos con $\mathcal{S}^O(S)$ o simplemente con $\mathcal{S}^O$.

En seguida definiremos las reglas de la deducción lógica.

2.2. **Reglas de la deducción lógica.** Sea $F \subset S$ un conjunto de oraciones atómicas. Los elementos de F deberán entenderse como las *oraciones atómicas verdaderas*. Construyamos, de allí, la lista de las oraciones verdaderas, denotada con $\mathcal{L}(F)$ o simplemente con $\mathcal{L}$.

Definición 2.3. El *cierre* (*lógico*) de $F \subset S$, denotado por $\mathcal{L}(F) \subset \mathcal{S}^O$, es el (único) conjunto que satisface las propiedades siguientes:[1]

$$\text{Si } a \in S, \text{ entonces } a \in \mathcal{L}(\mathcal{F}) \text{si y solamente si } a \in \mathcal{F}. \tag{L1}$$

$$a \text{ y } b \in \mathcal{L}(\mathcal{F}) \text{ si y solamente si } (a \wedge b) \in \mathcal{L}. \tag{L2}$$

$$a \in \mathcal{L}(\mathcal{F}) \text{ si y solamente si } \sim a \notin \mathcal{L}(F). \tag{L3}$$

Como todas las oraciones en $\mathcal{S}^O$ son obtenidas usando un número finito de veces las reglas R1, R2 y R3, para todo $a \in \mathcal{S}^O$ podemos decidir siempre, usando las reglas L1, L2 y L3, si $a \in \mathcal{L}(F)$ o $a \notin \mathcal{L}(F)$ y si una, y solamente una, de estas condiciones es válida. Por eso decíamos que hay un único conjunto que satisface las condiciones L1, L2 y L3. Es decir, $\mathcal{L}(F)$ queda bien definido. Tenemos, por otra parte, que $\mathcal{L}(F)$ es coherente, completo y cerrado, en los sentidos que definimos en seguida:

[1] Observemos que si $\mathcal{L} \subset \mathcal{S}^L$ es el cierre lógico de un conjunto $F \subset S$, entonces $F = \mathcal{L} \cap S$.

Definición 2.4. Una lista de oraciones $\mathcal{L} \subset \mathcal{S}$ se dice *coherente* si $\sim a \in \mathcal{L}$ implica que $a \notin \mathcal{L}$.[2]

Definición 2.5. Una lista de oraciones $\mathcal{L} \subset \mathcal{S}$ se dice *completa* si $a \notin \mathcal{L}$ implica que $\sim a \in \mathcal{L}$.

Definición 2.6. Una lista de oraciones $\mathcal{L} \subset \mathcal{S}$ se dice *cerrada* si $a, (a \Rightarrow b) \in \mathcal{L}$ implica que $b \in \mathcal{L}$, es decir, vale el *modus ponen*.

La definición del cierre lógico nos ahorra el trabajo de definir tabla-verdad y asignaciones, cómo es tradicional en Lógica. Es por eso que L1-L3 están escritos en la forma de si y solamente si, porque la información en las dos direcciones es lo que se obtiene de la tabla-verdad.

Es inmediato ver, de L3, que el cierre lógico $\mathcal{L}(F)$ es coherente y completo. Por otra parte, la propiedad siguiente justifica el nombre del cierre lógico:

Lema 2.1. *El cierre lógico $\mathcal{L}(\mathcal{F})$ es cerrado.*

Demostración. Si $a, (a \Rightarrow b) \in \mathcal{L}$, entonces $a \wedge \sim b \notin \mathcal{L}$, por la definición de $(a \Rightarrow b)$ y por L3. Como $a \in \mathcal{L}$, eso sólo puede ocurrir si $\sim b \notin \mathcal{L}$, por L2. Luego, nuevamente por L3, $b \in \mathcal{L}$. ☑

Ahora bien, es útil entender lo que significa la terminología “modelar” en nuestro contexto. Por eso damos la siguiente definición.

Definición 2.7. Dada una oración $a \in \mathcal{S}^O$, decimos que una *lista* $\mathcal{L} \subset \mathcal{S}$ *modela* a a o que $F = \mathcal{L} \cap \mathcal{S}$ *modela* a a si $a \in \mathcal{L}(F)$. En este caso, escribimos $F \models a$ o $\mathcal{L} \models a$. Una lista $\mathcal{L} \subset \mathcal{S}$ (o $F = \mathcal{L} \cap \mathcal{S}$) *modela un conjunto de oraciones objetivas* C si $\mathcal{L} \models a$, $\forall a \in C$. En este caso escribimos $\mathcal{L} \models C$.

Podemos ahora definir tautologías y contradicciones.

Definición 2.8. Dada S, una *tautología* es una oración objetiva $a \in \mathcal{S}^O(S)$ si para todo $F \subset S$, $a \in \mathcal{L}(F)$. En este caso, escribimos $\models a$.

Definición 2.9. Dada S, una *contradicción* es una oración a tal que $\sim a$ es una tautología. En este caso, escribimos $\perp a$.

O sea, una tautología es una oración que es verdad independientemente de cuál de las oraciones atómicas sea verdadera. Por ejemplo, $\sim a \vee a$ es una tautología, mientras $\sim a \wedge a$ es una contradicción.

[2]Algunos autores (AUMANN, por ejemplo) dicen que una lista de las oraciones $\mathcal{L} \subset \mathcal{S}$ es consistente si $a \in \mathcal{L}$ implica que $\sim a \notin \mathcal{L}$. Esta condición, sin embargo, es solamente la contrapositiva de: $\sim a \in \mathcal{L}$ implica que $a \notin \mathcal{L}$. Es decir una lista es consistente si y solo si es coherente.

2.3. **Verdades objetivas.** Al fijar un sistema de oraciones atómicas $F \subset S$, estamos fijando los hechos del mundo que son verdades: sólo los que están en F. Si $a \in S \setminus F$, a es falso. Sin embargo, las verdades objetivas no incluyen solamente a F, sino también a toda oración que se pueda lógicamente derivar de las oraciones en F. Es decir, tenemos la siguiente:

Definición 2.10. Dada una lista de oraciones atómicas verdaderas $F \subset S$, el conjunto de verdades objetivas es simplemente el *cierre lógico* de F, es decir, $\mathcal{L}(F)$.

Observemos que las oraciones atómicas son verdades si y solamente si están en F. Sin embargo, pueden existir oraciones subjetivas, como por ejemplo, que el individuo i conoce a, $k_i(a)$, tales que su veracidad no es deducible por implicaciones lógicas. Esto nos lleva al concepto de verdades subjetivas.

2.4. **Verdades y verdades subjetivas.** Definiremos ahora cuáles son los conjuntos que es razonable tener como oraciones verdaderas.

Definición 2.11. Un conjunto de oraciones $\mathcal{V} \subset \mathcal{S}$ es un *conjunto de verdades compatibles con* $F \subset S$ (o simplemente $\mathcal{V}$ *es un conjunto de verdades*) si las siguientes condiciones se satisfacen para todo $i \in I$:

$$\forall a \in S,\ a \in \mathcal{V} \quad \text{si y solamente si } a \in F. \tag{V1}$$

$$a \text{ y } b \in \mathcal{V} \quad \text{si y solamente si } (a \wedge b) \in \mathcal{V}. \tag{V2}$$

$$a \in \mathcal{V} \quad \text{si y solamente si } \sim a \notin \mathcal{V}. \tag{V3}$$

En este caso, diremos que $\mathcal{V} \subset \mathcal{S}$ es un *conjunto de verdades compatibles con el cierre de* F, $\mathcal{L}(F)$ (el cual determina y es determinado solamente por F).

Tenemos el resultado siguiente:

Teorema 2.12. *Un conjunto de oraciones $\mathcal{V} \subset \mathcal{S}$ es un conjunto de verdades si y solamente si es coherente, completo, cerrado y contiene a todas las tautologías.*[3]

Demostración. Sea $F = \mathcal{V} \cap S$. Si $\mathcal{V}$ es conjunto de verdades compatible con F, entonces es coherente, completo (por V3) y es cerrado por V2, V3 y por la demostración del Lema **??**. $\mathcal{V}$ contiene todas las tautologías porque contiene a $\mathcal{L}(F)$ (compárese V1, V2 y V3 con L1, L2 y L3).

Asumamos ahora que $\mathcal{V}$ sea coherente, completo, cerrado y que contenga todas las tautologías. Por la definición de F, vale V1. Como $\mathcal{V}$ es coherente y completo, vale V3. Verifiquemos que vale V2. Como $\mathcal{V}$ contiene a todas las tautologías, contiene a las tautologías $((a \wedge b) \Rightarrow a)$ y $((a \wedge b) \Rightarrow b)$. Entonces, si $(a \wedge b) \in \mathcal{V}$, dado que $\mathcal{V}$ es cerrado, $a, b \in \mathcal{V}$. Supongamos ahora que $a, b \in \mathcal{V}$. Si, por absurdo, $(a \wedge b) \notin \mathcal{V}$, entonces $\sim (a \wedge b) \in \mathcal{V}$, porque $\mathcal{V}$ es completo. La

[3]Observemos que las definiciones de coherente, completo y cerrado se aplican a las listas generales de oraciones.

tautología $\sim (a \wedge b) \Rightarrow \sim (a \wedge \sim\sim b)$, que por hipótesis está en $\mathcal{V}$, implica que $\sim (a \wedge \sim\sim b) \in \mathcal{V}$, porque $\mathcal{V}$ es cerrado. Observemos que la condición de que $\mathcal{V}$ sea cerrado es equivalente a la condición de que c y $\sim (c \wedge \sim d) \in \mathcal{V}$ impliquen $d \in \mathcal{V}$. Entonces, de la hipótesis $a \in \mathcal{V}$ se concluye que $\sim b \in \mathcal{V}$. Pero esto contradice la coherencia de $\mathcal{V}$ porque $b \in \mathcal{V}$. Esto prueba que V2 se satisface. ☑

Tenemos la siguiente:

Proposición 2.13. *Todas las consecuencias lógicas de las verdades atómicas, y solamente éstas, son las oraciones objetivas en $\mathcal{V}$, es decir, $\mathcal{L}(\mathcal{V} \cap S) = \mathcal{V} \cap \mathcal{S}^O(S)$.*

Demostración. La inclusión $\mathcal{L}(\mathcal{V} \cap S) \subset \mathcal{V} \cap \mathcal{S}^O(S)$ viene directamente de las condiciones V1, V2 y V3 (generalizaciones de L1, L2, L3). Por otro lado, si $a \in \mathcal{V} \cap \mathcal{S}^O(S) \diagdown \mathcal{L}(\mathcal{V} \cap S)$, entonces $\sim a \in \mathcal{L}(\mathcal{V} \cap S) \subset \mathcal{V} \cap \mathcal{S}^O(S)$, por L3. Esto contradice V2 arriba. ☑

Es útil definir "teoremas" en nuestro contexto. Esperamos que los enunciados demostrados en este artículo no sean confundidos con "teoremas".

Definición 2.14. Una oración $a \in \mathcal{S}$ es un *teorema* si a es una tautología o si $a \in \mathcal{V}$, para todo conjunto de verdades $\mathcal{V} \subset \mathcal{S}$. Si a es un teorema, escribimos $\models a$. El conjunto de teoremas será denotado por $\mathcal{T}$.

Podemos ampliar la definición anterior a la de contradicción:

Definición 2.15. Una *contradicción* es una oración a tal que $\sim a$ es un teorema. En este caso, escribimos $\perp a$.

Tenemos también la siguiente definición:

Definición 2.16. Dado un conjunto de verdades $\mathcal{V} \subset \mathcal{S}$, las *verdades subjetivas* son las oraciones $a \in \mathcal{V}$ que no son teoremas y que contienen un término en A_i para algún $i \in I$. El conjunto de las verdades subjetivas puede denotarse con $\mathcal{V}^S \subset \mathcal{V} \diagdown \mathcal{L}(\mathcal{V} \cap S)$.[4]

2.5. **Reglas del conocimiento.** Además de las reglas especificadas en las definiciones **??** y **??**, podemos establecer algunas reglas para los atributos subjetivos de las oraciones. Por ejemplo, es común especificar que si $k_i \in A_i$, entonces el conjunto de verdades $\mathcal{V}$ tendrá que satisfacer, además de V1, V2 y V3, también las características siguientes:

$$\text{Si } k_i(a) \in \mathcal{V}, \text{entonces } a \in \mathcal{V}. \tag{K1}$$

$$\text{Si } k_i(a \Rightarrow b) \in \mathcal{V}, \text{ entonces } (k_i(a) \Rightarrow k_i(b)) \in \mathcal{V}. \tag{K2}$$

$$\text{Si } k_i(a) \notin \mathcal{V}, \text{ entonces } k_i(\sim k_i(a)) \in \mathcal{V}. \tag{K3}$$

[4]En general, $\mathcal{V}^S \neq \mathcal{V} \diagdown \mathcal{L}(\mathcal{V} \cap S)$ porque puede haber teoremas formados con términos subjetivos. Por ejemplo, basta substituir algún término en una tautología por una fórmula con términos subjetivos.

El ítem K1 especifica que los individuos solo podrán "conocer" oraciones verdaderas. El ítem K2 dice que si un individuo conoce una implicación y conoce la premisa, entonces conoce la conclusión. En otras palabras, los individuos son capaces de llegar a conclusiones lógicas. K3 requiere que el individuo sea consciente de lo que no sabe, es decir, si él no sabe algo, entonces sabe que no lo sabe.[5]

Tenemos la siguiente:

Proposición 2.17. *Si se cumplen K1, K2 y K3, las afirmaciones siguientes son teoremas:*

(i) $k_i(a) \Rightarrow a$.

(ii) $k_i(a \Rightarrow b) \Rightarrow (k_i(a) \Rightarrow k_i(b))$.

(iii) $\sim k_i(a) \Rightarrow k_i(\sim k_i(a))$.

Demostración. (i), (ii) y (iii) son de la forma $f \Rightarrow g$ y corresponden a las afirmaciones K1, K2 y K3, si los escribimos como "si $f \in \mathcal{V}$ entonces $g \in \mathcal{V}$". Como K1, K2 y K3 son válidos para todo $\mathcal{V}$, entonces $\sim f \wedge \sim g$ no puede pertenecer a ningún $\mathcal{V}$. Luego, $\sim(\sim f \wedge \sim g) \in \mathcal{V}$ para todo $\mathcal{V}$, por K3. O sea, las oraciones $(f \Rightarrow g)$ están en todos los $\mathcal{V}$ y, por lo tanto, son teoremas. ☑

3. Modelo semántico

En la literatura de la economía y de la teoría de juegos los estados de la naturaleza y los estados del mundo se toman muchas veces como sinónimos. Sin embargo, algunos autores adoptan la distinción que en seguida haremos aquí.

Definición 3.1. Un *estado de la naturaleza* es cualquier conjunto de oraciones objetivas que sea coherente, completo, cerrado y que contenga a las tautologías. El sistema de los estados de la naturaleza lo denotaremos con $\mathcal{N}$.

Observemos que $\mathcal{N} = \{\mathcal{L}(F) : F \subset S\}$. Como un conjunto cerrado de oraciones objetivas define unívocamente las oraciones atómicas verdaderas y está determinado por éstas, podemos abusar de la terminología para llamar estado de la naturaleza también al conjunto correspondiente de las oraciones atómicas que son válidas en el mismo.

Definición 3.2. Un *estado del mundo* es cualquier conjunto de verdades compatible con algún estado de la naturaleza. Denotaremos al conjunto de los estados del mundo con Ω.

[5]Observemos que K3 no puede ser deducido de K1, por ejemplo. De hecho, aunque la premisa $k_i(a) \notin \mathcal{V}$ implique, por V3, que $\sim k_i(a) \in \mathcal{V}$, K1 no implica la conclusión de K3: $k_i(\sim k_i(a)) \in \mathcal{V}$.

La principal distinción de estos dos conceptos estriba en que los estados de la naturaleza se refieren solamente a los hechos objetivos. Los estados del mundo, por otra parte, toman en cuenta la actitud subjetiva de los individuos con relación a tales hechos.

Un comentario importante ahora es que podemos asociar con cada oración $a \in \mathcal{S}$, un evento $E_a \subset \Omega$, que es el conjunto de todos los estados del mundo en donde la oración a es verdadera, es decir, el conjunto de todos los $\omega \in \Omega$ que contienen a a. Matemáticamente:

$$E_a = \{\omega \in \Omega : a \in \omega\}. \tag{3.1}$$

Hay conjuntos $E \subset \Omega$ que no corresponden a ninguna oración (como nosotros mostraremos luego), pero existe un "isomorfismo" entre oraciones y subconjuntos de Ω que sean de la forma descrita arriba. Para precisar este hecho necesitaremos de algunas otras definiciones. Sea el conjunto de todos los subconjuntos de Ω que corresponden a una oración, es decir,

$$\mathcal{S}(\Omega) = \{E \subset \Omega : E = E_a \text{ para algún } a \in \mathcal{S}\}. \tag{3.2}$$

Recordemos que el sistema de todos los teoremas se denota con $\mathcal{T}$. Definimos la *clase de equivalencia de una oración* a como sigue:

$$[a] = \{b \in \mathcal{S} : (a \Leftrightarrow b) \in \mathcal{T}\}. \tag{3.3}$$

O sea, la clase de equivalencia de una oración a es el conjunto de todas las oraciones equivalentes a a. A continuación, definimos:

$$[\mathcal{S}] = \{[a] : a \in \mathcal{S}\} = \mathcal{S}|\mathcal{T}. \tag{3.4}$$

Denotemos el complemento de un conjunto E por $\sim E$. En particular, escribimos $\mathcal{S}(\Omega) \setminus \varphi([a])$ como $\sim\varphi([a])$. Tenemos entonces el siguiente:

Teorema 3.3. *La aplicación $\varphi : [\mathcal{S}] \to \mathcal{S}(\Omega)$ definida asociando a cada $[a] \in [\mathcal{S}]$ el conjunto E_a para algún $a \in [a]$, está bien definida y tiene las siguientes propiedades:*

(i) *es bijetivo;*

(ii) $\varphi([\sim a]) = \sim\varphi([a])$*;*

(iii) $\varphi([a \wedge b]) = \varphi([a]) \cap \varphi([b])$.

Demostración. En el apéndice. ☑

Quizás el lector se sienta más cómodo en la lectura de las afirmaciones (ii) y (iii) del anterior teorema usando sus correspondientes más directos: (ii) $E_{\sim a} = \sim E_a$; (iii) $E_{a\wedge b} = E_a \cap E_b$. Éstas se prueban fácilmente, como se muestra en el lema siguiente:

Lema 3.1. *Si $a, b \in \mathcal{S}$, entonces $E_{\sim a} = \sim E_a$; $E_{a\wedge b} = E_a \cap E_b$ y $E_{a\vee b} = E_a \cup E_b$.*

Demostración. De las características anteriores, tenemos las equivalencias siguientes (que, para no generar confusión con el símbolo $\Leftrightarrow$ de la sintaxis, escribimos como $\rightleftharpoons$ (se lee 'si y solamente si'):[6]

$$\omega \in E_{\sim a} \rightleftharpoons \sim a \in \omega \rightleftharpoons a \notin \omega \rightleftharpoons \omega \notin E_a \rightleftharpoons \omega \in \sim E_a.$$
$$\omega \in E_{a\wedge b} \rightleftharpoons a \wedge b \in \omega \rightleftharpoons a, b \in \omega \rightleftharpoons \omega \in E_a, \omega \in E_b \rightleftharpoons \omega \in E_a \cap E_b.$$

La primera línea demuestra que $E_{\sim a} = \sim E_a$ y la segunda que $E_{a\wedge b} = E_a \cap E_b$. La propiedad $E_{a\vee b} = E_a \cup E_b$ puede ser probada semejantemente o, con las anteriores, usando la definición de $a \vee b$:

$$\begin{aligned} E_{a\vee b} &= E_{\sim(\sim a \wedge \sim b)} = \sim E_{(\sim a\wedge \sim b)} \\ &= \sim (E_{\sim a} \cap E_{\sim b}) = \sim (\sim E_a \cap \sim E_b) = E_a \cup E_b. \quad ☑ \end{aligned}$$

El lector debe haber observado que incluimos en el lema la propiedad $E_{a\vee b} = E_a \cup E_b$, cuya correspondiente no aparece en el Teorema **??**. Sin embargo, tal correspondiente, es decir, $\varphi([a \vee b]) = \varphi([a]) \cup \varphi([b])$ es obviamente verdadera. Preferimos esta economía para guardar un cierto paralelismo con V2 y V3.

Otro resultado interesante es el siguiente:

Lema 3.2. *$E_a \subset E_b$ si y solamente si $a \Rightarrow b$ es un teorema.*

Demostración. Sea $a \Rightarrow b$ un teorema y sea $\omega \in E_a$. Por la definición de E_a, $a \in \omega$. Como ω contiene los teoremas y es cerrado, entonces $b \in \omega$, o sea, $\omega \in E_b$. Supongamos ahora que $E_a \subset E_b$. Queremos probar que $(a \Rightarrow b) \in \omega$ para todo $\omega \in \Omega$. Supongamos lo contrario. Entonces existe ω tal que $(a \wedge \sim b) \in \omega$. Eso significa que $a \in \omega \in E_a$ y que $b \notin \omega$, es decir, $\omega \notin E_b$. Eso contradice que $E_a \subset E_b$. ☑

Un corolario imediato de este lema es que $E_a = E_b$ si y solamente si $a \Leftrightarrow b$ es un teorema. (Otra consecuencia es que φ está bien definida en el Teorema **??**.)

Observemos que hemos demostrado la existencia del "isomorfismo" entre $[\mathcal{S}]$ y $\mathcal{S}(\Omega) = \{E \subset \Omega : \exists a \in \mathcal{S} \text{ tal que } E = E_a\}$. Como dijimos previamente, hay subconjuntos de Ω que no son de la forma E_a, es decir, no están en $\mathcal{S}(\Omega)$. Eso es verdad aún restringiéndonos a las oraciones atómicas, es decir, suponiendo que todos los conjuntos de atributos subjetivos A_i sean vacíos para todo individuo $i \in I$. De hecho, consideremos el siguiente:

Ejemplo 1. Suponga que $A_i = \varnothing$ para todo $i \in I$ y que S (el conjunto de las oraciones atómicas) es infinito: $S = \{a_1, a_2, ...\}$. Sea $E = \cap_{n\in\mathbb{N}} E_{a_n}$. Como $\mathcal{L}(S)$ es un estado del mundo (porque nuestro mundo no tiene partes subjetivas) y está contenido en E, esa intersección no es vacía. Afirmamos que no existe $s \in \mathcal{S}$ tal que $E = E_s$. Como s es una oración, solo puede contener

[6] El lector puede tomar como ejercicio la tarea simple de encontrar la justificación de cada de estas equivalencias.

un número finito de oraciones atómicas. Sea a_n una oración atómica que no esté en s. Entonces las dos posibilidades para a_n (ser verdadera o ser falsa) son compatibles con s. En otras palabras, existen $\omega, \omega' \in E_s = \{\omega \in \Omega : s \in \omega\}$ tales que $a_n \in \omega$ y $\sim a_n \in \omega'$. Luego, $E_s \neq E$, pues $a_n \notin \omega' \in E_s$ mientras $a_n \in \omega$, para todo $\omega \in E$.

El ejemplo anterior sugiere la pregunta de si realmente precisamos de la hipótesis de que S sea infinito. Para el caso de oraciones objetivas, la respuesta es negativa, como lo muestra el:

Lema 3.3. *Suponga que $A_i = \varnothing$ para todo $i \in I$ y S es finito. Entonces $\mathcal{S}(\Omega)$ comprende todos los subconjuntos de Ω.*

Demostración. Suponga que S tiene n elementos. Entonces hay a lo sumo 2^n diferentes subconjuntos $C \subset S$. Cada un de ellos define un estado del mundo (objetivo): $\omega = \mathcal{L}(C)$. Como los estados del mundo son los conjuntos coherentes, completos y cerrados que contienen a las tautologías, no hay ningún otro estado que no tenga esa forma. O sea, Ω tiene 2^n elementos y, por lo tanto, todos los subconjuntos de Ω pueden escribirse como unión finita de los conjuntos unitarios de los estados del mundo. Cada subconjunto $C \subset S$ determina una conjunción finita de fórmulas válidas y tales fórmulas determinan únivocamente un estado ω. Así, por el Lema **??**, todo subconjunto de Ω está asociado a una fórmula en $\mathcal{S}$. ☑

Sin embargo, cuando hay componentes subjetivos en el mundo, $\mathcal{S}(\Omega)$ no abarca todos los subconjuntos de Ω, aun en el caso en que S sea finito. La razón es que existen infinitas oraciones (construidas con las propiedades subjetivas) que pueden ser verdaderas o falsas y, así podemos usar el argumento del ejemplo.

Vamos ahora hacer una observación sobre $\mathcal{S}(\Omega)$ qué podría ser útil si quisiéramos definir probabilidades sobre Ω (lo que no haremos aquí). Para empezar, recordemos las siguientes definiciones:

Definición 3.4. Un *álgebra* $\mathcal{F}$ es una colección de subconjuntos de Ω que cumple:

(i) $\Omega \in \mathcal{F}$;

(ii) Si $E \in \mathcal{F}$ entonces $\sim E \in \mathcal{F}$;

(iii) Si $E_n \in \mathcal{F}$, $n \in N \subseteq \mathbb{N}$, con N finito, entonces $\cup_{n \in N} E_n \in \mathcal{F}$.

Tenemos el resultado siguiente:

Lema 3.4. $\mathcal{S}(\Omega)$ *es una álgebra.*

Demostración. Si a es un teorema, $a \in \omega$ para todo $\omega \in \Omega$, es decir, $E_a = \Omega$. Luego, $\Omega \in \mathcal{S}(\Omega)$. Si $E \in \mathcal{S}(\Omega)$, existe $a \in \mathcal{S}$ tal que $E = E_a$. Por el Lema **??**, $\sim E = \sim E_a = E_{\sim a} \in \mathcal{S}(\Omega)$. Si $E_{a_n} \in \mathcal{S}(\Omega)$, para $a_1, ..., a_n \in \mathcal{S}$, entonces

$a_1 \vee ... \vee a_n \in \mathcal{S}$. De nuevo por el Lema **??**, tenemos $\cup_{n=1}^{m} E_{a_n} = E_{a_1 \vee ... \vee a_n} \in \mathcal{S}(\Omega)$. ☑

El interés de este resultado proviene del hecho de que basta para definir una medida de probabilidad sobre un álgebra, de manera que ella se extienda de manera única a la sigma-álgebra generada (para estos conceptos de la teoría de la medida, vea libro especializado en el tema). Para comodidad de la lectura y para comparar con el concepto anterior, definimos:

Definición 3.5. Una *sigma-álgebra* $\mathcal{F}$ es una colección de subconjuntos de Ω que cumple:

(i) $\Omega \in \mathcal{F}$;

(ii) Si $E \in \mathcal{F}$ entonces $\sim E \in \mathcal{F}$;

(iii) Si $E_n \in \mathcal{F}$, $n \in N \subseteq \mathbb{N}$ entonces $\cup_{n \in N} E_n \in \mathcal{F}$.

En la próxima sección utilizaremos el concepto siguiente:

Definición 3.6. Un *álgebra universal* $\mathcal{F}$ es una colección de subconjuntos de Ω que cumple:

(i) $\Omega \in \mathcal{F}$;

(ii) Si $E \in \mathcal{F}$ entonces $\sim E \in \mathcal{F}$;

(iii) Si $E_n \in \mathcal{F}$, $n \in N$, N arbitrario, entonces $\cup_{n \in N} E_n \in \mathcal{F}$.

4. **Estados del mundo y conocimiento**

Los estados del mundo determinan lo que cada individuo conoce:

Notación 4.1. Dado un estado del mundo $\omega \in \Omega$, el conjunto de todas las oraciones que el individuo i conoce, es decir, las fórmulas que empiezan con k_i, lo denotaremos con $\kappa_i(\omega)$.

Una noción importante es el *conjunto de todos los estados del mundo que son indistinguibles para el jugador i* (desde el punto del conocimiento). Este conjunto de estados del mundo es tal que las fórmulas que i conoce son exactamente las mismas. Esta noción se define por el conjunto:

$$I_i(\omega) = \{\omega' \in \Omega : \kappa_i(\omega') = \kappa_i(\omega)\}. \tag{4.1}$$

Para entender el significado de $I_i(\omega)$, pensemos que el estado del mundo ω describe exactamente lo que es verdad. De todas las verdades que valen en el estado ω, el individuo i conoce apenas una parte. Así, algunas cosas que él no sabe pueden ser verdaderas o falsas. O sea, hay otros estados posibles del mundo donde el individuo sabe exactamente lo mismo. A partir de lo que sabe, el individuo es incapaz de distinguir el estado ω de esos otros estados (que están, por definición, en $I_i(\omega)$). Obviamente, tenemos que $\omega \in I_i(\omega)$.

Los conjuntos $I_i(\omega)$ definen una partición de Ω (una partición de un conjunto es una colección de subconjuntos disjuntos entre sí cuya unión es exactamente el conjunto original). De la definición (**??**), es fácil ver que dados los estados ω y ω', los conjuntos $I_i(\omega)$ y $I_i(\omega')$ o son disjuntos o son iguales. Además, todo $\omega \in \Omega$ está en por lo menos un $I_i(\omega)$, probando la afirmación de que $\mathcal{P}_i = \{I_i(\omega) : \omega \in \Omega\}$ es una partición de Ω. Estas dos observaciones inspiran la siguiente definición:

Definición 4.2. Una *función de información* es una función $I : \Omega \to \wp(\Omega)$ que cumple: (i) $\omega \in I(\omega)$ y (ii) para $\omega, \omega' \in \Omega$ arbitrarios, vale una de las de los condiciones: $I_i(\omega) \cap I_i(\omega') = \varnothing$ o $I_i(\omega) = I_i(\omega')$.

Una *partición de información* será simplemente la partición generada por una función de información.

Sea $\mathfrak{K}_i$ la menor álgebra universal (es decir, la que tenga menos conjuntos) que contiene a $\mathcal{P}_i$. (Puede probarse que existe.) Sea $\wp(\Omega)$ el conjunto de las partes de Ω, es decir, el conjunto de todos los subconjuntos de Ω. Definamos el operador $K_i : \wp(\Omega) \to \wp(\Omega)$ de la siguiente manera:

$$K_i(E) = \{\omega \in \Omega : I(\omega) \subset E\}. \tag{4.2}$$

Es fácil ver que $K_i(E)$ es la unión de todos los eventos de la partición que están contenidos en E. $K_i(E)$ puede aún caracterizarse como el mayor elemento de $\mathfrak{K}_i$ que está contenido en E. Vemos, pues, que $K_i(E) \in \mathfrak{K}_i$. Por otro lado, si $E \in \mathfrak{K}_i$, entonces para cada $\omega \in E$, $I(\omega) \subset E$. Luego, $E = K_i(E)$. Así, podemos escribir $\mathfrak{K}_i = K_i(\wp(\Omega))$.

Tenemos la siguiente definición:

Definición 4.3. Un *operador de conocimiento* es una función

$$K : \wp(\Omega) \to \wp(\Omega)$$

que satisface las siguientes propiedades:

(i) $K(E) \subset E$;

(ii) $E \subset F \Rightarrow K(E) \subset K(F)$;

(iii) $\sim K(E) = K(\sim K(E))$. [7]

Las propiedades que definen un operador de conocimiento pueden ser leídas de la siguiente forma: (i) para que se sepa algo, es necesario que ese algo sea verdad; (ii) si algunos hechos E implican los hechos F, entonces si se sabe E, se sabe F; (iii) si no se sabe algo, entonces se sabe que no se sabe. La condición (iii) podría ser llamada el *axioma socrático*, en homenaje al filósofo griego a quien se atribuye la famosa frase: “sé que nada sé”.

Como el lector ya debe haber imaginado, tenemos el siguiente:

Lema 4.1. *El operador $K_i : \wp(\Omega) \to \wp(\Omega)$ definido por (**??**) es un operador de conocimiento.*

[7] Podríamos haber pedido la condición más débil $\sim K(E) \subset K(\sim K(E))$. Sin embargo, si el ítem (i) vale, estas dos condiciones son equivalentes.

Demostración. Los ítemes (i) y (ii) de la definición de operador de conocimiento siguen directamente de (**??**), una vez que $\omega \in I_i(\omega)$. Para probar (iii), es suficiente establecer $\sim K_i(E) \subset K_i(\sim K_i(E))$, dado que la otra inclusión viene de (i). Sea $\omega \in \sim K_i(E)$, es decir, $\omega \notin K_i(E)$. Como $\omega \in I_i(\omega)$, entonces $I(\omega) \cap K_i(E) = \varnothing$, dado que el conjunto $K_i(E)$ está formado por conjuntos de la forma $I(\omega)$ y éstos son disjuntos o iguales. Entonces, $I(\omega) \subset \sim K_i(E)$. Eso implica que $\omega \in K_i(\sim K_i(E))$. ☑

El ítem (i) del Lema abajo significa que los teoremas son conocidos. El ítem (ii) significa que se conoce lo que se conoce.

Lema 4.2. *Si $K : \wp(\Omega) \to \wp(\Omega)$ es un operador de conocimiento, entonces vale el siguiente:*

(i) $K(\Omega) = \Omega$;

(ii) $K(E) = K(K(E))$.

Demostración. (i) Por el ítem (i) de la definición **??**, tenemos que $K(\varnothing) = \varnothing$ Entonces, $\Omega \supset K(\Omega) = K(\sim \varnothing) = K(\sim K(\varnothing))$. Por el ítem (iii) de la definición **??**, este último conjunto es lo mismo que $\sim K(\varnothing) = \sim \varnothing = \Omega$. Esto demuestra (i).

(ii) Haremos repetido uso del ítem (iii) de la definición **??**. Tenemos:

$$\sim K(E) = K(\sim K(E)) \ ,$$

o sea, $K(E) = \sim K(\sim K(E))$. De ahí, $K(K(E)) = K(\sim K(\sim K(E)))$ $= \sim K(\sim K(E)) = K(K(E))$. ☑

5. Conocimiento común

El concepto de conocimiento común es muy importante en teoría de juegos e implica algunas consecuencias interesantes, como mencionamos a continuación. Un evento $E \subset \Omega$ es de conocimiento común si todos los individuos saben que E es verdad, saben que todos saben esto, saben que todos saben que todos saben, etc., al infinito. Para definir formalmente tales eventos, vamos definir operadores de conocimiento mutuo de orden $m, K^m : \wp(\Omega) \to \wp(\Omega)$ recursivamente. El operador de orden 1 está definido por:

$$K^1(E) \equiv \bigcap_{i \in I} K_i(E) \, , \tag{5.1}$$

o sea, $K^1(E)$ representa el evento en donde todos los individuos saben que E es verdad, es decir, el conjunto de estados del mundo $\omega \in \Omega$ donde es mutuamente conocido que E es verdad. Ahora, definimos $K^m(E)$ recursivamente:

$$K^{m+1}(E) \equiv K^1(K^m(E)) \, . \tag{5.2}$$

Si $K^m(E) = E$ decimos que E es *mutuamente conocido hasta el orden m* o que *tiene nivel m de conocimiento común.* Finalmente, definimos el operador de conocimiento común

$$K^{\infty}(E) \equiv \bigcap_{m\in\mathbb{N}} K^{m}(E). \tag{5.3}$$

Hay otra construcción para el concepto de conocimiento común que quizá sea más intuitiva. Ella se basa en la noción de las particiones del conocimiento, mencionada previamente. En hecho, suponga que $\mathcal{P}_i$ sea la partición de conocimiento para el individuo $i \in I$. (En particular, definimos el operador de conocimiento K_i a partir de la partición, estableciendo que $K_i(E)$ es la unión de todos los elementos de la partición que están contenidos en E.[8])

Para llegar a la definición de *partición de conocimiento común*, necesitamos antes la noción de *refinamiento de una partición*. Decimos que una partición $\mathcal{P}'$ refina la partición $\mathcal{P}$ si para cada elemento $E \in \mathcal{P}'$, existe un $F \in \mathcal{P}$ tal que $E \subset F$. El concepto se puede entender fácilmente si pensamos que $\mathcal{P}'$ es obtenido refinando, es decir, haciendo una subpartición de cada elemento de $\mathcal{P}$.

Ahora definimos la *partición de conocimiento común* $\mathcal{P}$ como la partición más fina $\mathcal{P}_i$ que es un refinamiento de $\mathcal{P}$ para todo $i \in I$. Se puede mostrar que tal partición siempre existe.

Algunos ejemplos tornarán más claro el concepto. Supongamos que $\Omega = \{\omega_1, \omega_2, ..., \omega_6\}$ y consideremos $I = \{1, 2\}$ con las correspondientes particiones de conocimiento:

$$\begin{aligned} \mathcal{P}_1 &= \{\{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}, \{\omega_5\}, \{\omega_6\}\}; \\ \mathcal{P}_2 &= \{\{\omega_1, \omega_2, \omega_3\}, \{\omega_4\}, \{\omega_5, \omega_6\}\}. \end{aligned}$$

Observemos que $\mathcal{P}_1$ y $\mathcal{P}_2$ son refinamientos, respectivamente, de las particiones

$$\begin{aligned} \hat{\mathcal{P}}_1 &= \{\{\omega_1, \omega_2, \omega_6\}, \{\omega_3, \omega_4, \omega_5\}\}; \\ \hat{\mathcal{P}}_2 &= \{\{\omega_1, \omega_2, \omega_3\}, \{\omega_4, \omega_5, \omega_6\}\}. \end{aligned}$$

Sin embargo, $\mathcal{P}_1$ no es refinamiento de $\hat{\mathcal{P}}_2$ (por que $\{\omega_3, \omega_4\} \in \mathcal{P}_1$ no está contenido en ningún elemento de $\hat{\mathcal{P}}_2$) y $\mathcal{P}_2$ no es refinamiento de $\hat{\mathcal{P}}_1$ (porque $\{\omega_5, \omega_6\}$ no está contenido en ningún elemento de $\hat{\mathcal{P}}_1$). La partición de conocimiento común $\mathcal{P}$ (a veces también escrita como $\mathcal{P} = \mathcal{P}_1 \wedge \mathcal{P}_2$) es simplemente:[9]

$$\mathcal{P} = \{\{\omega_1, \omega_2, \omega_3, \omega_4\}, \{\omega_5, \omega_6\}\}.$$

Es útil hacer la interpretación de ese ejemplo. Supongamos que ω_1 es el verdadero estado del mundo. Entonces el individuo 1 sabe que $\{\omega_1, \omega_2\}$ es verdad (pero no sabe si de hecho el estado correcto es ω_1 o ω_2). El individuo 2,

[8]También el ejercicio contrario puede ser hecho: dado al operador K_i, los elementos de la partición se definen como los sistemas del tipo $\sim K_i(\sim \{\omega\})$. La verificación de que tales conjuntos forman una partición puede verse en AUMANN (1999a).

[9]Quizá el lector quiera observar que $\hat{\mathcal{P}} = \hat{\mathcal{P}}_1 \wedge \hat{\mathcal{P}}_2 = \{\{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}\}$.

por otra parte, sólo sabe $\{\omega_1, \omega_2, \omega_3\}$. Ambos saben, por lo tanto, que ω_4 no ocurre. Sin embargo, el individuo 1 sabe que el individuo 2 considera posible a ω_3. Entonces, el individuo 2 considera posible que el individuo 1 "sepa" que el estado correcto está en $\{\omega_3, \omega_4\}$. Entonces el individuo 2 piensa que el individuo 1 puede pensar que el individuo 2 "sabe" $\{\omega_4\}$. Felizmente, las cosas paran aquí, porque si el individuo 2 "sabe" $\{\omega_4\}$, entonces 1 considera posible $\{\omega_3, \omega_4\}$. Así, dado ω_1, lo que es de conocimiento común es $\{\omega_1, \omega_2, \omega_3, \omega_4\}$ (aunque $\{\omega_1, \omega_2, \omega_3\}$ sea de conocimiento mutuo). Observemos además que el evento $\{\omega_5, \omega_6\}$ es de conocimiento común mientras el evento $\{\omega_2, \omega_5\}$ no lo es. El menor evento de conocimiento común que contiene a $\{\omega_2, \omega_5\}$ es simplemente Ω, es decir, $\{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$.

Una cuestión interesante es que las dos construcciones llevan a conceptos equivalentes. Para demostrar esto, vamos a definir al operador K a partir de la partición de conocimiento común y probar que, de hecho, es igual al operador K^∞ definido anteriormente. Sea $\mathcal{P} = \wedge_{i \in I} \mathcal{P}_i$ la partición de conocimiento común y sea $\mathcal{P}(\omega)$ el elemento de la partición $\mathcal{P}$ que contiene $\omega \in \Omega$. Definamos:

$$K(E) = \bigcup_{\omega : \mathcal{P}(\omega) \subset E} \mathcal{P}(\omega).$$

Esta definición está de acuerdo con la noción dada anteriormente para el operador K_i, cuando dijimos que $K_i(E)$ era la unión de todos los elementos de la partición $\mathcal{P}_i$ que están contenidos en E. Tenemos la siguiente:

Proposición 5.1. *$K(E) = K^\infty(E)$ para todo $E \subset \Omega$.*

Demostración. Inicialmente probemos que $K(E) \subset K^\infty(E)$. Es suficiente verificar que si $\mathcal{P}(\omega) \subset E$ entonces $\mathcal{P}(\omega) \subset K^\infty(E)$. De hecho, como $\mathcal{P}_i(\omega) \subset \mathcal{P}(\omega)$ (por definición) y $K_i(E)$ es la unión de todos los conjuntos $\mathcal{P}_i(\omega) \subset E$, entonces $\mathcal{P}(\omega) \subset K_i(E)$ para todo i. Luego, $\mathcal{P}(\omega)$ está en todas las intersecciones que aparecen en la definición de K^∞, o sea $\mathcal{P}(\omega) \subset K^\infty(E)$. Para probar la otra inclusión, basta probar que si $\omega \in K^\infty(E)$, entonces $\mathcal{P}(\omega) \subset E$. Si eso no fuera verdad, habría algún i tal que $\mathcal{P}_i(\omega) \nsubseteq E$. Pero entonces, $\omega \notin K_i(E)$, lo que haría imposible que $\omega \in K^\infty(E)$. ☑

El lector interesado puede hallar otra axiomatización de operadores de conocimiento común (equivalente a esta) en Milgrom (1981).

5.1. **Ejemplo de conocimiento común.** En esta subsección, ilustraremos con un ejemplo que el conocimiento común tiene implicaciones curiosas. Éste es una variación de ejemplos famosos. (Véase Samuelson 2004, por ejemplo).

Ejemplo 2. Los lógicos incompetentes. En un departamento de investigación matemática, había un grupo de 3 lógicos que se consideraban unos a otros incompetentes. Cuando dos de ellos se reunían por separado, compartían su impresión sobre el la incompetencia del colega ausente. Por supuesto, ninguno de ellos escuchaba el comentario sobre su propia incompetencia.

En el departamento, había la tradición de que ningún investigador podría quedarse en el instituto cuando supiera que sus colegas lo consideraban incompetente.[10] Si el investigador llegara a conocer que se le consideraba incompetente, la tradición determinaba que presentaría su dimisión en la reunión de la comisión de investigadores, que se celebraba todos los meses. Como nadie hablaba directamente de la (in)competencia del otro, ninguno de ellos sabía que era considerado incompetente, y por lo tanto, no hacía la solicitud de dimisión. Así, habían hecho sus investigaciones por muchos años (aun siendo incompetentes).

Pero un día, el grupo recibió la visita de un investigador extranjero. El visitante habló individualmente con todos los tres lógicos y cada un de ellos le confió su opinión sobre los otros dos. Por supuesto el visitante se sintió incómodo con tal situación y, antes de partir, dijo en la reunión de la comisión de investigadores que había por lo menos un lógico incompetente en el departamento.

Eso no fue ninguna novedad para los lógicos, porque cada uno de ellos sabía que consideraba a los otros dos incompetentes. De modo que en esa reunión nada sucedió y el investigador extranjero se marchó. En la segunda reunión de la comisión, no hubo tampoco ningún pedido de dimisión y, entonces, los tres lógicos salieron preocupados porque, aunque eran incompetentes, por lo menos eran lógicos. En el tercer mes, todos presentaron su dimisión. El director del departamento los felicitó a los tres por sus decisiones, añadiendo que, después de todo, habían probado en esa ocasión que no eran tan incompetentes.[11]

¿Es capaz el lector de explicar lo que pasó en esa historia?, ¿Por qué ellos pidieron dimisión en la tercera reunión (y no antes o después)?

La explicación es como sigue. En la primera reunión, después del anuncio de que había un lógico incompetente, esta información pasó a ser de conocimiento común. Notemos que esto no era de conocimiento común, porque cada uno de ellos podría pensar que no era incompetente y, por lo tanto, no podría ser de conocimiento común la existencia de un incompetente. (Formalizaremos

[10]Para simplificar las cosas, vamos a hacer la hipótesis siguiente: un investigador es considerado incompetente si y solamente si es incompetente. Por otra parte, vamos a asumir que si un investigador es incompetente, todos comentan esto en su ausencia, pero nunca en su presencia.

[11]Una variación de esta historia que puede ser útil para las ilustraciones en clase es la siguiente: se ponen tres alumnos en un círculo. Se coloca sobre sus cabezas un sombrero con uno de entre dos símbolos posibles, por ejemplo, cuadrado o círculo. Cada de los alumnos puede ver el símbolo arriba de la cabeza de sus colegas, pero no el suyo. Se ponen los mismos símbolos, digamos el cuadrado, en la cabeza de todos ellos, de forma que cada uno ve dos cuadrados. Entonces el profesor dice: hay un alumno con símbolo cuadrado. El profesor pide: ¿alguien sabe el símbolo que tiene en su cabeza? Nadie debe contestar. El profesor repite la pregunta, siguiendo silencio otra vez. En la tercera vez que repite la pregunta, todos los tres deben contestar: cuadrado. Las experiencias hechas en sala de clase demuestran que los alumnos en general no son tan competentes como los lógicos de la historia.

mejor este punto abajo.) Además, la regla de pedir la dimisión tiene el siguiente efecto: si un investigador no sabe de ningún otro incompetente, entonces debe concluir inmediatamente que él mismo es incompetente. Así, si hubiera apenas un incompetente, este pediría su dimisión inmediatamente, al fin de la primera reunión. En el inicio de la segunda reunión, todos saben que hay por lo menos dos incompetentes. Si hubiera exactamente dos incompetentes, entonces cada uno de ellos sabría de apenas un otro. Entonces, cada uno de esos dos sabría que es incompetente y pediría su dimisión en la segunda reunión. No habiendo pedidos de dimisión en la segunda reunión, se vuelve conocimiento común que hay por lo menos tres incompetentes, lo que significa que todos son incompetentes. Entonces, todos piden su dimisión en la tercera reunión.

En la próxima subsección hacemos el modelamento matemático de esta historia.

5.2. **Modelamiento matemático de la historia.** Podemos modelar matemáticamente esa historia a través de la definición adecuada de los estados del mundo.

Cada uno de los lógicos puede ser incompetente (representado por 0) o competente (representado por 1). (Recordemos que identificamos las nociones de ser incompetente y de ser considerado competente, es decir, un investigador es incompetente si y solamente si es considerado incompetente — y, en este caso, es considerado incompetente por cada uno de sus pares.) Entonces tenemos que $\Omega = \{0,1\}^3$. Así, el elemento $\omega = (1,0,1)$ significa que apenas el segundo investigador es incompetente.

Las particiones de cada uno de los investigadores están dadas por:

$$\begin{aligned}
\mathcal{P}_1 &= \{\{(0,i,j),(1,i,j)\} : i,j \in \{0,1\}\};\\
\mathcal{P}_2 &= \{\{(i,0,j),(i,1,j)\} : i,j \in \{0,1\}\};\\
\mathcal{P}_3 &= \{\{(i,j,0),(i,j,1)\} : i,j \in \{0,1\}\}.
\end{aligned}$$

Naturalmente, la partición de conocimiento común es $\mathcal{P} = \{\Omega\}$. La afirmación del investigador extranjero tiene el efecto de modificar la partición de conocimiento común. De hecho, ella crea la siguiente partición:

$$\mathcal{P}^0 = \{\{(1,1,1)\},\{\Omega\setminus\{(1,1,1)\}\}\}.$$

La afirmación del investigador visitante implica también que el estado $(1,1,1)$ no es verdadero, es decir, que no es posible que todos sean competentes. Esto era de conocimiento mutuo (todos lo sabian), pero no era de conocimiento común.

A continuación, entra en acción la regla de pedir dimisión al comprobarse que se es incompetente. Al ser de conocimiento común que hay por lo menos un incompetente, de pronto, este incompetente debiera acusarse. Si al final de la primera reunión nadie lo hace, es que el estado verdadero no puede ser con apenas un incompetente. Así, la partición de conocimiento común modifícase una vez más. Para denotarla, vamos escribir como $\{\Omega\diagdown\text{anteriores}\}$ al conjunto obtenido de Ω menos los otros conjuntos que aparecen en la partición. Así, la partición al final de la primera reunión será:

$$\mathcal{P}^1 = \{\{(1,1,1)\},\{(1,1,0)\},\{(1,0,1)\},\{(0,1,1)\},\{\Omega\diagdown\text{anteriores}\}\}.$$

Al final de la segunda reunión, apréndese que no puede haber solo dos incompetentes. La partición de conocimiento común, por lo tanto, se torna la más fina posible:

$$\mathcal{P}^2 = \{\{(i,j,k)\} : i,j,k \in \{0,1\}\}.$$

El estado verdadero, $(0,0,0)$, se vuelve de conocimiento común y, por lo tanto, los lógicos son obligados a pedir la dimisión.

6. Juego de los mensajes electrónicos

Presentaremos ahora el ejemplo paradójico propuesto por RUBINSTEIN (1989), que es muy útil para ver que aun órdenes muy grandes de conocimiento mutuo no es la misma cosa que conocimiento común.[12]

El juego que vamos a formalizar puede modelar la siguiente situación: dos divisiones de un ejército rodean al ejército enemigo, por lados opuestos. Si las dos divisiones atacan en conjunto, lograrán derrotar al enemigo. Si atacan por separado, sin embargo, las dos serán derrotadas. Los generales de las divisiones sólo necesitan coordinarse para atacar en conjunto. Ellos tienen dos opciones: o bien atacan durante el día o bien durante la noche. Por una cierta razón militar, si hay lluvia, es mejor atacar durante el día, mientras que si no hay lluvia, es mejor atacar por la noche. Uno de los generales recibe el pronóstico del tiempo. Puede comunicarse con el otro general, su colega, a través del equipo de transmisión de mensajes electrónicos, pero hay una probabilidad positiva de que el mensaje no llegue al otro general. Para disminuir el problema de la ruptura de comunicación, ellos deciden enviar sucesivos mensajes que digan que recibieron el mensaje anterior. ¿Pero será que van a poder utilizar con ventaja el pronóstico perfecto del tiempo y la eficacia (casi absoluta) de su equipo de comunicación? La sorprendente contestación es que no.

Para modelar formalmente el juego descrito, suponga que hay dos jugadores, 1 y 2. Cada uno tiene que elegir una acción, A o B. Hay dos posibles estados de la naturaleza, a y b. Si el estado de la naturaleza es a, lo mejor es jugar A. Si el estado es b, lo mejor es jugar B. El estado de la naturaleza a ocurre con probabilidad $1-p$ y el estado b ocurre con probabilidad $p \in [0,1]$. Abajo mostramos

[12]En esta sección, utilizamos conceptos de la teoría de los juegos que no son necesarios para la porción restante de este artículo. De cualquier forma, definimos todos los conceptos que necesitaremos.

la matriz de pagos de los jugadores para cada estado y acción posibles.

Estado a	A	B
A	M, M	$0, -L$
B	$-L, 0$	$0, 0$

y

Estado b	A	B
A	$0, 0$	$0, -L$
B	$-L, 0$	M, M

Las aciones en las filas son aciones del jugador 1, mientras las de las columnas son las aciones del jugador 2. Los términos en la matriz representan el pago del jugador 1 y el pago del jugador 2. Por ejemplo, si el estado es a, el jugador 1 juega B y el jugador 2 juega A entonces el primero jugador sufre una pérdida de L (recibe $-L$) y el jugador 2 no recibe nada.

Asumamos que $L > M > 0$ y $p < \frac{1}{2}$, de modo que el estado a es el más probable.

Ocurre que el jugador 1 sabe cuál es el estado correcto (a o b) y dispone de un ordenador para enviar un mensaje electrónico con esa información para el jugador 2. (Ellos no pueden comunicarse directamente, solo mediante mensajes electrónicos.) Existe una probabilidad positiva que los mensajes electrónicos no lleguen a su destinatario. Supongamos que esa probabilidad es $\varepsilon > 0$.

Ellos establecen que el primer mensaje solo será enviado si el estado es b y los dos jugadores programan sus ordenadores para enviar mensajes electrónicos de confirmación de recibimiento automáticamente después de recibir un mensaje. Así, si el jugador 1 aprende que el estado de la naturaleza es b, su ordenador envía un mensaje electrónico para el ordenador del jugador 2. Si este mensaje no se pierde, el ordenador del jugador 2 inmediatamente responde que lo recibió y el ordenador del jugador 1, si recibe la confirmación, envía la confirmación de la confirmación y así en delante. El proceso sigue hasta que un mensaje se pierda, y es entonces cuando la comunicación se detiene. Al final de la comunicación, cada ordenador informa el número de mensajes enviados.

Vamos a modelar la situación. El conjunto de estados va ser descrito como el conjunto:

$$\Omega = \{(a, 0, 0)\} \cup \{(b, i, j) : i, j = 0, 1, 2, ... \text{ e } k = i \text{ o } k = i - 1\},$$

o sea, el conjunto de las triplas (x, i, j) donde x representa el estado de la naturaleza (a o b), i y j representan el número de mensajes que cada ordenador envió.

Si el ordenador del jugador 1 envió n mensajes, el jugador 1 no sabe si fue su enésimo mensaje el que se perdió o si fue el enésimo mensaje del ordenador 2. O sea, él no sabe si el estado correcto es $(b, n, n-1)$ o (b, n, n). Así, su partición de información es:

$$\begin{aligned}\mathcal{P}_1 = &\{\{(a, 0, 0)\}, \{(b, 1, 0), (b, 1, 1)\}, \{(b, 2, 1), (b, 2, 2)\},\\ &\{(b, 3, 2), (b, 3, 3)\}, \ldots\}.\end{aligned}$$

Por otro lado, el jugador 2 tiene como partición de conocimiento a:

$$\mathcal{P}_2 = \{\{(a, 0, 0), (b, 1, 0)\}, \{(b, 1, 1), (b, 2, 1)\}, \{(b, 2, 2), (b, 3, 2)\}, ...\}.$$

Es inmediato ver que la partición de conocimiento común es la trivial, $\mathcal{P} = \{\Omega\}$.[13]

Para describir lo que pasa en este juego, vamos a definir qué son las estrategias de los jugadores. Sea $\Delta = \{(\alpha, 1-\alpha) : \alpha \in [0,1]\}$. Una estrategia para el jugador i es una función $S_i : \mathcal{P}_i \to \Delta$ con el siguiente sentido: si $E \in \mathcal{P}_i$, entonces $S_i(E) = (\alpha, 1-\alpha)$ significa que el jugador i juega a estrategia A con probabilidad α y juega B con probabilidad $1-\alpha$. Si $\alpha = 0$ o $\alpha = 1$, apenas una acción (B o A, respectivamente) es elegida con seguridad.

Un equilibrio de Nash es un par de estrategias (S_1, S_2) tal que S_1 da el pago máximo al jugador 1 si el jugador 2 usa S_2 y S_2 da el pago máximo a jugador 2 si el jugador 1 usa S_1.

Un resultado importante, que será usado en la demostración abajo es que si una acción da un pago esperado estrictamente mayor que otra acción, entonces esta última acción es jugada con probabilidad cero en equilibrio.

Tenemos la siguiente:

Proposición 6.1. *Suponga que (S_1, S_2) es un equilibrio de Nash tal que*

$$S_1(\{(a,0,0)\}) = (1,0),$$

es decir, el jugador 1 juega A si sabe que el estado de la naturaleza es a . Entonces para todo $E \in \mathcal{P}_i$ y todo $i = 1,2$, tenemos $S_i(E) = (1,0)$, es decir, ambos jugadores juegan siempre A con seguridad.

Demostración. Denotemos los conjuntos en la partición del jugador 1 de la siguiente forma: $I_0^1 = \{(a,0,0)\}$, y $I_k^1 = \{(b,k,k-1), (b,k,k)\}$, para $k = 1, 2, ...$ De forma similar, definimos los conjuntos de la partición de 2 como $I_0^2 = \{(a,0,0), (b,1,0)\}$ y $I_k^2 = \{(b,k,k), (b,k+1,k)\}$. Inicialmente verifiquemos que $S_2\left(I_0^2\right) = (1,0)$. Como el jugador 2 no sabe en que estado está ($(a,0,0)$ o $(b,1,0)$), es decir, él no sabe si la razón de no recibir un mensaje fue porque no fue enviado o porque se extravió, él atribuye las siguientes probabilidades: $(a,0,0)$ ocurre con probabilidad $(1-p)$ y $(b,1,0)$ ocurre con probabilidad $p\varepsilon$. Luego, sin importar cuál sea $S_1\left(I_1^1\right)$, si se juega A, el jugador 2 recibirá el pago $\Pi_0^2(A)$ que cumple:

$$\Pi_0^2(A) \geq \frac{(1-p)}{(1-p)+p\varepsilon} M + \frac{p\varepsilon}{(1-p)+p\varepsilon} \cdot 0.$$

(Lo peor que puede ocurrir es que $S_1\left(I_1^1\right) = (0,1)$ que da al jugador 2 exactamente el pago a la derecha.) Si juega B, ganará $\Pi_0^2(B)$, que cumple:

$$\Pi_0^2(B) \leq \frac{(1-p)}{(1-p)+p\varepsilon}(-L) + \frac{p\varepsilon}{(1-p)+p\varepsilon} M,$$

[13] Observemos que si la probabilidad de la imperfección en la transmisión es cero, es decir, $\varepsilon = 0$, entonces hay solamente dos estados posibles: $\Omega = \{(a,0,0), (b,\infty,\infty)\}$ y la partición de los dos jugadores seria la misma: $\mathcal{P}_1 = \mathcal{P}_2 = \{\{(a,0,0)\}, \{(b,\infty,\infty)\}\}$.

sin importar $S_1\left(I_1^1\right)$. Luego, el pago de la jugada en A es mayor que el pago de la jugada en B, es decir, $\Pi_0^2(A) > \Pi_0^2(B)$, como el lector podrá fácilmente verificar. Luego, por el resultado antes mencionado en el enunciado de la proposición, tenemos que la jugada B es una jugada con probabilidad cero, es decir, $S_2\left(I_0^2\right) = (1,0)$, tal como queríamos establecer.

Concluiremos la prueba por inducción. Supongamos que

$$S_1\left(I_{k-1}^1\right) = S_2\left(I_{k-1}^2\right) = (1,0)$$

y probemos que $S_1\left(I_k^1\right) = S_2\left(I_k^2\right) = (1,0)$. Como $I_k^1 = \{(b,k,k-1),(b,k,k)\}$, entonces el jugador 1 no sabe si el jugador 2 envió k o $k-1$ mensajes. La probabilidad (condicional) que él atribuye a $\{(b,k,k-1)\}$ es

$$\frac{\varepsilon}{\varepsilon+(1-\varepsilon)\varepsilon} > \frac{1}{2}$$

y, por lo tanto, es menos probable el estado (b,k,k) que $(b,k,k-1)$. Como $S_2\left(I_{k-1}^2\right) = (1,0)$, entonces el jugador 2 juega A en el estado $(b,k,k-1) \in I_{k-1}^2$. Si el jugador 1 juega A, su pago va a ser $\Pi_k^1(A) = 0$, mientras que si juega B, su pago será $\Pi_k^1(B)$, el cual cumple

$$\Pi_k^1(B) \leq \frac{\varepsilon}{\varepsilon+(1-\varepsilon)\varepsilon}(-L) + \frac{(1-\varepsilon)\varepsilon}{\varepsilon+(1-\varepsilon)\varepsilon}M < 0,$$

porque $L > M > 0$. Luego, es estrictamente mejor elegir A. Es decir, tenemos que $S_1\left(I_k^1\right) = (1,0)$. Un argumento similar establece que $S_2\left(I_k^2\right) = (1,0)$, tal como queríamos. ☑

Este resultado es tan sorprendente que es casi una paradoja, porque va en contra nuestra forma natural de pensar. Como siempre es mejor jugar A si el estado es a, entonces tiene sentido que el jugador 1 juegue A si sabe que el estado es a. Es intuitivo suponer, también, que los jugadores van cambiar su jugada para B después de un número suficientemente grande de confirmaciones. Por tanto, es natural esperar que, delante de un problema como este, la mayoría de las personas usaría estrategias como las siguientes:

$$\begin{aligned} S_1(\{(a,0,0)\}) &= (1,0)\text{, es decir, juega } A \text{ con certeza;} \\ S_1(\{(b,k,k-1),(b,k,k)\}) &= \begin{cases} (1,0) & (\text{juega } A), \quad \text{si } k \leq n_1 \\ (0,1) & (\text{juega } B), \quad \text{si } k > n_1 \end{cases} \end{aligned}$$

y

$$S_2(\{(b,k,k),(b,k+1,k)\}) = \begin{cases} (1,0) & (\text{juega } A), \quad \text{si } k \leq n_2 \\ (0,1) & (\text{juega } B), \quad \text{si } k > n_2\,, \end{cases}$$

donde n_1 y n_2 son números pequeños. Por ejemplo $n_1 = n_2 = 5$ o aun $n_1 = n_2 = 3$. La proposición muestra, sin embargo, que eso nunca es equilibrio para los jugadores, aun si $n_1 = n_2 = 1000$!

7. Aplicaciones a la economía

El concepto de conocimiento común fue formalizado inicialmente por Aumann (1976). En ese artículo, él también prueba un resultado simple, pero bastante interesante que dice que dos individuos no pueden "ponerse de acuerdo en discrepar". Más precisamente, si dos individuos tienen las mismas creencias originales (las mismas probabilidades sobre el conjunto de estados del mundo Ω), entonces en todos los eventos de conocimiento común ellos tienen que concordar sobre la probabilidad que atribuyen a tal evento. (Véase Aumann (1976) para mayores detalles.)

Una consecuencia de ese resultado es que agentes en un mercado pueden no llegar a negociar, como muestra un teorema de *no-trade* de Milgrom & Stokey (1982).

El modelamiento de conocimiento se usa extensamente en finanzas, principalmente en la forma semántica, como lo presentamos en la sección **??**. En general, se supone que el conocimiento de los negociadores de activos financieros está dado por una sigma-álgebra generada a partir de la partición de conocimiento descrita en la sección **??**. A medida que el tiempo pasa, el agente obtiene más información y entonces cambian sus particiones de información. Se asume entonces que existe una filtración de sigmas-álgebras, $\{\mathcal{F}_t\}_{t\geq 0}$, cada vez más finas, significando que el conocimiento aumenta con el paso del tiempo. Se asume también que las negociaciones de los agentes son mensurables con respecto a la sigma-álgebra correspondiente. Esta condición es análoga a la que usamos en el juego de los mensajes electrónicos (seción **??**), cuando exigimos que la acción del jugador sea constante dentro de un átomo de su partición de conocimiento.

Hay además una relación profunda de la noción de conocimiento y conocimiento común con la teoría de juegos. Una vez que en situaciones estratégicas, la información que cada jugador detiene es fundamental para sus decisiones, el modelamiento y análisis propio de teoría de juegos depende profundamente de los conceptos presentados aquí.

Apéndice. Prueba del teorema ??

Para probar que φ está bien definida, tenemos que verificar que si $b \in [a]$, entonces $E_a = E_b$ (es trivial que $a \in [a]$ porque $a \Leftrightarrow a$ es un teorema). Supongamos que exista $\omega \in E_a \setminus E_b$. Esto significa que $a \in \omega$ y $b \notin \omega$. Como ω es coherente, entonces $\sim b \in \omega$. Como $b \in [a]$ significa que $a \Leftrightarrow b$ es un teorema y ω contiene todos los teoremas, tenemos que $\sim a \in \omega$, lo que contradice la coherencia de ω. El caso $\omega \in E_b \setminus E_a$ es análogo.

Supongamos ahora que $[a] \neq [b]$, o sea, $a \Leftrightarrow b$ no es un teorema. Por la definición de teorema, eso significa que existe $\omega \in \Omega$ tal que $(a \Leftrightarrow b) \notin \omega$. Por la completitud de ω, eso implica que $\sim (a \Leftrightarrow b) \in \omega$. Es decir,

$((a \wedge \sim b) \vee (\sim a \wedge b)) \in \omega$. En este caso, sabemos que $(a \wedge \sim b) \in \omega$ o $(\sim a \wedge b) \in \omega$. (No probamos que podemos hacer esa dedución, mas el lector no tendrá dificultad para probarla usando la definición de $\vee$ y las reglas V1, V2 y V3.) En el primero caso, $a \in \omega$ y $b \notin \omega$, mientras que en el segundo, $a \notin \omega$ y $b \in \omega$. En cualquiera de estas situacines, encontramos un ω que no puede estar al mismo tiempo en E_a y en E_b. O sea, φ es inyectiva.

Sea $E \in \mathcal{S}(\Omega)$. Por definición, $E = E_a$ para algún $a \in \mathcal{S}$. Tome ese $a \in \mathcal{S}$. Entonces $\varphi([a]) = E$ y φ es sobreyectiva.

Del Lema **??**, sabemos que $E_{\sim a} = \sim E_a$ y $E_{a \wedge b} = E_a \cap E_b$. Por los resultados anteriores, sabemos que $\varphi([\sim a]) = E_{\sim a}$; $\sim\varphi([a]) = \sim E_a$; $\varphi([a \wedge b]) = E_{a \wedge b}$; $\varphi([a]) = E_a$; $\varphi([b]) = E_b$. Luego, los ítemes (ii) y (iii) se siguen del Lema **??**. ☑

Bibliografía

[1] Aumann, Robert J.,*Agreeing to disagree*, Ann.Statist., **4** (1976), 1236–1239.

[2] Aumann, Robert J., *Interactive Epistemology I: Knowledge*, Intern. Journ. of Game Theory, **28** (1999a), 263–300.

[3] Aumann, Robert J., *Interactive Epistemology II: Probability*, Intern. Journ. of Game Theory, **28** (1999b), 301–314.

[4] Milgrom, Paul , *An Axiomatic Characterization of Common Knowledge.* In Econometrica, **49** (1) (1981), 219–222.

[5] Milgrom, Paul and Nancy Stokey, *Information, trade and common knowledge*, Journ. of Econom. Theory, **26** (1) (1982), 17–27.

[6] Rubinstein, Ariel, *The electronic mail game: strategic behavior under almost common knowledge*, Amer.Econ.Rev., **79** (1989), 385–391.

[7] Samuelson, Larry , *Modeling Knowledge in Economic Analysis*, Journ. of Econom. Literat., **42** (2) (2004), 367–403.

(Recibido en abril de 2006. Aceptado para publicación en agosto de 2006)

Departmento de Economía
Universidad Carlos III de Madrid
Madrid, España
e-mail: decastro.luciano@gmail.com