

# Un Modelo Abstracto de Diálogo Sobre Creencias para Sistemas Multiagente

M. Julieta Marcos   Marcelo A. Falappa   Guillermo R. Simari

Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET),

Laboratorio de Investigación y Desarrollo en Inteligencia Artificial,

Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur,

Avenida Alem 1253,(B8000BCP), Bahía Blanca, Argentina

Tel: (0291) 459-5135 / Fax: (0291) 459-5136

Email: {mjm,mfalappa,grs}@cs.uns.edu.ar

## Resumen

Este trabajo muestra una relación entre dos áreas de investigación en Inteligencia Artificial: el modelamiento de *Diálogos en Sistemas Multiagente* por un lado, y la *Teoría de Cambio de Creencias* por el otro. Presentamos un modelo abstracto de diálogo sobre creencias, basado en operadores de cambio no priorizado. La abstracción se refiere tanto al sistema de razonamiento interno utilizado por los agentes, como al tipo particular de diálogo que se quiera modelar.

Básicamente, vemos al diálogo como un proceso mediante el cual los agentes provocan sucesivos cambios sobre una base de conocimiento pública (que representa el *estado* del diálogo). Los agentes tienen *metas* que dictan qué conocimiento exponer en determinado momento.

El modelo impone ciertas restricciones, como por ejemplo que las bases de conocimiento privadas de los agentes no se modifiquen durante el diálogo, y que todos los agentes tengan el mismo grado de credibilidad o autoridad.

**Palabras Clave:** sistemas multiagente, diálogos entre agentes, cambio de creencias.

## 1. INTRODUCCIÓN

En un sistema multi-agente los agentes necesitan comunicarse, por diferentes motivos: resolver diferencias de opinión o intereses en conflicto, cooperar para resolver dilemas o encontrar pruebas, o simplemente informarse uno a otro sobre hechos pertinentes. En muchos casos no alcanza con intercambiar mensajes aislados, sino que los agentes necesitan entablar *diálogos* (secuencias de mensajes sobre el mismo tema) [10]. Además, se puede mejorar la calidad de la interacción si los agentes exponen los *argumentos* que justifican lo que dicen [11], es decir, entablan diálogos *basados en argumentación*.

Existe una gran variedad de interacciones con características diferentes que podrían quererse modelar. Una posible tipología, teniendo en cuenta el objetivo común del diálogo y las metas particulares de cada participante, es la siguiente [12]:

**Diálogo de Búsqueda de Información.** Un agente busca la respuesta a una pregunta en el conocimiento de otro agente. Se supone que este último conoce la respuesta.

**Diálogo de Investigación.** Todos los agentes colaboran para encontrar la respuesta a una pregunta. Se supone que ninguno de ellos conoce la respuesta.

**Diálogo Persuasivo.** Un agente trata de convencer a otro para que se adhiera a cierta creencia o punto de vista.

**Negociación.** Los agentes tratan de llegar a un acuerdo aceptable sobre la división de recursos escasos. Cada uno trata de maximizar su ganancia. La meta del diálogo puede estar en conflicto con las metas individuales de los agentes.

**Diálogo Deliberativo.** Los agentes colaboran para decidir que acción realizar en cierta situación.

Se puede hacer una distinción entre diálogos *colaborativos* y diálogos *no-colaborativos*. En un diálogo colaborativo los agentes no tienen metas individuales más allá de la meta común del diálogo; por lo tanto, todos colaboran en aras del mismo fin. En un diálogo no-colaborativo, en cambio, los agentes tienen metas individuales que podrían estar en conflicto. Por ejemplo, un diálogo de investigación es colaborativo (el único objetivo compartido por todos los agentes investigadores es descubrir la verdad) pero una negociación no lo es (cada agente negociador tiene como meta maximizar su propia ganancia). Un diálogo deliberativo podría ser colaborativo o no (dependiendo de si los agentes tienen algún interés particular por tomar cierto curso de acción). Un diálogo persuasivo puede ser visto como un diálogo *semi-colaborativo*, donde el agente que persuade tiene una meta particular, pero el persuadido podría no tenerla. Los *agentes colaborativos* expondrán toda la información que consideren relevante, mientras que los *agentes no-colaborativos* podrían ocultar información, sabiendo que es relevante, porque no favorece el cumplimiento de sus metas individuales.

Otra posible diferenciación es entre aquellos diálogos que son *sobre creencias*, y aquellos que no lo son. En un diálogo sobre creencias los participantes hablan sobre la verdad de cierta proposición. A esta categoría corresponden los primeros tres tipos de diálogo: búsqueda de información, investigación y diálogo persuasivo. Sin embargo, un diálogo persuasivo no es *necesariamente* sobre creencias (podría ser, por ejemplo, sobre acciones). Este trabajo está dedicado principalmente a *diálogos sobre creencias*.

Se han realizado varios trabajos con el objetivo de modelar formalmente estas interacciones. Sin embargo, las soluciones propuestas son *ad hoc* y carecen de una fundamentación teórica sólida. En [11], por ejemplo, se investiga un tipo particular de diálogo: la *negociación basada en argumentación*, identificando y describiendo elementos necesarios para su modelamiento (tanto internos como externos a los agentes). En [10], por otro lado, se concentran en *diálogos de investigación*, de *búsqueda de información* y *persuasivos*. Definen un conjunto de *locuciones* para que los agentes puedan intercambiar argumentos, un conjunto de *actitudes* que marcan una relación entre los argumentos que puede construir un agente y las locuciones que puede realizar (intuitivamente, los agentes “menos atrevidos” sólo afirman proposiciones soportadas por “buenos argumentos”), y definen también un conjunto de *protocolos* para llevar a cabo los diálogos. Creemos que una falencia del trabajo citado es que construyen el modelo de diálogo sobre la base de un sistema argumentativo particular.

El objetivo de nuestro trabajo es mostrar como puede construirse un modelo (limitado) de diálogo con un mayor nivel de abstracción, en cuanto al sistema de razonamiento interno utilizado por los agentes. Para ello, utilizaremos el basamento teórico que brinda la *Teoría de Cambio de Creencias*. La misma estudia la *dinámica del conocimiento* en agentes o mundos, es decir, los cambios provocados en una base de conocimiento por el arribo de nueva información. De esta manera, modelaremos el diálogo como un proceso mediante el cual los agentes provocan sucesivos cambios sobre una base de conocimiento pública (que representa el *estado* del diálogo).

El trabajo está estructurado de la siguiente manera. La Sección 2 repasa conceptos básicos de la Teoría de Cambio de Creencias. En la Sección 3 presentamos un modelo parametrizable

de diálogo, basado en operadores de cambio. Se presenta un algoritmo de diálogo, y se analizan algunos aspectos importantes y sus limitaciones. Finalmente, en la Sección 4, se muestra cómo puede utilizarse el algoritmo propuesto para modelar diferentes tipos de interacción.

## 2. TEORÍA DE CAMBIO DE CREENCIAS

La *Teoría de Cambio de Creencias* estudia la dinámica del conocimiento, esto es, la forma en que se actualiza el conocimiento de un agente después de que recibe información nueva. Un *estado epistémico* es una representación del conocimiento de un agente en un momento del tiempo. Existen, principalmente, dos alternativas para representar estados epistémicos: *conjuntos de conocimiento* [1] o *bases de conocimiento* [6]. Los primeros son conjuntos clausurados bajo algún operador de consecuencia lógica. Los segundos son conjuntos no clausurados, y son los que utilizaremos en este trabajo. Una *actitud epistémica* describe el estado de varios elementos del conocimiento que están contenidos en un estado epistémico. En los modelos clásicos de la Teoría de Cambio se consideran tres actitudes epistémicas: *aceptación*, *rechazo* e *indeterminación*. Una *entrada epistémica* es una pieza de información externa que puede producir *cambios* en un estado epistémico. Las operaciones de *cambio epistémico* que utilizaremos en este trabajo son: *expansión* [1] notada con “+”, *mezcla* [4] notada con “o” y *consolidación* [8] notada con “!”. El significado intuitivo de cada una de ellas es el siguiente:

**Expansión.** Se incorpora conocimiento sin importar si el estado resultante es consistente.

**Mezcla.** Se combinan dos estados de conocimiento, buscando que el resultado sea consistente.

**Consolidación.** Se eliminan inconsistencias de un estado de conocimiento.

La *expansión* es la operación más simple. Cuando el estado epistémico se representa con bases, una expansión consiste en una simple unión de conjuntos. Si  $K$  es una base de creencias y  $\alpha$  una entrada epistémica, entonces la expansión se define como  $K + \alpha = K \cup \{\alpha\}$  [6].

La *consolidación* es, en realidad, un caso particular de otra operación: la *contracción* [1]. La operación de contracción elimina una creencia de un estado de conocimiento. En una consolidación la creencia a eliminar es  $\perp$  (la contradicción). Entre varios tipos de contracciones, nos enfocaremos en: *Partial Meet Contraction* [1] y *Kernel Contraction* [7]. En base a éstas se definen dos formas de consolidación [8]: *Partial Meet Consolidation* y *Kernel Consolidation*.

Las consolidaciones de tipo *Partial Meet* se basan en *subconjuntos maximales consistentes* y *funciones de selección*. Sea  $K$  la base a consolidar. Una *función de selección* selecciona uno o más de todos los subconjuntos maximales consistentes de  $K$ . Luego, se define la *consolidación partial meet* como la intersección de todos los subconjuntos elegidos por la función de selección.

Las consolidaciones de tipo *Kernel* se basan en *subconjuntos minimales inconsistentes* y *funciones de incisión*. Sea  $K$  la base a consolidar. Una *función de incisión* selecciona una o más creencias para eliminar de cada subconjunto minimal inconsistente de  $K$ , buscando restaurar la consistencia del mismo. Luego, se define la *consolidación kernel* como la base original sin las creencias seleccionadas por la función de incisión.

Existen algunas propiedades intuitivas que deberían ser satisfechas por un operador de consolidación [8]:

- *Inclusión.* Para toda base  $K$ , debe ser  $K! \subseteq K$ .
- *Consistencia.*  $K!$  debe ser consistente.

- *Relevancia y Retención de Núcleo.* Buscan captar (en menor y mayor medida, respectivamente) la noción de no eliminar de más, es decir, no eliminar creencias que no contribuyen a que la base sea inconsistente.

Luego, pueden definirse las operaciones de consolidación *partial meet* y *kernel* en función de las propiedades anteriores [8]:

- $!$  es un operador de *partial meet consolidation* si y solo si, para toda base  $K$ ,  $!$  satisface *inclusión, consistencia y relevancia*.
- $!$  es un operador de *kernel consolidation* si y solo si, para toda base  $K$ ,  $!$  satisface *inclusión, consistencia y retención de núcleo*.

La propiedad de *relevancia* implica la propiedad de *retención de núcleo*. Por esta razón, todo operador de *partial meet consolidation* es también un operador de *kernel consolidation*. No es cierta la relación inversa.

La operación de *mezcla* puede definirse en términos de la operación de consolidación, y viceversa. Si deseamos combinar una base  $K$  con una nueva base  $H$ , se define la mezcla como  $K \circ H = (K \cup H)!$  [4]. Notemos que no hay inconveniente en hacer una mezcla de bases previamente inconsistentes. También puede definirse la consolidación de  $K$ , en términos de una mezcla, como  $K! = (K \circ K) = (K \circ \emptyset)$  [4].

La mezcla es un operador de cambio *no priorizado*. Esto significa que no se asigna ninguna prioridad especial al conocimiento nuevo, por lo que podría o no pertenecer al resultado. Otros operadores de cambio no priorizado son, por ejemplo, la *semi-revisión* definida en [8], la *revisión por conjuntos de sentencias* definida en [2], la *revisión selectiva* definida en [3], la *screened revision* definida en [9], etc.. Existe otro tipo de operadores, llamados operadores de cambio *priorizado*, mediante los cuales se asegura que el nuevo conocimiento pertenecerá al estado resultante, como por ejemplo el operador de *revisión* definido en [1]. Estos últimos no se presentarán porque no serán utilizados en este trabajo.

### 3. UN MODELO DE DIÁLOGO BASADO EN OPERADORES DE CAMBIO

En el modelo que presentaremos, el diálogo se desarrolla en torno a una base de conocimiento pública (nos referimos a ésta como el *estado* del diálogo) que contiene todo el conocimiento expuesto por los agentes hasta el momento. Una *consolidación* del estado del diálogo representa el consenso alcanzado en un momento dado (nos referimos a ésta como el *estado consensuado* del diálogo). Los agentes exponen conocimiento modificando el estado actual, y tienen *metas* que dictan qué conocimiento exponer para lograr el efecto deseado en el estado consensuado del diálogo. Asumiremos los siguientes elementos para construir el modelo:

1. Un *lenguaje de representación de conocimiento*  $L$  (al menos proposicional), junto con una *noción de consistencia* de una base de conocimiento  $K \subseteq L$ , y un *mecanismo de inferencia* ( $\vdash$ ) para derivar conclusiones a partir de subconjuntos de  $L$ .
2. *Operadores de cambio* definidos sobre bases de conocimiento  $K \subseteq L$ : un operador “+” de *expansión*, un operador “!” de *kernel consolidation*, y el operador “o” de *mezcla* asociado.
3. Un *conjunto de metas*  $S_G$  (la noción de *meta* es tratada de manera abstracta).

A continuación introducimos algunas nociones preliminares:

**Definición 1 (Agente de Diálogo)** Un *agente de diálogo* es un par  $A = (K, G)$ , donde  $K \subseteq L$  es la base de conocimiento privada del agente (asumimos que  $K$  es consistente) y  $G \in S_G$  es la meta del agente en el diálogo.

**Definición 2 (Estado del Diálogo)** Un *estado del diálogo* es un conjunto  $E \subseteq L$  posiblemente inconsistente.

**Definición 3 (Criterio de Éxito)** Un *criterio de éxito* es una función booleana  $\varphi$  que toma como entrada un agente de diálogo  $A = (G, K)$  y un estado  $E$ . Decimos que la meta del agente  $A$  *se cumple según*  $\varphi$  en el estado  $E$  si  $\varphi(A, E) = \text{verdadero}$ , y que *no se cumple* si  $\varphi(A, E) = \text{falso}$ .

Cuando no haya lugar a dudas sobre el criterio de éxito usado, diremos directamente que la meta del agente  $A$  *se cumple* o *no se cumple* en el estado  $E$ .

**Definición 4 (Entorno de Diálogo)** Un *entorno de diálogo* es un par  $(S_A, \varphi)$  donde  $S_A$  es un conjunto de agentes de diálogo y  $\varphi$  es un criterio de éxito.

La idea básica es que un agente  $A_i$  buscará en su base de conocimiento privada un subconjunto minimal capaz de expandir el estado actual  $E$  y obtener un nuevo estado en el que su meta  $G_i$  se cumpla. El diálogo termina cuando cada agente o bien alcanza su meta o bien descubre que no tiene medios para alcanzarla (es decir, no existe un subconjunto de su conocimiento capaz de hacer una expansión exitosa). El Algoritmo 1 describe como se desarrollaría un diálogo entre los agentes  $A_1 \dots A_n$  con bases de conocimiento  $K_1 \dots K_n$  y metas  $G_1 \dots G_n$ .

---

**Algoritmo 1 : Diálogo entre los Agentes  $\{A_1 = (K_1, G_1), \dots, A_n = (K_n, G_n)\}$**

---

- 1:  $E \leftarrow \emptyset$
  - 2: **Repetir**
  - 3:  $A_i \leftarrow$  algún agente del conjunto  $\{A_1 \dots A_n\}$  tal que  $G_i$  no se cumple en  $E$ , pero existe  $X \subseteq K_i$  ( $X$  minimal) tal que  $G_i$  se cumple en  $E + X$
  - 4:  $E \leftarrow E + X$
  - 5: **Hasta** que no exista tal agente  $A_i$
  - 6: **Retornar**  $E!$
- 

Podemos ver que hay elementos sin especificar en este algoritmo. La noción de *meta* y el *criterio de éxito* de una meta en un estado son tratados en forma abstracta porque dependen del tipo de diálogo que se quiera modelar. En la sección siguiente mostraremos una forma de definir estos elementos para modelar algunos tipos de diálogos.

El operador de consolidación adecuado también depende de las características particulares de la interacción que se quiera modelar, por eso su implementación no es especificada en el modelo. El modelo sí especifica que el cambio es *no priorizado*. Sin embargo, existen algunas situaciones en las que sería más adecuado un operador de cambio *priorizado*. Supongamos que un agente  $A$  tienen más autoridad o credibilidad que otro agente  $B$ , entonces el conocimiento que expone  $A$  debería tener prioridad con respecto al conocimiento que expone  $B$ . Por otro lado, supongamos que un agente rectifica su conocimiento durante el diálogo porque recibió percepciones (provenientes de una fuente externa al diálogo) más acertadas, entonces las nuevas

creencias expuestas por el agente deberían tener más prioridad que las creencias previas expuestas por ese mismo agente. Por lo tanto, el uso de cambio no priorizado en este trabajo se justifica en las siguientes suposiciones:

1. Todos los agentes tienen el mismo grado de autoridad (o credibilidad).
2. Los agentes no reciben nuevas percepciones (externas al diálogo) mientras dialogan. Es decir, sus bases de conocimiento internas no se modifican durante el diálogo.

Otro aspecto observable en este algoritmo es que es *no determinístico*, por dos razones: pueden existir varios agentes  $A_i$ , y varios conjuntos  $X \subseteq K_i$ , que verifican la condición del paso 3. Diferentes caminos pueden conducir a resultados distintos. Si bien los agentes no se turnan estrictamente para hablar, se puede asegurar que un agente nunca hablará dos veces consecutivas (ya que su meta se cumple inmediatamente después de hablar).

La ejecución termina cuando se alcanza un estado en el que cada agente, o bien cumple su meta o bien descubre que no puede cumplirla.

**Definición 5 (Estado Final)** Un estado  $E$  es un *estado final* para un entorno de diálogo  $(S_A, \varphi)$  si y solo si para cada agente  $A_i = (K_i, G_i) \in S_A$  se verifica alguna de las condiciones siguientes:

1.  $G_i$  se cumple en  $E$ , o bien
2. No existe  $X \subseteq K_i$  tal que  $G_i$  se cumpla en  $E + X$

Es fácil ver que el algoritmo siempre termina. En un caso extremo, los agentes exponen la totalidad de su conocimiento, alcanzando un estado final. El resultado devuelto por el algoritmo es la consolidación del último estado alcanzado, y esto debe interpretarse como el consenso al que llegaron los agentes mediante el diálogo. La siguiente definición relaciona un entorno de diálogo con un posible resultado devuelto por el algoritmo.

**Definición 6 (Diálogo)** Un *diálogo* es un par  $(\Gamma, E)$  donde  $\Gamma$  es un entorno de diálogo y  $E$  es un estado final para  $\Gamma$ .

El estado del diálogo mantiene *todo* el conocimiento que ha sido expuesto (notemos que en el paso 4 del algoritmo el estado  $E$  se modifica mediante una *expansión*). Hacer *expansiones* en lugar de *mezclas* brinda dos ventajas: (1) los agentes pueden hacer uso implícito del conocimiento publicado por otros con anterioridad, y (2) se controla implícitamente que los agentes no cometan *falacias* [5] en el diálogo.

Con respecto al estado inicial del diálogo (llamémoslo  $E_0$ ), si bien en el algoritmo anterior asumimos  $E_0 = \emptyset$ , esto podría ser de otra manera si resultara más adecuado. Supongamos, por ejemplo, que los agentes recuerdan diálogos pasados. Entonces podrían comenzar el diálogo con un conjunto no vacío de conocimiento público, proveniente de diálogos anteriores entre esos mismos agentes.

Finalmente, podemos destacar algunas restricciones impuestas por el modelo:

1. El *conocimiento público* es un subconjunto del *conocimiento privado* de los agentes. Es decir, que los agentes no dicen nada que no forme parte explícitamente de sus bases de conocimiento privadas.
2. Los agentes no pueden retractarse arbitrariamente sobre locuciones pasadas, sino sólo a través del operador *no priorizado* de mezcla.

3. No se modifican las bases de conocimiento privadas de los agentes. El conocimiento que expone un agente sólo se ve reflejado en el estado público del diálogo, pero no afecta a las bases privadas de otros agentes.

Creemos que estas restricciones no impiden modelar con naturalidad la mayoría de los diálogos sobre creencias, pero podrían ser demasiado fuertes para otros tipos de diálogo, como la negociación o el diálogo deliberativo.

## 4. MODELANDO DIFERENTES TIPOS DE DIÁLOGO

El Algoritmo 1 intenta capturar el esquema general de cualquier tipo de diálogo: en un momento dado (un estado del diálogo) un participante considera que debe exponer cierto conocimiento, entonces lo hace y se produce un cambio de estado. Creemos que la diferencia entre distintos tipos de diálogo radica, en parte, en el criterio usado por los participantes para determinar si tienen algo para decir (y para elegir qué decir); y usamos la noción de *meta* para representar ese criterio. Veremos como pueden modelarse algunos tipos de diálogo de los presentados en la Sección 1 definiendo de manera adecuada las metas de los agentes. Por ejemplo, en un diálogo persuasivo la meta del agente que persuade podría ser una sentencia  $\alpha \in L$  y el criterio de cumplimiento, para un estado  $E$ , podría ser  $E! \vdash \alpha$ . En este trabajo usaremos la siguiente representación de metas:

**Definición 7 (Meta)** Una *meta* es un par  $(\alpha, actitud)$ , con  $\alpha \in L$  y  $actitud \in \{\oplus, \ominus, ?\}$ , donde  $\oplus$  representa una actitud *a favor* de  $\alpha$ ,  $\ominus$  representa una actitud *en contra* de  $\alpha$ , y  $?$  representa una actitud *imparcial* con respecto a  $\alpha$ .

De esta manera queda definido un conjunto  $S_G$  de metas para un lenguaje  $L$ . En lo que resta del trabajo, asumiremos siempre este conjunto de metas.

No deben confundirse estas actitudes de los agentes en el diálogo con las *actitudes epistémicas* mencionadas en la Sección 2. Si bien existe cierta relación entre las actitudes epistémicas de aceptación/rechazo y las metas  $(\alpha, \oplus)/(\alpha, \ominus)$ , no sucede lo mismo con la actitud epistémica de indeterminación y la meta  $(\alpha, ?)$ . La siguiente definición clarifica el significado de las actitudes mencionadas en la Definición 7.

**Definición 8 (Primer Criterio de Exito)** Definimos el siguiente criterio  $\varphi_1$  de cumplimiento de metas:

- una meta  $G=(\alpha, \oplus)$  se cumple en un estado  $E \subseteq L$  si y solo si  $E! \vdash \alpha$
- una meta  $G=(\alpha, \ominus)$  se cumple en un estado  $E \subseteq L$  si y solo si  $E! \vdash \neg\alpha$
- una meta  $G=(\alpha, ?)$  de un agente  $A = (K, G)$  se cumple en un estado  $E \subseteq L$  si y solo si se verifican las siguientes condiciones:
  1.  $E! \vdash \alpha \Leftrightarrow (E \circ K) \vdash \alpha$
  2.  $E! \vdash \neg\alpha \Leftrightarrow (E \circ K) \vdash \neg\alpha$

**Observación 1** Dada una sentencia  $\alpha$ , una meta  $(\alpha, ?)$  es equivalente a una meta  $(\neg\alpha, ?)$ , y una meta  $(\alpha, \ominus)$  es equivalente a una meta  $(\neg\alpha, \oplus)$ .

Intuitivamente, si un agente tiene una meta  $(\alpha, \oplus)$  significa que tiene una inclinación particular por concluir que la sentencia  $\alpha$  es verdadera. Por el contrario, si un agente tiene una meta  $(\alpha, \ominus)$  significa que tiene una inclinación particular por concluir que la sentencia  $\alpha$  es falsa. Por último, si un agente tiene una meta  $(\alpha, ?)$  significa que su actitud en el diálogo es imparcial y su único objetivo es averiguar la verdad sobre la sentencia  $\alpha$ . Las metas  $(\alpha, \oplus)$  y  $(\alpha, \ominus)$  son adecuadas para agentes no-colaborativos, mientras que las metas  $(\alpha, ?)$  son adecuadas para agentes colaborativos.

Nos concentraremos en diálogos en los cuales todos los agentes tienen metas referidas a la misma sentencia  $\alpha$ , como se define a continuación.

**Definición 9 (Entorno de Diálogo Sobre  $\alpha$ )** Sea  $\alpha \in L$ .  $(S_A, \varphi)$  es un *entorno de diálogo sobre  $\alpha$*  si y solo si  $G_i = (\alpha, actitud_i)$  para todo agente  $A_i \in S_A$ .

**Definición 10 (Diálogo Sobre  $\alpha$ )**  $(\Gamma, E)$  es un *diálogo sobre  $\alpha$*  si y solo si  $\Gamma$  es un entorno de diálogo sobre  $\alpha$ .

Las metas  $(\alpha, ?)$  apuntan (aunque no siempre lo logran, como veremos más adelante) a obtener diálogos con una propiedad especial: que las conclusiones resultantes coincidan con lo que se concluiría de la unión consolidada de las bases de conocimiento privadas de los agentes. Formalizamos esta propiedad de los diálogos con la siguiente definición.

**Definición 11 (Diálogo Completo)** Un diálogo sobre  $\alpha$   $(\Gamma, E)$ , con  $\Gamma = (\{A_1 = (K_1, G_1), \dots, A_n = (K_n, G_n)\}, \varphi)$ , es un *diálogo completo* si y solo si se verifican las siguientes condiciones:

1.  $E! \vdash \alpha \Leftrightarrow (\bigcup_{1 \leq i \leq n} \{K_i\})! \vdash \alpha$
2.  $E! \vdash \neg\alpha \Leftrightarrow (\bigcup_{1 \leq i \leq n} \{K_i\})! \vdash \neg\alpha$

Ahora podemos definir formalmente algunos tipos de diálogo:

**Definición 12 (Diálogo de Investigación)** Sea  $\alpha \in L$ . Un *Diálogo de Investigación* (sobre  $\alpha$ ) es un diálogo sobre  $\alpha$  en el cual todos los agentes participantes tienen la misma meta  $(\alpha, ?)$ .

**Definición 13 (Diálogo de Búsqueda de Información)** Un *Diálogo de Búsqueda de Información* (sobre  $\alpha$ ) es un Diálogo de Investigación (sobre  $\alpha$ ) en el cual existe por lo menos un agente participante  $A_i = (K_i, G_i)$  tal que: o bien  $K_i \vdash \alpha$  o bien  $K_i \vdash \neg\alpha$ .

**Definición 14 (Diálogo Persuasivo)** Sea  $\alpha \in L$ . Un *Diálogo Persuasivo* (sobre  $\alpha$ ) es un diálogo sobre  $\alpha$  en el cual existe por lo menos un agente participante  $A_i = (K_i, G_i)$  tal que  $G_i = (\alpha, actitud_i)$  y  $actitud_i \neq ?$ .

Esta definición no respeta exactamente la caracterización previa (Sección 1) de Diálogo de Investigación, en la cual se menciona que ningún participante conoce la respuesta a la pregunta en cuestión. Decidimos adoptar esta visión más general y ver al Diálogo de Búsqueda de Información como un caso particular de Investigación.

Las metas y diálogos así definidos son solamente ejemplificaciones sobre como podrían modelarse algunas interacciones con el modelo propuesto. A continuación, se ilustra con ejemplos de Diálogos de Investigación y Diálogos Persuasivos el funcionamiento del Algoritmo 1, y se muestran algunos problemas que podrían surgir. Por simplicidad consideramos diálogos entre dos agentes y utilizamos *Lógica Clásica Proposicional* como lenguaje de representación de



conocimiento. Las consolidaciones son realizadas de manera arbitraria, dado que esta simplificación no afecta la ilustratividad de los ejemplos. La representación de metas es de acuerdo a las Definiciones 7 y 8.

**Ejemplo 1** Supongamos un *Diálogo de Investigación* en un entorno  $(\{A_1 = (K_1, G_1) \text{ y } A_2 = (K_2, G_2)\}, \varphi_1)$  con:

$$K_1 = \{a, a \rightarrow c, b \rightarrow c, c \rightarrow d\} \quad K_2 = \{a \wedge c \rightarrow \neg d, b \wedge c \rightarrow d\} \quad G_1 = G_2 = (d, ?)$$

Supongamos además que el operador ! de *partial meet consolidation* se comporta de la manera especificada a continuación, cuando es aplicado al siguiente estado:

$$\{a, a \rightarrow c, c \rightarrow d, a \wedge c \rightarrow \neg d\}! \vdash \neg d \quad (1)$$

De acuerdo a esto, se muestra a continuación una posible ejecución del algoritmo. En cada paso indicamos qué agente interviene y qué conocimiento publica (respetando una posible elección del paso 3 del Algoritmo 1), y también indicamos el efecto en el estado consensuado del diálogo. Notaremos con  $E_i$  el estado del diálogo en la iteración  $i$ :

1.  $E_1 = \emptyset$
2. El agente  $A_1$  dice:  $\{a, a \rightarrow c, c \rightarrow d\}$   
 $E_2! \vdash d$
3. El agente  $A_2$  dice:  $\{a \wedge c \rightarrow \neg d\}$   
 $E_3! \vdash \neg d$
4. Termina el diálogo en el estado  $E_3$

El diálogo termina en el estado  $E_3$  porque ambos agentes cumplen sus metas. En general, podemos decir que:

**Observación 2** En un diálogo de investigación todos los agentes cumplen sus metas al terminar el diálogo (es fácil ver que si una meta  $G_i = (\alpha, ?)$  de un agente  $A_i$  no se cumple en  $E$  entonces debe existir  $X \subseteq K_i$  tal que  $G_i$  se cumple en  $E + X$ ).

En este caso se concluye, por (1), que la sentencia  $d$  es falsa. Esta misma conclusión se obtiene de  $(K_1 \cup K_2)!$  (el diálogo es completo). El Ejemplo 2 muestra que esto no siempre es así. Hay casos en los que los agentes quedan en una situación de *bloqueo* sin poder exponer toda la información relevante (porque no advierten que es relevante).

El Ejemplo 3 muestra un diálogo de investigación en el que, a diferencia de este, la conclusión final contradice la opinión individual de todos los agentes.

**Ejemplo 2** Supongamos que modificamos el entorno del Ejemplo 1 agregando sentencias a ambas bases:

$$K_1 = \{a, a \rightarrow c, b \rightarrow c, c \rightarrow d, f \wedge g \rightarrow \neg d\} \quad K_2 = \{a \wedge c \rightarrow \neg d, b \wedge c \rightarrow d, f, f \rightarrow g, g \rightarrow d\}$$

$$G_1 = G_2 = (d, ?)$$

Supongamos además que el operador ! de *partial meet consolidation* se comporta de la manera especificada a continuación, cuando es aplicado a cada uno de los estados siguientes:

$$\{a, a \rightarrow c, c \rightarrow d, a \wedge c \rightarrow \neg d\}! \vdash \neg d \quad (1)$$

$$\{a, a \rightarrow c, c \rightarrow d, a \wedge c \rightarrow \neg d, f, f \rightarrow g, g \rightarrow d\}! \vdash d \quad (2)$$

$$\{a, a \rightarrow c, c \rightarrow d, a \wedge c \rightarrow \neg d, f, f \rightarrow g, g \rightarrow d, f \wedge g \rightarrow \neg d\}! \vdash \neg d \quad (3)$$

De acuerdo a esto, se muestra a continuación una posible ejecución del algoritmo.

1.  $E_1 = \emptyset$
2. El agente  $A_1$  dice:  $\{a, a \rightarrow c, c \rightarrow d\}$   
 $E_2! \vdash d$
3. Termina el diálogo en el estado  $E_2$

En este caso podemos ver que el diálogo resulta en la aceptación de la sentencia  $d$ , pero si ambos agentes expusieran todo su conocimiento se obtendría una conclusión diferente. Notemos que el agente  $A_2$  considera, en el estado  $E_2$ , que no tiene nada relevante para decir. Esto se debe a que dicho agente advierte, por (2), que aún aportando todo su conocimiento no cambiaría la conclusión sobre  $d$ . Lo que el agente  $A_2$  no alcanza a advertir es que, por (3), la conclusión sí cambiaría en una iteración posterior, luego de la intervención del agente  $A_1$ . Esta situación (la llamamos situación de *bloqueo*) es poco deseable para agentes con metas  $(\alpha, ?)$ , ya que estos agentes buscan idealmente diálogos completos. Observemos que, por (1), existe en este caso un subconjunto propio del conocimiento privado del agente  $A_2$  capaz de cambiar la conclusión pública sobre  $d$ . Una forma de reducir las situaciones de bloqueo es redefinir el criterio de cumplimiento de las metas  $(\alpha, ?)$  de la siguiente manera:

**Definición 15 (Redefinición del Criterio de Éxito de Metas  $(\alpha, ?)$ )** Una meta  $G=(\alpha, ?)$  de un agente  $A = (K, G)$  se cumple en un estado  $E \subseteq L$  si y solo si, para todo  $X \subseteq K$ , se verifican las siguientes condiciones:

1.  $E! \vdash \alpha \Leftrightarrow (E \circ X) \vdash \alpha$
2.  $E! \vdash \neg\alpha \Leftrightarrow (E \circ X) \vdash \neg\alpha$

Llamaremos  $\varphi_2$  el criterio de éxito  $\varphi_1$  modificado según la Definición 15. El cumplimiento de una meta  $(\alpha, ?)$  según  $\varphi_2$  implica trivialmente el cumplimiento de una meta  $(\alpha, ?)$  según  $\varphi_1$ . La diferencia entre un diálogo de investigación con el criterio  $\varphi_1$  y uno con el criterio  $\varphi_2$  es que en el segundo caso los agentes publicarán más información, evitando algunas situaciones de bloqueo.

**Ejemplo 3** Reconsideremos el Ejemplo 2, pero ahora usando el criterio  $\varphi_2$  en lugar de  $\varphi_1$ :

$$K_1 = \{a, a \rightarrow c, b \rightarrow c, c \rightarrow d, f \wedge g \rightarrow \neg d\} \quad K_2 = \{a \wedge c \rightarrow \neg d, b \wedge c \rightarrow d, f, f \rightarrow g, g \rightarrow d\}$$

$$G_1 = G_2 = (d, ?)$$

A continuación se muestra una posible ejecución del algoritmo, asumiendo el mismo operador de consolidación que en el Ejemplo 2:

1.  $E_1 = \emptyset$
2. El agente  $A_1$  dice:  $\{a, a \rightarrow c, c \rightarrow d\}$   
 $E_2! \vdash d$

3. El agente  $A_2$  dice:  $\{a \wedge c \rightarrow \neg d, f, f \rightarrow g, g \rightarrow d\}$   
 $E_3! \vdash d$
4. El agente  $A_1$  dice:  $\{f \wedge g \rightarrow \neg d\}$   
 $E_4! \vdash \neg d$
5. Termina el diálogo en el estado  $E_4$

En este caso el diálogo termina concluyendo que la sentencia  $d$  es falsa, a pesar de que cada agente creía individualmente lo contrario. El resultado coincide con lo que se concluiría de  $(K_1 \cup K_2)!$  (el diálogo es completo).

Veremos con el Ejemplo 4 que en algunos casos la situación de bloqueo persiste a pesar de la redefinición de criterio del éxito.

**Ejemplo 4** Supongamos un caso de *Diálogo de Investigación* extremadamente simple:

$$K_1 = \{a\} \quad K_2 = \{a \rightarrow b\} \quad G_1 = G_2 = (b, ?)$$

Sería deseable concluir que la sentencia  $b$  es verdadera. Sin embargo, el modelo de diálogo propuesto no logra obtener esta conclusión (ya sea con el criterio de éxito  $\varphi_1$  o  $\varphi_2$ ). Se puede ver fácilmente que se genera un diálogo vacío en el que ninguno de los agentes puede empezar a hablar.

**Ejemplo 5** Por último, veremos un ejemplo de *Diálogo Persuasivo*. Tomaremos el entorno del Ejemplo 3 pero modificaremos la meta del agente  $A_1$ , de la siguiente manera:

$$K_1 = \{a, a \rightarrow c, b \rightarrow c, c \rightarrow d, f \wedge g \rightarrow \neg d\} \quad K_2 = \{a \wedge c \rightarrow \neg d, b \wedge c \rightarrow d, f, f \rightarrow g, g \rightarrow d\}$$

$$G_1 = (d, \oplus) \quad G_2 = (d, ?)$$

A continuación se muestra una posible ejecución del algoritmo:

1.  $E_1 = \emptyset$
2. El agente  $A_1$  dice:  $\{a, a \rightarrow c, c \rightarrow d\}$   
 $E_2! \vdash d$
3. El agente  $A_2$  dice:  $\{a \wedge c \rightarrow \neg d, f, f \rightarrow g, g \rightarrow d\}$   
 $E_3! \vdash d$
4. Termina el diálogo en el estado  $E_3$

La conclusión alcanzada es que la sentencia  $d$  es verdadera. Podemos ver que el diálogo no es completo. Sin embargo, la incompletitud no es un problema en este caso, sino que es causada intencionalmente por el agente  $A_1$  para cumplir su meta individual.

## 5. CONCLUSIONES Y TRABAJO FUTURO

El problema abordado en este artículo es el modelamiento de diálogos en sistemas multi-agente. Propusimos un algoritmo abstracto no determinístico que simula una interacción entre dos o más participantes. Creemos que los siguientes son *aspectos positivos* de la solución propuesta: (1) el uso de *Operadores de Cambio* nos permite abstraernos de la teoría lógica subyacente (lenguaje y mecanismo de inferencia), y (2) el uso de la noción abstracta de *meta* nos permite

parametrizar las actitudes de los agentes en el diálogo, abstrayéndonos, en cierta medida, del tipo de diálogo.

Identificamos también algunas *limitaciones* del modelo propuesto: (1) está principalmente orientado a *diálogos sobre creencias*, (2) todos los agentes tienen el mismo grado de autoridad (o credibilidad), (3) los agentes no reciben nuevas percepciones (externas al diálogo) mientras dialogan, y (4) el trabajo brinda un aporte teórico, pero no práctico, dada la complejidad computacional del algoritmo propuesto.

Nuestro trabajo futuro estará orientado a: (1) investigar el modelamiento de otros tipos de diálogo (como la negociación y el diálogo deliberativo), (2) analizar el uso de *lógicas argumentativas* como lenguaje de representación de conocimiento (creemos que esto podría facilitar el hallazgo del conjunto  $X$  del paso 3 del Algoritmo 1), (3) buscar estrategias para solucionar las situaciones de bloqueo mencionadas en la Sección 4, y (4) profundizar sobre implementaciones adecuadas del operador de consolidación en diferentes tipos de diálogo.

## REFERENCIAS

- [1] C. Alchourrón, P. Gärdenfors, and D. Makinson. *On the Logic of Theory Change: Partial Meet Contraction and Revision Functions*. *The Journal of Symbolic Logic*, 50:510-530, 1985.
- [2] M. A. Falappa, G. Kern-Isberner, and G. R. Simari. *Belief Revision, Explanations and Defeasible Reasoning*. *Artificial Intelligence Journal*, 141:1-28, 2002.
- [3] E. L. Fermé and S. O. Hansson. *Selective Revision*. *Studia Logica*, 63:331-342, 1998.
- [4] A. Fuhrmann. *An Essay on Contraction*. *Studies in Logic, Language and Information, CSLI Publications, Stanford, California*, 1997.
- [5] C. L. Hamblin. *Fallacies*. *Methuen and Co Ltd, London*, 1970.
- [6] S. O. Hansson. *Belief Base Dynamics*. *PhD thesis, Uppsala University, Department of Philosophy, Uppsala, Sweden*, 1991.
- [7] S. O. Hansson. *Kernel Contraction*. *The Journal of Symbolic Logic*, 59:845-859, 1994.
- [8] S. O. Hansson. *Semi-Revision*. *Journal of Applied Non-Classical Logic*, 7:151-175, 1997.
- [9] D. Makinson. *Screened Revision*. *Theoria: Special Issue on Non-Prioritized Belief Revision*, 1997.
- [10] S. Parsons, M. Wooldridge, and L. Amgoud. *Properties and Complexity of Some Formal Inter-agent Dialogues*. *Journal of Logic and Computation*, 13:347-376, 2003.
- [11] I. Rahwan, S. D. Ramchurn, N. R. Jennings, P. McBurney, S. Parsons, and L. Sonenberg. *Argumentation-based negotiation*. *The Knowledge Engineering Review*, 18:343-375, 2003.
- [12] D. N. Walton and E. C. W. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. *State University of New York Press, Albany, NY*, 1995.