Kwansei Gakuin University

信念改竄によるばれない嘘の生成
Perfect Lie Generation by Belief Alteration

奥野　健一[*1]　高橋　和子[*2]
Kenichi Okuno　Kazuko Takahashi

[*1]関西学院大学大学院理工学研究科
Graduate School of Science and Technology, Kwansei Gakuin University

[*2]関西学院大学理工学部
School of Science and Technology, Kwansei Gakuin University

This paper discusses *the perfect lie*, a lie that is not revealed to be false logically to a specific person at the present instant. First, we construct a belief system that models human reasoning, and consider the perfect lie on it. We define a human's memory as a belief world using a set of logical formulas, the operations of addition/deletion of formulas to/from the belief world and the condition for the perfect lie based on these definitions. Then, we propose an algorithm which derives the sequence of operations that makes an addition of a formula to succeed as the perfect lie to the belief world of the person to be deceived. The algorithm uses the framework of planning. Moreover, we show the conditions for belief alteration that judge whether a lie can be added as the perfect one, and prove the correctness of the conditions.

1. 序論

論理的にばれない嘘とはどのような性質をもった嘘なのだろうか．嘘に関する研究は様々な分野で行われているが [8]，どのような嘘がばれない嘘なのかを論理的な観点から追及したような研究は見当たらない．そこで本研究では，ばれない嘘の性質とその生成方法について，次の 2 点を仮定した上で，議論する．

- ばれない嘘を「ある特定の相手に現時点で論理的にはばれない嘘」とする．
- ある嘘がばれない嘘であるための条件を「相手がその嘘を信念に追加し，それに伴う推論を行っても，信念内で矛盾が生じない，あるいは論理的には矛盾が生じていてもその矛盾に気付かないこと」とする．

嘘をつく状況として，次のような例を考える．「クリスマスの夜に太郎の枕元にプレゼントを置いたのは父である」と思って拗ねている太郎に対し，太郎の父は「クリスマスの夜に太郎の枕元にプレゼントを置いたのは父ではない」と嘘をつきたい．この場合，単に「クリスマスの夜に太郎の枕元にプレゼントを置いたのは父ではない」と言うだけでは，太郎は受け入れない．何故なら，太郎がもし父の言葉を受け入れたなら，太郎の信念世界（信じている世界）の中で矛盾（相反する事柄を信じている状態）が生じるからである．そのような矛盾を生じさせないためには，太郎がもともと信じている「クリスマスの夜に太郎の枕元にプレゼントを置いたのは父である」を信じさせなくする必要がある．しかし，これを信じさせなくすることも素直にはいかない．「クリスマスの夜に太郎の枕元にプレゼントを置いたのは父である」を信じるに至った演繹パスが存在す

NII-Electronic Library Service

Kwansei Gakuin University

る可能性があるし，また，このことを唯一の説明とするような事実も存在するかもしれないからである．例えば，太郎が「サンタクロースはこの世に存在しない」，「サンタクロースがこの世に存在しないならば，クリスマスの夜に太郎の枕元にプレゼントを置いたのは父である」という2つのことを信じたことが「クリスマスの夜に太郎の枕元にプレゼントを置いたのは父である」と推論するに至らせた要因であるなら，この2つがある限り太郎の信念世界から「クリスマスの夜に太郎の枕元にプレゼントを置いたのは父である」を取り除くことはできない．また，「クリスマスの夜に父は夜更かしをしていた」，「クリスマスの夜に太郎の枕元にプレゼントを置いたのは父であるなら，クリスマスの夜に父は夜更かしをしていた」という2つのことを信じており，「クリスマスの夜に父は夜更かしをしていた」ということを説明できるものがこの「クリスマスの夜に太郎の枕元にプレゼントを置いたのは父である」しかない場合も，取り除くことはできない．もちろん，以上のような状況でなくとも，頭ごなしに信じていることを放棄せよと言うだけでは受け入れられない．そこで，背理法によってその事実を信じることの不当性を示すことも必要になってくる．例えば，太郎が「父は太郎の願いが何であるかを知らなかった」と信じているとすると，もともと信じている「クリスマスの夜に太郎の枕元にプレゼントを置いたのが父であるなら，プレゼントを購入したのは父である」に加えて，「プレゼントを購入したのが父であるなら，父は太郎の願いが何であるかを知っていた」を信じさせることにより，「クリスマスの夜に太郎の枕元にプレゼントを置いたのは父である」の不当性が背理法によって証明可能になる．本研究では，以上のような考えを基に理論を構築する．

本研究では，人間の推論モデルとして信念システムを構築し，信念システム上でのばれない嘘について考える．具体的には，まず，人間のもつ記憶領域を，論理式集合の組からなる信念世界として定義する．そして，信念世界への操作（追加，除去）に伴う人間の推論も定義することにより信念システムを構築した上で，ばれない嘘の条件を定義する．本研究ではさらに，つきたい嘘をばれない嘘として相手の信念世界に追加成功するに至る手順を導き出すアルゴリズムを，プラニングの考えを用いる形で提案する．そしてさらに，つきたい嘘をばれない嘘にすることが可能であるかを判定するための信念改竄条件を示し，その妥当性を証明する．

以下，2節では人間の推論モデルとして信念システムを定義し，そのモデル上でのばれない嘘を定義する．3節では，つきたい嘘をばれない嘘として相手の信念世界に追加するための操作手順を導き出すアルゴリズムを提案する．4節では，信念システムに関する定理を提示し，その証明を行う．5節では既存の研究との関係を議論し，6節でまとめる．

2. 信念システム

この節では，人間の推論モデルとして信念システム (belief system) を定義する．信念システムは，人間の記憶領域としての信念世界 (belief world) と，信念変更処理としてのオペレータ (operator) の集合によって構成される．以下では，まず信念世界を定義する．そして，人間の推論能力についての仮定を示した上で，信念世界に対する追加オペレータ，除去オペレータをそれぞれ定義し，それらを用いて信念システムを構築する．

2.1 信念世界

本研究では，全ての事柄は論理式 (formula) として記述されるものと仮定する．論理式は以下のように定義する．

$$
\begin{aligned}
\text{formula} &::= \text{fact} \mid \text{rule} \\
\text{fact} &::= \text{atom} \mid \neg\, \text{atom} \\
\text{rule} &::= \text{fact} \rightarrow \text{fact} \mid \neg\, (\text{fact} \rightarrow \text{fact}) \\
\text{atom} &::= \text{命題}
\end{aligned}
$$

NII-Electronic Library Service

Kwansei Gakuin University

論理式全体の集合は $\mathbf{L}$ とする．なお，以下で 'fact' や 'rule' という単語を用いた場合，上の意味でのものとする．また，'positive'，'negative' はそれぞれ，接頭辞が $\neg$ でない，接頭辞が $\neg$ である，を意味するものとする．また， positive なものの接頭辞に $\neg$ をつけたもの，あるいは negative なものの接頭辞を取り除いたものを，その 'complement' と呼び，もとのものを ψ とすると $\bar{\psi}$ のように記述する．

信念は，対象の人間が信じている事柄であり，その信じ方によって弱い信念 (weak belief)，強い信念 (strong belief)，固定された信念 (fixed belief)，の 3 種類に分類される．weak belief はその人が信じている事柄の中で変更可能性のあるもので， strong belief は強く信じられている事柄であるために変更不可能なものである．なお，weak belief と strong belief の和集合を belief とする．fixed belief はある特定の相手から変更を促されて変更した信念で，その相手からは二度とその信念の変更は受け付けないような信念，つまり信念に関する信念（メタ信念）である．そして，これら belief, strong belief, fixed belief の 3 つ組で構成されるものを信念世界と呼ぶ．よって，信念世界 $\mathbf{b}$ は以下のように定義される．

$$\mathbf{b} = \langle \mathbf{b.n}, \mathbf{b.s}, \mathbf{b.f} \rangle$$

$\mathbf{b.n}$ ： 信念 (belief) の集合
$\mathbf{b.s}$ ： 強い信念 (strong belief) の集合
$\mathbf{b.f}$ ： 固定された信念 (fixed belief) の集合

$$(\text{ただし，} \mathbf{b.n} \subseteq \mathbf{L}\ ,\ \{\psi \mid \psi \in \mathbf{b.n}, \bar{\psi} \in \mathbf{b.n}\} = \emptyset\ ,\ \mathbf{b.s} \subseteq \mathbf{b.n}\ ,\ \mathbf{b.f} \subseteq \mathbf{L})$$

なお，$\{\psi \mid \psi \in \mathbf{b.n}, \bar{\psi} \in \mathbf{b.n}\} = \emptyset$ は信念の無矛盾性を意味するものである．以後，無矛盾という言葉を用いた場合，これと同様の意味であるものとする．

相手の信じていることと論理式との対応関係の基準については，以下のように定義する．

- 論理的全知の問題 [3] の扱い

  信念世界はその人間が実際に信じていること全てを書き下したものとする．つまり，「論理的には推論可能であっても，その人がそのことに気付いていないのであれば，それは信念世界にも含まれない」ということである．一般にこのような仮定のことを「論理的全知を仮定しない」と言うが，本研究でもそのような仮定を設ける．

- $\rightarrow$ の意味

  論理式における $\rightarrow$ はいわゆる実質含意ではなく，2 つの引数の関係性を表現するための論理結合子である．人間の記憶領域を信念世界に書き下す際，この $\rightarrow$ を含む論理式に対応するものは，その人間が「その第一引数が真であるなら第二引数も真であると言える」と信じているという事実にすぎない．つまり，条件関係を示すものであろうと因果関係を示すもの [2] であろうと等しく $\rightarrow$ を用いて記述されるということである．ここでは，$\rightarrow$ を含む論理式の役割は，あくまでその推論の妥当性を「相手が信じているかどうか」を示すためだけにあり，そこに厳密な定義は必要ない，という立場をとっている．これは，人間の推論を扱うがゆえの特徴である．

2.2　人間の推論能力

推論の種類としては演繹とアブダクションのみを仮定する．演繹とは論理式 ψ と論理式 $\psi \rightarrow \chi$ から χ を導くような推論で，アブダクションとは論理式 χ と論理式 $\psi \rightarrow \chi$ から ψ を導くような推論とする．また，上のアブダクションに関して，$\psi \rightarrow \chi$ であるような ψ が 1 つしか存在しない場合，そのようなアブダクションを「決定的なアブダクション」と呼ぶことにする．ただし，実際に新たな論理式を推論によって捻出・信念化

NII-Electronic Library Service

Kwansei Gakuin University

するために使用される推論は，必然的推論である演繹のみとする．アブダクションはあくまで信念世界に対する妥当でない変更を防ぐためだけに用いられる．

推論が実行されるタイミングは，信念世界に変更が促された際のみとする．つまり，自発的に思索に耽るなどの現象は考慮しない．

推論の深さについては，論理的全知を仮定しないので，どの程度の深さまで行うかという基準が必要となる．一般に，そのような合理的な基準は存在しないとされているが，本研究では，「推論の過程で既に気付いている fact に到達した場合，それ以後の推論は行わない」という基準を設けることにする．この基準の意図は，「既に信じている事柄を起点とする推論は，過去（それが信念世界に追加された際）に行っているので，ここで再度行うことはない」ということである．ここで，制限付きの演繹を簡潔に表現するために，次のような表記法を定義しておく．（ψ_i は fact , s は論理式集合）

notation	meaning
$s, \psi_0 \rightsquigarrow \psi_m$	$(\psi_0 \rightarrow \psi_1), (\psi_1 \rightarrow \psi_2), \ldots, (\psi_{m-1} \rightarrow \psi_m) \in s$
$s, \psi_0 \rightsquigarrow_{\mathsf{T}} \psi_m$	$s, \psi_0 \rightsquigarrow \psi_m$ **and** $\psi_0, \psi_1, ..., \psi_m \in s$
$s, \psi_0 \rightsquigarrow_{\mathsf{NT}} \psi_m$	$s, \psi_0 \rightsquigarrow \psi_m$ **and** $\psi_0, \psi_1, ..., \psi_m \notin s$
$s, (\psi_0 \rightarrow \psi_1) \rightsquigarrow_{\mathsf{NT}} \psi_m$	$(\psi_0 \rightarrow \psi_1) \notin s$ **and** $\psi_0 \in s$ **and** $(s, \psi_1 \rightsquigarrow_{\mathsf{NT}} \psi_m$ **or** $(m = 1$ **and** $\psi_1 \notin s)$ $)$

$\rightsquigarrow_{\mathsf{T}}$ は対象となる人間が過去に行ったであろう推論を表現するために用いる．論理的全知でないことは「たとえ論理的には推論可能な事であっても，その人の現在の信念世界にその事が存在しないのであれば，その人はその事に気付いていない（推論できていない）.」と解釈できる．それゆえ，そのような人間に可能であった推論は $\rightsquigarrow_{\mathsf{T}}$ に対応している．

$\rightsquigarrow_{\mathsf{NT}}$ は対象となる人間の信念に変更が加えられた際にその人の信念世界が（その人の推論によって）どのように変化するのかを表現するために用いる．人間が信念を得た際にそこから新たな論理式を捻出・信念化する際の推論に対して本研究では上述の仮定を設けているので，その際の推論は $\rightsquigarrow_{\mathsf{NT}}$ に対応している．

2.3 信念システム

人間の推論モデルとして信念システムを以下のように定義する．

$$\langle \mathbf{L}, \mathbf{B}, \{\oplus, \ominus\} \rangle$$

$\mathbf{L}$: 論理式全体の集合
$\mathbf{B}$: 信念世界全体の集合
$\oplus$: $\mathbf{B} \times \mathbf{L} \mapsto \mathbf{B}$
$\ominus$: $\mathbf{B} \times \mathbf{L} \mapsto \mathbf{B}$

$\oplus, \ominus$ はそれぞれ信念世界に対する論理式の追加，除去を行うための演算子で．以下のように定義する．なお，以後，「失敗」という言葉は $\mathbf{b} \oplus \psi = \mathbf{b}$ あるいは $\mathbf{b} \ominus \psi = \mathbf{b}$ を，「成功」はそうでない場合を意味するものとする．

$\mathbf{b} \oplus \psi =$
　if(
　　$\mathbf{b}' = \langle (\mathbf{b.n} \cup \{\psi\} \cup \{\chi \mid \mathbf{b.n}, \psi \rightsquigarrow_{\mathsf{NT}} \chi\}), \mathbf{b.s}, (\mathbf{b.f} \cup \{\psi\}) \rangle$ **and**
　　$\{\chi \mid \chi \in \mathbf{b}'.\mathbf{n}, \bar{\chi} \in \mathbf{b}'.\mathbf{n}\} = \emptyset$ **and**

NII-Electronic Library Service

Kwansei Gakuin University

$(\mathbf{b}'.\mathbf{n} \setminus \mathbf{b}.\mathbf{n}) \cap \mathbf{b}.\mathbf{f} = \emptyset$
)then $\mathbf{b}'$.
else $\mathbf{b}$.

$\mathbf{b} \ominus \psi =$
if(
(
(ψ is a negative formula , $\mathbf{b}' = \langle(\mathbf{b}.\mathbf{n} \setminus \{\psi\}), \mathbf{b}.\mathbf{s}, (\mathbf{b}.\mathbf{f} \cup \{\psi\})\rangle$) **or**
(ψ is a positive formula , $\mathbf{b}' = \langle(\mathbf{b}.\mathbf{n} \setminus \{\psi\}), \mathbf{b}.\mathbf{s}, (\mathbf{b}.\mathbf{f} \cup \{\psi\})\rangle \oplus \neg\psi, \mathbf{b}'.\mathbf{n} \neq (\mathbf{b}.\mathbf{n} \setminus \{\psi\})$
)
) **and** ・・・(0)
(
($\psi \notin \mathbf{b}.\mathbf{n}$) **or**
(
$\psi \notin \mathbf{b}.\mathbf{f}$ **and** ・・・(1)
$\psi \notin \mathbf{b}.\mathbf{s}$ **and** ・・・(2)
$\{\chi \mid (\mathbf{b}'.\mathbf{n} \cup \{\psi\}), \chi \rightsquigarrow_{\mathsf{T}} \psi\} = \emptyset$ **and** ・・・(3)
$\{\chi \mid (((\psi \rightarrow \chi) \in \mathbf{b}.\mathbf{n})$ **or** $(\psi = (\gamma \rightarrow \chi), \gamma \in \mathbf{b}.\mathbf{n})), \chi \in \mathbf{b}.\mathbf{s}, \{\omega \mid (\omega \rightarrow \chi) \in \mathbf{b}.\mathbf{n}, \omega \in \mathbf{b}'.\mathbf{n}, \omega \neq \psi\} = \emptyset\} = \emptyset$ **and** ・・・(4)
($(\psi$ is not a fact) **or** $(\{\chi \mid \mathbf{b}.\mathbf{n}, \psi \rightsquigarrow \chi, \bar{\chi} \in \mathbf{b}'.\mathbf{n}\} \neq \emptyset)$) ・・・(5)
)
)
)then $\mathbf{b}'$.
else $\mathbf{b}$.

$\mathbf{b} \oplus \psi$ は，信念世界 $\mathbf{b}$ に対する論理式 ψ の追加を意味するオペレータである．もし $\psi \notin \mathbf{b}.\mathbf{n}$ であるなら，$\mathbf{b}'$（オペレータの実行結果のビジョン）は $\mathbf{b}.\mathbf{n}$ に ψ を加えただけのものではなく，$\mathbf{b}.\mathbf{n}$ において ψ を起点として $\rightsquigarrow_{\mathsf{NT}}$ できるものも加えたものである．これは，改竄をきっかけに推論を行うこと，信念の捻出に用いられる推論は演繹のみであること，既に気付いている事実に到達した場合はそれ以後の推論は行わないこと，という仮定を反映するためのものである．そして，この結果として得られる集合が無矛盾で，かつ固定された信念に関しては変更されていないのであれば，オペレータの実行結果はその得られた結果 $\mathbf{b}'$ である $(\mathbf{b} \oplus \psi = \mathbf{b}')$．そうでないなら，追加は受け入れられなかったこととなり，このオペレーションは $\mathbf{b}$ に何も変更を及ぼさないまま終了（失敗）する．$(\mathbf{b} \oplus \psi = \mathbf{b})$．

$\mathbf{b} \ominus \psi$ は，信念世界 $\mathbf{b}$ に対する論理式 ψ の除去を意味するオペレータである．除去の成否は，相手の推論による妨げを回避できるか否かによって決まる．まず $\mathbf{b}'$（オペレータの実行結果のビジョン）を設定する．ψ が negative formula であれば $\mathbf{b}'.\mathbf{n}$ は $\mathbf{b}'.\mathbf{n} \setminus \{\psi\}$ であるが，ψ が positive formula の場合は $\neg\psi$ を $\oplus$ することも必要となる．何故なら，否定文の否定は肯定文とは言い切れないが，肯定文の否定は確実に否定文となるからである．もしもこの $\oplus$ が失敗するのであれば，$\mathbf{b} \ominus \psi$ は失敗である．そうでないなら，この $\mathbf{b}'$ も用いて，除去可能性のチェックが行われる．まず，(1),(2) より，ψ が固定された信念，あるいは強い信念であれ

NII-Electronic Library Service

Kwansei Gakuin University

ば失敗である．(3) は ψ を $\rightsquigarrow_{\mathsf{T}}$ できる仕組みが存在してはならないことを示している．これは，過去に自ら合理的と判断して導いたものでかつ，今もなおその演繹パスが存在するものに対する除去を容易に受け入れるはずがない，ということを意味する．(4) は ψ を決定的にアブダクションできる仕組みが存在してはならないことを意味し，ψ を唯一の説明としているような fact が信念世界に存在した場合を考慮するためのものである．(5) は背理法によって ψ の否定形を証明できるような仕組みが存在しなくてはならないことを意味し，相手の納得感を考慮するものである．そして，これら全てのチェックを通過するか，あるいはもともと ψ が $\mathbf{b.n}$ に含まれていなかったなら，$\mathbf{b}'$ がこのオペレーションの結果となる．($\mathbf{b} \ominus \psi = \mathbf{b}'$)．そうでないなら，除去は受け入れられなかったことになり，このオペレーションは $\mathbf{b}$ に何も変更を及ぼさないまま終了（失敗）する ($\mathbf{b} \ominus \psi = \mathbf{b}$)．

以上により，「ばれない嘘」は次のように定義できる．

定義　「論理式 φ が（嘘をつきたい相手に現時点で）ばれない嘘である．」**iff**「嘘をつきたい相手の現在の記憶領域が読み込まれている信念世界 $\mathbf{b}$ をもつ信念システムにおいて，$\mathbf{b} \oplus \varphi$ が成功する．」

3. アルゴリズム

前節のばれない嘘の定義により，つきたい嘘 φ をばれない嘘にするためには，現在の相手の信念世界 $\mathbf{b}$ を $\mathbf{b} \oplus \varphi$ が成功するような $\mathbf{b}$ へと変化させればよい．信念システムのオペレータで $\mathbf{b}$ を変化させることができるものは $\oplus, \ominus$ の 2 つである．よって，この 2 つを適切な引数と共に適切な回数，適切な順序で $\mathbf{b}$ に対して実行することにより，$\mathbf{b} \oplus \varphi$ が成功するような $\mathbf{b}$ へと変化させることができる．以下では，この $\mathbf{b} \oplus \varphi$ を成功するに至らせるオペレータのシーケンスを求めるアルゴリズムを，プラニングの考えを用いる形で提案する．

3.1 プラニング

本論文でのプラニングは，人工知能分野における基本的なプラナーである STRIPS [12] に基づくものを考える．STRIPS のプラニングでは，入力としてオペレータ集合，初期状態（論理式集合），目標状態（論理式集合），を受け取り，与えられた初期状態でワーキングメモリを初期化し，目標状態がワーキングメモリの部分集合になるようにワーキングメモリを変化させるようなオペレータのシーケンスを求める．オペレータは，条件リスト（論理式集合），追加リスト（論理式集合），除去リスト（論理式集合）の 3 つから構成され，条件リストに含まれる全ての論理式がワーキングメモリに存在する場合に，追加リストの全ての要素をワーキングメモリに追加し，除去リストの全ての要素をワーキングメモリから除去する，という操作を表現するものである．

次小節では，信念システムにおけるオペレータをプラニング用のオペレータに変換する．そして，初期状態を現在の相手の信念世界，目標状態を「相手の信念世界に嘘が含まれている」というメタ的な命題のみを含む集合，ワーキングメモリを相手の信念世界，とそれぞれ置き換えると，つきたい嘘をばれない嘘として相手の信念世界に追加成功するに至るオペレータのシーケンスをプラニングの方法で求めることができる．

ただし，本論文でのオペレータの条件リストは論理式の集合ではなく，$\varphi \in \mathbf{b}$（ただし φ は論理式，$\mathbf{b}$ は信念世界）のように $\mathbf{b}$ を含むメタ的な命題の集合であり，また，追加リストと除去リストについてもメタ的に記述されるので，STRIPS のプラニングアルゴリズムをそのまま適用することはできない．$\mathbf{b}$ の状態に依存して各オペレータの条件リスト，追加リスト，除去リストの具体的な内容が変化する，つまり，あるオペレータの採用を検討する際，そのオペレータの各リスト中の $\mathbf{b}$ は，初期状態からその検討時点に至るまでのオペレータを初期状態に対して順次実行した結果ということである．このことは，1 つのオペレータを生成・プランに追加すると，その副作用でプラン中の他のオペレータのそれまで成り立っていた条件リスト中の条件が成り立

NII-Electronic Library Service

Kwansei Gakuin University

たなくなる可能性があることを示す．よって，このアルゴリズムでは，目標状態に至らせるためのオペレータを後向き推論により順に生成し，1 つ生成するたびにこれまで生成した全てのオペレータの正当性を前向き推論により検証する，という一連の処理を繰り返す．これにより，図 1 の例のように，生成済み部分プランの先頭へのオペレータの追加による不都合な副作用を検知することができる．この例では，(3) のプランでは追加されない $\neg\psi_3$ が (4) のプランでは追加されることにより，(4.3) において矛盾が生じている $(\psi_3, \neg\psi_3 \in \mathbf{b}.\mathbf{n})$．つまり，add($\varphi$) の条件リストが，(3) のプランでは満たされているにもかかわらず，その先頭に add($\neg\psi_2$) を追加した (4) のプランにおいては満たされていない．

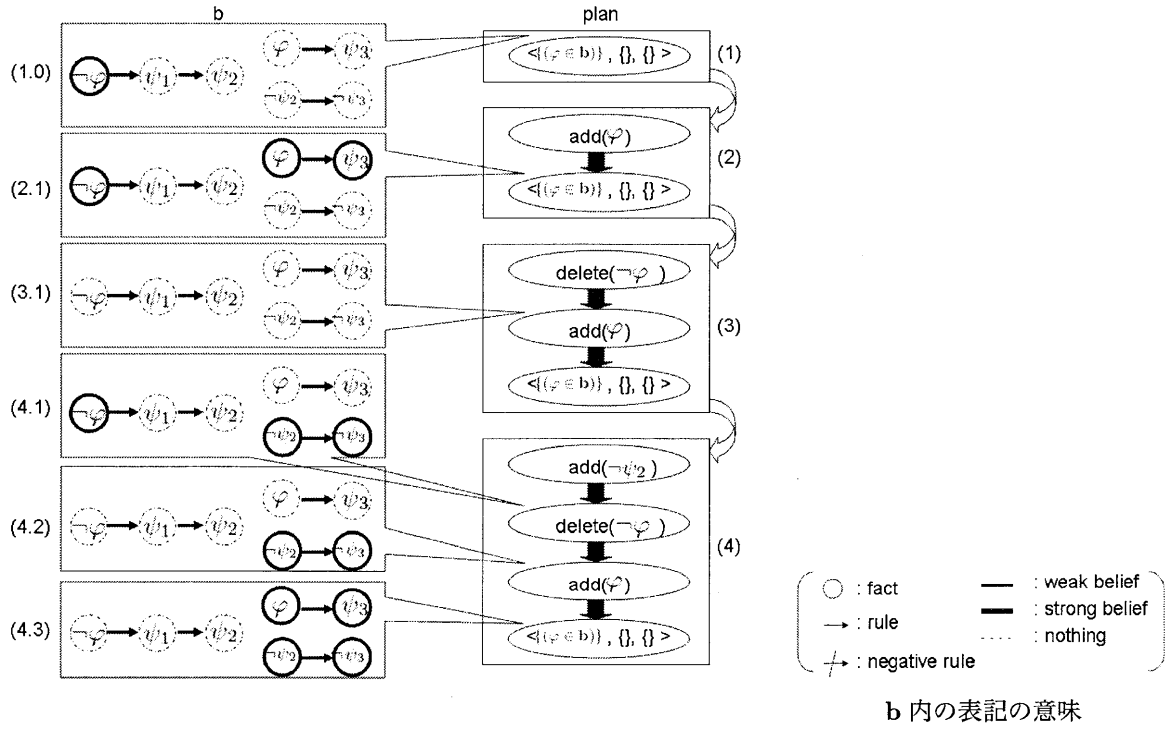

図 1　プランの生成過程と b の変化

3.2　オペレータ

まず，前節で提案したオペレータ $\oplus, \ominus$ をそれぞれ STRIPS のオペレータの形式に変換する．ただし，本論文では，条件リストはメタ的な命題の集合とし，追加リストと除去リストは信念世界と同じ構成の 3 つ組とする．なお，以下の $\mathbf{b}' \setminus \mathbf{b}$ という記述は $\langle \mathbf{b}'.n \setminus \mathbf{b}.n, \mathbf{b}'.s \setminus \mathbf{b}.s, \mathbf{b}'.f \setminus \mathbf{b}.f \rangle$ を意味するものとする．

NII-Electronic Library Service

Kwansei Gakuin University

オペレータ *add*(*Formula* ψ)

BeliefWorld $\mathbf{b}' = \langle(\mathbf{b.n} \cup \{\psi\} \cup \{\chi \mid \mathbf{b.n}, \psi \rightsquigarrow_{\mathsf{NT}} \chi\}), \mathbf{b.s}, (\mathbf{b.f} \cup \{\psi\})\rangle$ とすると,

- 条件リスト (prerequisite list)

  $\{\chi \mid \chi \in \mathbf{b}'.\mathbf{n}, \bar{\chi} \in \mathbf{b}'.\mathbf{n}\} = \emptyset$ **and**
  $(\mathbf{b}'.\mathbf{n} \setminus \mathbf{b.n}) \cap \mathbf{b.f} = \emptyset$

- 削除リスト (delete list)

  なし.

- 追加リスト (add list)

  $\mathbf{b}' \setminus \mathbf{b}$

オペレータ *delete*(*Formula* ψ)

ψ is a positive formula の時は *BeliefWorld* $\mathbf{b}' = \langle(\mathbf{b.n} \setminus \{\psi\}), \mathbf{b.s}, (\mathbf{b.f} \cup \{\psi\})\rangle \oplus \neg\psi$ とし,
ψ is a negative formula の時は *BeliefWorld* $\mathbf{b}' = \langle(\mathbf{b.n} \setminus \{\psi\}), \mathbf{b.s}, (\mathbf{b.f} \cup \{\psi\})\rangle$ とすると,

- 条件リスト (prerequisite list)

  (ψ is a negative formula) **or** $(\mathbf{b}' \neq \langle(\mathbf{b.n} \setminus \{\psi\}), \mathbf{b.s}, (\mathbf{b.f} \cup \{\psi\})\rangle)$ **and** ・・・(0)
  (
  $\psi \notin \mathbf{b.n}$ **or**
  (
  $\psi \notin \mathbf{b.f}$ **and** ・・・(1)
  $\psi \notin \mathbf{b.s}$ **and** ・・・(2)
  $\{\chi \mid (\mathbf{b}'.\mathbf{n} \cup \{\psi\}), \chi \rightsquigarrow_{\mathsf{T}} \psi\} = \emptyset$ **and** ・・・(3)
  $\{\chi \mid (((\psi \rightarrow \chi) \in \mathbf{b.n})$ **or** $(\psi = (\gamma \rightarrow \chi), \gamma \in \mathbf{b.n})), \chi \in \mathbf{b.s}, \{\omega \mid (\omega \rightarrow \chi) \in$
  $\mathbf{b.n}, \omega \in \mathbf{b}'.\mathbf{n}, \omega \neq \psi\} = \emptyset\} = \emptyset$ **and** ・・・(4)
  ((ψ is not a fact) **or** $\{\chi \mid \mathbf{b.n}, \psi \rightsquigarrow \chi, \bar{\chi} \in \mathbf{b}'.\mathbf{n}\} \neq \emptyset$) ・・・(5)
)
)

- 削除リスト (delete list)

  $\mathbf{b} \setminus \mathbf{b}'$

- 追加リスト (add list)

  $\mathbf{b}' \setminus \mathbf{b}$

NII-Electronic Library Service

Kwansei Gakuin University

3.3 アルゴリズム

プラニングアルゴリズムを提案する．以下に示すアルゴリズムでは，関数 *planning* をメインの関数とし，実行する際はまず， *planning* (**b**, {($\varphi \in$ **b**)}, {add(ψ),delete(ψ) | $\psi \in$ **L**}) を呼び出す．初期状態 initial は現在の相手の信念世界 **b** である．目標状態 goal は，つきたい嘘が相手の信念世界に含まれていさえすればいいので，{($\varphi \in$ **b**)} である．os はオペレータ全体の集合である．

関数 *planning* はまず，現時点でのプラン（オペレータのシーケンス） plan と初期信念世界 initial に対して，*check* (plan, initial) を呼び出す．次に，その結果 p と plan, initial, os に対して，*resolve* (p, plan, initial, os) を呼び出す．関数 *check* は， plan に含まれる全てのオペレータの全ての条件リストの要素の内，満たされていないものがあればその内の 1 つを選択して返し，そうでない場合は null を返す．関数 *check* の内部では， plan に含まれるオペレータ o, plan, initial に対して *decideBeliefWorld* (o, plan, initial) を呼び出す．この関数は，initial に対して plan を o の直前まで実行した結果得られる信念世界を返す．この結果 **b** と， plan 中のオペレータの条件リストの要素 p に対して，*interpretation* (p, **b**) を呼び出す．この関数は，受け取った **b** に対して p が満たされているかを判定し，結果の真偽値を返す．そして，関数 *resolve* は，p を満たすようにするオペレータのシーケンスを探し， plan の先頭に挿入する．なお，ここで挿入するオペレータが単独ではなくシーケンスとなっているのは，オペレータ delete の条件リストの要素 (4) あるいは (5) について，それを満たさせる単独のオペレータは存在しないがシーケンスであれば存在する場合があるからである．

以上の呼び出し関係により， plan が順次更新されていく．しかし， plan が更新されると，更新前の *check* の結果が意味をなさなくなる部分があるので，この一連の処理を関数 *planning* において繰り返している．

ただし，このアルゴリズムにおいて選択を行っている部分にはいくつかの選択肢があるので，ある選択の結果失敗した場合はバックトラックを行ってその部分を選択し直すものとする．全ての選択において失敗して初めて，失敗として終了する．

定義

- Operator
  - Operator o = ⟨ Propositions pl , Formulas al , Formulas dl ⟩
    - pl: prerequisite list
    - al: add list
    - dl: delete list
- Plan
  - Plan plan = [Operator o_1, Operator o_2, ..., Operator o_n]

NII-Electronic Library Service

Kwansei Gakuin University

アルゴリズム

```
Plan planning (BeliefWorld initial, Propositions goal, Operators os){
  Plan plan := [ ⟨ goal , {} , {} ⟩ ] ;
  while(check (plan, initial) ≠ null ){
    if((plan := resolve (check (plan, initial), plan, initial, os)) = null){
      return null;
    }
  }
  return plan;
}

Proposition check (Plan plan, BeliefWorld initial){
  while(全ての o∈plan について){
    while(全ての p∈o.pl について){
      if(interpretation (p, decideBeliefWorld (o, plan, initial)) = false){
        return p;
      }
    }
  }
  return null;
}

BeliefWorld decideBeliefWorld (Operator o, Plan plan, BeliefWorld initial){
  BeliefWorld b := initial;
  while( o'∈plan を前から順に o の手前まで){
    b := ⟨ ((b.n∪ o'.al ) \ o'.dl ), b.s , (b.f∪ o'.al ∪ o'.dl ) ⟩;
  }
  return b;
}

boolean interpretation (Proposition p, BeliefWorld b){
  return p 中の全ての b の出現をこの引数の b に置き換えて評価した結果;
}

Plan resolve (Proposition p, Plan plan, BeliefWorld initial, Operators os){
  while(Operator o を plan 内で p をもつオペレータの直前のオペレータから順に， 先頭のオペレータまで){
    if(b が decideBeliefWorld (o, plan, initial) のとき， Operator のシーケンスで p を満たすようにするものが
存在する){
      plan := plan の o の手前にそのシーケンスを挿入したもの;
      return plan;
    }
  }
  return null;
}
```

NII-Electronic Library Service

Kwansei Gakuin University

以下，上記アルゴリズムの大まかな実行の流れを例によって示す．(図 2)

初期化

$\mathbf{b} = \langle\{\neg\varphi, \psi_0, \neg\psi_2, \psi_3, (\psi_0 \to \neg\varphi), (\neg\varphi \to \psi_1), (\neg\varphi \to \psi_3), (\varphi \to \neg\psi_0)\}, \{\psi_3, (\neg\varphi \to \psi_3)\}, \emptyset\rangle$
φ: 「父がプレゼントを置いたのではない．」
ψ_0: 「サンタは存在しない．」
ψ_1: 「父がプレゼントを買った．」
ψ_2: 「父は太郎の願いを知っていた．」
ψ_3: 「父は夜更かしをしていた．」
ψ_4: 「父はサンタを出迎えた．」
os $= \{add(\psi), delete(\psi) \mid \psi \in \mathbf{L}\}$
initial $= \mathbf{b}$
goal $= \{(\varphi \in \mathbf{b})\}$

実行の流れ

- plan = { ⟨ goal, {}, {} ⟩ }
- goal に含まれる $(\varphi \in \mathbf{b})$ は満たされていないので，満たさせるオペレータ add(φ) を plan の先頭に追加する．
- plan に含まれる全てのオペレータの条件リストが満たされているかを，先頭から順に調べる．add(φ) の条件リストの要素 $\{\chi \mid \chi \in \mathbf{b}'.\mathbf{n}, \bar{\chi} \in \mathbf{b}'.\mathbf{n}\} = \emptyset$ が満たされていないので，満たさせるオペレータ delete($\neg\varphi$) を plan の先頭に追加する．
- 同様の処理を繰り返し，
- delete($(\psi_0 \to \neg\varphi)$) を plan の先頭に追加する．
- delete($(\psi_0 \to \neg\varphi)$) の条件リストは満たされているので，次に delete($\neg\varphi$) の条件リストを調べる．このとき，delete($\neg\varphi$) の条件リストの要素中に現れる $\mathbf{b}$ は， initial に delete($(\psi_0 \to \neg\varphi)$) を実行した結果として得られる信念世界である．
- 同様の処理を繰り返し，最後に，
- plan に含まれる全てのオペレータの条件リストが満たされ，得られた plan を返して終了する．

実行結果（得られたプラン）

$add(\psi_4 \to \psi_3), add(\psi_4), add(\psi_1 \to \psi_2), delete(\psi_0 \to \neg\varphi), delete(\neg\varphi), add(\varphi).$

得られたこのプランを実行した結果改竄された信念世界では，父がプレゼントを置いたのではない (φ) という嘘がばれない嘘として追加されている．すなわち，サンタが存在しないからといって父がプレゼントを置いたというわけではない，と信じさせ，父が夜更かしをしていたことに対し，プレゼントを置くためでなくサンタを出迎えるためだという理由がついており，さらに，プレゼントを置いたのが父だ ($\neg\varphi$) と仮定すると，父

NII-Electronic Library Service

Kwansei Gakuin University

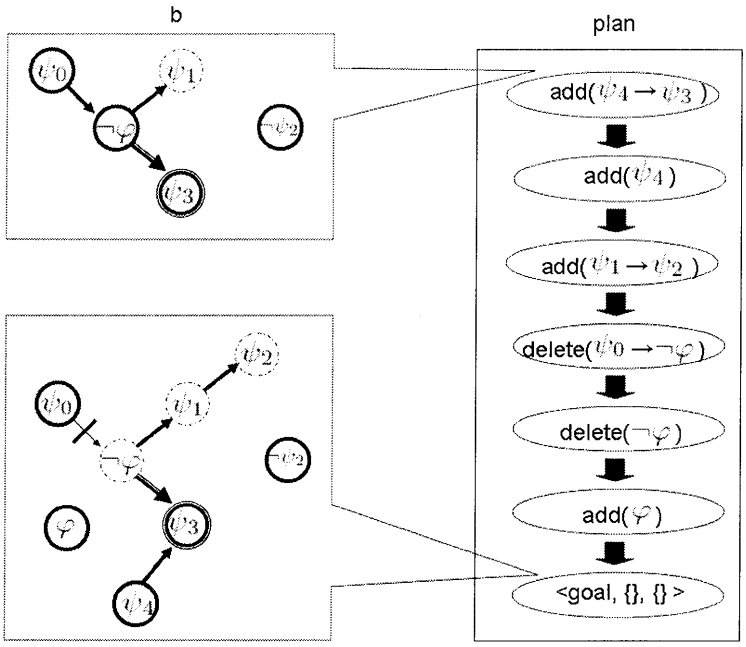

図 2　例に対する実行結果

は太郎の願いを知っていた (ψ_2) ことになり，太郎の信念 ($\neg\psi_2$) と矛盾することから，太郎の納得感を増している．

4.　信念システムの性質

前節で示したアルゴリズムによって，$\mathbf{b}$ に φ を追加するのを成功させるオペレータのシーケンスが求められ，それを $\mathbf{b}$ に実行した結果として $\mathbf{b}'$ が得られるとする．本節では，$\mathbf{b}$ と $\mathbf{b}'$ の間に成り立つ条件について議論する．具体的には，2 つの信念世界 $\mathbf{b}$ と $\mathbf{b}'$ に対して信念改竄条件を定義し，信念改竄条件が満たされることとオペレータのシーケンスが存在することとが一致することを証明する．

4.1　信念改竄条件

信念世界 $\mathbf{b}$ と信念集合 $\mathbf{b}'.\mathbf{n}$ が与えられたとき，追加集合 $\mathbf{D}^+ = \mathbf{b}'.\mathbf{n} \setminus \mathbf{b}.\mathbf{n}$，除去集合 $\mathbf{D}^- = \mathbf{b}.\mathbf{n} \setminus \mathbf{b}'.\mathbf{n}$ とすると，信念改竄条件は，この $\mathbf{b}, \mathbf{b}', \mathbf{D}^+, \mathbf{D}^-$ について，以下の 7 つの条件全てを満たすことである．

条件 **1.** $\{ \psi \mid \psi \in \mathbf{D}^-, \psi \in \mathbf{b}.\mathbf{s}\} = \emptyset$.
(強い信念は除去されない．)

条件 **2.** $\{ \psi \mid \psi \in \mathbf{D}^-, \{ \chi \mid (\chi \rightarrow \psi) \in \mathbf{b}'.\mathbf{n}, \chi \in \mathbf{b}'.\mathbf{n}\} \neq \emptyset\} = \emptyset$.
(ある信念が除去されているなら，その信念を $\rightsquigarrow_{\mathsf{T}}$ できる仕組みも除去されているか，あるいはもともとそのような仕組みが存在せずかつ追加もされないか，である．)

条件 **3.** $\{ \psi \mid \psi \in \mathbf{D}^-, \{ \chi \mid (((\psi \rightarrow \chi) \in \mathbf{b}.\mathbf{n}) \textbf{ or } (\psi = (\gamma \rightarrow \chi), \gamma \in \mathbf{b}.\mathbf{n})), \chi \in \mathbf{b}.\mathbf{s}, \{ \omega \mid \omega \neq \psi, ((\omega \rightarrow \chi) \in \mathbf{b}.\mathbf{n} \textbf{ or } (\omega \rightarrow \chi) \in \mathbf{D}^+), (\omega \in \mathbf{b}.\mathbf{n} \textbf{ or } \omega \in \mathbf{D}^+)\} = \emptyset\} \neq \emptyset\} = \emptyset$.

NII-Electronic Library Service

Kwansei Gakuin University

(ある信念が除去されているなら，その信念を決定的にアブダクションできる仕組み（その信念を唯一の説明とする強い信念を構成要素として持つ）が存在していたならそれも除去されている.)

条件 **4.** $\{\ \psi \mid \psi \in \mathbf{D}^-, \psi \text{ is a fact }, \{\ \chi \mid ((\mathbf{b.n}, \psi \rightsquigarrow \chi) \textbf{ or } ((\mathbf{b.n}, \psi \not\rightsquigarrow \chi) \text{ を } \psi \rightsquigarrow \chi \text{ とするもの} \in \mathbf{D}^+)), ((\bar{\chi} \in \mathbf{b.n}) \textbf{ or } (\bar{\chi} \in \mathbf{D}^+))\} = \emptyset\} = \emptyset.$
(ある fact である信念が除去されているなら，その complement を背理法によって証明できる仕組みが改竄後に存在する.)
ただし，「$\psi \rightsquigarrow \chi$ とするもの」とは，$(\psi \rightarrow \psi_1), (\psi_1 \rightarrow \psi_2), ..., (\psi_n \rightarrow \chi)$ の中の 1 つ以上が belief に含まれていない状況において，その含まれていないもの全てのことである.

条件 **5.** $\{\ \psi \mid \psi \in \mathbf{D}^-, \psi \text{ is a positive }, \neg\psi \notin \mathbf{D}^+\} = \emptyset.$
(ある positive な信念が除去されているなら，その complement が改竄後に含まれている.)

条件 **6.** $\{\ \psi \mid \psi \in \mathbf{D}^+, \{\ \chi \mid (\mathbf{b}'.\mathbf{n} \setminus \{\psi\}), \psi \rightsquigarrow_{\mathsf{NT}} \chi, (\chi \notin \mathbf{b}'.\mathbf{n})\} \neq \emptyset\} = \emptyset.$
(ある信念が追加されているなら，それを起点とした $\rightsquigarrow_{\mathsf{NT}}$ によって導出されるものも追加されている.)

条件 **7.** $\{\ \psi \mid \psi \in \mathbf{D}^+, \bar{\psi} \in \mathbf{b}'.\mathbf{n}\} = \emptyset.$
(ある信念が追加されているなら，その complement が改竄後に含まれていない.)

4.2　信念システムの性質

定理 1. 命題 P を「信念世界 $\mathbf{b}(= \langle \mathbf{b.n}, \mathbf{b.s}, \mathbf{b.f}(= \emptyset)\rangle)$ と信念集合 $\mathbf{b}'.\mathbf{n}$ が与えられたとき，$\mathbf{b}$ に対して信念改竄オペレーションからなるあるシーケンスを実行することによって $\mathbf{b}'.\mathbf{n}$ をもつ信念世界 $\mathbf{b}'$ を作り出すことができる.」，命題 Q を「$\mathbf{b}$ と $\mathbf{b}'.\mathbf{n}$ について，信念改竄条件が満たされている.」とするとき，P **iff** Q.

以下，定理 1 を証明する．まず，¬Q⇒ ¬P を証明する．これは， Q でない，すなわち信念改竄条件が 1 つでも満たされていないならば P でないことを証明することによって示す（補題 1〜補題 7)．次に，Q⇒P を証明する（補題 8)．

補題 1. $\mathbf{b}$ と $\mathbf{b}'$ について，信念改竄条件中の条件 1 が満たされていない ⇒P ではない.

証明．補題 1 を証明する.
仮定より，「$\psi \in \mathbf{D}^-$ かつ $\psi \in \mathbf{b.s}$ であるような ψ が存在する」なので，このような ψ が存在するとき，$\psi \in \mathbf{D}^-$ より，$\mathbf{b} \ominus \psi$ として成功するはずである．これは $\psi \notin \mathbf{b.s}$ でないと成功しないが，strong belief を改竄できるオペレーションは存在しないので，$\psi \notin \mathbf{b.s}$ にすることは不可能である．つまり，上のような ψ を実現する信念改竄オペレーションのシーケンスは存在しない.
以上により，補題 1 が証明された. □

補題 2. $\mathbf{b}$ と $\mathbf{b}'$ について，信念改竄条件中の条件 2 が満たされていない ⇒P ではない.

証明．補題 2 を証明する.
仮定より，「$\psi \in \mathbf{D}^-$ かつ『$(\chi \rightarrow \psi) \in \mathbf{b}'.\mathbf{n}$ かつ $\chi \in \mathbf{b}'.\mathbf{n}$ であるような χ が存在する』であるような ψ が存在する」なので，このような ψ が存在するとき，$\psi \in \mathbf{D}^-$ より，$\mathbf{b} \ominus \psi$ として成功するはずである．ここで，$\chi \in \mathbf{b}'.\mathbf{n}$ とは，χ がもともと $\mathbf{b.n}$ の要素でありかつ除去されないか，あるいは χ は $\mathbf{b.n}$ の要素でなくかつ追

NII-Electronic Library Service

Kwansei Gakuin University

加されるかのどちらかである．これと同様のことが $(\chi \to \psi) \in \mathbf{b}'.\mathbf{n}$ についても言えるので，$(\chi \to \psi) \in \mathbf{b}'.\mathbf{n}$ かつ $\chi \in \mathbf{b}'.\mathbf{n}$ の部分は，次の 4 つの場合に分けられる．

- $(\chi \to \psi) \in \mathbf{b.n}, (\chi \to \psi) \notin \mathbf{D}^-, \chi \in \mathbf{b.n}, \chi \notin \mathbf{D}^-$
  $(\chi \to \psi) \in \mathbf{b.n}, \chi \in \mathbf{b.n}$ により，$\mathbf{b} \ominus \psi$ は失敗する．この失敗を回避するには，ψ の除去の前に $\chi \to \psi$ か χ のどちらかを除去しておく必要がある．つまり，$(\mathbf{b} \ominus (\chi \to \psi)) \ominus \psi$ あるいは $(\mathbf{b} \ominus \chi) \ominus \psi$ とする必要があるが，これではそれぞれ $(\chi \to \psi) \notin \mathbf{b}'.\mathbf{n}$, $\chi \notin \mathbf{b}'.\mathbf{n}$ となってしまい，$(\chi \to \psi) \in \mathbf{b}'.\mathbf{n}$, $\chi \in \mathbf{b}'.\mathbf{n}$ にそれぞれ反する．これをも回避するには，$\chi \to \psi$ を除去した後に再び $\chi \to \psi$ を追加，あるいは χ を除去した後に再び χ を追加する必要がある．つまり，それぞれ $((\mathbf{b} \ominus (\chi \to \psi)) \ominus \psi) \oplus (\chi \to \psi)$, $((\mathbf{b} \ominus \chi) \ominus \psi) \oplus \chi$ とする必要があるが，それぞれの $(\chi \to \psi)$, χ は $\ominus$ された際に fixed belief となっているので，この $\oplus$ は失敗する．このように，失敗を回避する方法として失敗の要因となるものを『一旦除去して後で追加』(あるいは『一旦追加して後で除去』) を考えてみても，fixed belief の仕組みにより必ず失敗に終わる．
- $(\chi \to \psi) \in \mathbf{b.n}, (\chi \to \psi) \notin \mathbf{D}^-, \chi \notin \mathbf{b.n}, \chi \in \mathbf{D}^+$
  $\mathbf{b} \ominus \psi$ は成功する．しかし，$\chi \in \mathbf{D}^+$ より，$(\mathbf{b} \ominus \psi) \oplus \chi$ としなければならないが，これは $(\mathbf{b} \ominus \psi), \chi \rightsquigarrow_{\mathsf{NT}} \psi$ であるので，副作用的に ψ も追加されることになる．しかし，ψ は fixed belief なのでこの ψ の追加は失敗し，結果として $(\mathbf{b} \ominus \psi) \oplus \chi$ も失敗に終わる．$(\chi \to \psi) \notin \mathbf{D}^-$ と fixed belief の仕組みにより，この失敗は回避不可能である．
- $(\chi \to \psi) \notin \mathbf{b.n}, (\chi \to \psi) \in \mathbf{D}^+, \chi \in \mathbf{b.n}, \chi \notin \mathbf{D}^-$
  $\mathbf{b} \ominus \psi$ は成功する．しかし，$(\chi \to \psi) \in \mathbf{D}^+$ より，$(\mathbf{b} \ominus \psi) \oplus (\chi \to \psi)$ としなければならないが，これは $(\mathbf{b} \ominus \psi), (\chi \to \psi) \rightsquigarrow_{\mathsf{NT}} \psi$ であるので，副作用的に ψ も追加されることになる．しかし，ψ は fixed belief なのでこの ψ の追加は失敗し，結果として $(\mathbf{b} \ominus \psi) \oplus (\chi \to \psi)$ も失敗に終わる．$\chi \notin \mathbf{D}^-$ と fixed belief の仕組みにより，この失敗は回避不可能である．
- $(\chi \to \psi) \notin \mathbf{b.n}, (\chi \to \psi) \in \mathbf{D}^+, \chi \notin \mathbf{b.n}, \chi \in \mathbf{D}^+$
  $\mathbf{b} \ominus \psi$ は成功する．しかし，$(\chi \to \psi) \in \mathbf{D}^+$, $\chi \in \mathbf{D}^+$ より，$((\mathbf{b} \ominus \psi) \oplus (\chi \to \psi)) \oplus \chi$ あるいは $((\mathbf{b} \ominus \psi) \oplus \chi) \oplus (\chi \to \psi)$ としなければならず，どちらの場合も上と同様に失敗は回避不可能である．

よって，上のような ψ を実現する信念改竄オペレーションのシーケンスは存在しない．
以上により，補題 2 が証明された． □

補題 3. $\mathbf{b}$ と $\mathbf{b}'$ について，信念改竄条件中の条件 3 が満たされていない $\Rightarrow$P ではない．

証明． 補題 3 を証明する．
仮定より，「$\psi \in \mathbf{D}^-$ で，『$((\psi \to \chi) \in \mathbf{b.n}$ **or** $(\psi = (\gamma \to \chi), \gamma \in \mathbf{b.n})), \chi \in \mathbf{b.s}$ かつ『$\omega \neq \psi, ((\omega \to \chi) \in \mathbf{b.n}$ **or** $(\omega \to \chi) \in \mathbf{D}^+), (\omega \in \mathbf{b.n}$ **or** $\omega \in \mathbf{D}^+)$ であるような ω は存在しない』であるような χ が存在する』であるような ψ が存在する」なので，このような ψ が存在するとき，$\psi \in \mathbf{D}^-$ より，$\mathbf{b} \ominus \psi$ として成功するはずである．ここで，『$\omega \neq \psi, ((\omega \to \chi) \in \mathbf{b.n}$ **or** $(\omega \to \chi) \in \mathbf{D}^+), (\omega \in \mathbf{b.n}$ **or** $\omega \in \mathbf{D}^+)$ であるような ω は存在しない』の部分は，『全ての ω は $\omega = \psi$ **or** $((\omega \to \chi) \notin \mathbf{b.n}, (\omega \to \chi) \notin \mathbf{D}^+)$ **or** $(\omega \notin \mathbf{b.n}, \omega \notin \mathbf{D}^+)$ である』...(1) と同値である．一方，$\ominus$ の定義における $\{\chi \mid (((\psi \to \chi) \in \mathbf{b.n})$ **or** $(\psi = (\gamma \to \chi), \gamma \in \mathbf{b.n})), \chi \in \mathbf{b.s}, \{\omega \mid (\omega \to \chi) \in \mathbf{b.n}, \omega \in \mathbf{b}'.\mathbf{n}, \omega \neq \psi\} = \emptyset\} = \emptyset$ の部分により，「$(((\psi \to \chi) \in \mathbf{b.n})$ **or** $(\psi = (\gamma \to \chi), \gamma \in \mathbf{b.n})), \chi \in \mathbf{b.s}$,『全ての ω は $(\omega \to \chi) \notin \mathbf{b.n}$ **or** $\omega \notin \mathbf{b}'.\mathbf{n}$ **or** $\omega = \psi$ である』であるような χ が存在する」である場合に $\mathbf{b} \ominus \psi$ は失敗に終わる．ここで，(1) は次のように場合分けできる．

NII-Electronic Library Service

Kwansei Gakuin University

- $\omega = \psi$ の場合
- $((\omega \to \chi) \notin \mathbf{b.n}, (\omega \to \chi) \notin \mathbf{D}^+)$ の場合
- $(\omega \notin \mathbf{b.n}, \omega \notin \mathbf{D}^+)$ の場合

いずれの場合も $\mathbf{b} \ominus \psi$ は失敗し，補題 2 の証明と同様に，この失敗は fixed belief の仕組みにより回避不可能なものである．よって，上のような ψ を実現する信念改竄オペレーションのシーケンスは存在しない．
以上により，補題 3 が証明された．　□

補題 4. $\mathbf{b}$ と $\mathbf{b}'$ について，信念改竄条件中の条件 4 が満たされていない $\Rightarrow$P ではない．

証明．補題 4 を証明する．
仮定より，「$\psi \in \mathbf{D}^-$ で，ψ は fact で，『$((\mathbf{b.n}, \psi \rightsquigarrow \chi)$ **or** $((\mathbf{b.n}, \psi \not\rightsquigarrow \chi)$ を $\psi \rightsquigarrow \chi$ とするもの $\in \mathbf{D}^+))$, $((\bar{\chi} \in \mathbf{b.n})$ **or** $(\bar{\chi} \in \mathbf{D}^+))$ であるような χ は存在しない』であるような ψ が存在する」である．このような ψ が存在するとき，$\psi \in \mathbf{D}^-$ より，$\mathbf{b} \ominus \psi$ として成功するはずである．ここで，『$((\mathbf{b.n}, \psi \rightsquigarrow \chi)$ **or** $((\mathbf{b.n}, \psi \not\rightsquigarrow \chi)$ を $\psi \rightsquigarrow \chi$ とするもの $\in \mathbf{D}^+))$, $((\bar{\chi} \in \mathbf{b.n})$ **or** $(\bar{\chi} \in \mathbf{D}^+))$ であるような χ は存在しない』の部分は，『全ての χ は $((\mathbf{b.n}, \psi \not\rightsquigarrow \chi), ((\mathbf{b.n}, \psi \not\rightsquigarrow \chi)$ を $\psi \rightsquigarrow \chi$ とするもの $\notin \mathbf{D}^+))$ **or** $((\bar{\chi} \notin \mathbf{b.n}), (\bar{\chi} \notin \mathbf{D}^+))$ である』...(2) と同値である．一方，$\ominus$ の定義における $((\psi$ is a rule$)$ **or** $(\psi$ is a fact, $\{\chi \mid \mathbf{b.n}, \psi \rightsquigarrow \chi, \bar{\chi} \in \mathbf{b}'.\mathbf{n}\} \neq \emptyset))$ の部分により，「ψ is a fact,『全ての χ は $\mathbf{b.n}, \psi \not\rightsquigarrow \chi$ **or** $\bar{\chi} \notin \mathbf{b}'.\mathbf{n}$ である』であるような ψ が存在する」である場合に $\mathbf{b} \ominus \psi$ は失敗に終わる．ここで，(2) は次のように場合分けできる．

- $((\mathbf{b.n}, \psi \not\rightsquigarrow \chi), ((\mathbf{b.n}, \psi \not\rightsquigarrow \chi)$ を $\psi \rightsquigarrow \chi$ とするもの $\notin \mathbf{D}^+))$ の場合
- $((\bar{\chi} \notin \mathbf{b.n}), (\bar{\chi} \notin \mathbf{D}^+))$ の場合

いずれの場合も $\mathbf{b} \ominus \psi$ は失敗し，補題 2 の証明と同様に，この失敗は fixed belief の仕組みにより回避不可能なものである．よって，上のような ψ を実現する信念改竄オペレーションのシーケンスは存在しない．
以上により，補題 4 が証明された．　□

補題 5. $\mathbf{b}$ と $\mathbf{b}'$ について，信念改竄条件中の条件 5 が満たされていない $\Rightarrow$P ではない．

証明．補題 5 を証明する．
仮定より，「$\psi \in \mathbf{D}^-$ で ψ が positive な論理式で，$\neg\psi \notin \mathbf{D}^+$ であるような ψ が存在する」なので，このような ψ が存在するとき，$\psi \in \mathbf{D}^-$ より，$\mathbf{b} \ominus \psi$ として成功するはずである．しかし，$\mathbf{b} \ominus \psi$ が成功する場合は，「ψ が positive な論理式」により $(\mathbf{b} \setminus \{\psi\}) \oplus \neg\psi$ となり，$\neg\psi \notin \mathbf{D}^+$ に反する．よって，上のような ψ を実現する信念改竄オペレーションのシーケンスは存在しない．
以上により，補題 5 が証明された．　□

補題 6. $\mathbf{b}$ と $\mathbf{b}'$ について，信念改竄条件中の条件 6 が満たされていない $\Rightarrow$P ではない．

証明．補題 6 を証明する．
仮定より，「$\psi \in \mathbf{D}^+$ で，『$(\mathbf{b}'.\mathbf{n} \setminus \{\psi\}), \psi \rightsquigarrow_{\mathsf{NT}} \chi$ かつ $\chi \notin \mathbf{b}'.\mathbf{n}$ であるような χ が存在する』であるような ψ が存在する」なので，このような ψ が存在するとき，$\psi \in \mathbf{D}^+$ より，$\mathbf{b} \oplus \psi$ として成功するはずである．ここで，$(\mathbf{b}'.\mathbf{n} \setminus \{\psi\}), \psi \rightsquigarrow_{\mathsf{NT}} \chi$ かつ $\chi \notin \mathbf{b}'.\mathbf{n}$ の部分は，次のように場合分けされる．

- $\mathbf{b.n}, \psi \rightsquigarrow_{\mathsf{NT}} \chi, (\mathbf{b}'.\mathbf{n} \setminus \{\psi\}), \psi \rightsquigarrow_{\mathsf{NT}} \chi$
  $\mathbf{b.n} \oplus \psi$ が成功の場合，$\mathbf{b.n}, \psi \rightsquigarrow_{\mathsf{NT}} \chi$ により $\chi \in (\mathbf{b.n} \oplus \psi)$ となるので，$(\mathbf{b}'.\mathbf{n} \setminus \{\psi\}), \psi \rightsquigarrow_{\mathsf{NT}} \chi$ と

NII-Electronic Library Service

Kwansei Gakuin University

矛盾する．補題 2 の証明と同様に，これは fixed belief の仕組みにより回避不可能なものである．

- $\mathbf{b.n}, \psi \not\rightsquigarrow_{\mathsf{NT}} \chi, (\mathbf{b}'.\mathbf{n} \setminus \{\psi\}), \psi \rightsquigarrow_{\mathsf{NT}} \chi, \chi \notin \mathbf{b.n}, \chi \notin \mathbf{D}^+$（つまり，$\psi$ の追加と，$\psi \rightsquigarrow_{\mathsf{NT}} \chi$ とするような追加・除去）
  - $\mathbf{b.n} \oplus \psi$ は χ に関与しないが，$(\mathbf{b}'.\mathbf{n} \setminus \{\psi\}), \psi \rightsquigarrow_{\mathsf{NT}} \chi$ とさせようとすると必ず χ が副作用として追加される．もちろん，$\ominus\chi$ を考えても $\rightsquigarrow_{\mathsf{T}} \chi$ により失敗．
  - $\psi \rightsquigarrow_{\mathsf{NT}} \chi$ とさせた後に $\oplus\psi$ という順番で行うと，$(\mathbf{b}'.\mathbf{n} \setminus \{\psi\}), \psi \rightsquigarrow_{\mathsf{NT}} \chi$ に反し，これを回避しようとしても $\rightsquigarrow_{\mathsf{T}} \chi$ に阻まれて失敗．
- $\mathbf{b.n}, \psi \not\rightsquigarrow_{\mathsf{NT}} \chi, (\mathbf{b}'.\mathbf{n} \setminus \{\psi\}), \psi \rightsquigarrow_{\mathsf{NT}} \chi, \chi \in \mathbf{b.n}, \chi \in \mathbf{D}^-$（つまり，$\psi$ の追加と χ の除去と，$\psi \rightsquigarrow_{\mathsf{NT}} \chi$ とするような追加・除去）
  - 上記と同様にどの順番で行っても失敗する．

よって，上のような ψ を実現する信念改竄オペレーションのシーケンスは存在しない．
以上により，補題 6 が証明された． □

補題 7. $\mathbf{b}$ と $\mathbf{b}'$ について，信念改竄条件中の条件 7 が満たされていない $\Rightarrow$P ではない．

証明．補題 7 を証明する．
仮定より，「$\psi \in \mathbf{D}^+$ かつ $\bar{\psi} \in \mathbf{b}'.\mathbf{n}$ であるような ψ が存在する」である．
このような ψ が存在するとき，$\psi \in \mathbf{D}^+$ より，$\mathbf{b} \oplus \psi$ として成功するはずである．ここで，$\bar{\psi} \in \mathbf{b}'.\mathbf{n}$ の部分は，次の 2 つの場合に分けられる．

- $\bar{\psi} \in \mathbf{b.n}, \bar{\psi} \notin \mathbf{D}^-$
- $\bar{\psi} \notin \mathbf{b.n}, \bar{\psi} \in \mathbf{D}^+$

どちらの場合も，明らかに失敗する．よって，上のような ψ を実現する信念改竄オペレーションのシーケンスは存在しない．
以上により，補題 7 が証明された． □

補題 8. Q$\Rightarrow$P.

証明．補題 8 を証明する．
信念世界 $\mathbf{b}, \mathbf{b}'$ を考える．（追加集合 $\mathbf{D}^+ = \mathbf{b}'.\mathbf{n} \setminus \mathbf{b.n}$，除去集合 $\mathbf{D}^- = \mathbf{b.n} \setminus \mathbf{b}'.\mathbf{n}$ とし，$\mathbf{b}$ と $\mathbf{b}'$ の間には $\mathbf{b}'.\mathbf{s} = \mathbf{b.s}$，$\mathbf{b}'.\mathbf{f} = \mathbf{b.f} \cup \mathbf{D}^+ \cup \mathbf{D}^-$ という関係が満たされているものとする）．ここではさらに，$\mathbf{D}^+ = \{p_0, p_1, ..., p_n\}$, $\mathbf{D}^- = \{q_0, q_1, ..., q_m\}$ とおく．そして，$\mathbf{D}^+$ の要素を順次 $\oplus$ していき，次いで $\mathbf{D}^-$ の要素を $\ominus$ していくことを考える．$\mathbf{b} \oplus p_0$ が成功するなら，$\mathbf{b} := \mathbf{b} \oplus p_0, \mathbf{D}^+ := \{p_1, p_2, ..., p_n\}$（$\mathbf{D}^+$ の要素はもっと減っている可能性もあるが）の様に操作を繰り返せばよく，$\mathbf{D}^-$ については $\mathbf{b} \ominus q_0$ が成功するなら同様の操作を繰り返せばよい（$\mathbf{D}^+ = \mathbf{D}^- = \emptyset$ で終了）ので，以下では $\mathbf{b} \oplus p_0$ と $\mathbf{b} \ominus q_0$ が成功することのみを示す．

● $\mathbf{b} \oplus p_0$ を考える．
$\mathbf{b} \oplus p_0$ は，$\{p_0\} \cup \{\chi \mid \mathbf{b.n}, p_0 \rightsquigarrow_{\mathsf{NT}} \chi\}$ の要素を $\mathbf{b.n}$ に加えた集合が無矛盾でかつその加えられたものが $\mathbf{b.f}$ の要素でないなら，成功する．この加えられるもの全てが $\mathbf{D}^+$ の要素であるなら，条件 7 より，加えられた集合は無矛盾となる．ここで，条件 6 と $p_0 \in \mathbf{D}^+$ より，$(\mathbf{b}'.\mathbf{n} \setminus \{p_0\}), p_0 \rightsquigarrow_{\mathsf{NT}} \chi$ かつ $\chi \notin \mathbf{b}'.\mathbf{n}$ であるような

NII-Electronic Library Service

Kwansei Gakuin University

χ は存在しないので，任意の χ は $(\mathbf{b}'.\mathbf{n} \setminus \{p_0\}), p_0 \not\rightsquigarrow_{\mathsf{NT}} \chi$ **or** $\chi \in \mathbf{b}'.\mathbf{n}$ である．これを以下のように場合分けし，それぞれの場合について，加えられるものが $\mathbf{D}^+$ の要素でかつ $\mathbf{b}.\mathbf{f}$ の要素でないことを示す．

- $(\mathbf{b}'.\mathbf{n} \setminus \{p_0\}), p_0 \rightsquigarrow_{\mathsf{NT}} \chi$ の場合
  $(\mathbf{b}'.\mathbf{n} \setminus \{p_0\}), p_0 \not\rightsquigarrow_{\mathsf{NT}} \chi$ **or** $\chi \in \mathbf{b}'.\mathbf{n}$ より，$\chi \in \mathbf{b}'.\mathbf{n}$ である．しかし，$(\mathbf{b}'.\mathbf{n} \setminus \{p_0\}), p_0 \rightsquigarrow_{\mathsf{NT}} \chi$ と $\chi \in \mathbf{b}'.\mathbf{n}$ は矛盾する．よって，この場合は考えなくてよい．
- $(\mathbf{b}'.\mathbf{n} \setminus \{p_0\}), p_0 \not\rightsquigarrow_{\mathsf{NT}} \chi$ の場合
  - $\mathbf{b}.\mathbf{n}, p_0 \rightsquigarrow_{\mathsf{NT}} \chi$ の場合
    $\chi \notin \mathbf{b}.\mathbf{n}$ なので，$\chi \notin \mathbf{D}^-$ である．
    * $\chi \in \mathbf{b}'.\mathbf{n}$ の場合
      $\chi \notin \mathbf{b}.\mathbf{n}$ より，$\chi \in \mathbf{D}^+$ である．χ が fixed belief であるとすると，$\oplus\chi$ か $\ominus\chi$ が実行済みであることになるが，$\oplus\chi$ は今試みているところであるので，可能性としては $\ominus\chi$ のみが考えられる．しかし，$\chi \notin \mathbf{D}^-$ により $\ominus\chi$ は実行されていないことが保証される．よって，χ は fixed belief ではない．
    * $\chi \notin \mathbf{b}'.\mathbf{n}$ の場合
      $\chi \notin \mathbf{b}.\mathbf{n}$ より，$\chi \notin \mathbf{D}^+$ である．そこで，χ を加えられるものでなくさせることを考える．χ の追加は，$p_0 \rightsquigarrow_{\mathsf{NT}} \chi$ である場合に $\oplus p_0$ の副作用として行われるので，$\oplus p_0$ を実行する前に $p_0 \not\rightsquigarrow_{\mathsf{NT}} \chi$ にしておけばよい．$p_0 \not\rightsquigarrow_{\mathsf{NT}} \chi$ にするには，$\rightsquigarrow_{\mathsf{NT}}$ の定義により，その時点の信念世界の belief を s とすると，$(\psi_0 \rightarrow \psi_1), (\psi_1 \rightarrow \psi_2), \ldots, (\psi_{m-1} \rightarrow \psi_m) \in s$ でなくさせるか，あるいは $\psi_0, \psi_1, ..., \psi_m \notin s$ でなくさせるかのどちらかである．ここで，以上の事実と $\mathbf{b}.\mathbf{n}, p_0 \rightsquigarrow_{\mathsf{NT}} \chi$, $(\mathbf{b}'.\mathbf{n} \setminus \{p_0\}), p_0 \not\rightsquigarrow_{\mathsf{NT}} \chi$ により，$(p_0 \rightarrow p_1), (p_1 \rightarrow p_2), \ldots, (p_x \rightarrow \chi)$ 中の 1 つ以上の rule が $\mathbf{D}^-$ の要素である場合と，$p_0, p_1, p_2, \ldots, p_x, \chi$ 中の 1 つ以上の fact が $\mathbf{D}^+$ の要素である場合が考えられる．しかし，後者の fact を $\oplus$ する方法では，その fact から χ を $\rightsquigarrow_{\mathsf{NT}}$ できるため，副作用として χ も追加されてしまう．この副作用の過程で，χ を $\rightsquigarrow_{\mathsf{T}}$ する仕組みができあがり，その結果 χ は除去不可能となり，$\chi \notin \mathbf{D}^+$ に反する．よって，前者の rule を $\ominus$ する方法が適当である．つまり，その rule は $\mathbf{D}^-$ の要素である．$\mathbf{D}^-$ の要素の $\ominus$ は必ず成功することは下で証明している．以上により，χ を加えられるものでなくさせることができる．
  - $\mathbf{b}.\mathbf{n}, p_0 \not\rightsquigarrow_{\mathsf{NT}} \chi$ の場合
    $\oplus p_0$ の実行の副作用として χ が追加されることはない．

よって，加えられるもの全てが $\mathbf{D}^+$ の要素であり，その加えられたものは $\mathbf{b}.\mathbf{f}$ の要素でないので，$\mathbf{b} \oplus p_0$ は必ず成功する．

● $\mathbf{b} \ominus q_0$ を考える．
オペレータ $\mathbf{b} \ominus q_0$ の条件リストを全て満たすことを以下に示す．

(0) (ψ is a negative formula) **or** (オペレータ $\langle(\mathbf{b}.\mathbf{n} \setminus \{\psi\}), \mathbf{b}.\mathbf{s}, (\mathbf{b}.\mathbf{f} \cup \{\psi\})\rangle \oplus \neg\psi$ の条件リストを満たす)
- q_0 が negative の場合，
  明らかに (0) を満たす．

NII-Electronic Library Service

Kwansei Gakuin University

– q_0 が positive の場合,

条件 5, $q_0 \in \mathbf{D}^-$, q_0 が positive であること, により $\neg q_0 \in \mathbf{D}^+$ である. そして, $\mathbf{D}^+$ の要素の $\oplus$ が必ず成功することは上で証明した通りであるので, (0) は満たされる.

- $q_0 \notin \mathbf{b.n}$ を満たしている場合は, この時点で成功.
- $q_0 \in \mathbf{b.n}$ の場合,

(1) $q_0 \notin \mathbf{b.f}$ について

もし $q_0 \in \mathbf{b.f}$ であるなら $\oplus q_0$ が実行済みであるはずだが, $q_0 \notin \mathbf{D}^+$ より, $\oplus q_0$ は実行済みであり得ない. よって, (1) は満たされる.

(2) $q_0 \notin \mathbf{b.s}$ について

条件 1 と $q_0 \in \mathbf{D}^-$ から, (2) は満たされる.

(3) $\{\chi \mid (\mathbf{b}'.\mathbf{n} \cup \{q_0\}), \chi \rightsquigarrow_{\mathsf{T}} q_0\} = \emptyset$ について

条件 2 と $q_0 \in \mathbf{D}^-$ から, $\{\chi \mid (\chi \rightarrow q_0) \in \mathbf{b}'.\mathbf{n}, \chi \in \mathbf{b}'.\mathbf{n}\} = \emptyset$ である. つまり,「全ての χ について, $(\chi \rightarrow q_0) \notin \mathbf{b}'.\mathbf{n}$ **or** $\chi \notin \mathbf{b}'.\mathbf{n}$ である」. $(\chi \rightarrow q_0) \notin \mathbf{b}'.\mathbf{n}$ は $(\chi \rightarrow q_0) \in \mathbf{b.n}, (\chi \rightarrow q_0) \in \mathbf{D}^-$ あるいは $(\chi \rightarrow q_0) \notin \mathbf{b.n}, (\chi \rightarrow q_0) \notin \mathbf{D}^+$ であるが, 前者の場合は $\ominus(\chi \rightarrow q_0)$ を $\ominus q_0$ より先に実行しておけば (3) を満たし, 後者の場合は $\chi \rightsquigarrow_{\mathsf{T}} q_0$ でなくまたどんな rule も副作用的には追加されないので (3) を満たす. 一方, $\chi \notin \mathbf{b}'.\mathbf{n}$ は $\chi \in \mathbf{b.n}, \chi \in \mathbf{D}^-$ あるいは $\chi \notin \mathbf{b.n}, \chi \notin \mathbf{D}^+$ であるが, 前者の場合は $\ominus\chi$ を $\ominus q_0$ より先に実行しておけば (3) を満たし, 後者の場合は $\chi \rightsquigarrow_{\mathsf{T}} q_0$ でなくまた副作用的に χ は追加されないので (3) を満たす. (何故なら, 副作用的に χ が追加されたと仮定すると, $\chi \notin \mathbf{D}^+$ より, $\ominus\chi$ が実行・成功されたことになるが, $\ominus\chi$ は $\gamma \rightsquigarrow_{\mathsf{T}} \chi$ であるような γ が存在するために必ず失敗する. よって, そもそもの仮定が間違っていることになる). よって, (3) は満たされる.

(4) $\{\chi \mid ((q_0 \rightarrow \chi) \in \mathbf{b.n}$ **or** $(q_0 = (\gamma \rightarrow \chi), \gamma \in \mathbf{b.n})), \chi \in \mathbf{b.s}, \{\omega \mid (\omega \rightarrow \chi) \in \mathbf{b.n}, \omega \in \mathbf{b}'.\mathbf{n}, \omega \neq q_0\} = \emptyset\} = \emptyset$ について

条件 3 と $q_0 \in \mathbf{D}^-$ から, $\{\chi \mid ((q_0 \rightarrow \chi) \in \mathbf{b.n}$ **or** $(q_0 = (\gamma \rightarrow \chi), \gamma \in \mathbf{b.n})), \chi \in \mathbf{b.s}, \{\omega \mid q_0 \neq \omega, ((\omega \rightarrow \chi) \in \mathbf{b.n}$ **or** $(\omega \rightarrow \chi) \in \mathbf{D}^+), (\omega \in \mathbf{b.n}$ **or** $\omega \in \mathbf{D}^+)\} = \emptyset\} = \emptyset$ である. この式中で (4) と異なっているのは, $\{\omega \mid q_0 \neq \omega, ((\omega \rightarrow \chi) \in \mathbf{b.n}$ **or** $(\omega \rightarrow \chi) \in \mathbf{D}^+), (\omega \in \mathbf{b.n}$ **or** $\omega \in \mathbf{D}^+)\} = \emptyset$ の部分である. ここで, 次のように場合分けできる.

* 任意の χ について, $((q_0 \rightarrow \chi) \in \mathbf{b.n}$ **or** $(q_0 = (\gamma \rightarrow \chi)$ **and** $\gamma \in \mathbf{b.n}))$ **and** $\chi \in \mathbf{b.s}$ を満たさない場合

  (4) は満たされる.

* そうでない場合

  条件 3 より, $\{\omega \mid q_0 \neq \omega, ((\omega \rightarrow \chi) \in \mathbf{b.n}$ **or** $(\omega \rightarrow \chi) \in \mathbf{D}^+), (\omega \in \mathbf{b.n}$ **or** $\omega \in \mathbf{D}^+)\} = \emptyset$ である. この時, $(\omega \rightarrow \chi)$, ω はそれぞれ, もともと $\mathbf{b.n}$ の要素であるか, そうでなければ $\mathbf{D}^+$ の要素であるので $\ominus q_0$ の前に $\oplus$ しておくことができるので $\oplus$ しておけば, (4) を満たす.

よって, (4) は満たされる.

(5) (q_0 is not a fact) **or** $\{\chi \mid \mathbf{b.n}, q_0 \rightsquigarrow \chi, \bar{\chi} \in \mathbf{b}'.\mathbf{n}\} \neq \emptyset$ について

条件 4 と $q_0 \in \mathbf{D}^-$ から, q_0 is a rule **or** $\{\chi \mid ((\mathbf{b.n}, q_0 \rightsquigarrow \chi)$ **or** $((\mathbf{b.n}, q_0 \not\rightsquigarrow \chi), q_0 \rightsquigarrow \chi$ とするもの $\in \mathbf{D}^+)), ((\bar{\chi} \in \mathbf{b.n})$ **or** $(\bar{\chi} \in \mathbf{D}^+))\} \neq \emptyset$ である. $q_0 \rightsquigarrow \chi$ とするもの, $\bar{\chi}$ をそれぞれ $\ominus q_0$ より先に $\oplus$ しておくことができるので $\oplus$ しておけば, (5) は満たされる.

NII-Electronic Library Service

Kwansei Gakuin University

以上により，$\mathbf{b}$ と $\mathbf{b}'.\mathbf{n}$ について信念改竄条件が満たされているとき，$\mathbf{b}$ から $\mathbf{b}'$ を作り出すような信念改竄オペレーションからなるシーケンスが必ず存在する．よって，補題 8 が証明された．□

証明．定理 1 を証明する．
補題 1〜7 により信念改竄条件中のどの 1 つの条件を満たさなくても P は満たされないので，¬Q⇒ ¬P つまり P⇒Q である．
補題 8 により Q⇒P である．
よって，P **iff** Q である．□

定理 1 により，ある特定の改竄が可能であるかの判定と，もし可能であるならどのような改竄になるのかを，信念改竄条件から導くことができる．これにより，つきたい嘘をばれない嘘にすることが可能かどうか，そして最終的に相手の信念世界がどのようになるのか，という議論を簡潔に行うことができる．

5. 関連研究

既存の研究との比較を行う．

- Belief Revision との比較

  信念改竄は信念修正 (Belief Revision)[1][5][6][7][9][11] の変種と考えることができる．本研究の特徴は，演繹だけでなくアブダクションに対する考慮も行っている点，人間に何かを信じさせることの難しさを緩和させる試みとして背理法による証明を義務付けている点，人間の行う推論の深さに明確な基準を設けている点，などである．なお，信念修正の目的は信念を正しい方向に向かわせるための整合化であり，本研究のように間違いだとわかっていることを正当化するためのものではない．このような理由で，本研究では「改竄」という言葉を用いている．

- Argumentation との比較

  Argumentation[15] における undercut の概念は，除去対象を演繹できるような仕組みを破壊するという点で，本研究と似ている．しかし，本研究では背理法による証明やアブダクションを考慮している点等で，一般的な undercut の概念とは異なっている．また， undercut は信念データベースへの追加等の変更を伴わないが，本研究は信念データベースそのものを変更しており，より複雑な問題を扱っていると考えられる．

- Awareness との比較

  信念と論理的全知の問題を扱った研究はいくつか行われている．Lebesque は，論理的全知でないことを適切に表現するために， explicit belief の概念を提案している [10]．その一方，気付いている事柄と気付いていない事柄を明示的に記述する awareness の概念なども提案されている [4]．本研究における論理的全知の問題への対処は awareness の概念に類似のものといえる．

6. 結論

人間の推論モデルとして信念システムを構成し，その上でのばれない嘘を定義した．そして，つきたい嘘をばれない嘘として相手の信念世界に追加するための手順を導き出すアルゴリズムを，プラニングの手法を用いて提案した．また，つきたい嘘をばれない嘘にすることが可能であるかを判定するための条件を示し，その妥当性を証明した．

NII-Electronic Library Service

Kwansei Gakuin University

本研究はエージェント間の交渉，特に相手を納得させることが必要であるような交渉への応用が考えられる．その他にも，人間らしいロボットの会話システムに導入するなど，様々な応用が考えられる．

今後の課題としては，信念に確信度の概念を導入することなどが挙げられる．

参考文献

[1] Carlos E. Alchourrón, Peter Gärdenfors, David Makinson: On the Logic of Theory Change: Partial Meet Contraction and Revision Functions. The Journal of Symbolic Logic, Vol. 50, No. 2 (1985) 510–530

[2] Alexander Bochman: A Logic For Causal Reasoning. IJCAI-03 (2003) 141–146

[3] Ronald Fagin, Joseph Y. Halpern, Yoram Moses and Moshe Y. Vardi: Reasoning About Knowledge. The MIT Press (1995)

[4] Ronald Fagin and Joseph Y. Halpern: Belief, Awareness, and Limited Reasoning. Artificial Intelligence, Vol. 34 (1988) 39–76

[5] Peter Gärdenfors: Belief Revision: An introduction. Belief Revision. Cambridge University Press (1992) 1–20

[6] Peter Gärdenfors and Hans Rott: Belief Revision. in D. M. Gabbay, C. J. Hogger, J. A. Robinson (eds.). Handbook of Logic in Articial Intelligence and Logic Programming, Vol. 4. Oxford UniversityPress (1995) 35–132

[7] Gösta Grahne, Alberto O. Mendelzon and Raymond Reiter: On the Semantics of Belief Revision Systems. Theoretical Aspects of Reasoning about Knowledge: Proceedings of the Fourth Conference (1992) 132–142

[8] 箱田裕司, 仁平義明: 嘘とだましの心理学 戦略的なだましからあたたかい嘘まで. 有斐閣 (2006)

[9] Hirofumi Katsuno and Hideki Isozaki: Simplified semantic structures for representing belief states in multi-agent environments. IEICE Transactions on Information and Systems, Vol. E84-D, No. 1 (2001) 129–141

[10] Hector J. Levesque: A logic of implicit and explicit belief. In Proceedings of the National Conference on Artificial Intelligence (1984) 198–202

[11] Bernhard Nebel: A Knowledge Level Analysis of Belief Revision. In Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning (1989) 301–311

[12] Nils J. Nilsson and Richard E. Fikes: STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving. Artificial Intelligence, Vol. 2, (1971) 189–208

[13] 奥野健一，高橋和子: 信念改竄によるばれない嘘の生成. In Proceedings of the 21st Annual Conference of the JSAI, 3E9-1 (2007)

[14] Erik. J. Olsson: Making Beliefs Coherent. Journal of Logic, Language, and Information, Vol 7 (1998) 143–163

[15] Michael J. Wooldridge: An Introduction to Multiagent Systems. Wiley (2002)

NII-Electronic Library Service