4 LÓGICAS *BDI*

¿Cómo podemos razonar acerca del comportamiento de los agentes racionales y su relación con los estados Intencionales? Tradicionalmente, la alternativa ha sido definir una lógica que nos permita razonar sobre las actitudes proposicionales de los agentes, sus propiedades, sus interrelaciones y la manera en que cambian en el tiempo. Diversas lógicas han sido propuestas en la literatura [102, 127], con base en los compromisos ontológicos asumidos. Por tanto, he decidido presentar una formalización cercana a los fundamentos filosóficos expuestos en la teoría de razonamiento práctico propuesta por Bratman [22], conocida genericamente como **Lógicas BDI**. Esta familia de lógicas fue propuesta originalmente por Rao [151]. Rao y Georgeff [150] presentan una axiomatización consistente y completa de estas lógicas, así como procedimientos de decisión para determinar la satisfacción y validez de sus fórmulas. Como suele ser el caso con estos formalismos, además de los operadores modales para creencias (B), deseos (D) e intenciones (I), se hará uso de otros operadores, debido a que:

Lógicas BDI

- Estas lógicas pueden verse como extensiones de las lógicas temporales conocidas como **lógicas computacionales arborescentes** CTL y $CTL*$ [56, 57, 55]. El apéndice 1, en su página 327, ofrece una introducción más detallada a estos formalismos. También se ofrece una recapitulación de las lógicas proposicional, de predicados y modal, que pueden verse como antecedentes de este capítulo.

  Lógica temporal

- Estas lógicas pueden incluir un componente de acción para representar los eventos registrados por los agentes y sus acciones. Este componente se basa normalmente en la **lógica dinámica** [95]; o se define usando fórmulas de estado para expresar la ocurrencia de eventos, que es el enfoque utilizado aquí.

  Lógica dinámica

El capítulo se organiza como sigue. Primero definiremos la sintáxis de las lógicas BDI_{CTL} y BDI_{CTL*}; luego definiremos su semántica con base en el principio de mundos posibles, de uso común en las lógicas modales. Con la semántica de los operadores primitivos definida, procederemos a axiomatizar lo que se conoce como un Sistema Básico I, para lo cual axiomatizaremos los componentes BDI, las relaciones entre ellos y las acciones de los agentes. Para terminar, a manera de ejemplo del uso de estas lógicas, analizaremos las posibles estrategias de compromiso que un Sistema Básico I puede adoptar y algunas de las consecuencias de tal adopción.

4.1 SINTAXIS

El lenguaje de las Lógicas BDI es el de la lógica proposicional (Ecuación 4.1), extendido con los operadores para representar las actitudes proposicionales

de los agentes (Ecuación 4.2) y los operadores temporales de las lógicas CTL y $CTL*$ (Ecuación 4.3):

Definición 4.1. *Las fórmulas bien formadas (fbf) de la lógica BDI_{CTL*} son las definidas por la siguiente forma Backus Naur (BNF):*

$$\phi ::= \bot \mid p \mid \neg\phi \mid \phi \wedge \phi \mid \tag{4.1}$$
$$\mathsf{BEL}(\phi) \mid \mathsf{DES}(\phi) \mid \mathsf{INT}(\phi) \mid \tag{4.2}$$
$$\bigcirc\phi \mid \phi \,\mathsf{U}\, \phi \mid \mathsf{E}\phi \tag{4.3}$$

donde p denota cualquier fórmula atómica proposicional.

La sintáxis se ha definido tomando como operadores primitivos la contradicción, la negación y la conjunción, así como los operadores Intencionales para creencia (BEL), deseo (DES) e intención (INT) y los operadores temporales siguiente ($\bigcirc$), hasta (U) y opcionalmente (E). Esta elección es arbitraria, en el sentido que podríamos haber optado por la disyunción en lugar de la conjunción. Dada nuestra elección, los otros operadores comunes de la lógica proposicional se definen como sigue:

Definición 4.2 (Otros operadores proposicionales)**.** *La disyunción, la implicación y la equivalencia material se definen respectivamente como de costumbre:*

- $\phi \vee \psi \stackrel{def}{=} \neg(\neg\phi \wedge \neg\psi)$.
- $\phi \rightarrow \psi) \stackrel{def}{=} (\neg\phi \vee \psi)$.
- $\phi \leftrightarrow \psi) \stackrel{def}{=} ((\phi \rightarrow \psi) \wedge (\psi \rightarrow \phi))$.
- $false \stackrel{def}{=} \bot$
- $true \stackrel{def}{=} \neg\bot$

Definición 4.3 (Otros operadores temporales)**.** *Eventualmente, siempre, e inevitablemente se definen respectivamente como de costumbre:*

- $\Diamond\phi \stackrel{def}{=} true \,\mathsf{U}\, \phi$.
- $\Box\phi \stackrel{def}{=} \neg(\Diamond\neg\phi)$.
- $\mathsf{A}\phi \stackrel{def}{=} \neg(\mathsf{E}\neg\phi)$.

Ejemplo 4.1. $\mathsf{A}\Diamond\mathsf{BEL}(\neg crisis)$ *es una fbf que expresa que en todo futuro posible, eventualmente creeremos que no hay crisis.* $\mathsf{A}(\mathsf{BEL}(crisis) \rightarrow \neg\mathsf{INT}(viaje))$ *expresa que inevitablemente si creo que hay crisis entonces no intento ir de viaje.*

4.1.1 Sintaxis de la lógica BDI_{CTL*}

Al igual que en las lógicas temporales arborescentes, las fbf de las Lógicas BDI se clasifican en **fórmulas de estado**, que son evaluadas con respecto a un mundo posible en particular, en un estado dado; y las **fórmulas de camino**, que son evaluadas con respecto al camino formado por una serie de transiciones entre mundos posibles:

Fórmulas de estado
Fórmulas de camino

Definición 4.4 (Fórmulas de estado)**.** *Las fbfs de estado en la lógica BDI_{CTL*} se definen inductivamente como sigue:*

1. *Toda fbf proposicional (Ecuación 4.1) es una fbf de estado.*
2. *Si ϕ es una fbf de estado, entonces $\mathsf{BEL}(\phi)$, $\mathsf{DES}(\phi)$ e $\mathsf{INT}(\phi)$ son fbf de estado.*
3. *Si ϕ es una fbf de camino, entonces $\mathsf{E}\phi$ (opcional) y $\mathsf{A}\phi$ (inevitable) son fbfs de estado.*

Las fbfs de camino en BDI_{CTL*} pueden contener cualquier combinación arbitraria de fórmulas temporales lineales, conteniendo negaciones, disyunciones y los operadores temporales:

Definición 4.5 (Fórmulas de camino)**.** *Las fbf de camino en la lógica BDI_{CTL*} se define inductivamente como sigue:*

1. *Cualquier fbf de estado es una fbf de camino.*
2. *Si ϕ y ψ son fbfs de camino, entonces $\neg\phi$ y $\phi \wedge \psi$ también lo son.*
3. *Si ϕ y ψ son fbfs de camino, entonces $\bigcirc\phi$ y $\phi \,\mathsf{U}\, \psi$ también lo son.*

Observen que las fbfs de estado son a su vez fbfs de camino, un camino de longitud uno.

4.1.2 Sintaxis de la lógica BDI_{CTL}

La lógica restringida BDI_{CTL} se obtiene al prohibir combinaciones booleanas y anidamiento de los operadores temporales en las fórmulas de camino. Formalmente, substituimos la definición 4.5 por:

Definición 4.6 (Fórmulas de camino restringidas)**.** *Si ϕ y ψ son fbfs de estado, entonces $\bigcirc\phi$ y $\phi \,\mathsf{U}\, \psi$ son fbfs de camino.*

Ejemplo 4.2. *$\Box\Diamond$crisis, que expresa que siempre eventualmente hay una crisis, es una fbf de camino en BDI_{CTL*}, pero no lo es en BDI_{CTL}.*

4.1.3 Fórmulas opcionales e inevitables

El uso de cuantificadores de camino introduce una clasificación de las fbfs. Las fórmulas opcionales, denotadas como **O-fórmulas**, son aquellas que no contienen ocurrencias de A (o de la negación de E) fuera del alcance de los operadores BEL, DES e INT. Las fórmulas inevitables, denotadas como **I-fórmulas**, son fórmulas que no contiene ocurrencias de E (o negaciones de A) fuera del alcance de los operadores intencionales BEL, DES, e INT.

Fórmulas opcionales

Fórmulas inevitables

4.2 SEMÁNTICA

Las Lógicas BDI tienen una semántica basada en **mundos posibles**. Cada mundo posible se define como una estructura de árbol donde los nodos son

Mundos posibles

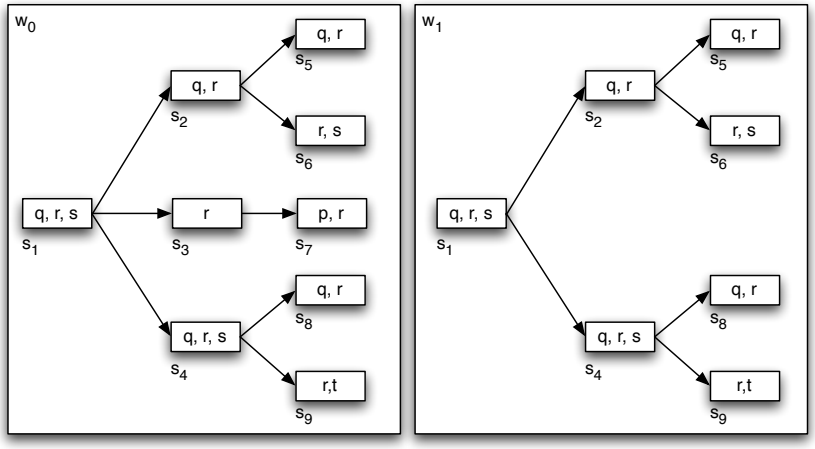

Figura 4.1: Sea $W = \{w_0, w_1\}$ un conjunto de mundos posibles. Los estados del mundo w_0, se denotan por $S_{w_0} = \{s_1, \ldots, s_9\}$. $R_{w_0} = \{(s_1, s_2), (s_1, s_3), (s_1, s_4), (s_2, s_5), (s_2, s_6), \ldots\}$. $L(w_0, s_1)q = true$, pero $L(w_0, s_3)q = false$. Ademas $w_1 \sqsubseteq w_0$.

conjuntos de fórmulas proposicionales, y los arcos representan transiciones en el tiempo. El pasado es único, pero el futuro es arborescente, para representar posibles cursos de acción. Una vez que el agente actúa, el pasado se vuelve lineal, debido a que solo uno de los cursos de acción posibles se ha llevado a cabo. La Figura 4.1 muestra dos de estos mundos posibles. Los estados s_i funcionan como índices. Una relación de accesibilidad para las creencias establece que mundos son creíbles a partir de un mundo y un estado dados. Las funciones de accesibilidad para deseos e intenciones funcionan de manera similar.

Estructura de Kripke

La semántica de las Lógicas BDI se define usualmente sobre una **estructura de Kripke**:

Definición 4.7 (Estructura de Kripke)**.** *Una estructura de Kripke se define como la tupla* $M = \langle W, \{S_w : w \in W\}, \{R_w : w \in W\}, L, \mathcal{B}, \mathcal{D}, \mathcal{I}\rangle$*, donde* W *es un conjunto de mundos posibles;* S_w *es el conjunto de estados para cada mundo* $w \in W$*;* R_w *es una relación binaria total sobre los estados del mundo* w*,* $R_w \subseteq S_w \times S_w$*;* L *es una función de asignación de verdad para las proposiciones primitivas en cada mundo* $w \in W$ *en cada estado* $s \in S_w$*, por ejemplo* $L(w, s) : \phi \to \{true, false\}$*; y* $\mathcal{B}$*,* $\mathcal{D}$*,* $\mathcal{I}$ *son relaciones sobre los mundos y sus estados, por ejemplo* $\mathcal{B} \subseteq W \times S_w \times W$*. Lo mismo para* $\mathcal{D}$ *e* $\mathcal{I}$*.*

Sub-mundos

Observen que S_w y R_w definen cada mundo como una estructura de árbol. Algunas veces R_w es forzada a ser lineal hacía atrás, de forma que el pasado sea único. Se dice que un mundo es **sub-mundo** de otro si contiene menos ramas que el otro, pero es idéntico en todo lo demás. Por ejemplo, en la Figura 4.1 w_1 es un sub-mundo de w_0. Formalmente:

Definición 4.8 (Sub-mundos)**.** *Un mundo* w' *es un sub-mundo de* w*, denotado por* $w \sqsubseteq w'$*, si y sólo si:*

1. $S_w \subseteq S'_w$;
2. $R_w \subseteq R'_w$;

3. $\forall s \in S_w, L(w,s) = L(w',s)$; *y*

4. $\forall s \in S_w, (w,s,v) \in \mathcal{B}$ *si y sólo si* $(w',s,v) \in \mathcal{B}$. *Lo mismo para* $\mathcal{D}$ *e* $\mathcal{I}$.

Un mundo w' *es un sub-mundo estricto de* w, *denotado por* $w' \sqsubset w$, *si y sólo si* $w' \sqsubseteq w$ *y* $w \not\sqsubseteq w'$. *Si* w' *es un sub-mundo de* w, *entonces* w *es un super-mundo de* w'. *Se dice que* w' *es estructuralmente equivalente a* w, *denotado por* $w' \approx w$, *si y sólo si* $w' \sqsubset w$ *y* $w \sqsubset w'$.

4.2.1 Satisfacción

La satisfacción de las fbfs se denota por $\models$, y se define con respecto a la estructura M, un mundo w y un estado s. La expresión $M, w_s \models \phi$ se lee como "la estructura M en el mundo w y el estado s satisface ϕ". Una notación alternativa para esta expresión es $M \models_{w_s} \phi$. Un camino $s_0, s_1, \ldots$, en el mundo w es denotado por $(w_{s_0}, w_{s_1}, \ldots)$.

Definición 4.9 (Semántica BDI_{CTL*})**.** *Las reglas semánticas para las fbfs de estado se definen como sigue:*

s1. $M, w_s \models \phi$, *si y sólo si (ssi)* $\phi \in L(w,s)$, *donde* ϕ *es una proposición atómica.*

s2.
- *a)* $M, w_s \models \neg\phi$, *ssi* $M, w_s \not\models \phi$.
- *b)* $M, w_s \models \phi \vee \psi$, *ssi* $M, w_s \models \phi$ *ó* $M, w_s \models \psi$.

s3.
- *a)* $M, w_{s_0} \models \mathsf{E}\phi$, *ssi existe un camino* $w_{s_0}, w_{s_1}, \ldots$, *t.q.* $M, (w_{s_0}, w_{s_1}, \ldots) \models \phi$.
- *b)* $M, w_{s_0} \models \mathsf{A}\phi$, *ssi para todo camino* $w_{s_0}, w_{s_1}, \ldots$, $M, (w_{s_0}, w_{s_1}, \ldots) \models \phi$.

s4.
- *a)* $M, w_s \models \mathsf{BEL}(\phi)$, *ssi existe un* v *t.q.* $(w,s,v) \in \mathcal{B}$, $M, v_s \models \phi$.
- *b)* $M, w_s \models \mathsf{DES}(\phi)$, *ssi para todo* v *t.q.* $(w,s,v) \in \mathcal{D}$, $M, v_s \models \phi$.
- *c)* $M, w_s \models \mathsf{INT}(\phi)$, *ssi para todo* v *t.q.* $(w,s,v) \in \mathcal{I}$, $M, v_s \models \phi$.

Las reglas semánticas para las fbfs de camino se definen como sigue:

p1. $M, (w_{s_0}, w_{s_1}, \ldots) \models \phi$, *ssi* $M, w_{s_0} \models \phi$.

p2.
- *a)* $M, (w_{s_0}, w_{s_1}, \ldots) \models \neg\psi$, *ssi* $M, w_{s_0} \not\models \psi$.
- *b)* $M, (w_{s_0}, w_{s_1}, \ldots) \models \phi \vee \psi$, *ssi* $M, (w_{s_0}, w_{s_1}, \ldots) \models \phi$ *ó* $M, (w_{s_0}, w_{s_1}, \ldots) \models \psi$.

p3.
- *a)* $M, (w_{s_0}, w_{s_1}, \ldots) \models \bigcirc\psi$, *ssi* $M, (w_{s_1}, \ldots) \models \psi$.
- *b)* $M, (w_{s_0}, w_{s_1}, \ldots) \models \phi \,\mathsf{U}\, \psi$, *ssi i)* $\exists k, k \geq 0$ *t.q.* $M, (w_{s_k}, \ldots) \models \psi$ *y* $\forall 0 \leq j < k, M, (w_{s_j}, \ldots) \models \phi$; *ó ii)* $\forall j \geq 0, M, (w_{s_j}, \ldots) \models \phi$.

Validez

La validez y satisfacción de una fórmula se define de manera estándar. Una fbf se dice **válida** si y sólo si es verdadera en todo estado, de todo mundo, de toda estructura. La validez y satisfacción con respecto a una familia de estructuras también puede definirse. Rao y Georgeff [154] consideran dos clases de estructuras con respecto a las cuales evaluar validez y satisfacción:

i) $\mathcal{M}$ que requiere que $\mathcal{R}$ sea total, sin imponer ninguna restricción sobre los operadores Intencionales; y ii) $\mathcal{R}^{est}$, que requiere que $\mathcal{R}$ sea total; $\mathcal{B}$ serial, transitiva y euclidiana; y, $\mathcal{D}$ e $\mathcal{I}$ sean seriales. Este modelo subyace en la lógica identificada como ${B^{KD45}D^{KD}I^{KD}}_{CTL}$. La teoría de correspondencia [1] hace obvia esta nomenclatura.

Definición 4.10 (Tipos de Relaciones)**.** *Las definiciones de relación total a la izquierda, serial, transitiva y euclidiana son las estándares. Formalmente:*

TOTAL A LA IZQUIERDA: $\forall w \forall s \exists t \quad (s,t) \in R_w$.

SERIAL: $\forall w \forall s \exists v \quad (w,s,v) \in \mathcal{B}$.

TRANSITIVA: $\forall w,v,x,s$ *Si* $(w,s,v) \in \mathcal{B} \wedge (v,s,x) \in \mathcal{B}$, *entonces* $(w,s,x) \in \mathcal{B}$.

EUCLIDIANA: $\forall w,v,x,s$ *Si* $(w,s,v) \in \mathcal{B} \wedge (w,s,x) \in \mathcal{B}$, *entonces* $(v,s,x) \in \mathcal{B}$.

En este contexto, que $\mathcal{R}$ sea total, significa que para todo estado de un mundo dado, existe un estado sucesor. Que una relación Intencional, digamos $\mathcal{B}$ sea serial, quiere decir que todo estado de un mundo, tiene al menos un mundo creíble.

4.3 AXIOMATIZACIÓN DE LOS COMPONENTES BDI

En lo que sigue, sólo consideraremos a la lógica BDI_{CTL}, por lo tanto, todos sus axiomas y reglas de inferencia [55] son adoptados. Puesto que los componentes BDI son modelados como sistemas modales normales, el **axioma K** de la lógica modal se adopta para las creencias, los deseos y las intenciones: *Axioma K*

Axioma 4.1 (BK)**.** $\mathsf{BEL}(\phi) \wedge \mathsf{BEL}(\phi \rightarrow \psi) \rightarrow \mathsf{BEL}(\psi)$;

Axioma 4.2 (DK)**.** $\mathsf{DES}(\phi) \wedge \mathsf{DES}(\phi \rightarrow \psi) \rightarrow \mathsf{DES}(\psi)$;

Axioma 4.3 (IK)**.** $\mathsf{INT}(\phi) \wedge \mathsf{INT}(\phi \rightarrow \psi) \rightarrow \mathsf{INT}(\psi)$;

La regla de **generalización** se adopta para las creencias, los deseos y las intenciones. Esta regla formula que toda fbf válida es creída, deseada e intentada. A la lógica resultante se le identifica como BDI^{K}_{CTL}. La lógica es consistente y completa con respecto a la familia de estructuras $\mathcal{M}$ [154]. *Generalización*

Regla de Inferencia 1 (B-Gen)**.** *Si* $\vdash \phi$ *entonces* $\vdash \mathsf{BEL}(\phi)$;

Regla de Inferencia 2 (D-Gen)**.** *Si* $\vdash \phi$ *entonces* $\vdash \mathsf{DES}(\phi)$;

Regla de Inferencia 3 (I-Gen)**.** *Si* $\vdash \phi$ *entonces* $\vdash \mathsf{INT}(\phi)$;

El sistema modal estándar $KD45$, también conocido como $S5$-débil, es adoptado para las creencias. El axioma D expresa la consistencia de las creencias; y los axiomas 4 y 5 expresan introspección positiva y negativa. Los axiomas son las siguientes: *KD45*

Axioma 4.4 (BD)**.** $\mathsf{BEL}(\phi) \rightarrow \neg\mathsf{BEL}(\neg\phi)$;

1 La teoría de correspondencia en la lógica modal se aborda en la página 326.

Axioma 4.5 (B4). $\mathsf{BEL}(\phi) \rightarrow \mathsf{BEL}(\mathsf{BEL}(\phi))$;

Axioma 4.6 (B5). $\neg\mathsf{BEL}(\phi) \rightarrow \mathsf{BEL}(\neg\mathsf{BEL}(\phi))$.

Para los deseos y las intenciones se adopta además el axioma D para expresar consistencia entre los deseos y las intenciones.

Axioma 4.7 (DD). $\mathsf{DES}(\phi) \rightarrow \neg\mathsf{DES}(\neg\phi)$;

Axioma 4.8 (ID). $\mathsf{INT}(\phi) \rightarrow \neg\mathsf{INT}(\neg\phi)$;

A la lógica resultante se le conoce como ${B^{KD45}D^{KD}I^{KD}}_{CTL}$. Es consistente y completa con respecto a la familia de estructuras $\mathcal{M}^{est}$. Para estas dos lógicas BDI basadas en CTL, Rao y Georgeff [154, 150] han propuesto métodos de decisión para verificar las satisfacción y validez de sus fbf.

4.4 AXIOMATIZACIÓN BDI MULTI MODAL

La relación deseada entre las creencias, los deseos y las intenciones puede examinarse en dos dimensiones: con respecto al conjunto de relaciones entre estas actitudes, representada por las relaciones de accesibilidad; y con respecto a las estructuras de mundos posibles.

Dados dos conjuntos S y R, las siguientes relaciones entre ellos pueden darse: $S \subseteq R$, $R \subseteq S$, $S \cap R \neq \{\}$, y $S \cap R = \{\}$. Aunque no todas estas relaciones entre mundos accesibles son significativas en el contexto BDI, es posible caracterizar aquellas que son significativas semántica y axiomaticamente. Tenemos tres casos de interés:

- El conjunto de mundos deseados es un subconjunto de aquellos creídos posibles, lo que ocurre cuando el agente puede creer un mundo sin desear estar en él;
- Los mundos creídos son un subconjunto de los deseados, intuitivamente esto significa que hay ciertos mundos deseados que no son creídos por el agente; y
- Algunos mundos que son deseados no son creídos y viceversa. Como una combinación de los dos primeros casos.

Dados dos mundos w y v, las siguientes relaciones pueden mantenerse entre ellos: w puede ser un sub-mundo de v, v un sub-mundo de w, v y w pueden ser idénticos, o v y w pueden ser totalmente diferentes. Relaciones similares pueden establecerse entre mundos accesibles por creencias e intenciones, y para mundos accesibles por deseos e intenciones. Los casos significativos son: i) Un mundo accesible por deseos que es sub-mundo de uno accesible por creencias. De manera intuitiva, esto significa que de todos los caminos que el agente cree que puede elegir, sólo desea algunos de entre ellos; ii) Un mundo accesible por creencias es un sub-mundo de otro que es accesible por deseos, lo que significa que todos los caminos que el agente desea, sólo cree que algunos de ellos pueden ser logrados. Los casos anteriores pueden combinarse dando como resultado que el agente cree que puede lograr todos los caminos que desea y viceversa.

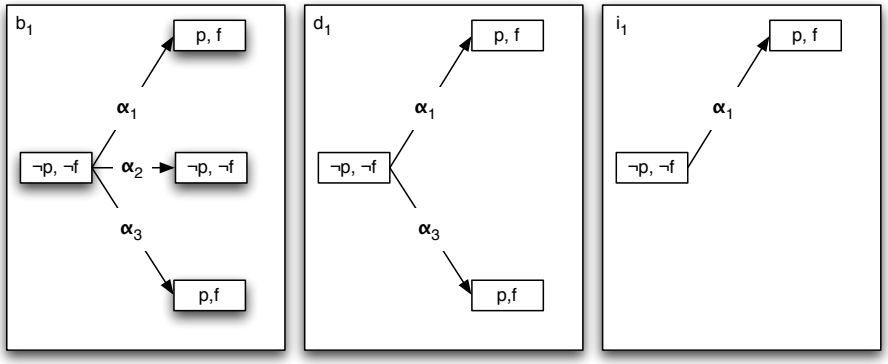

Figura 4.2: Un ejemplo de mundos posibles BDI, el escenario del dentista. El agente tiene como opciones las acciones α_1 y α_2 ir al dentista y α_3 no ir al dentista. Las proposiciones son p dolor, f diente tapado.

Rao y Georgeff [149] ejemplifican una de estas relaciones, tal y como se muestra en la Figura 4.2. Consideren un agente que desea ir al dentista para taparse un diente. Consideren que el mundo b_1 es accesible por creencia a partir de la situación actual del agente, por ej., el mundo w_0 en su estado s_0. La proposición p se interpreta como tener dolor. La proposición f se interpreta como diente tapado. Las etiquetas entre los estados son acciones del agente. En este caso α_1 denota ir al dentista 1, α_2 denota ir al dentista 2 y α_3 denota ir de compras. El mundo d_1 es accesible por deseos desde el mundo w_0 en el estado s_0. El mundo d_1 es un sub-mundo del mundo b_1. El mundo i_1 es accesible por intención desde el mundo w_0 en el estado s_0. El mundo i_1 es un sub-mundo de d_1. Observen que d_1 e i_1 representan elecciones selectivas incrementales a partir de b_1[2].

El conjunto de relaciones estructurales puede combinarse para obtener una variedad de estructuras de mundos posibles diferentes. Es posible obtener nueve relaciones entre los mundos creídos y los mundos deseados. De manera similar, hay nueve relaciones posibles entre los mundos deseados e intentados, y entre los mundos creídos e intentados. Tres de estas relaciones han sido consideradas en la literatura bajo los términos de realismo [40], realismo fuerte [149], y realismo débil [153].

4.4.1 Realismo fuerte

Bajo el realismo fuerte, el conjunto de mundos accesibles por creencias es un subconjunto de los mundos accesibles por los deseos; y cada mundo accesible por creencias es un subconjunto de los mundos deseados. Como resultado, si el agente desea opcionalmente lograr una proposición, entonces cree que la proposición es una opción que, si es elegida, se logra. El realismo fuerte también puede aplicarse a los deseos e intenciones. Como resultado, si el agente intenta opcionalmente lograr una proposición, entonces también desea opcionalmente lograr esa proposición.

Los diferentes mundos accesibles por creencias, deseos e intenciones, representan diferentes posibles escenarios para el agente. Intuitivamente, el

2 Recuerden los conceptos de estándar de relevancia y filtro de admisibilidad en la teoría de Bratman.

agente cree que el mundo actual es uno de sus mundos accesibles por creencias; si sucede que estuviera en el mundo accesible por creencias b_1, entonces sus deseos (con respecto a b_1) serían un mundo accesible por deseos, por ej., d_1; y sus intenciones un mundo accesible por intenciones, por ej., i_1. Los mundos d_1 e i_1 representan incrementalmente opciones selectivas desde b_1 acerca de los deseos por una opción y opciones de posibles cursos de acción futuros. Si ϕ es una o-fórmula, las condiciones anteriores se expresan con los axiomas de **realismo fuerte**: *Realismo fuerte*

Axioma 4.9 (BD-Realismo fuerte)**.** $\mathsf{DES}(\phi) \rightarrow \mathsf{BEL}(\phi)$*;*

Axioma 4.10 (DI-Realismo fuerte)**.** $\mathsf{INT}(\phi) \rightarrow \mathsf{DES}(\phi)$.

Los axiomas anteriores expresan que si el agente tiene la intención hacía $E(\psi)$, también desea que $E(\psi)$, esto es, existe al menos un camino en el que en todos los mundos accesibles por deseo ψ es verdadera. Esto se asegura porque $\mathcal{B}$ y $\mathcal{D}$ tienen que ser seriales y compatibles. Las condiciones semánticas para el realismo fuerte se expresan como:

DB–REALISMO FUERTE $\forall w \forall s \forall v$ si $(w, s, v) \in \mathcal{B}$ entonces $\exists v', (w, s, v') \in \mathcal{D}$ y $v \sqsubseteq v'$ (or $\mathcal{B} \subseteq_{sub} \mathcal{D}$);

DI–REALISMO FUERTE $\forall w \forall s \forall v$ si $(w, s, v) \in \mathcal{D}$ entonces $\exists v', (w, s, v') \in \mathcal{I}$ y $v \sqsupseteq v'$ (or $\mathcal{D} \subseteq_{sup} \mathcal{I}$).

4.4.2 Realismo

Cohen y Levesque [40] consideran una estructura donde el conjunto de mundos accesibles por intención es un subconjunto del conjunto de mundos accesibles por creencias y las estructuras de creencia e intención son idénticas (una línea de tiempo). Esta restricción se conoce como realismo y tiene como efecto que si un agente cree una proposición también tendrá una intención con respecto a esa proposición. Los axiomas del **realismo** son; *Realismo*

Axioma 4.11 (BD-Realismo)**.** $\mathsf{BEL}(\phi) \rightarrow \mathsf{DES}(\phi)$*;*

Axioma 4.12 (DI-Realismo)**.** $\mathsf{DES}(\phi) \rightarrow \mathsf{INT}(\phi)$.

Estos axiomas corresponde a las siguientes restricciones multi-modales:

DB–REALISMO $\forall w \forall s \forall v$ si $(w, s, v) \in \mathcal{D}$ entonces $(w, s, v) \in \mathcal{B}$ (o $\mathcal{D} \subseteq \mathcal{B}$);

ID–REALISMO $\forall w \forall s \forall v$ si $(w, s, v) \in \mathcal{I}$ entonces $(w, s, v) \in \mathcal{D}$ (o $\mathcal{I} \subseteq \mathcal{D}$).

Esta forma de realismo induce a un agente a desear todas las proposiciones que cree posibles e intentarlas, lo cual resulta demasiado entusiasta. En cambio, el realismo fuerte induce agentes que son demasiado cautos.

4.4.3 Realismo débil

Es posible obtener un balance entre los dos enfoques anteriores si los agentes no desean aquellas proposiciones cuya negación es creída; no intentan proposiciones cuya negación es deseada; y no intentan proposiciones cuya negación es creída. A esta propiedad se le conoce como **realismo débil** y se especifica como los siguientes axiomas: *Realismo débil*

Axioma 4.13 (DB-Realismo débil). $\mathsf{DES}(\phi) \rightarrow \neg\mathsf{BEL}(\neg\phi)$;

Axioma 4.14 (IB-Realismo débil). $\mathsf{INT}(\phi) \rightarrow \neg\mathsf{BEL}(\neg\phi)$;

Axioma 4.15 (ID-Reralismo débil). $\mathsf{INT}(\phi) \rightarrow \neg\mathsf{DES}(\neg\phi)$.

Estos axiomas corresponde a la versión multi modal de la condición serial:

DB-REALISMO DÉBIL $\forall w \forall s \exists v(w,s,v) \in \mathcal{D}$ si y sólo si $(w,s,v) \in \mathcal{B}$ (o $\mathcal{B} \cap \mathcal{D} \neq \{\}$);

IB-REALISMO DÉBIL $\forall w \forall s \exists v(w,s,v) \in \mathcal{I}$ si y sólo si $(w,s,v) \in \mathcal{B}$ (o $\mathcal{B} \cap \mathcal{I} \neq \{\}$);

ID-REALISMO DÉBIL $\forall w \forall s \exists v(w,s,v) \in \mathcal{I}$ si y sólo si $(w,s,v) \in \mathcal{D}$ (o $\mathcal{D} \cap \mathcal{I} \neq \{\}$).

4.4.4 Otras relaciones multi-modales

Si un agente tiene una intención, éste cree que tiene tal intención:

Axioma 4.16 (I-BI). $\mathsf{INT}(\phi) \rightarrow \mathsf{BEL}(\mathsf{INT}(\phi))$

I-BI $\forall w \forall s \forall w' \forall w''$ si $(w,s,w') \in \mathcal{B}$ and $(w,s,w'') \in \mathcal{I}$ entonces $(w',s,w'') \in \mathcal{B}$.

Si un agente tiene un deseo, éste cree que tiene tal deseo:

Axioma 4.17 (D-BD). $\mathsf{DES}(\phi) \rightarrow \mathsf{BEL}(\mathsf{DES}(\phi))$

D-BD $\forall w \forall s \forall w' \forall w''$ si $(w,s,w') \in \mathcal{B}$ and $(w,s,w'') \in \mathcal{D}$ entonces $(w',s,w'') \in \mathcal{B}$.

Si un agente tiene una intención, debe desear tal intención:

Axioma 4.18 (I-DI). $\mathsf{INT}(\phi) \rightarrow \mathsf{DES}(\mathsf{INT}(\phi))$

I-DI $\forall w \forall s \forall w' \forall w''$ si $(w,s,w') \in \mathcal{D}$ y $(w,s,w'') \in \mathcal{I}$ entonces $(w',s,w'') \in \mathcal{D}$.

Si la relación de equivalencia ($\leftrightarrow$) es usada en este axioma en lugar de la implicación $\rightarrow$, las modalidades anidadas se colapsan.

4.4.5 Tesis de asimetría

Bratman [22] argumenta que es irracional para un agente intentar algo y también creer que ese algo no hará ese algo. Esto puede verse como una restricción que prohibe la inconsistencia entre intenciones y creencias. Por ejemplo, un robot que intenta servir una cerveza, pero también cree que no la servirá es irracional. Por otra parte, es racional (o menos irracional) intentar algo que no creo que no vaya a hacer. La incompletez entre intenciones y creencias si que es permitida. Es racional para nuestro robot intentar servir la cerveza y no creer que la va a servir (podría estar ocupado en ese momento). A esto se le conoce como **Tesis de asimetría** (Ver capítulo 2, página 34). La tesis puede generalizarse a otros operadores Intencionales y expresarse mediante los siguientes principios:

Tesis de asimetría

AT1 $\models \mathsf{INT}(\phi) \rightarrow \neg\mathsf{BEL}(\neg\phi)$

AT2 $\not\models \mathsf{INT}(\phi) \rightarrow \mathsf{BEL}(\phi)$

AT3 $\not\models \mathsf{BEL}(\phi) \rightarrow \mathsf{INT}(\phi)$

AT4 $\models \mathsf{INT}(\phi) \rightarrow \neg\mathsf{DES}(\neg\phi)$

AT5 $\not\models \mathsf{INT}(\phi) \rightarrow \mathsf{DES}(\phi)$

AT6 $\not\models \mathsf{DES}(\phi) \rightarrow \mathsf{INT}(\phi)$

AT7 $\models \mathsf{DES}(\phi) \rightarrow \neg\mathsf{BEL}(\neg\phi)$

AT8 $\not\models \mathsf{DES}(\phi) \rightarrow \mathsf{BEL}(\phi)$

AT9 $\not\models \mathsf{BEL}(\phi) \rightarrow \mathsf{DES}(\phi)$

Rao y Georgeff [150] demuestran que diferentes sistemas BDI satisfacen diferentes principios de la tesis de asimetría. Para un sistema $B^{KD45}D^{KD}I^{KD}$ bajo diferentes formas de realismo, tenemos:

	AT1	AT2	AT3	AT4	AT5	AT6	AT7	AT8	AT9
Realismo fuerte	✓		✓	✓	✓	✓	✓		
Realismo	✓	✓		✓	✓		✓	✓	
Realismo débil	✓	✓	✓	✓	✓	✓	✓	✓	✓

4.4.6 Posposición infinita

Si un agente forma una intención, entonces inevitablemente en algún momento futuro la abandonará. Esto se conoce como **posposición finita** (*no infinite deferral*) y es un requisito de la racionalidad acotada: *Posposición finita*

Axioma 4.19 (Posposición finita)**.** $\mathsf{INT}(\phi) \rightarrow \mathsf{A}\Diamond(\neg\mathsf{INT}(\phi))$.

4.5 EVENTOS

Para poder describir la conducta de un agente, es necesario describir la ocurrencia de acciones, aquí llamados **eventos**. Las extensiones necesarias en este sentido incluyen permitir que las fbf expresen el éxito y fracaso de los eventos. Si e es un evento primitivo, $succeeded(e)$ denota la ocurrencia exitosa de e en el pasado inmediato; $failed(e)$ denota el fracaso de e en el pasado inmediato; $done(e)$ denota la ocurrencia de e en el pasado inmediato (con éxito o fracaso). De manera similar, $succeeds(e)$, $fails(e)$, y $does(e)$ se usan para denotar ocurrencias futuras de e. *Eventos*

4.5.1 Sintaxis de los eventos

Primero, necesitamos un conjunto de símbolos para identificar a los **eventos primitivos**, por ej., E. Ahora, la definición de fórmula de estado debe extenderse con la inclusión de la siguiente fórmula: *Eventos primitivos*

Definición 4.11 (Sintaxis de los Eventos)**.** *Si $e \in E$, entonces $succeeds(e)$, $fails(e)$, $does(e)$, $succeeded(e)$, $failed(e)$, y $done(e)$ son fórmulas de estado.*

4.5.2 Semántica de los eventos

La estructura M en la que basamos la interpretación debe redefinirse para incluir los siguientes elementos: E es un conjunto de tipos de evento primitivos; $SE_w : S_w \times S_w \mapsto E$ y $FE_w : S_w \times S_w \mapsto E$ que representan ocurrencias con éxito y fracaso de los eventos. Observen que SE_w y FE_w son disjuntos.

La semántica de los eventos se define que sigue:

Definición 4.12 (Semántica de los Eventos)**.** *La semántica de los eventos se define por las siguientes dos reglas:*

1. $M, w_{s_1} \models succeeded(e)$ *si y sólo si* $SE_w(s_0, s_1) = e$
2. $M, w_{s_1} \models failed(e)$ *si y sólo si* $FE_w(s_0, s_1) = e$

El resto de los eventos se define como sigue:

$$\begin{aligned} done(e) &\stackrel{\text{def}}{=} succeeded(e) \vee failed(e) \\ succeeds(e) &\stackrel{\text{def}}{=} \mathsf{A} \bigcirc (succeeded(e)) \\ fails(e) &\stackrel{\text{def}}{=} \mathsf{A} \bigcirc (failed(e)) \\ does(e) &\stackrel{\text{def}}{=} \mathsf{A} \bigcirc (done(e)) \end{aligned}$$

4.5.3 Axiomatización de los eventos

Es necesario definir un axioma que capture el carácter volitivo del compromiso subyacente en las intenciones. Este axioma debe expresar que un agente actuará si tiene una intención dirigida hacía un tipo de evento primitivo:

Axioma 4.20. $\mathsf{INT}(does(e)) \rightarrow does(e)$.

Un agente debe ser consciente (debe creer) de todos los tipos de eventos primitivos que ocurren en su medio ambiente:

Axioma 4.21. $done(e) \rightarrow \mathsf{BEL}(done(e))$.

Al elegir uno de los realismos y adoptar el resto de los axiomas introducidos con anterioridad, configuramos lo que Rao y Georgeff [149] llaman un **Sistema Básico I**. *Sistema Básico I*

4.6 COMPROMISO COMO AXIOMAS DE CAMBIO

¿Cómo es que las intenciones determinan el compromiso futuro de un agente hacía sus acciones? Necesitamos especificar formalmente como las intenciones actuales se relacionan con las intenciones futuras. Una alternativa consiste en pensar en esta relación como un proceso de mantenimiento y revisión de intenciones o una estrategia de compromiso. Tres estrategias son bien conocidas en la literatura multi agentes: los compromisos: ciego, racional, y emocional [3]. Eligiendo una de estas tres estrategias y asumiendo el Basic I System, se configuran diferentes agentes básicos.

3 Originalmente estos compromisos se llaman *blind*, *single-minded* y *open-minded* respectivamente. He decido los cambios de denotación en español en base a que la segunda es una

4.6.1 Compromiso ciego (blind)

Compromiso ciego

Un agente se compromete **ciegamente** si mantiene sus intenciones hasta que cree que las ha logrado satisfacer. Formalmente:

Teorema 4.1. $\mathsf{INT}(\mathsf{A}\Diamond\phi) \rightarrow \mathsf{A}(\mathsf{INT}(\mathsf{A}\Diamond\phi)\ \mathsf{U}\ \mathsf{BEL}(\phi))$

Observen que este axioma se define para I-fórmulas. Nada se dice sobre la intención de un agente por lograr opcionalmente algún medio o fin particular. Resulta evidente que esta estrategia es demasiado fuerte. Para un agente básico comprometido ciegamente, resulta inevitable eventualmente creer que ha logrado sus medios (o fines). Esto se debe a que el teorema 4.1 sólo permite caminos futuros en los cuales o bien el objeto de la intención es creído, o la intención se mantiene para siempre. Sin embargo, debido al axioma 4.19 (posposición finita), este último tipo de caminos no está permitido, lo que nos lleva a obtener agentes que creen eventualmente que han logrado sus intenciones:

Teorema 4.2. $\mathsf{INT}(\mathsf{A}\Diamond\phi) \rightarrow \mathsf{A}\Diamond(\mathsf{BEL}(\phi))$

4.6.2 Compromiso racional (single-minded)

Compromiso racional

Se puede definir una estrategia de **compromiso racional** relajando la estrategia ciega, de manera que el agente mantenga sus intenciones en tanto considere que siguen siendo una opción viable. Formalmente:

Teorema 4.3. $\mathsf{INT}(\mathsf{A}\Diamond\phi) \rightarrow \mathsf{A}(\mathsf{INT}(\mathsf{A}\Diamond\phi)\ \mathsf{U}\ (\mathsf{BEL}(\phi) \vee \mathsf{BEL}(\mathsf{E}\Diamond\phi)))$

En tanto el agente crea que sus intenciones se pueden lograr, no abandonará sus intenciones y seguirá comprometido.

Un agente básico inflexible, de manera inevitable, eventualmente creerá que ha logrado satisfacer sus medios (o fines) sólo si continua creyendo que, hasta creer que sus medios (o fines) se han satisfecho, que el objeto de sus intenciones sigue siendo una opción. Formalmente:

Teorema 4.4. $\mathsf{INT}(\mathsf{A}\Diamond\phi) \wedge \mathsf{A}(\mathsf{BEL}(\mathsf{E}\Diamond\phi))\ \mathsf{U}\ \mathsf{BEL}(\phi)) \rightarrow \mathsf{A}\Diamond(\mathsf{BEL}(\phi))$.

4.6.3 Compromiso emocional (open-minded)

Compromiso emocional

Un agente bajo el **compromiso emocional** mantiene sus intenciones mientras éstas sigan siendo deseadas. Formalmente:

Teorema 4.5. $\mathsf{INT}(\mathsf{A}\Diamond\phi) \rightarrow \mathsf{A}(\mathsf{INT}(\mathsf{A}\Diamond\phi)\ \mathsf{U}\ \ (\mathsf{BEL}(\phi) \vee \neg\mathsf{DES}(\mathsf{E}\Diamond\phi)))$

Un agente básico con compromiso emocional, de manera inevitable, eventualmente creerá que ha logrado satisfacer sus medios (o fines) si mantiene sus intenciones como deseos, hasta que cree que ha logrado satisfacerlos. Formalmente:

Teorema 4.6. $\mathsf{INT}(\mathsf{A}\Diamond\phi) \wedge \mathsf{A}(\mathsf{DES}(\mathsf{E}\Diamond\phi))\ \mathsf{U}\ \mathsf{BEL}(\phi) \rightarrow\ \mathsf{A}\Diamond(\mathsf{BEL}(\phi))$.

estrategia creencia-intención y la tercera es deseo-intención. Creo que los nombre de racional y emocional enfatizan mejor el fundamento de la reconsideración, que las opciones de inflexible y flexible.

4.7 OTROS RESULTADOS

Rao y Georgeff [149] consideran también el caso de un **agente competente** que satisface el siguiente axioma: *Agente competente*

Axioma 4.22 (Competencia)**.** $\mathsf{BEL}(\phi) \rightarrow \phi$

Bajo cada una de las estrategias de compromiso mencionadas, el agente competente logrará sus medios (o fines) en lugar de sólo creer que los ha logrado. Sin embargo tener creencias verdaderas siempre suele ser complicado. Esta forma de omnisciencia se puede relajar si se consideran solo las creencias actuales del agente, o bien si ϕ se restringe a ser una fórmula de acción primitiva.

Teorema 4.7. *Bajo las mismas condiciones que el Sistema Básico I y los teoremas 4.2, 4.4 y 4.6, un agente competente llega a la conclusión de que* $\mathsf{A}\Diamond\phi$ *para todos los tipos de estrategia de compromiso descritas en la sección anterior.*

Otro agente a considerar es aquel que siempre lleva a cabo exclusivamente acciones intencionales. Esto puede reforzarse al requerir que el agente intente un tipo de acción primitiva en cada instante de tiempo (estado de los mundos posibles). En un medio ambiente libre de sorpresas el agente mantendría sus creencias tras ejecutar la acción intentada, es decir, no olvidará sus creencias previas a la ejecución.

Podemos afirmar que un agente preservará una creencia γ sobre acerca de una acción intencional (evento) e si y sólo si intenta hacer e y cree que γ será el caso luego de hacer e y entonces luego de hacer γ efectivamente es el caso que e. Formalmente:

Teorema 4.8. $\mathsf{INT}(does(e)) \wedge (\mathsf{BEL}(\mathsf{E}\bigcirc(done(e) \wedge \gamma)) \rightarrow \mathsf{E}\bigcirc(\mathsf{BEL}(done(e) \wedge \gamma))$

Un agente con compromiso racional que intenta inevitablemente que ϕ sea verdadera en el futuro, inevitablemente creerá ϕ , si sólo ejecuta acciones intencionales y preserva sus creencias acerca de ϕ sobre el resultado de sus acciones. Si el agente además es competente, logrará ϕ. Un agente básico con compromiso racional que preserve sus creencias, satisface la siguiente propiedad:

Teorema 4.9. $\mathsf{INT}(\mathsf{A}\Diamond\phi) \wedge \mathsf{A}\Box(\exists e(\mathsf{INT}(does(e)) \wedge (\Diamond\phi))) \rightarrow$
$(\mathsf{BEL}(\mathsf{E}\bigcirc(done(e) \wedge (\Diamond\phi)))) \rightarrow \mathsf{A}\Diamond(\mathsf{BEL}(\phi))$

Un agente competente con compromiso racional que preserva sus creencias, satisface la siguiente propiedad:

Teorema 4.10. $\mathsf{INT}(\mathsf{A}\Diamond\phi) \wedge \mathsf{A}\Box(\exists e(\mathsf{INT}(does(e)) \wedge (\Diamond\phi))) \rightarrow$
$(\mathsf{BEL}(\mathsf{E}\bigcirc(done(e) \wedge (\Diamond\phi)))) \rightarrow \mathsf{A}\Diamond(\phi)$.

4.8 CASO DE ESTUDIO: EL CANDIDATO

Rao y Georgeff [148] han extendido las lógicas BDI_{CTL*} para introducir probabilidades y recompensas subjetivas en el proceso de deliberación, en un

intento de reconciliar la toma de decisión clásica con la semántica BDI basada en mundos posibles. Retomaré un ejemplo de este trabajo, para ilustrar como es que un proceso de decisión sugiere los mundos posibles de un agente.

Árboles de decisión

La teoría de la decisión suele representar los procesos de deliberación como **árboles de decisión** compuestos por tres tipos de nodos:

- Nodos de decisión. Representan la elección de las acciones del agente.
- Nodos de posibilidad. Representan la incertidumbre en el ambiente.
- Nodos de utilidad. Representan la recompensa de la decisión.

Chucho es un buen agente que actualmente es diputado que cree que puede reelegirse en la cámara baja (dip); bien presentarse al senado (sen); o bien retirarse (ret). En realidad, Chucho no considera seriamente la opción de retirarse y está seguro de reelegirse en la cámara de diputados. La decisión que Chucho debe tomar es si recurrir o no a una encuesta para saber si se presenta o no al senado. El resultado de la encuesta puede ser que la ciudadanía aprueba que Chucho se presente al senado (si) o (no). El resultado de competir para llegar al senado ser ganador ($gano$) o perdedor ($\neg gano$). El árbol de decisión resultante, obviando el hecho de que Chucho no quiere retirarse, se muestra en la figura 4.3. Los nodos terminales expresan que Chucho obtiene una utilidad de 300 si llega al senado, 200 si se mantiene como diputado y 100 si pierde en su afán de ser senador. Las probabilidades subjetivas y condicionales pueden verse también ahí, por ejemplo la probabilidad de Chucho de llegar al senado, dado que la encuesta arrojó un si, es de 0.571.

Del árbol de decisión a mundos posibles

Estos árboles de decisión pueden **convertirse** en conjuntos de mundos posibles de la siguiente manera. Partiendo del nodo raíz del árbol de decisión, se recorren todos los arcos. Para cada estado único etiquetando un arco que emana de un nodo de posibilidad, creamos un nuevo árbol de decisión que es idéntico al original, exceptuando: a) El nodo de posibilidad es removido del árbol; b) el arco que incidía en el nodo de posibilidad es conectado con su sucesor. Este proceso se continua recursivamente, hasta que eliminar todos los nodos de posibilidad.

El árbol completo tiene una probabilidad $\alpha = 1$, conforme se va procesando el árbol de decisión, los nuevos mundos creados tienen una probabilidad igual al producto ponderado de las probabilidades de los nodos de posibilidad que se han procesado. Los mundos resultantes en la primera llamada de este proceso se muestran el la figura 4.4.

La figura 4.5 muestra, adornos más, adornos menos, los cuatro mundos posibles creídos por Chucho a partir de su situación inicial w_0. Esta transformación tiene la propiedad de conservar la información del árbol de decisión original, en la interpretación obtenida basada en mundos posibles. Si asumimos, por ejemplo, el realismo fuerte, se pueden obtener los mundos posibles deseados (por definición submundos de los creídos) e intentados (por definición submundos de los deseados).

De esta manera es posible saber que la fbf $\mathsf{BEL}(\Diamond done(sen))$ es satisfacible en el estado inicial del agente. De hecho, la extensión propuesta por

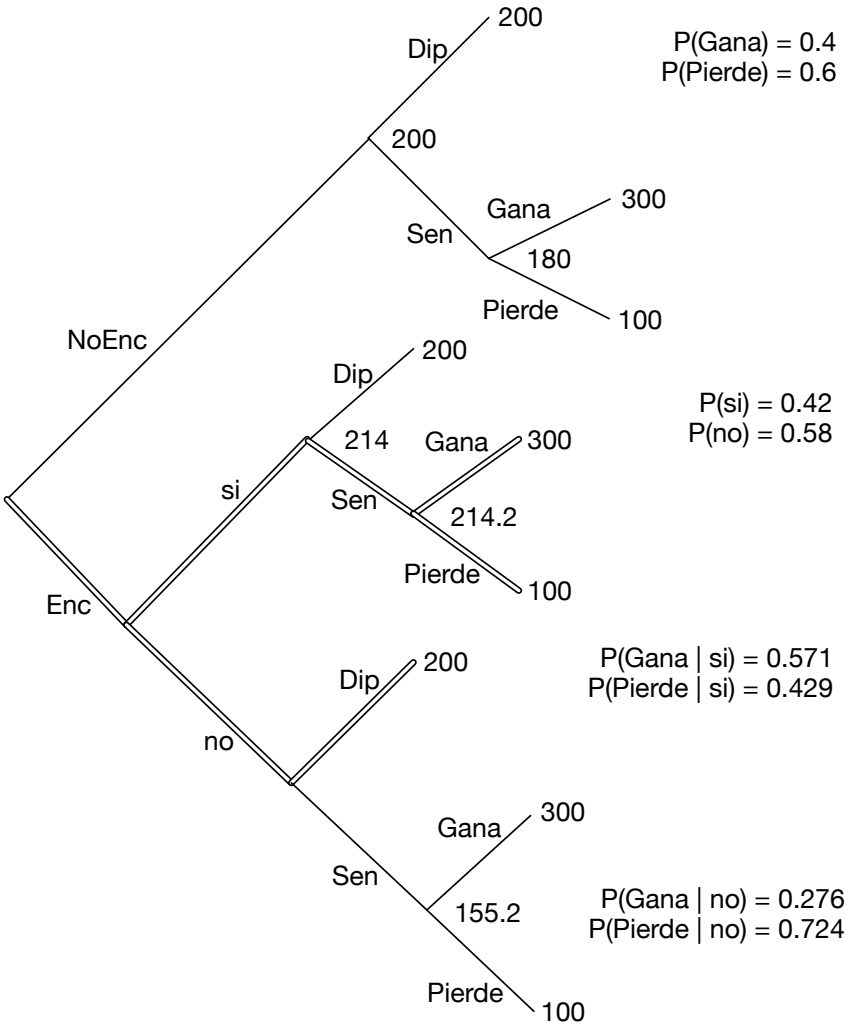

Figura 4.3: El árbol de decisión del agente político con sus probabilidades y recompensas subjetivas incluídas.

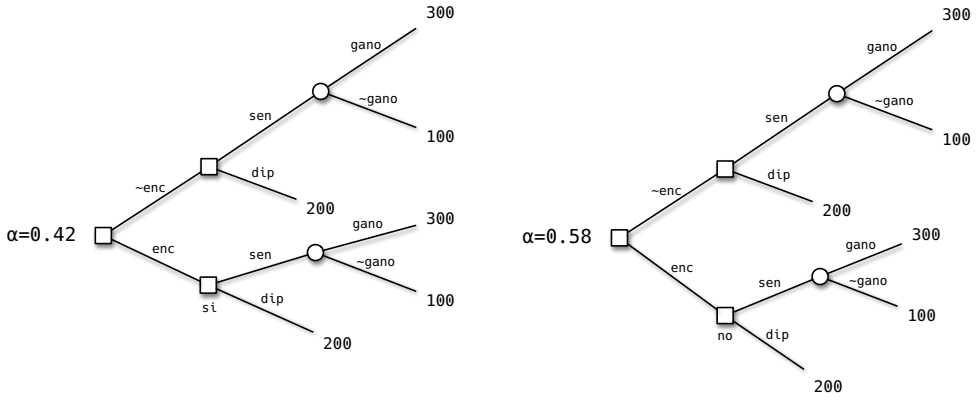

Figura 4.4: Los primeros dos mundos posibles obtenidos a partir del árbol de decisión original.

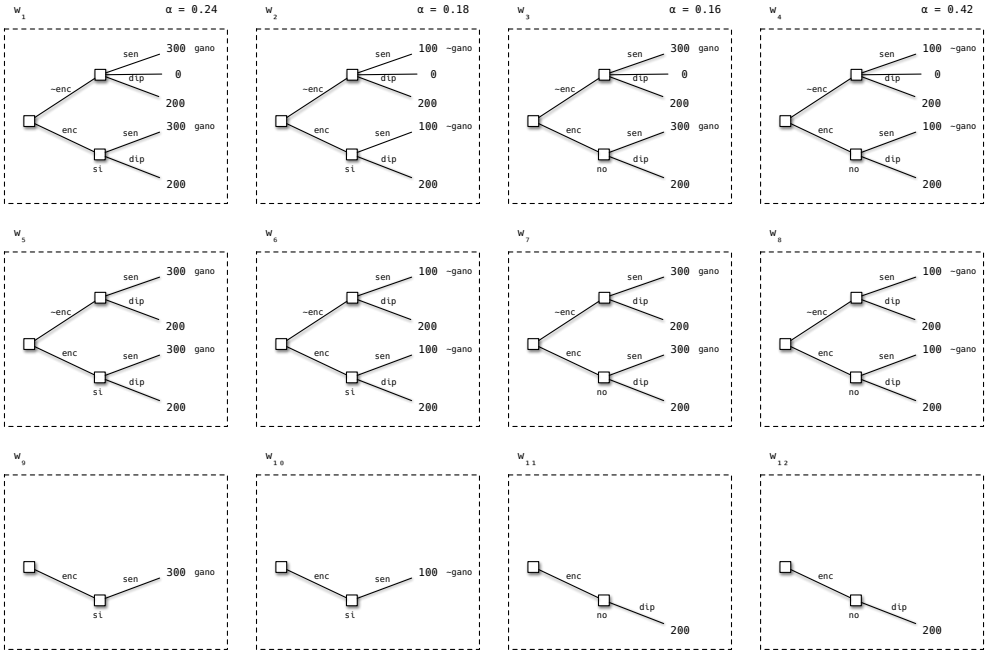

Figura 4.5: Los mundos posibles creídos, deseados e intentados a partir del árbol de decisón original, bajo realismo fuerte.

Rao y Georgeff permite saber que $PROB(\mathsf{E}(\Diamond si)) = 0{,}42$, esto es la probabilidad de evantualmente recbir un si en la encuesta es de 0,42. También es posible expresar que la recompensa de llegar a ser senador es de $PAYOFF(\Diamond(done(sen) \wedge gano)) = 300$. O que el agente desea ser senador $\mathsf{DES}(\Diamond sen)$.

4.9 LECTURAS Y EJERCICIOS SUGERIDOS

El trabajo fundacional sobre formalismos para razonar acerca de las intenciones se debe a Cohen y Levesque [40], sin embargo, ellos optan por una teoría de intención bajo el reduccionismo creencias-deseos [4]. Otro formalismo reportado en la literatura es LORA, la lógica BDI propuesta por Wooldridge [185]. Singh, Rao y Georgeff [175] ofrecen una recapitulación de los métodos formales aplicados a los Sistemas Multi-Agentes. Van der Hoek y Wooldridge [101, 102] hacen una revisión de diferentes lógicas BDI desde la perspectiva de la representación del conocimiento. Más recientemente, Meyer, Broersen y Herzig [127] presenta una revisión de las lógicas BDI, incluyendo su concepción *à la* Bratman, el enfoque de Cohen y Levesque, el de Rao y Georgeff, KARO y STIT.

4 Una crítica más detallada de este trabajo, puede encontrarse en la nota de Singh [174]. En lo que a nosotros respecta, basta observar que tal aproximación a las intenciones no sigue los principios de racionalidad práctica expuestos por Bratman [22]. Cabe señalar que el trabajo en cuestión es impecable formalmente, y es por ello que ha sido posible revisarlo.

Ejercicios

Ejercicio 4.1. *Escriba un ejemplo de fbf de BDI_{CTL*} que no lo sea en BDI_{CTL}.*

Ejercicio 4.2. *Usando el escenario del caso de estudio, de un ejemplo de fbf opcional e inevitable en BDI_{CTL}.*

Ejercicio 4.3. *Ejemplifique con esquemas las relaciones totales, seriales, transitivas y euclidianas.*

Ejercicio 4.4. *De un ejemplo de realismo, realismo fuerte y realismo débil usando el vocabulario del caso de estudio.*

Ejercicio 4.5. *De un ejemplo del tipo de razonamiento al que llegaría el agente del caso de estudio bajo las diferentes formas de compromiso: ciego, racional, emocional.*

Ejercicio 4.6. *Usando la figura 4.5, demuestre que $w_1 \sqsubseteq w_5$.*

Ejercicio 4.7. *Demuestre que el agente del caso de estudio, no satisface la fbf* $\mathsf{DES}(\Box(done(sen) \wedge gano))$.