
認識について考える 2
認識の認識

認識について考える 2-- 認識の認識 –

今回のマルゼミは、前回のマルレク「認識について考える」の続編です。

前回は、主要に、自然認識とそれを可能とする条件の変化の歴史を振り返ったのですが、今回は、「認識の認識」という構造に注目しようと思います。認識の対象となるのは、自然ではなく、認識そのものです。

以下、その概略を紹介しようと思います。

認識の形式的理論

具体的な自然認識を対象とした前回からは、少し飛躍があるのですが、今回のセミナーでは、「認識の認識」を、「認識の形式的・数学的認識」として捉えるというアプローチを取ってみようと思います。

認識のいくつかの特徴を、形式的・数学的に把握することが可能であるという立場をとります。

認識の発展のモデル

人間の認識は、変化・発展するのですが、「認識の発展」の数学的モデルが存在します。今回のセミナーでは、そのいくつかを紹介しようと思います。

- **Part I** ：Grzegorczykの「科学の探究」モデル
- **Part II** ：Kripke の「可能的世界」モデル
- **Part III**: Bayesian推論と相対エントロピー
- **Part IV** ：Jaynesの「最大エントロピー原理」

Part I

Grzegorczykの「科学の探究」モデル

Part I

Grzegorczykの「科学の探究」モデル

1. 科学的探究とは何か？

2. 情報の順序関係

3. Forcing Method

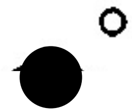

Grzegorczykの
「科学の探究」モデル　1
科学的探究とは何か？

科学的探究とは？

「科学的探求（たとえば，実験的研究）とは，我々の探求の方法によって得られた，新たに確立された諸事実によって，データの集合を継続的に豊富化することに存する。」

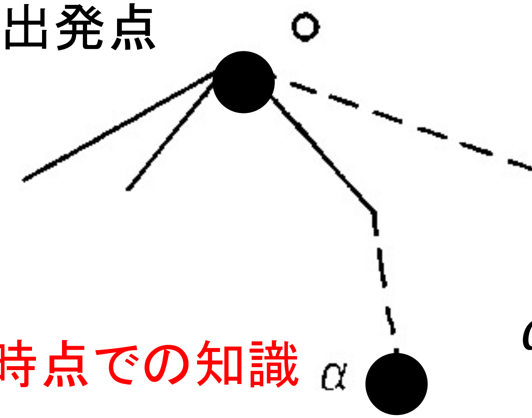
科学的探究＝
新たな事実のデータを継続的に豊富化すること

探究の出発点



oは，科学的探求に先立って，我々に与えられている情報である。地点oから，実験等の試行を繰り返し，新しいデータをえながら，我々は枝にそって進む。何回かの試行を繰り返しながら，我々は，経験的なデータの集まりαを，我々の知識として得たとする。

探究の出発点



現時点での知識

$\alpha = (A_1, A_2, \ldots, A_n)$

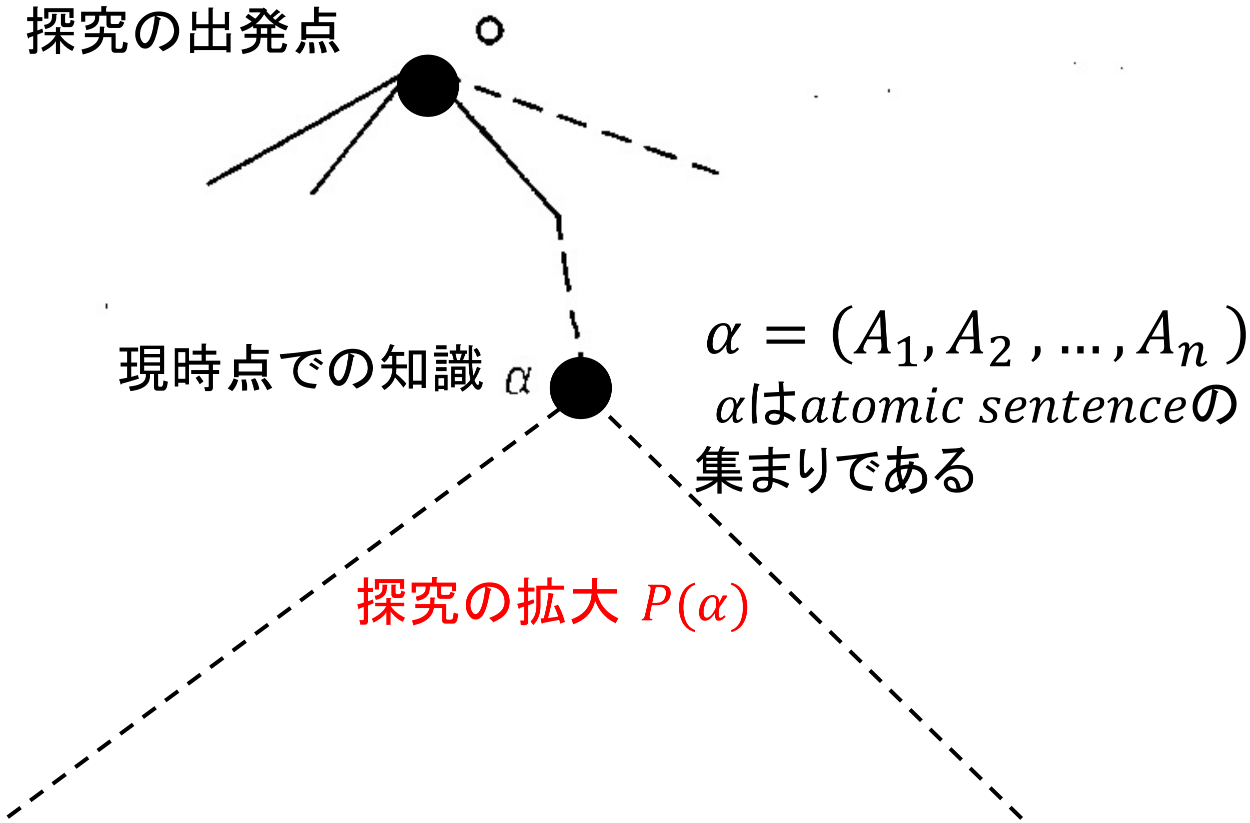
αは$atomic\ sentence$の集まりである

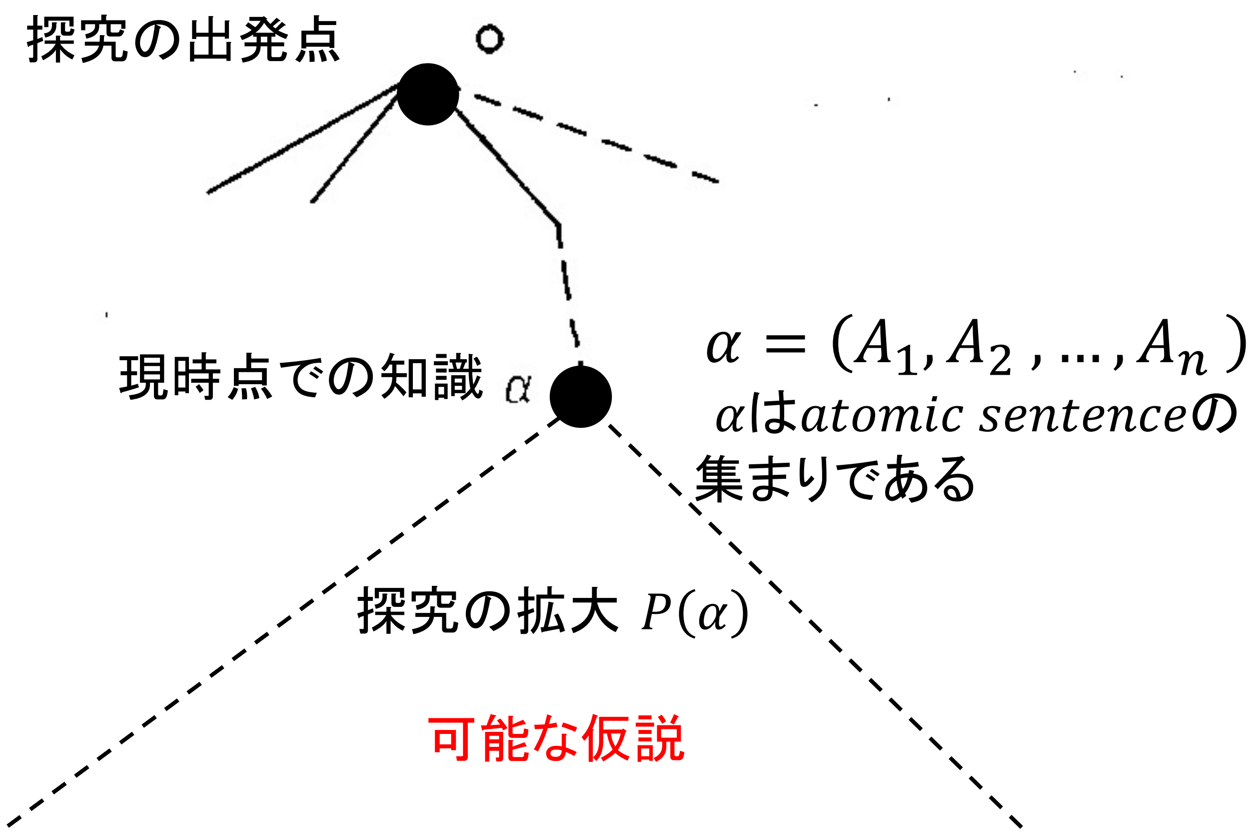
この段階でも我々は探求を続ける。

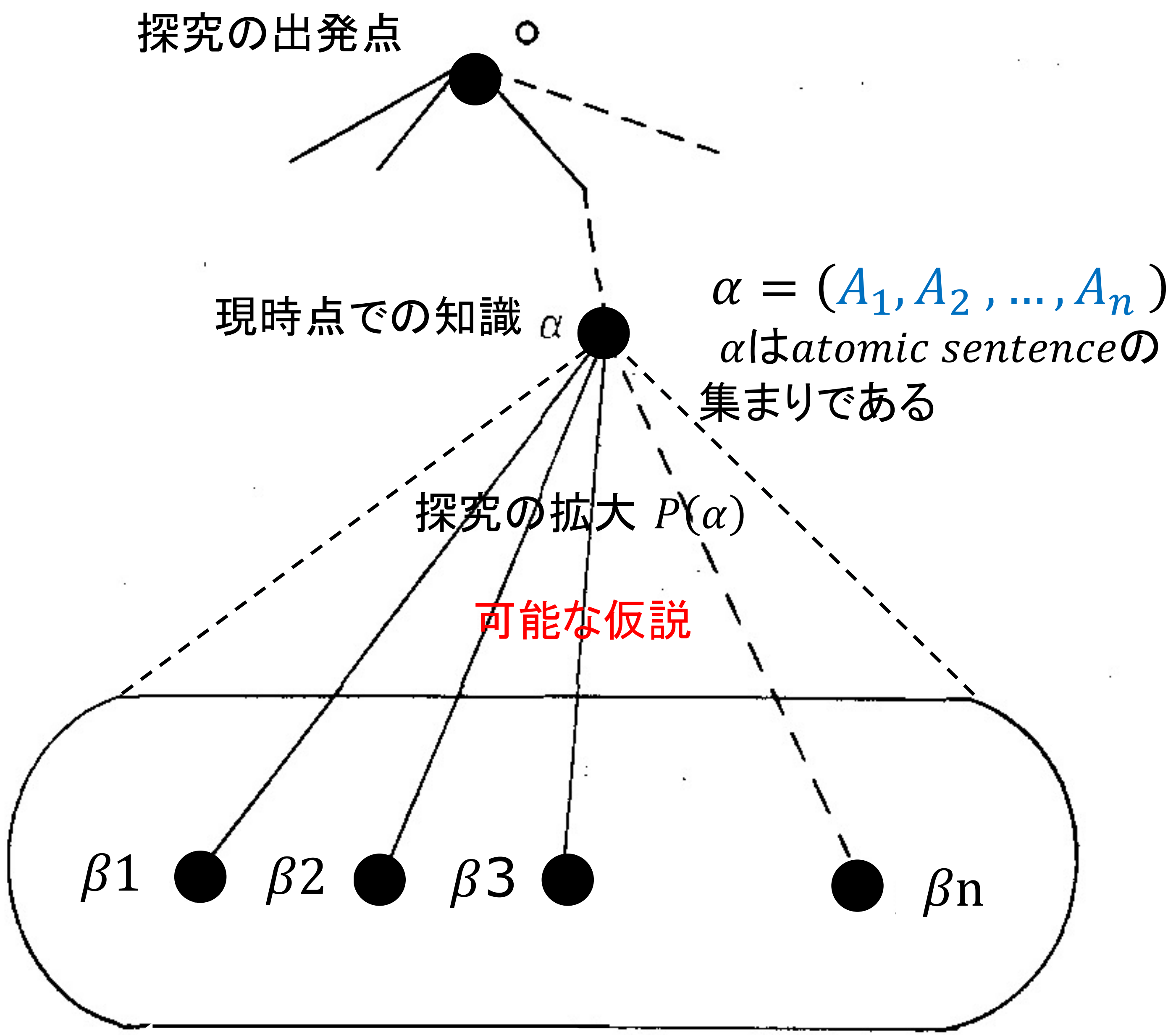
彼の表現を借りると,「我々は，自然に可能な解答の集合を提供する。」

この情報αの段階での「可能な解答の集合」がP(α)である。

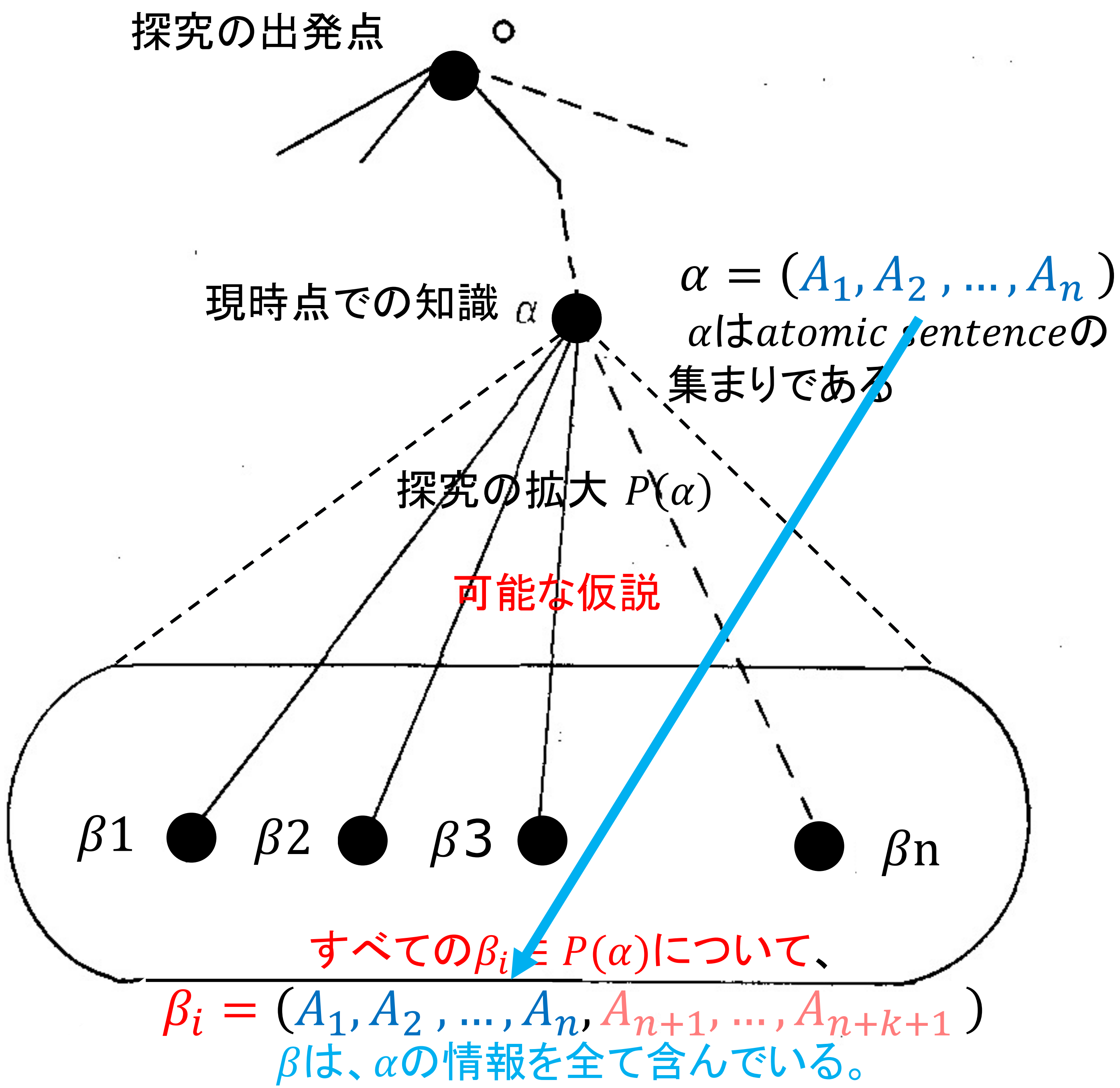
その要素 $\beta 1, \beta 2, \ldots$ の中から、「自然はひとつを選ぶ。」

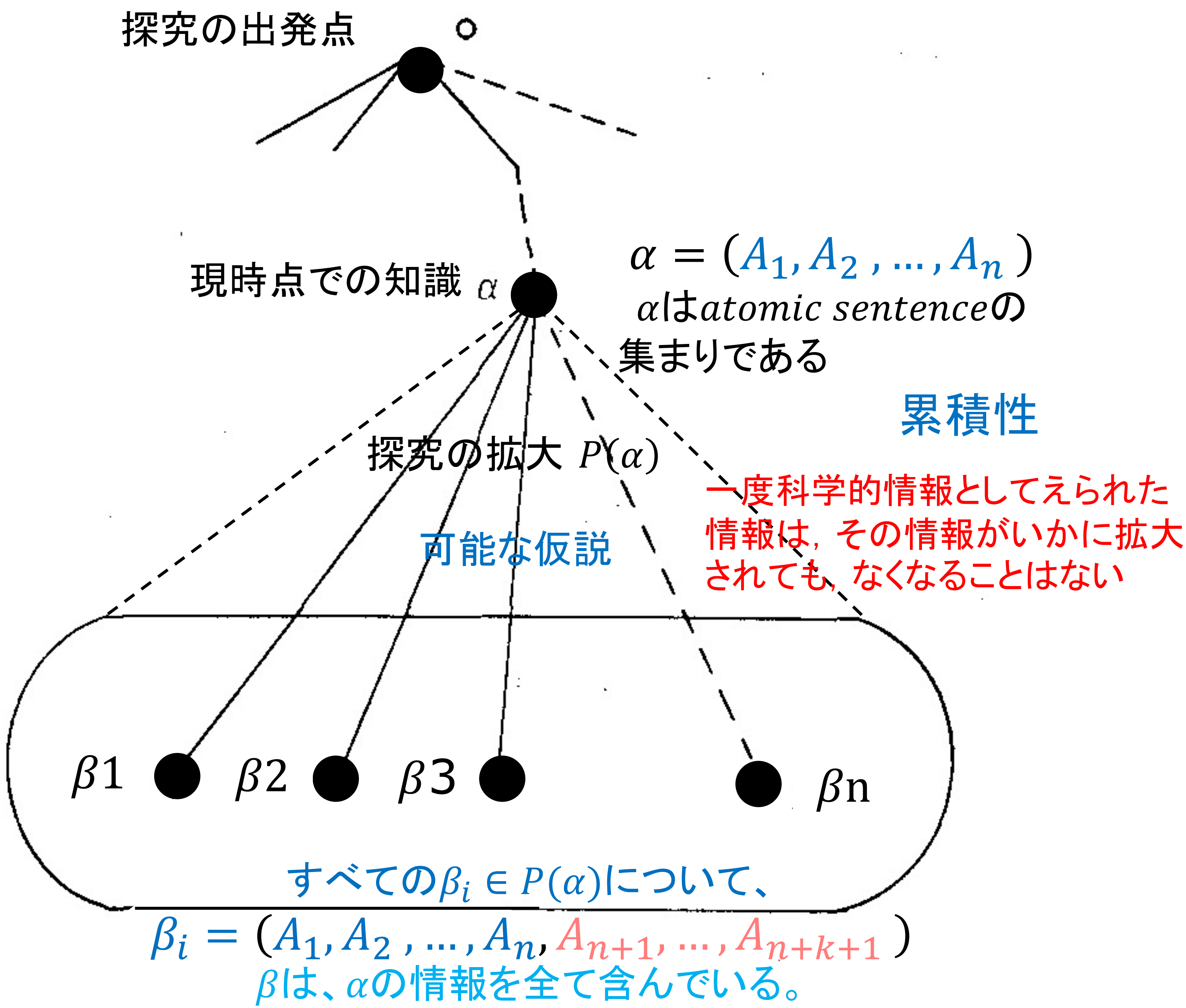

探究の出発点
現時点での知識 α
$\alpha = (A_1, A_2, \ldots, A_n)$
αは$atomic\ sentence$の
集まりである
探究の拡大 $P(\alpha)$

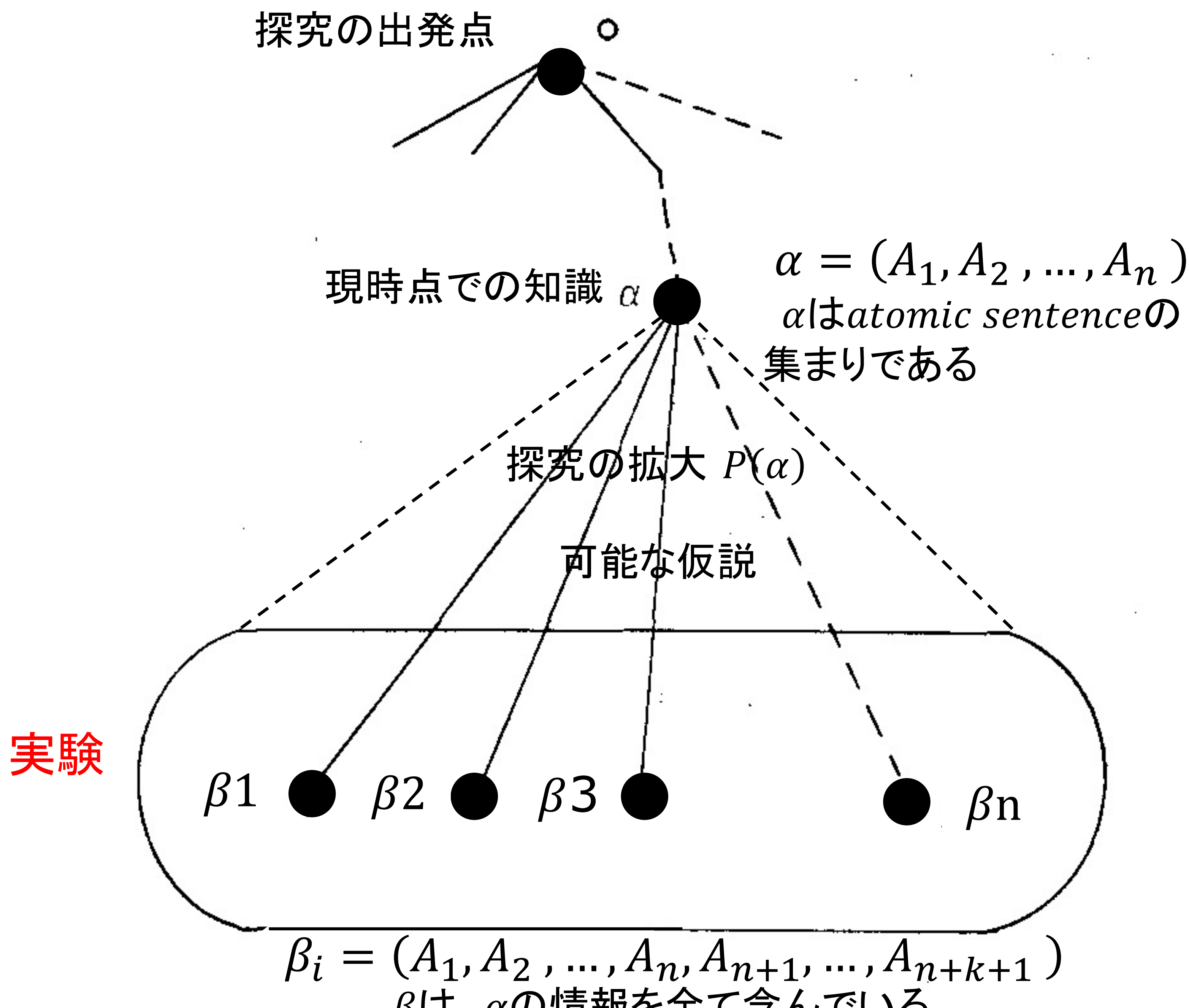

探究の出発点
現時点での知識 α
$\alpha = (A_1, A_2, \ldots, A_n)$
αはatomic sentenceの集まりである
探究の拡大 $P(\alpha)$
可能な仮説

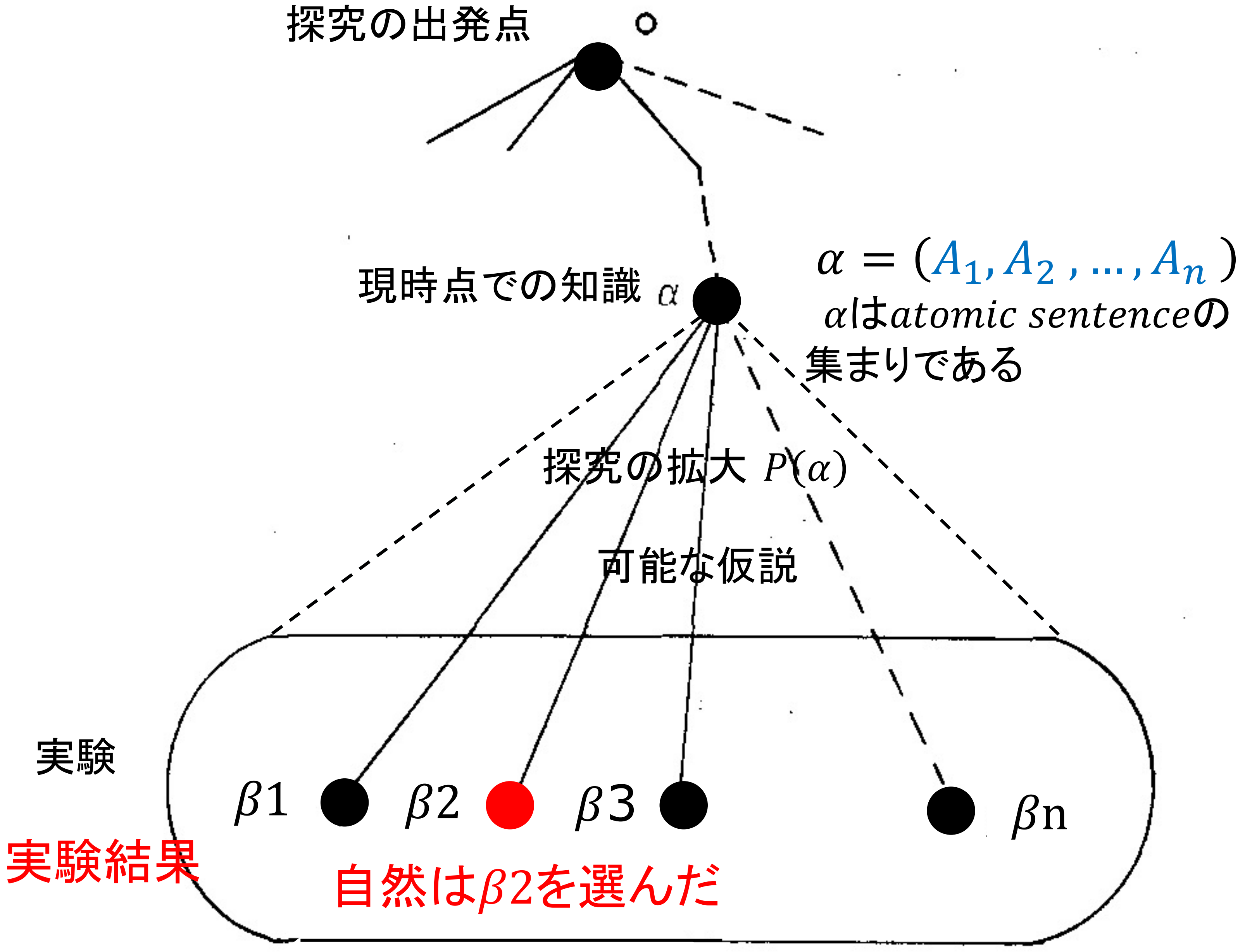


探究の出発点
現時点での知識 α
$\alpha = (A_1, A_2, \ldots, A_n)$
αは$atomic\ sentence$の集まりである
探究の拡大 $P(\alpha)$
可能な仮説
$\beta 1$
$\beta 2$
$\beta 3$
βn

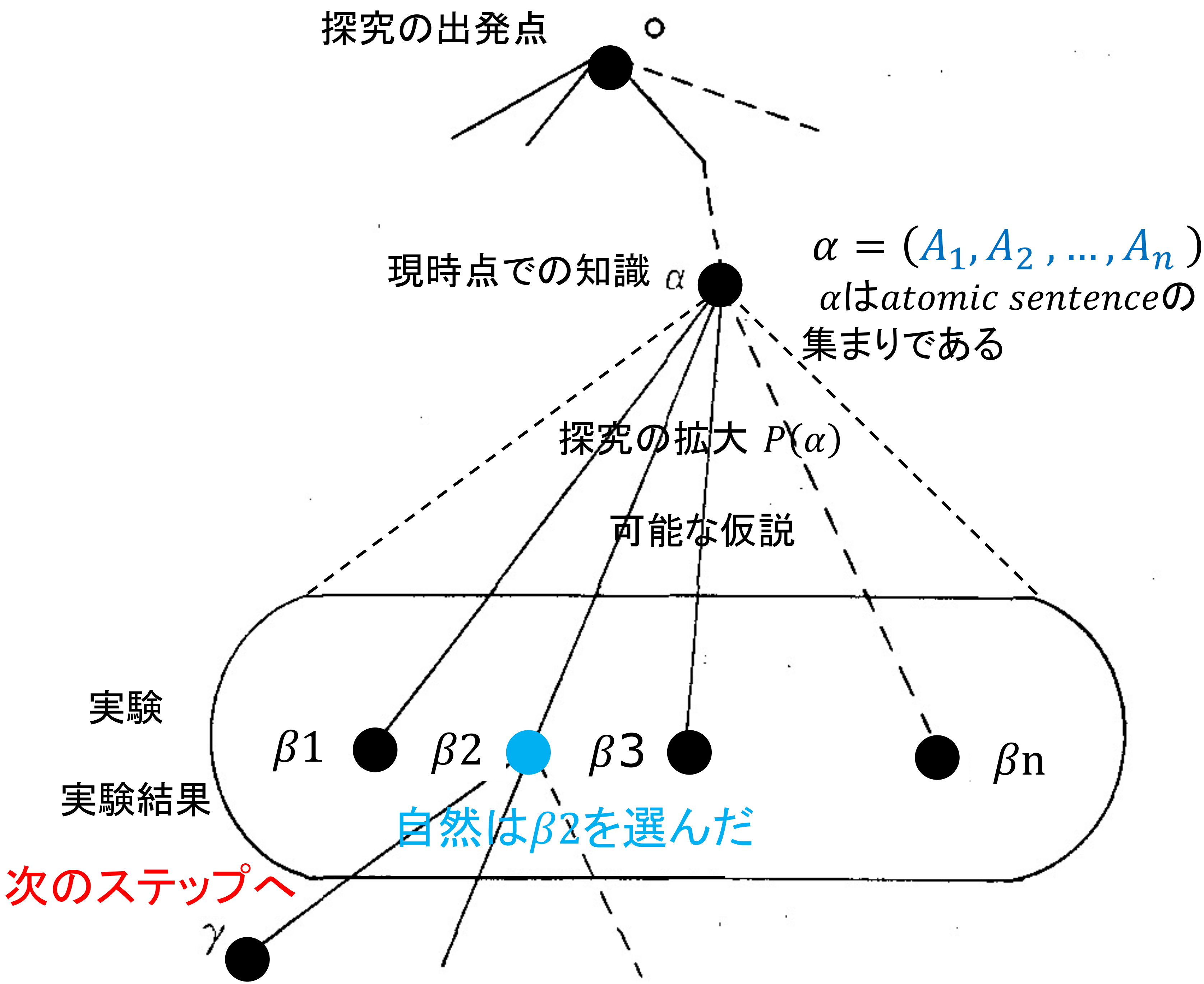

探究の出発点
現時点での知識 α
$\alpha = (A_1, A_2, \ldots, A_n)$
αは$atomic\ sentence$の
集まりである
探究の拡大 $P(\alpha)$
可能な仮説
$\beta 1$
$\beta 2$
$\beta 3$
βn
すべての$\beta_i \in P(\alpha)$について、
$\beta_i = (A_1, A_2, \ldots, A_n, A_{n+1}, \ldots, A_{n+k+1})$
βは、αの情報を全て含んでいる。


探究の出発点
現時点での知識 α
$\alpha = (A_1, A_2, \ldots, A_n)$
αはatomic sentenceの集まりである
累積性
探究の拡大 $P(\alpha)$
一度科学的情報としてえられた情報は，その情報がいかに拡大されても，なくなることはない
可能な仮説
$\beta 1$
$\beta 2$
$\beta 3$
βn
すべての$\beta_i \in P(\alpha)$について、
$\beta_i = (A_1, A_2, \ldots, A_n, A_{n+1}, \ldots, A_{n+k+1})$
βは、αの情報を全て含んでいる。


探究の出発点
現時点での知識 α
$\alpha = (A_1, A_2, \ldots, A_n)$
αは*atomic sentence*の集まりである
探究の拡大 $P(\alpha)$
可能な仮説
実験
$\beta 1$
$\beta 2$
$\beta 3$
βn
$\beta_i = (A_1, A_2, \ldots, A_n, A_{n+1}, \ldots, A_{n+k+1})$
βは、αの情報を全て含んでいる。



探究の出発点
現時点での知識 α
α = (A1, A2 , … , An)
αはatomic sentenceの
集まりである
探究の拡大 P(α)
可能な仮説
実験
β1
β2
β3
βn
実験結果
自然はβ2を選んだ



探究の出発点
現時点での知識 α
α = (A1, A2 , … , An)
αはatomic sentenceの
集まりである
探究の拡大 P(α)
可能な仮説
実験
実験結果
β1
β2
β3
βn
自然はβ2を選んだ
次のステップへ
γ

Grzegorczykの
「科学の探究」モデル 2
-- 情報の順序関係 --

情報と論理との関係

Grzegorczykは、科学的に得られた情報と科学的な論理との関係を次のように考えます。

ある情報がある論理式の成立を「強制」する

こうした考えは、実は、「連続体仮説の独立性」を証明したJ.P.Cohenが導入した独創的な「強制法」 Forcing Methodによるものです。

Grzegorczykは、CohenのForcing Method を「科学の探求の論理」として、解釈できることを示そうとしました。

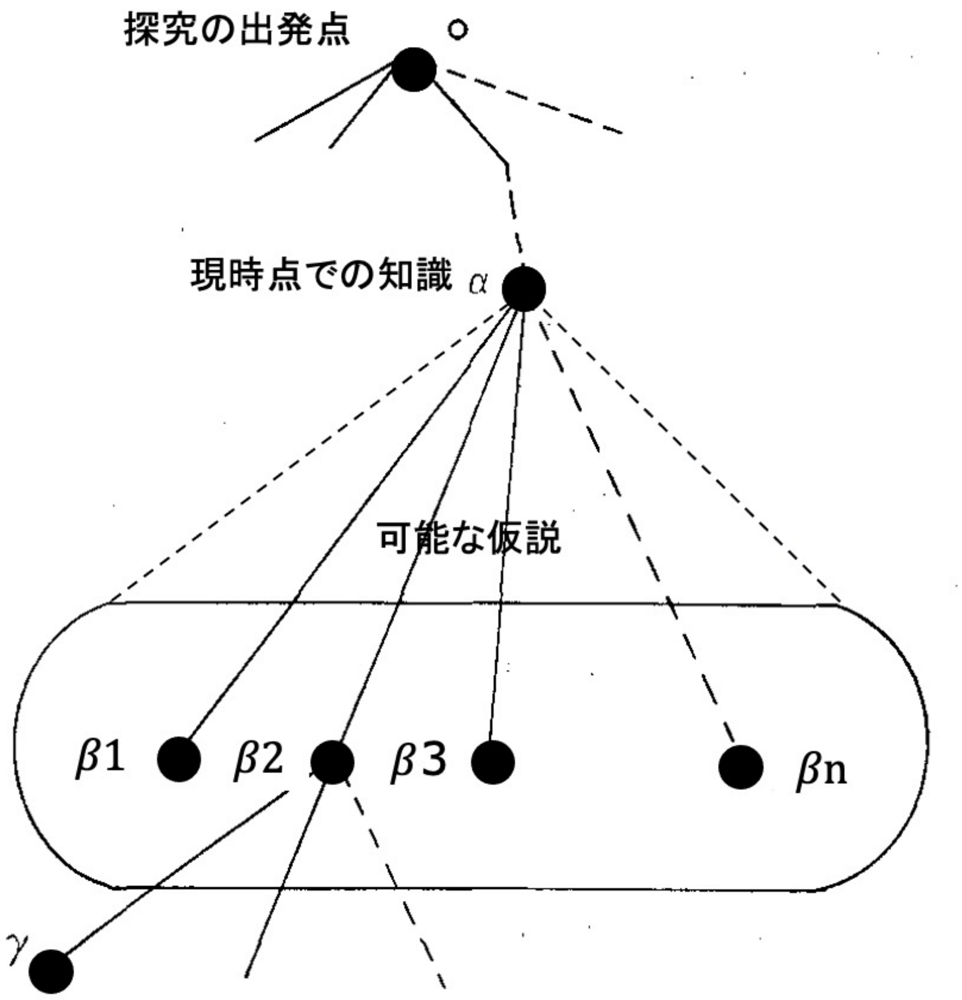
探究R上の情報 $\alpha, \beta, \ldots$ に順序$\succ_R$を定義する

そのためには、少し準備が必要です。
まず、科学の探究R上の情報 α, β,… に順序$\succ_R$を定義して（今回はここまでを扱います）、その上で、情報と論理式の「強制関係」を定義します。

科学の探究R上の情報 α, β,… は、探究が進むにつれ、ツリー上に枝分かれするかもしれません。探究R上の任意の情報 α, βについて、$\alpha \succ_R \beta$ あるいは$\beta \succ_R \alpha$ のどちらかが成り立つとは限りません。こうした順序関係を「半順序」といいます。

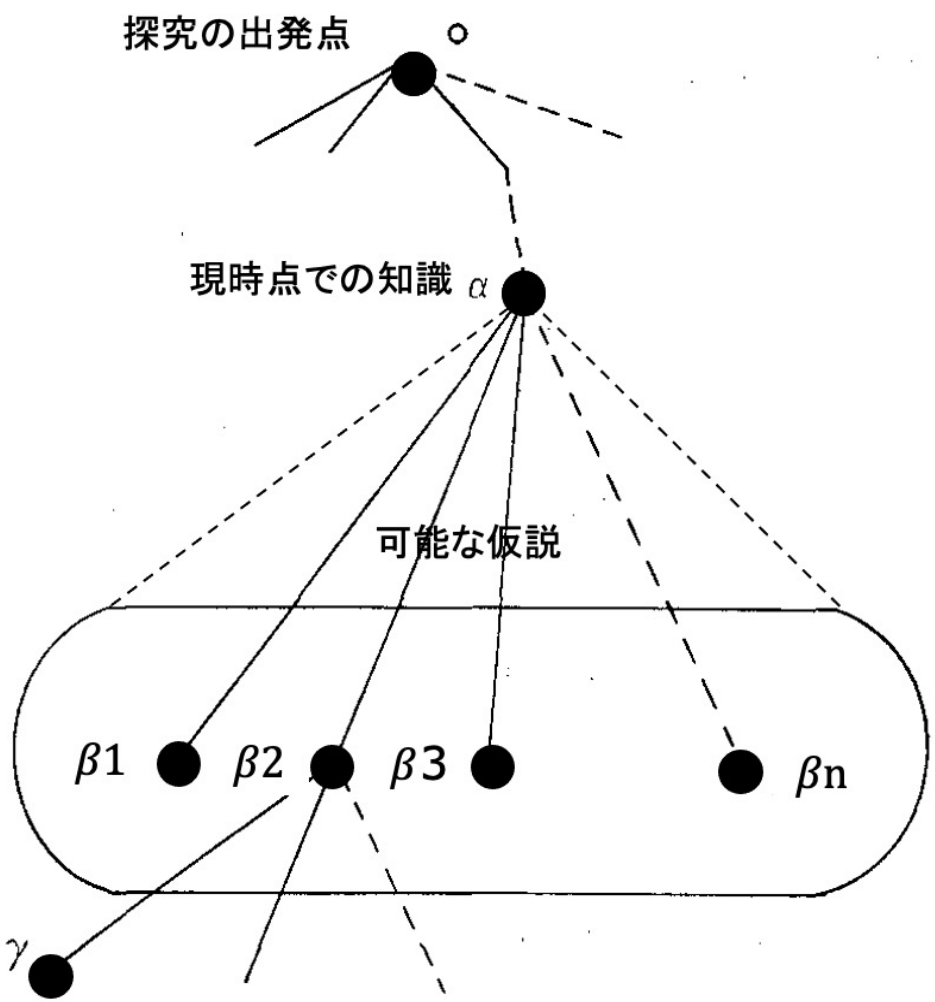
ただし、ある情報 α から出発した探究の枝の上の情報βは、$\beta \succ_R \alpha$ で、αの情報を全て含んでいます。

探究R上の情報 $\alpha, \beta, \ldots$ に順序$\succ_R$を定義する



$\beta \succ_0 \alpha \leftrightarrow \beta = \alpha$

探究R上の情報 $\alpha, \beta, \ldots$ に順序$\succ_R$を定義する



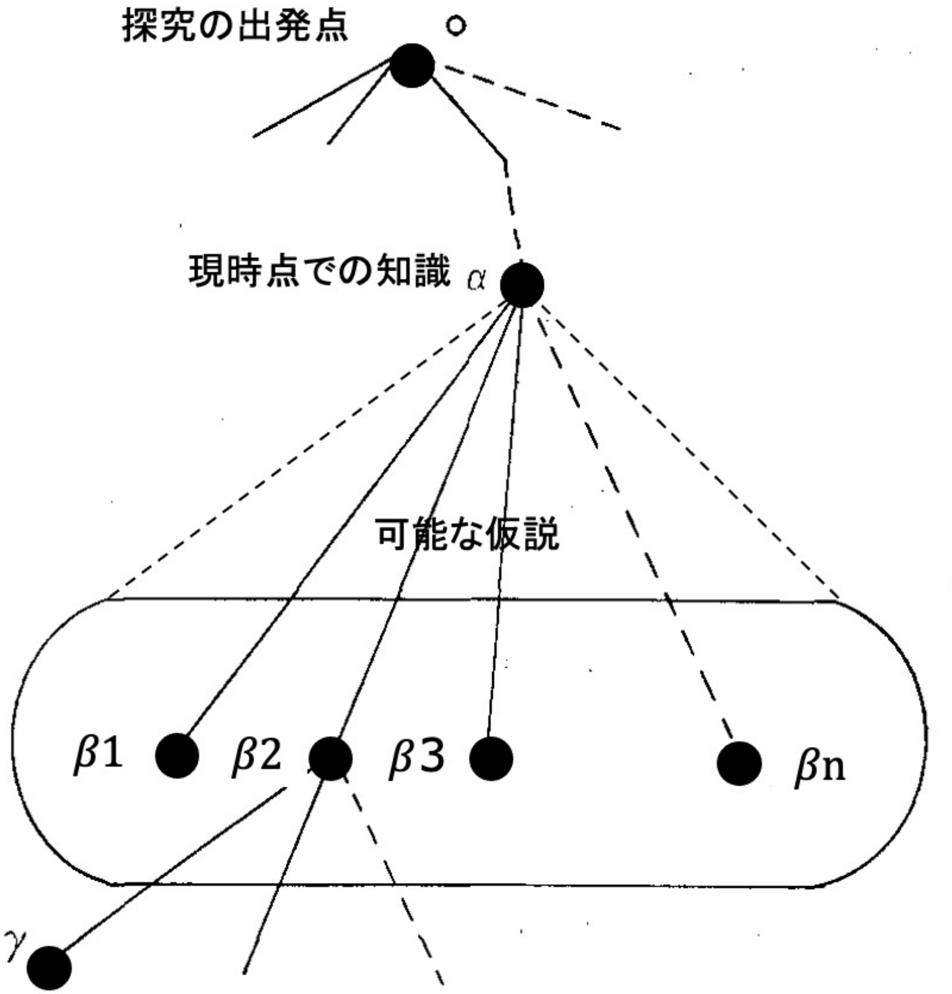
$$\beta \succ_0 \alpha \leftrightarrow \beta = \alpha$$

$$\alpha \succ_0 \alpha$$

$$\beta \succ_0 \beta$$

探究R上の情報 $\alpha, \beta, \ldots$ に順序$\succ_R$を定義する

$\alpha \succ_0 \alpha$ ：自分自身との距離は0

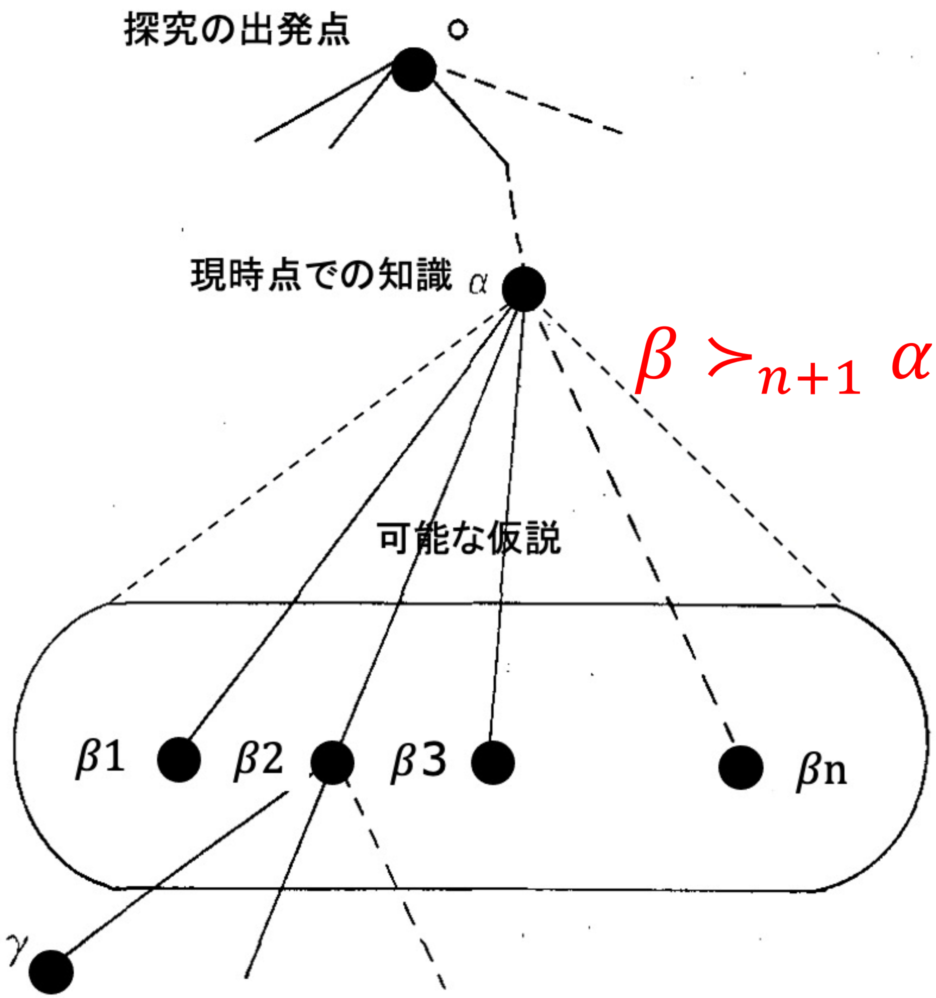


$\beta \succ_0 \alpha \leftrightarrow \beta = \alpha$

$\alpha \succ_0 \alpha$

$\beta \succ_0 \beta$

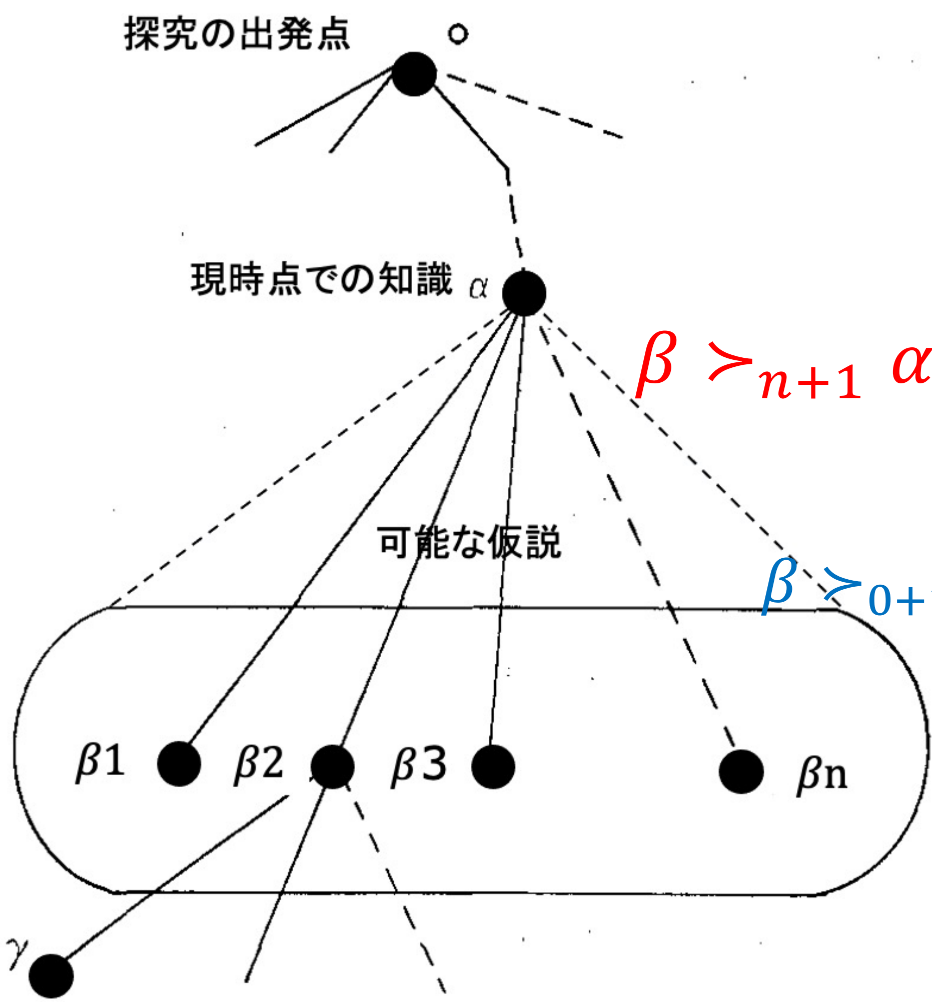
探究R上の情報 $\alpha, \beta, \ldots$ に順序$\succ_R$を定義する



$$\beta \succ_0 \alpha \leftrightarrow \beta = \alpha$$

$$\beta \succ_{n+1} \alpha \leftrightarrow \exists\gamma(\gamma \succ_n \alpha \wedge \beta \in P(\gamma))$$

探究R上の情報 $\alpha, \beta, \ldots$ に順序$\succ_R$を定義する



$$\beta \succ_0 \alpha \leftrightarrow \beta = \alpha$$

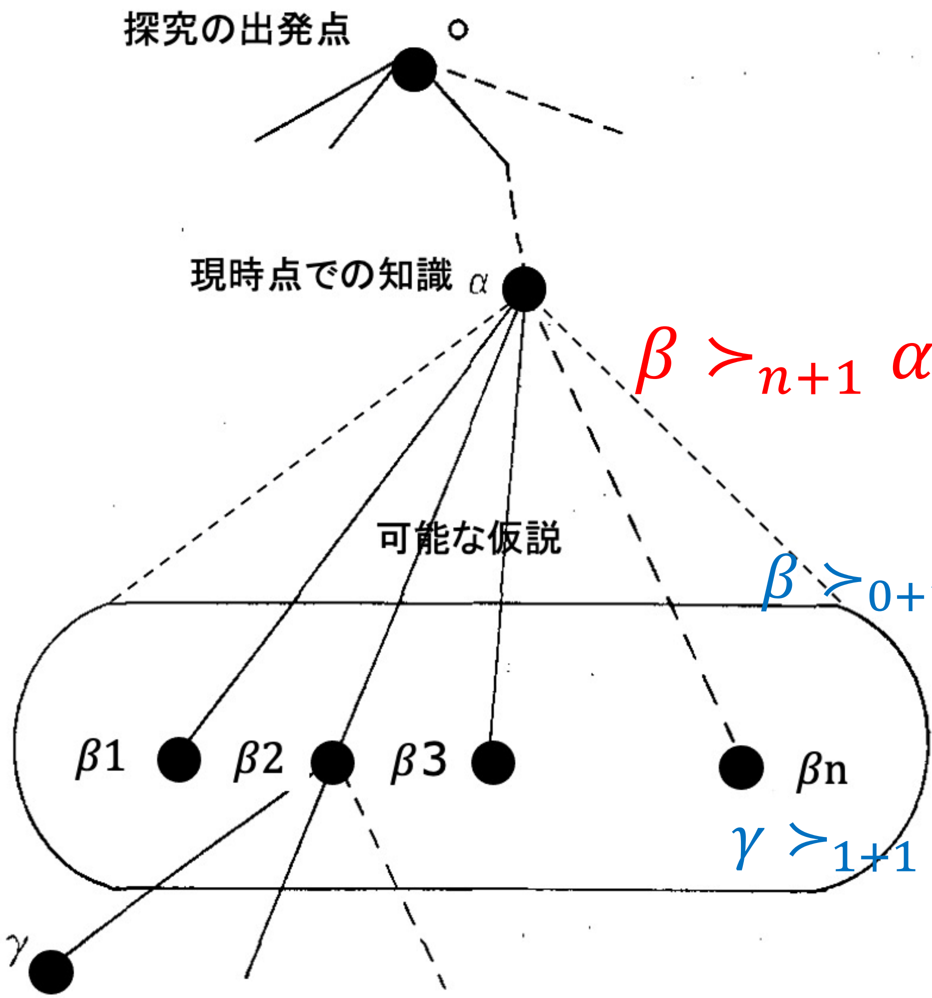
$$\beta \succ_{n+1} \alpha \leftrightarrow \exists\gamma(\gamma \succ_n \alpha \land \beta \in P(\gamma))$$

$$\alpha \succ_0 \alpha$$

$$\beta \succ_{0+1} \alpha \leftrightarrow \exists\alpha(\alpha \succ_0 \alpha \land \beta \in P(\alpha))$$

$$\beta_i \succ_1 \alpha$$

探究R上の情報 $\alpha, \beta, \ldots$ に順序$\succ_R$を定義する



$$\beta \succ_0 \alpha \leftrightarrow \beta = \alpha$$

$$\beta \succ_{n+1} \alpha \leftrightarrow \exists\gamma(\gamma \succ_n \alpha \land \beta \in P(\gamma))$$

$$\alpha \succ_0 \alpha$$

$$\beta \succ_{0+1} \alpha \leftrightarrow \exists\alpha(\alpha \succ_0 \alpha \land \beta \in P(\alpha))$$

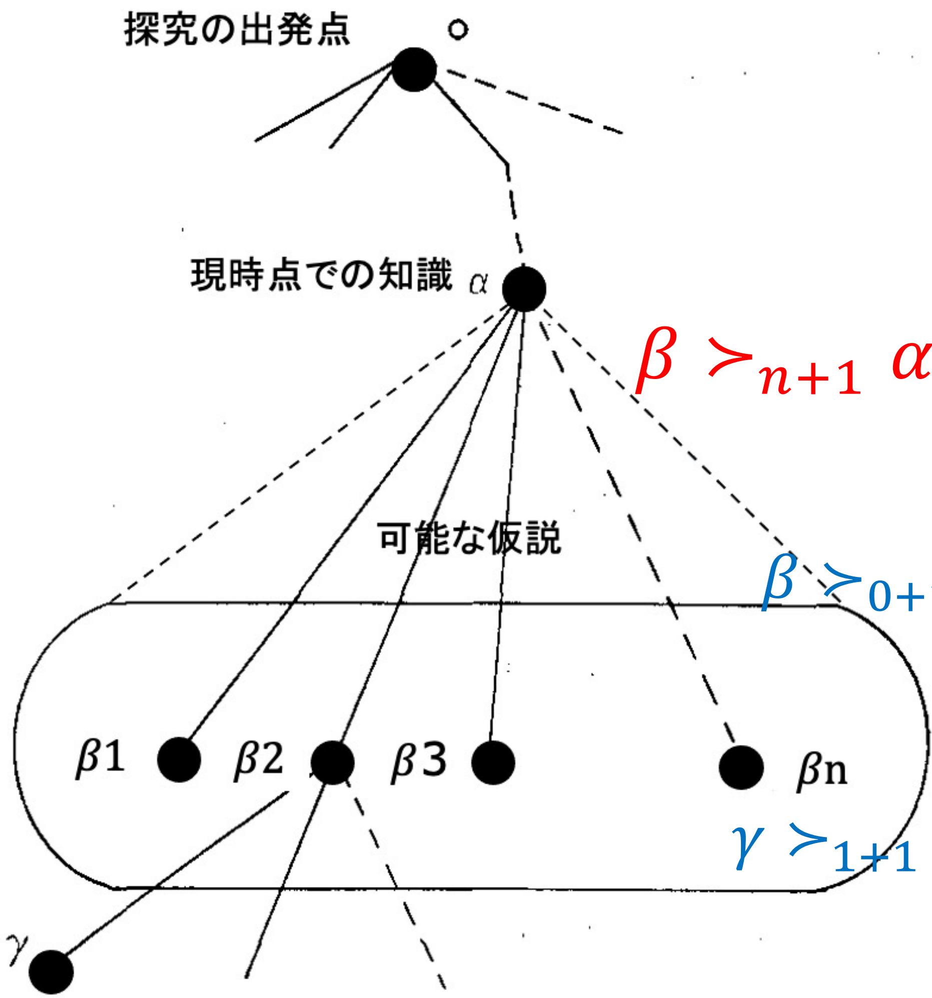
$$\beta_i \succ_1 \alpha$$

$$\gamma \succ_{1+1} \alpha \leftrightarrow \exists\beta_2(\beta_2 \succ_1 \alpha \land \gamma \in P(\beta_2))$$

$$\gamma \succ_2 \alpha$$

探究R上の情報 $\alpha, \beta, \ldots$ に順序$\succ_R$を定義する

$\beta \succ_n \alpha$:二点の経路上の距離は n



$$\beta \succ_0 \alpha \leftrightarrow \beta = \alpha$$

$$\beta \succ_{n+1} \alpha \leftrightarrow \exists\gamma(\gamma \succ_n \alpha \land \beta \in P(\gamma))$$

$$\alpha \succ_0 \alpha$$

$$\beta \succ_{0+1} \alpha \leftrightarrow \exists\alpha(\alpha \succ_0 \alpha \land \beta \in P(\alpha))$$

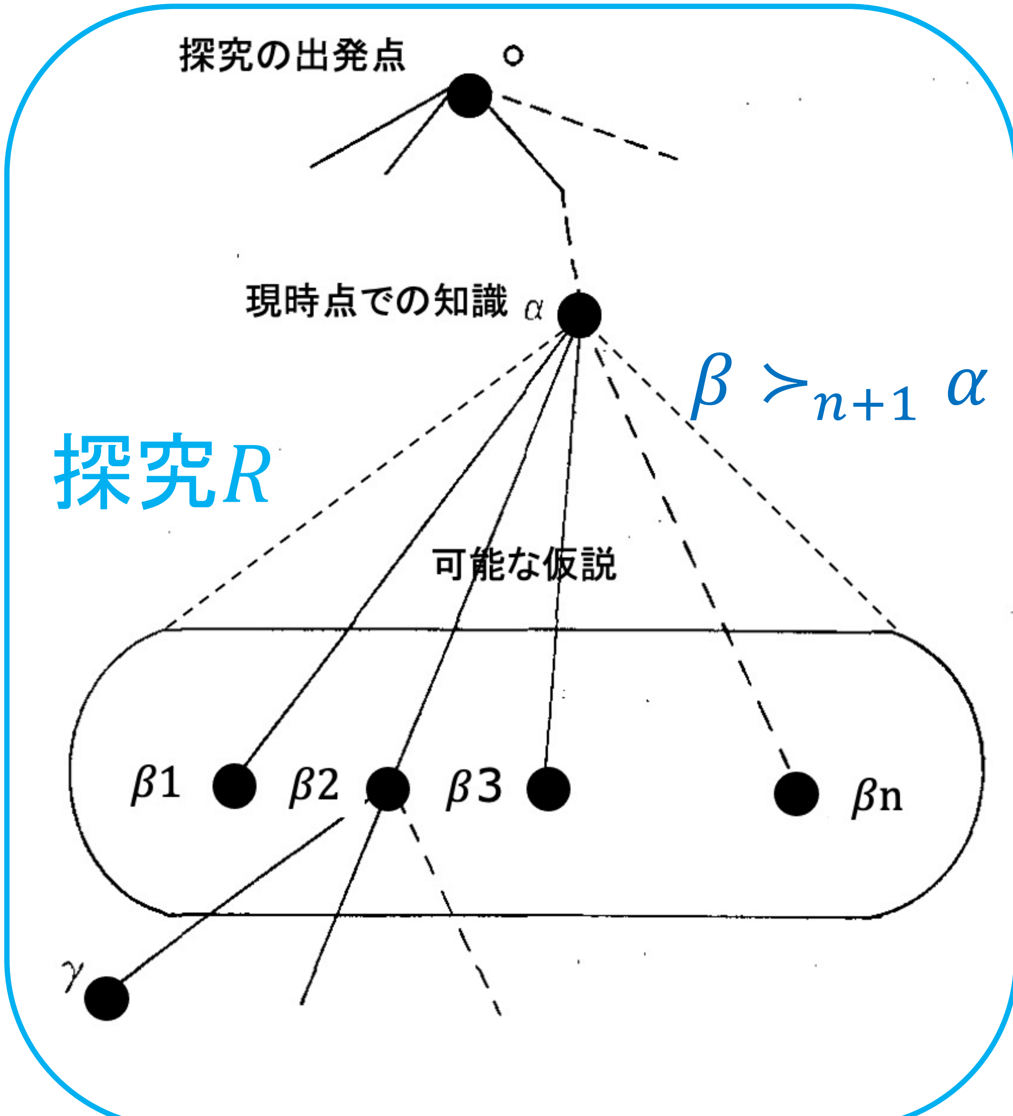
$$\beta_i \succ_1 \alpha$$

$$\gamma \succ_{1+1} \alpha \leftrightarrow \exists\beta_2(\beta_2 \succ_1 \alpha \land \gamma \in P(\beta_2))$$

$$\gamma \succ_2 \alpha$$

探究R上の情報 $\alpha, \beta, \ldots$ に順序$\succ_R$を定義する

$\beta \succ_R \alpha$:二点はRの経路上にある



$$\beta \succ_0 \alpha \leftrightarrow \beta = \alpha$$

$$\beta \succ_{n+1} \alpha \leftrightarrow \exists \gamma (\gamma \succ_n \alpha \wedge \beta \in P(\gamma))$$

$$\beta \succ_R \alpha \leftrightarrow \exists n (\beta \succ_n \alpha)$$

順序$\succ_R$の定義

$$\beta \succ_0 \alpha \leftrightarrow \beta = \alpha$$

自分自身との距離は0

$$\beta \succ_{n+1} \alpha \leftrightarrow \exists\gamma(\gamma \succ_n \alpha \land \beta \in P(\gamma))$$

二点の経路上の距離は n+1

$$\beta \succ_R \alpha \leftrightarrow \exists n(\beta \succ_n \alpha)$$

二点はRの経路上にある

Grzegorczykの「科学の探究」モデル 3 -- Forcing Method --

情報αが論理式ϕを成り立たせる「αはϕを強制する」

情報の集合の上に「半順序」を定義できたので、どのような情報が、論理式を成り立たせるのかを見ていきましょう。

情報の集合 α といっても、アトミックな論理式である命題で表現された情報＝命題の集合です。

ϕがアトミックな論理式の場合、情報 αが論理式ϕを成り立たせるのは、$\phi \in \alpha$ が成り立つ場合だけにしましょう。このことを、次のように表現します。

$$\alpha \rhd \phi \leftrightarrow \phi \in \alpha$$

左辺の$\alpha \rhd \phi$を、「αはϕを強制する」と読みます。

Forcing　Method

もっとも基本的なアトミックな論理式に、情報が論理式を成立させるという「強制」概念を定義できたので、今度は、一般の論理式に対して、「強制」概念を定義しましょう。

基本的には、論理記号で構成される論理式を、部分式に分解して、その部分式についての「強制」の定義を使っていけば、最終的には、先に見たアトミックな論理式の「強制」の定義に帰着します。

以下に見る、「強制」の定義は、基本的には、J.P.Cohenが導入した"Forcing Method"と、ほぼ同一のものです。

論理式が、∨あるいは∧を含む場合

論理式が、∨あるいは∧を含む場合、「強制」関係は次のように定義されます。

$$\alpha \rhd \lceil \phi \vee \psi \rceil \leftrightarrow (\alpha \rhd \phi \vee \alpha \rhd \psi)$$

αが$\lceil \phi \vee \psi \rceil$を強制するのは、$\alpha$が$\phi$を強制するか、あるいは、$\alpha$が$\psi$を強制する場合です。

$$\alpha \rhd \lceil \phi \wedge \psi \rceil \leftrightarrow (\alpha \rhd \phi \wedge \alpha \rhd \psi)$$

αが$\lceil \phi \wedge \psi \rceil$を強制するのは、$\alpha$が$\phi$を強制し、かつ、$\alpha$が$\psi$を強制する場合です。

論理式が、~あるいは→ を含む場合

これらの場合は、少し注意が必要です。

$$\alpha \rhd [\sim\phi] \leftrightarrow \forall\beta\left(\beta \succ_R \alpha \rightarrow \sim(\beta \rhd \phi)\right)$$

αが$[\sim\phi]$を強制するのは、αを拡大した全てのβが、ϕを強制しない場合です。

$$\alpha \rhd [\phi \rightarrow \psi] \leftrightarrow \forall\beta(\beta \succ_R \alpha \rightarrow (\beta \rhd \phi \rightarrow \beta \rhd \psi))$$

αが$[\phi \rightarrow \psi]$を強制するのは、αを拡大した全てのβ
において、β がϕを強制すれば、βがψを強制する場合です。

論理式が、量化記号を含む場合

$$\alpha \rhd [\,\exists x\phi(x)\,] \leftrightarrow \alpha \rhd \phi(a_i)\text{なる}a_i\text{が存在する}$$

αが$[\exists x\phi(x)\,]$を強制するのは、αが$\phi(a_i)$を強制するようなa_iが存在する場合です。

$$\alpha \rhd [\,\forall x\phi(x)\,] \leftrightarrow \text{全ての}a_i\text{について}\alpha \rhd \phi(a_i)$$

αが$[\forall x\phi(x)\,]$を強制するのは、全てのa_iについてαが$\phi(a_i)$を強制する場合です。

これらの量化記号の解釈は、Cohenのとは少し違っています。

科学の探求の論理は、直観主義論理

通常の古典論理では 排中律 $A \vee \sim A$ が成り立つのですが、Grzegorczykの論理ではそうなりません。

なぜなら、情報αが与えられている時点で$\alpha \rhd (A \vee \sim A)$ がなりたつことを示すためには、$\alpha \rhd A$あるいは$\alpha \rhd \sim A$を示さなければなりません。

ただ、先の否定の解釈によれば、$\alpha \rhd \sim A$ が成り立つことを示すためにはαを拡大した全ての情報βにおいて、$\sim A$ が成り立つことを示さなければなりません。それは、自明ではありません。

Grzegorczykの構成によれば、科学の探求の論理は、直観主義論理に従うということになります。

情報が論理式を強制する
強制関係の定義

atomicな論理式の場合

$$\alpha \rhd \phi \leftrightarrow \phi \in \alpha$$

atomicではない論理式の場合

$$\alpha \rhd \ulcorner \phi \lor \psi \urcorner \leftrightarrow (\alpha \rhd \phi \lor \alpha \rhd \psi)$$
$$\alpha \rhd \ulcorner \phi \land \psi \urcorner \leftrightarrow (\alpha \rhd \phi \land \alpha \rhd \psi)$$

$$\alpha \rhd \ulcorner \sim\phi \urcorner \leftrightarrow \forall\beta\big(\beta \succ_R \alpha \rightarrow \sim(\beta \rhd \phi)\big)$$
$$\alpha \rhd \ulcorner \phi \rightarrow \psi \urcorner \leftrightarrow \forall\beta(\beta \succ_R \alpha \rightarrow (\beta \rhd \phi \rightarrow \beta \rhd \psi))$$

$$\alpha \rhd \ulcorner \exists x\phi(x) \urcorner \leftrightarrow \alpha \rhd \phi(a_i)\text{なる}a_i\text{が存在する}$$
$$\alpha \rhd \ulcorner \forall x\phi(x) \urcorner \leftrightarrow \text{全ての}a_i\text{について}\alpha \rhd \phi(a_i)$$

Part II

Kripkeの「可能的世界」

Part II

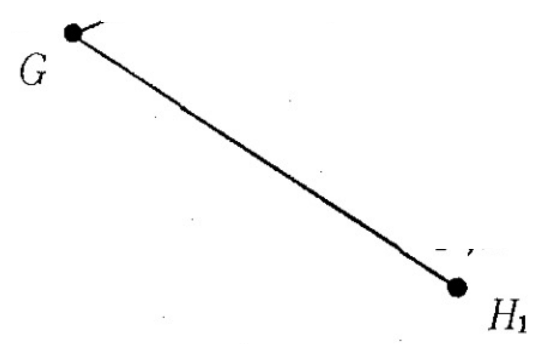
Kripkeの「可能的世界」1

1. Kripkeの「可能的世界」
2. 「可能的世界」と情報
3. 「可能的世界」へのジャンプ
4. 認識の特徴を表現するモデル


Kripkeの「可能的世界」1

Kripkeの「可能的世界」

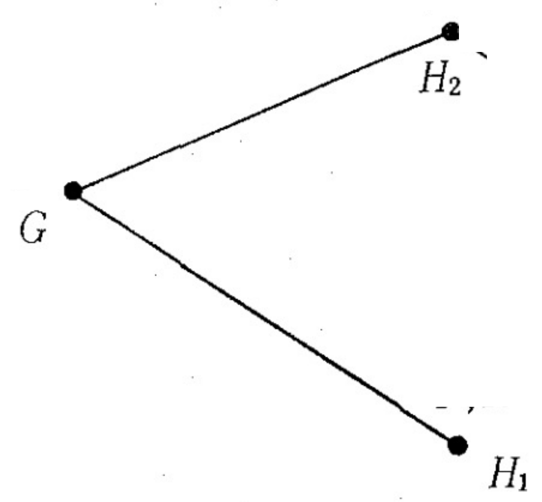
Kripkeは、ある世界から可能な世界を考えます。
下の図は、ある世界Gから世界H_1が「可能」であることを表しています。



Kripkeの「可能的世界」

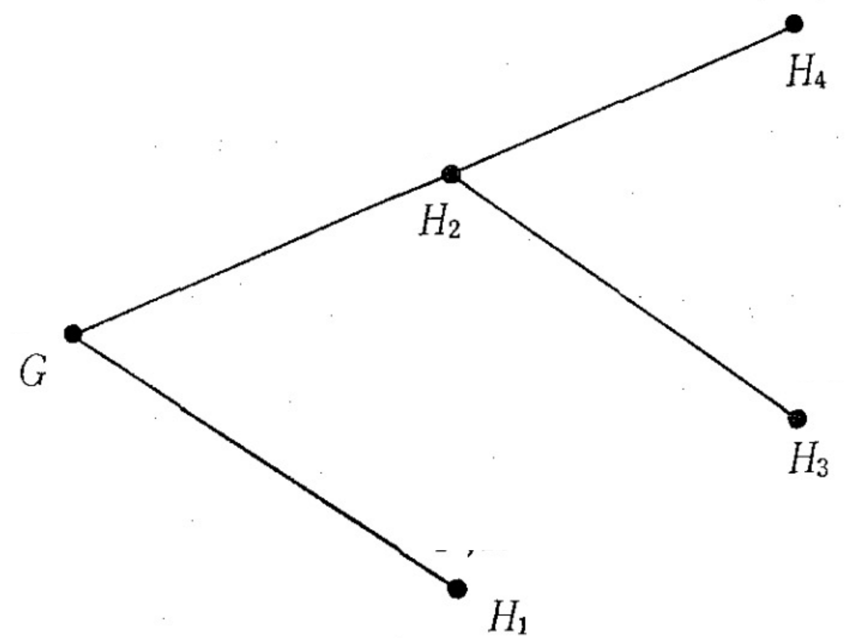
ただ、世界Gから「可能」な、もう一つの世界H_2があるかもしれません。

下の図は、ある世界Gから世界H_1と世界H_2が「可能」であることを表しています



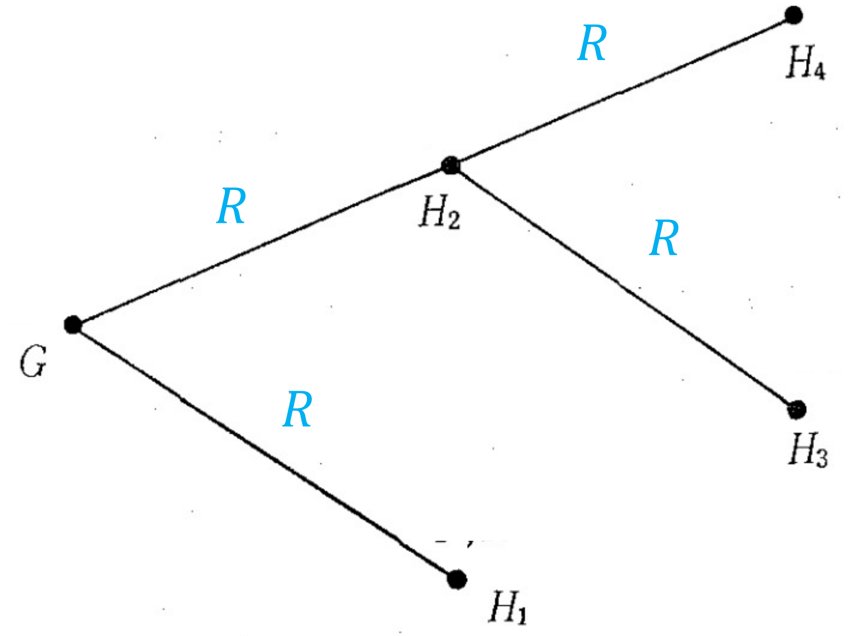
Kripkeの「可能的世界」

同様に、世界H_2から世界H_3と世界H_4が可能かもしれません。下の図は、ある世界Gから世界H_1と世界H_2が「可能」であり、世界H_2から世界H_3と世界H_4が「可能」であることを表しています。



「可能的世界」を結ぶ関係

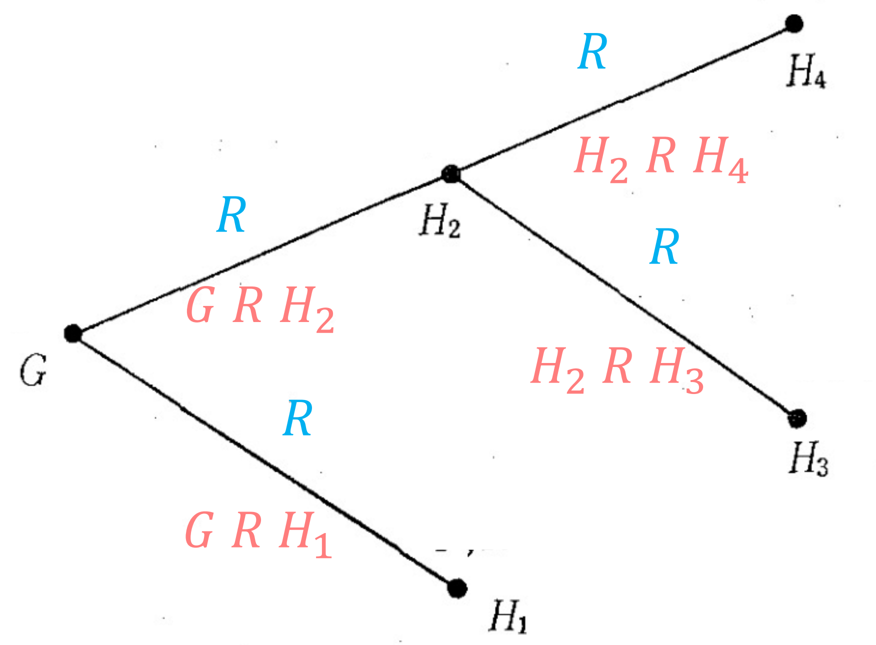
Kripkeは、ある世界からみてある世界が「可能である」時、そのある世界とある世界は、ある関係で結ばれていると考えます。この関係をRで表しましょう。



「可能的世界」を結ぶ関係

Kripkeは、ある世界からみてある世界が「可能である」時、そのある世界とある世界は、ある関係で結ばれていると考えます。この関係をRで表しましょう。

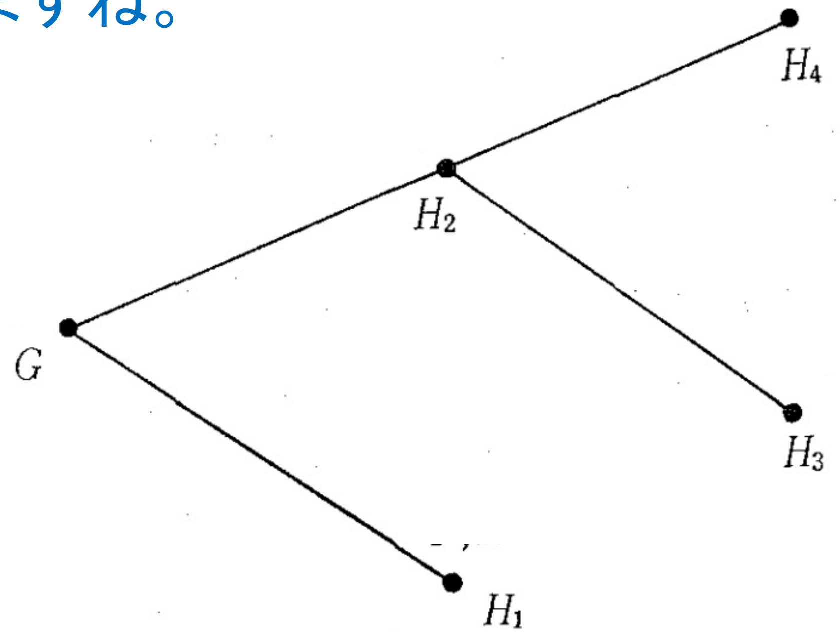
具体的には、この図では次のような関係が成り立っています。



「可能的世界」を結ぶ関係

Kripkeは、ある世界からみてある世界が「可能である」時、そのある世界とある世界は、ある関係で結ばれていると考えます。この関係をRで表しましょう。

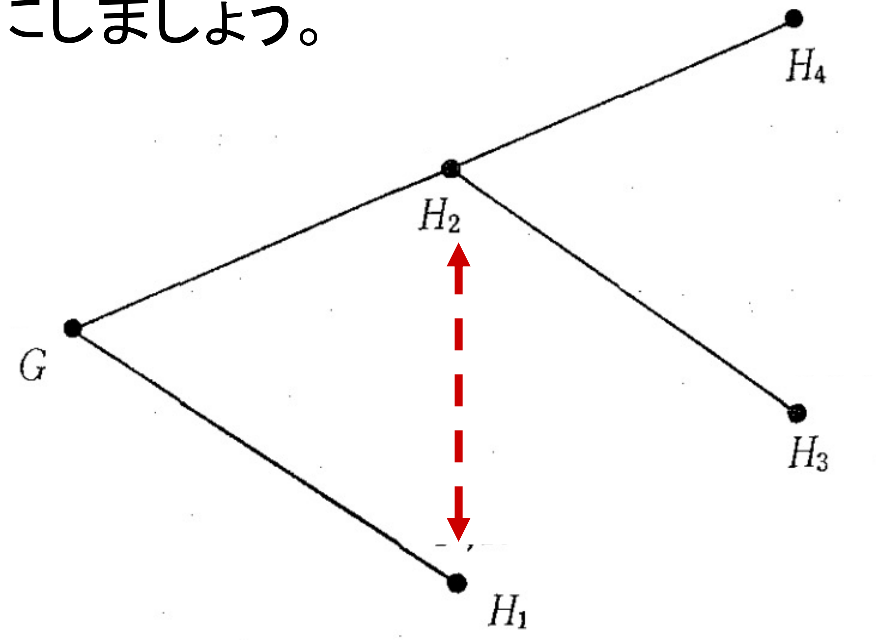
もっとも、こういう関係があることは、図に書き込まなくとも図を見れば分かりますね。



「可能的世界」を結ぶ関係

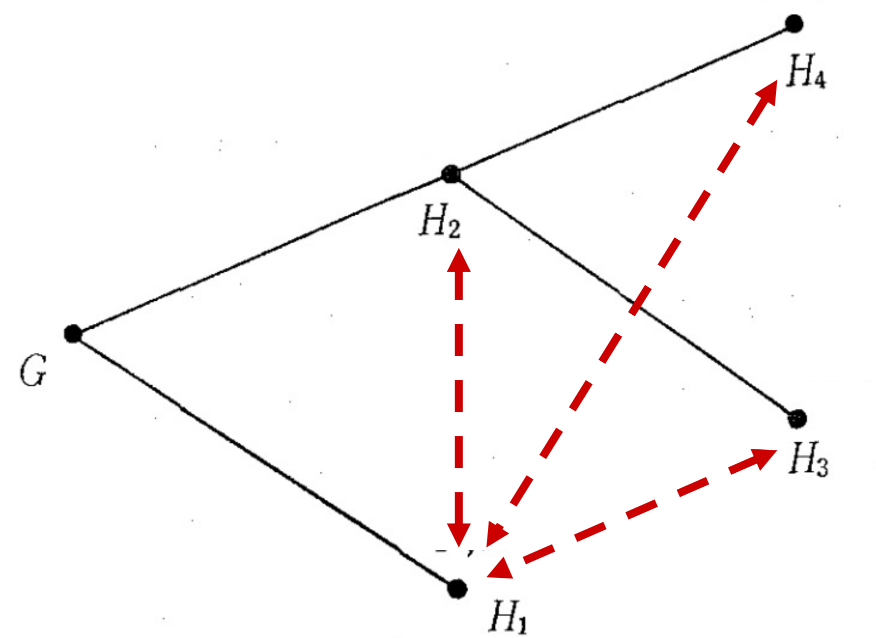
それでは、世界H_1と世界H_2の関係を考えてみましょう。

先の図では、世界H_1と世界H_2のあいだは線で結ばれていませんでした。このことは、世界H_1と世界H_2のあいだの可能的関係があるかどうかは、「分からない」ことを意味するというように考えることにしましょう。



「可能的世界」を結ぶ関係

H_1から見れば、H_2だけでなくH_3と世界H_4との関係も、可能的関係があるかは、「分からない」ということになります。



「可能的世界」を結ぶ関係Rの性質

ただ、図の上で二つの世界が線で結ばれていないということは、これらの世界が無関係であることを意味するわけではありません。それは関係があるかどうか「分からない」というだけです。

Kripkeは、この関係Rの性質によって、ある図形で表現される複数の世界のあいだの「可能的世界」という関係が変化することに気付きます。

Kripkeは、関係Rの性質に注目することによって、複数の様相論理の体系の関係を、見事に整理しました。残念ながら、彼のこの仕事を、今回は紹介できません。

「推移的」な関係

関係Rが、次の性質を持つ時、関係Rは「推移的」といいます。

$A\ R\ B$ で B R Cの時、A R C が成り立つ

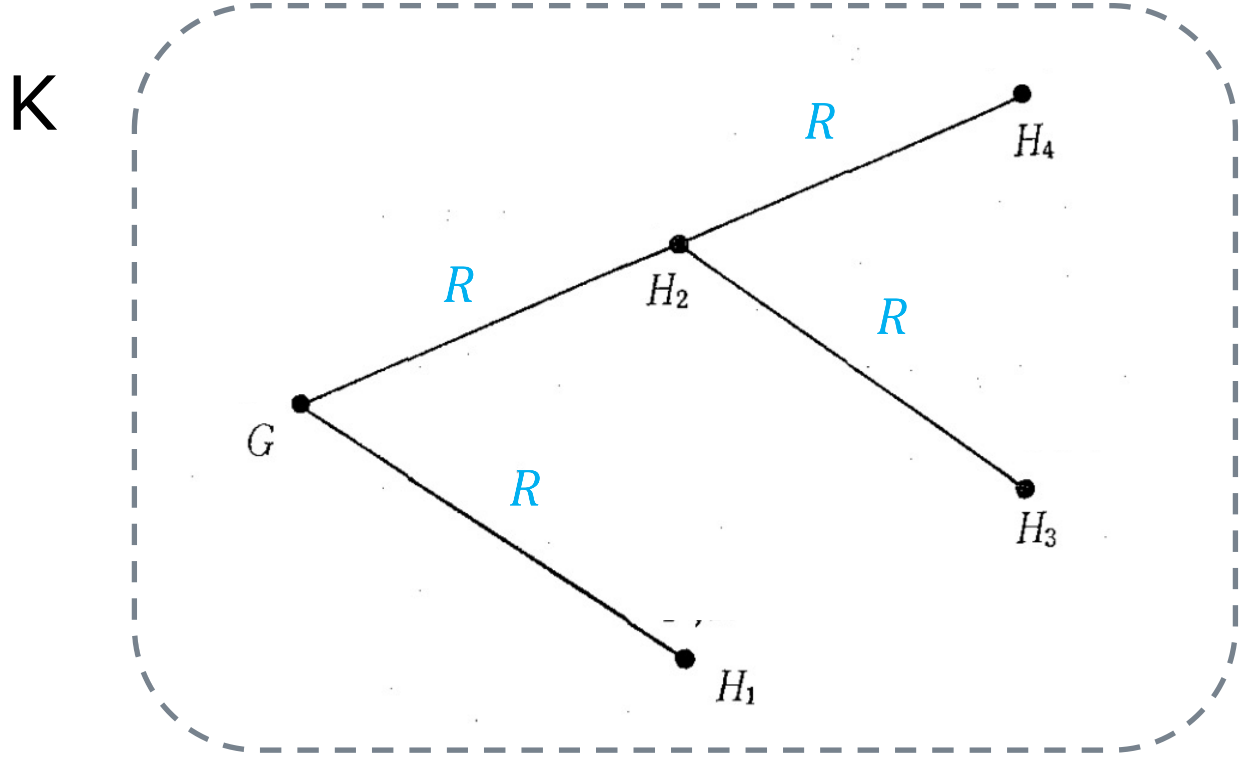
例えば、数の間の大小関係 < は、

$A < B$で B<C の時、$A < C$ が成り立つ

ので、推移的な関係です。

「可能的世界」を結ぶ関係Rが推移的とする

可能的な世界の関係Rを表した図形全体をKとしましょう。
今、Kの中の世界の関係Rが推移的だとします。



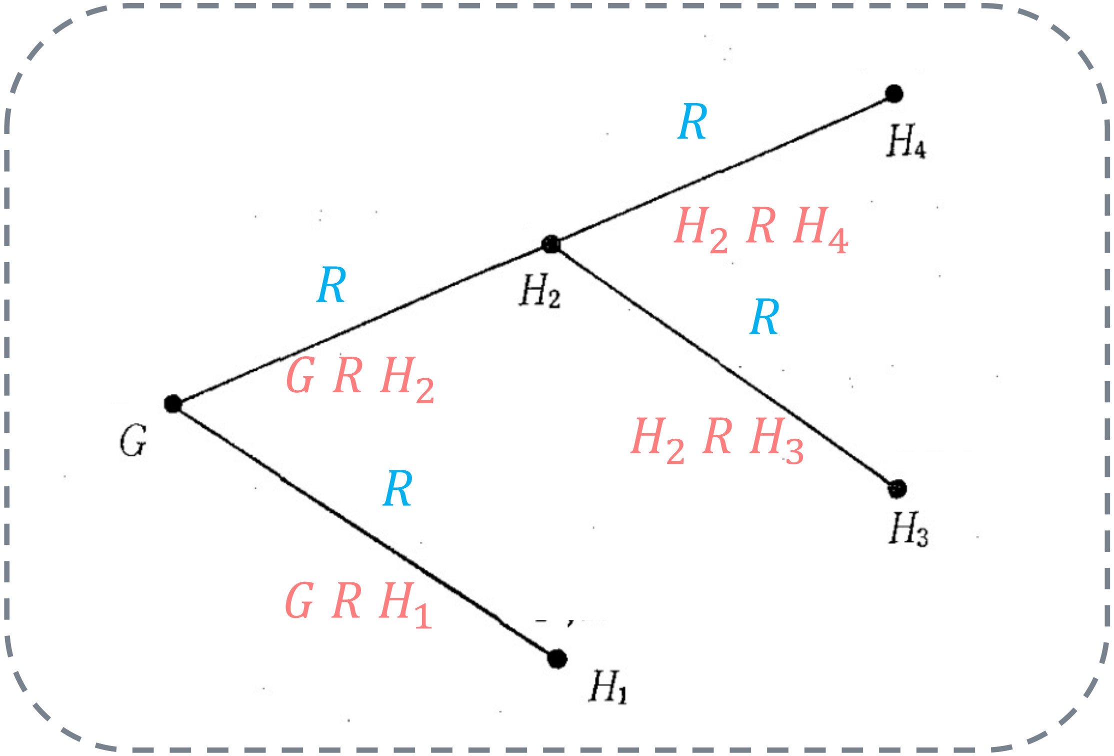
「可能的世界」を結ぶ関係Rが推移的とする

この時、次のことがRの推移性から分かります。

$G\ R\ H_2$ と $H_2\ R\ H_4$ から $G\ R\ H_4$が分かり、

$G\ R\ H_2$ と $H_2\ R\ H_3$ から $G\ R\ H_3$が分かります。

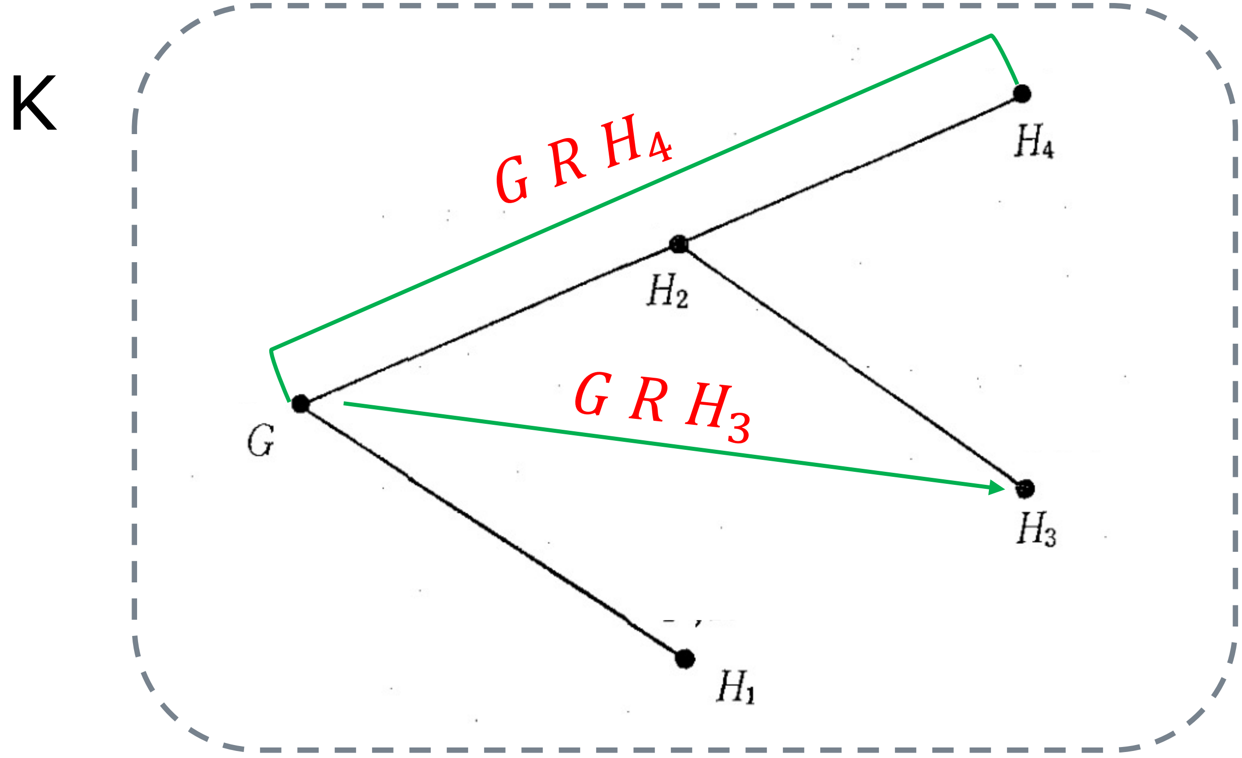
K



「可能的世界」を結ぶ
関係Rが推移的とする

要するにKの中では、図中で明示的に示された関係以外に、$G\ R\ H_4$ と $G\ R\ H_3$ という関係があることが分かります。

Rが推移的なら、H_3もH_4もGの可能的世界になります。

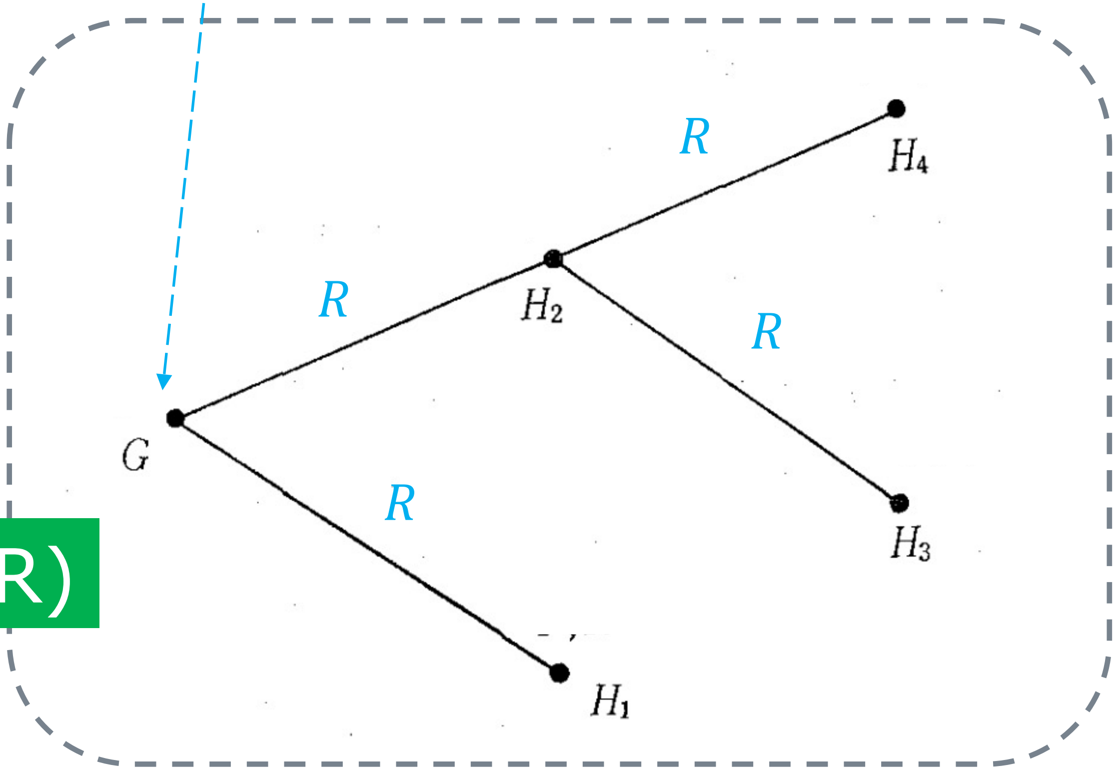



Kripkeの「可能的世界」２
-- 可能的世界と情報 --

「可能的世界」全体を考える

可能的な世界の関係Rを表したツリー構造をKとします。
この構造全体を (G, K, R) で表しましょう。
Gは、この構造の 「始点」にあたります。

K



(G, K, R)

それぞれの世界を特徴づけるものは、それぞれの世界の「情報」である

ノードGは、特別な「世界」です。なぜなら、そこは架空の世界ではなく、我々が生きる現実の世界だからです。この現実の世界で、我々は正しいと信じる情報を持っています。

「可能的世界」のそれぞれの世界は、時間の流れの中で、我々がそこでさまざまな情報を受け取る世界を表現していると考えることができます。「可能的世界」の枠組みは、基本的には、我々の認識は発展することを表現しています。

また、「可能的世界」のそれぞれの世界を、区別し特徴づけるものは、それぞれの世界が持つ「情報」なのです。

「ある命題Aを証明する十分な情報」を考える

ある世界を特徴付ける「情報」の表現として、「世界Xで、ある論理式Aを証明する十分な情報が存在する」ということを、次の式で表現してみましょう。

$$\phi(A, X) = T$$

もしも、こうした情報が欠けている時、

$$\phi(A, X) = F$$

としましょう。この式は、「世界Xで、論理式Aが偽である」ことを主張しているわけではありません。

「世界」が持つ情報を、単純な例で考える

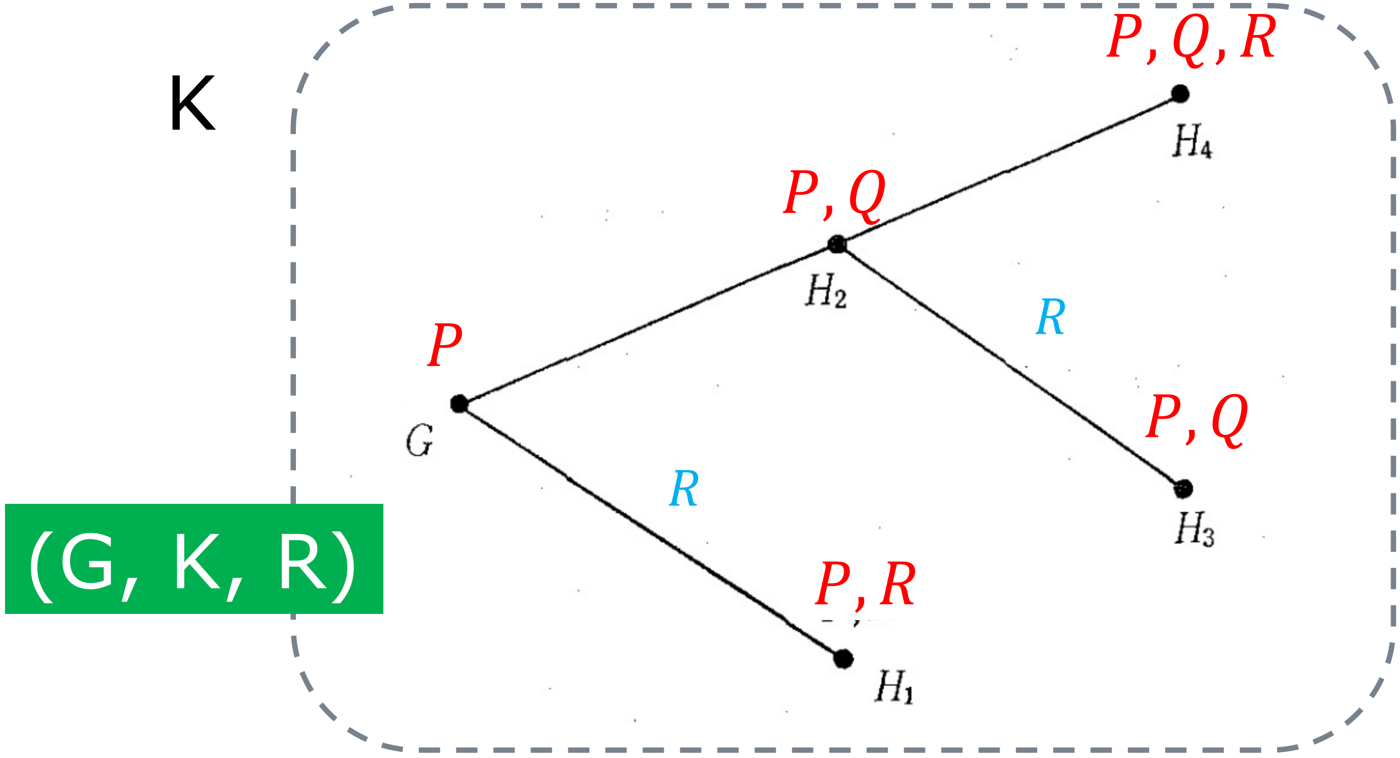
ϕの役割を明確にするために、状況を次のように、少し単純化して説明してみたいと思います。ある論理式Aが、三つのatomic formula P, Q, R から構成されているとします。

$\phi(P,G)=T$　なら、命題Pを証明する十分な情報が、世界Gに存在するということですので、Pは世界Gが持つ情報と考えることができます。この時、世界Gの上に、Pを書き加えます。

$\phi(Q,G)=F$　なら、命題Qを証明する十分な情報が、世界Gに存在しないということですので、Qは世界Gが持つ情報とみなすことはできません。この時、世界Gの上に、Qを書き加えることはしません。

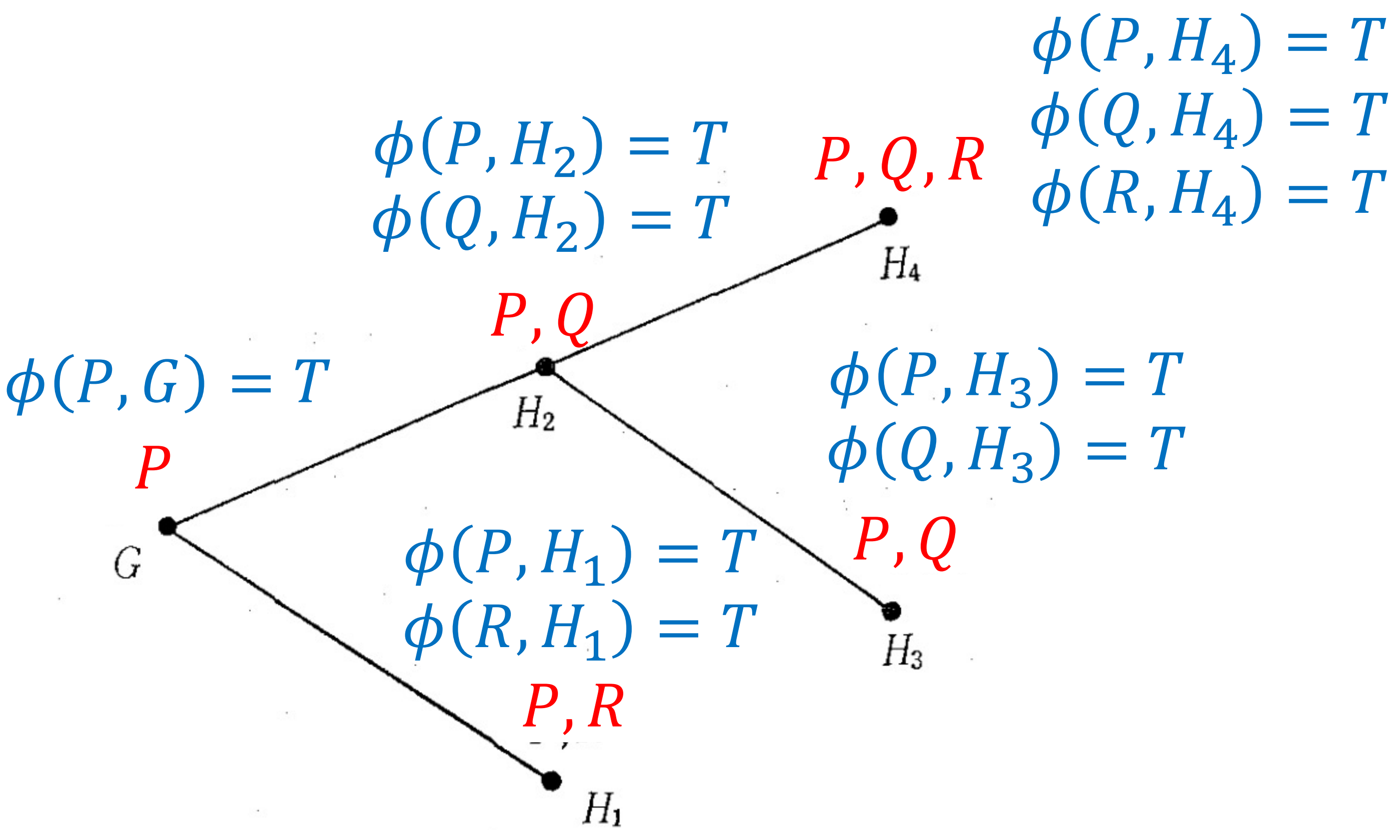
一つの単純な例として「世界」に「情報」を書き加えてみる

可能的な世界の構造を表した (G, K, R) に、それぞれの世界を持つ情報を、一つの例として、書き加えてみました。



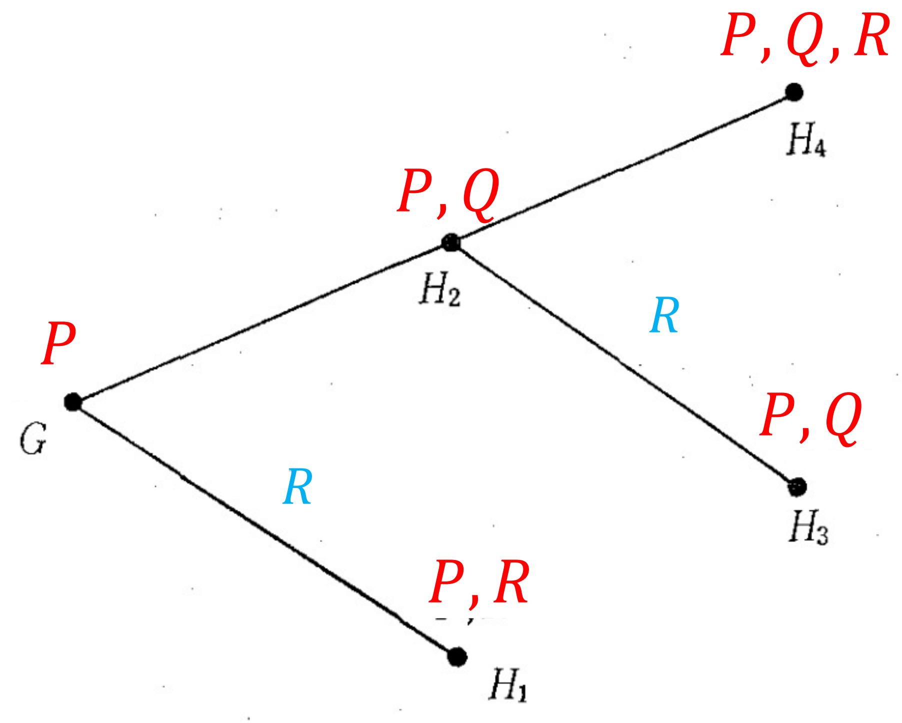
この例で成り立っている ϕについての式

この世界で成り立っている、ϕについての式も追加しました。



この例の場合の「世界」間の論理的関係？

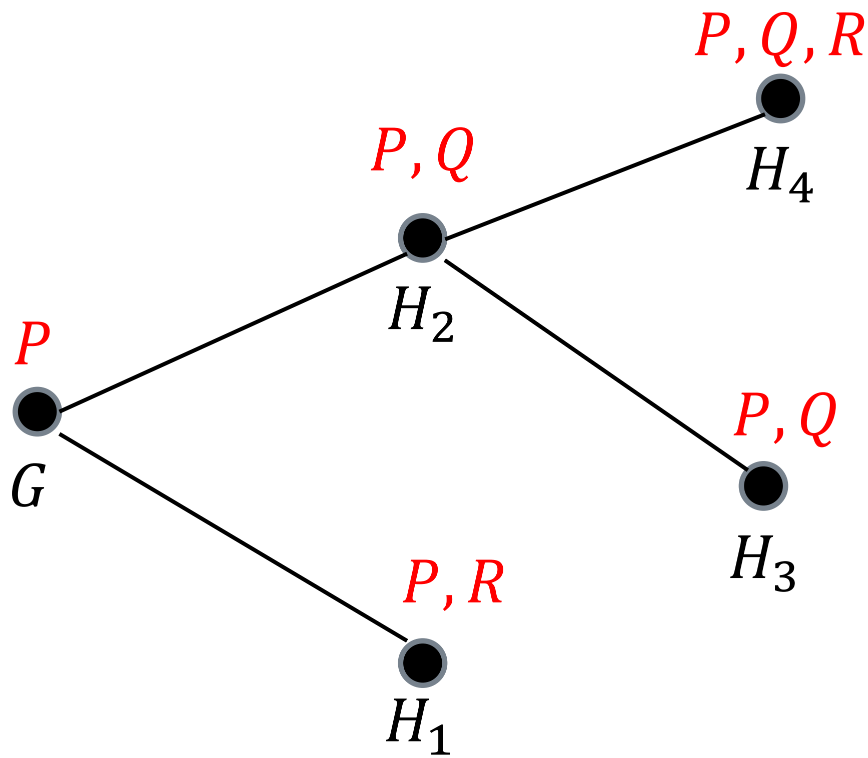
この例の場合の「世界」間の関係は、どのようなものになるのでしょう？それについては、次回、少し丁寧に説明しようと思います。




Kripkeの「可能的世界」３
「可能的世界」へのジャンプ

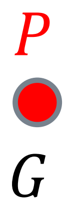
ある世界から別の世界へジャンプする

ひとつの例として、次の図のように、世界間の関係と、それぞれの世界が持つ情報が与えられているとします。この時、世界から世界へ、我々がジャンプした時、何が起きるかをみてみましょう。



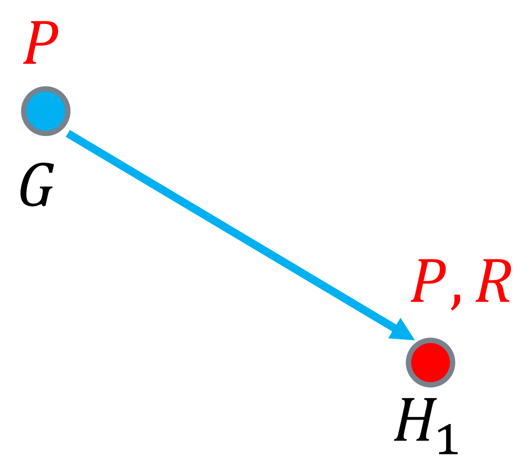
同じ世界Gにとどまり続ける

G ー我々の現在の世界ーでは我々はPを情報として持っています。我々がそれに満足するなら、何も別の世界にジャンプする必要はありません。ジャンプをしないという選択もあり得ます。我々の知るすべてのことによっても，何も新しい情報が我々にもたらされないなら，我々は，いくらでも長いあいだじっとGにとどまるかもしれません。



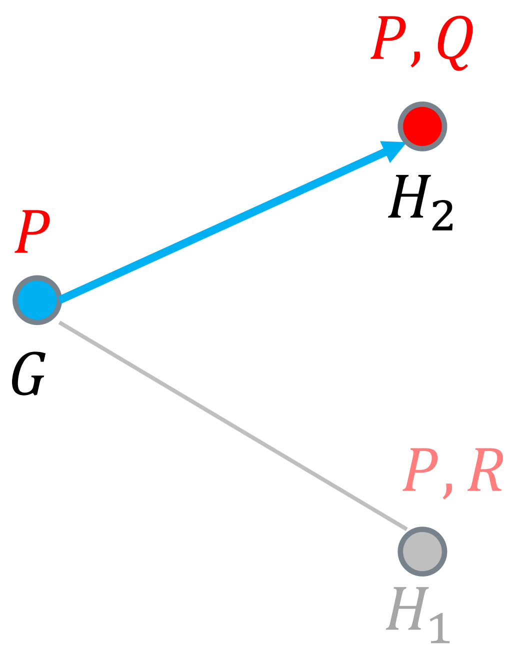
世界Gから世界 H_1 にジャンプする

我々が、世界Gから世界 H_1 にジャンプしたとします。世界 H_1 では，Pに加えてRの証明 – Rの情報を我々は得ます。



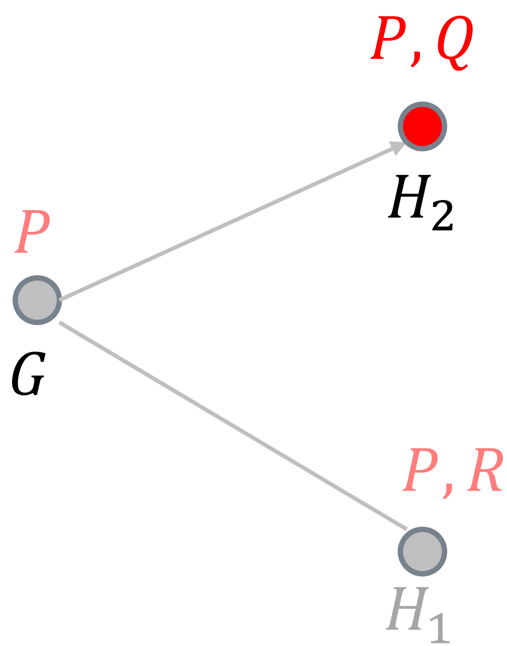
世界Gから世界 H_2にジャンプする

世界Gからジャンプ可能な世界は、H_1だけではありません。我々が、世界Gから世界 H_2にジャンプしたとします。この時、世界 H_2 では, Pに加えてQの証明 – Qの情報を我々は得ます。



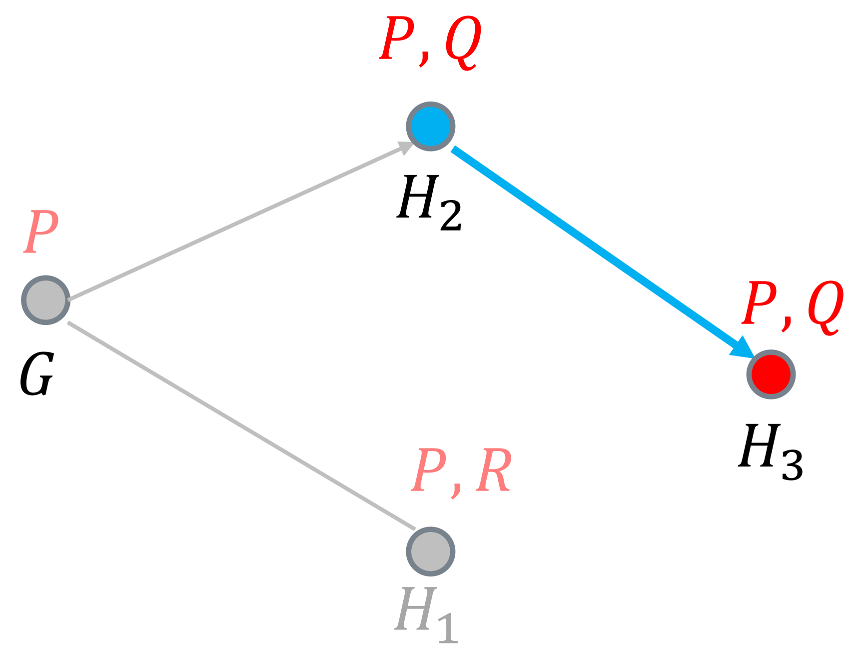
世界 H_2にとどまり続ける

我々が、世界 H_2 にジャンプして、PとQ両方の情報を得るのですが、その後、世界 H_2 とどまり続けるかもしれません。



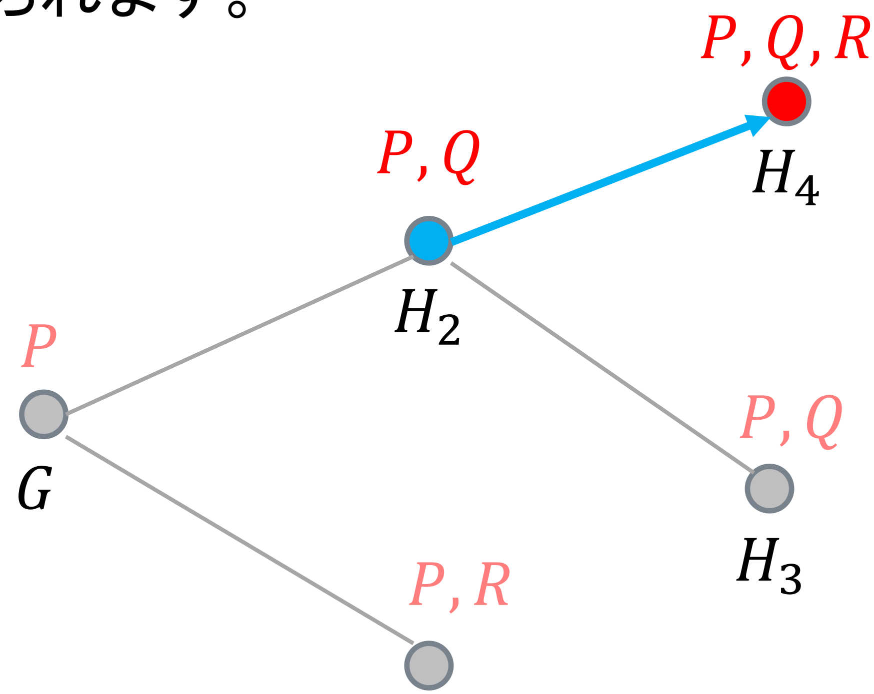
世界 H_2 から世界 H_3 へジャンプする

世界 H_2 には、世界 H_3 あるいは世界 H_4 にジャンプする可能性があります。ここでは、世界 H_3 にジャンプした場合を考えましょう。世界 H_3 では、世界 H_2 と同じく PとQ両方の情報を維持しているのが分かります。それでは、世界 H_2にとどまるのと、世界 H_3 にジャンプするのとは同じことなのでしょうか？



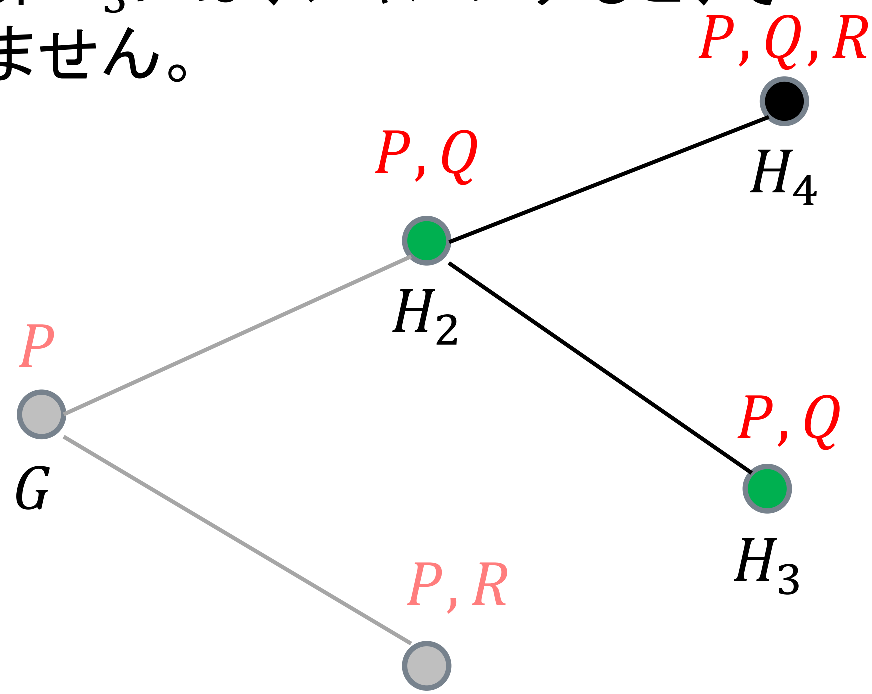
世界 H_2から世界 H_4へジャンプする

先の、世界 H_2 と世界 H_2 とは同じ世界かという疑問に答えるために、世界 H_2のもうひとつの可能性である 世界H_4へのジャンプを考えてみましょう。 世界H_4では、情報P, Qに加えて、情報Rが得られます。



世界 H_2と世界 H_3は同じものか？

確かに、世界 H_2 と世界 H_3 は同じ情報 P,Qを持っています。ただ、世界 H_2には、そこにとどまる限り、いつか世界 H_4にジャンプして情報Rを獲得する可能性があります。それに対して、ひとたび世界 H_3には、ジャンプすると、そこでは情報Rを得ることはできません。



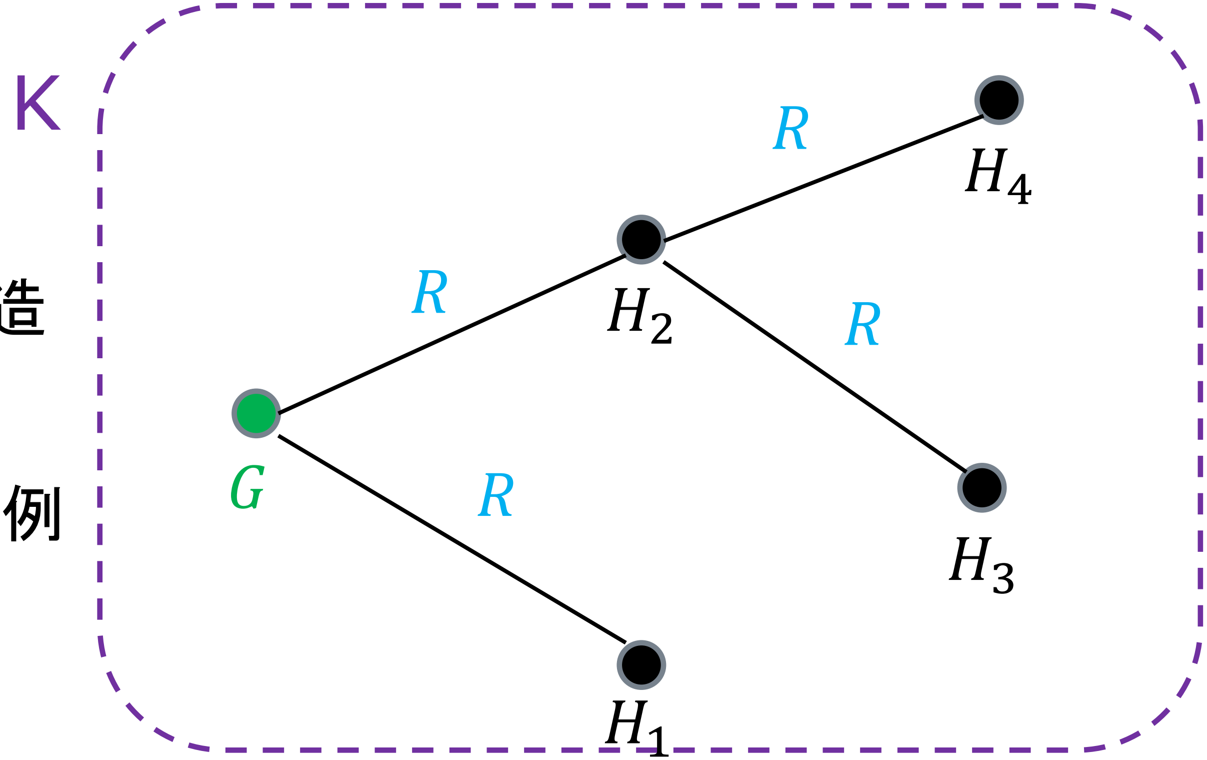

Kripkeの「可能的世界」４
認識の特徴を表現するモデル

モデル構造 (G,K,R)

可能的な世界全体の関係Rを表したツリー構造をKとして、この構造全体を、モデル構造 (G,K,R) と呼びます。

ツリーKの始点Gは、「現在の世界」として解釈されます。

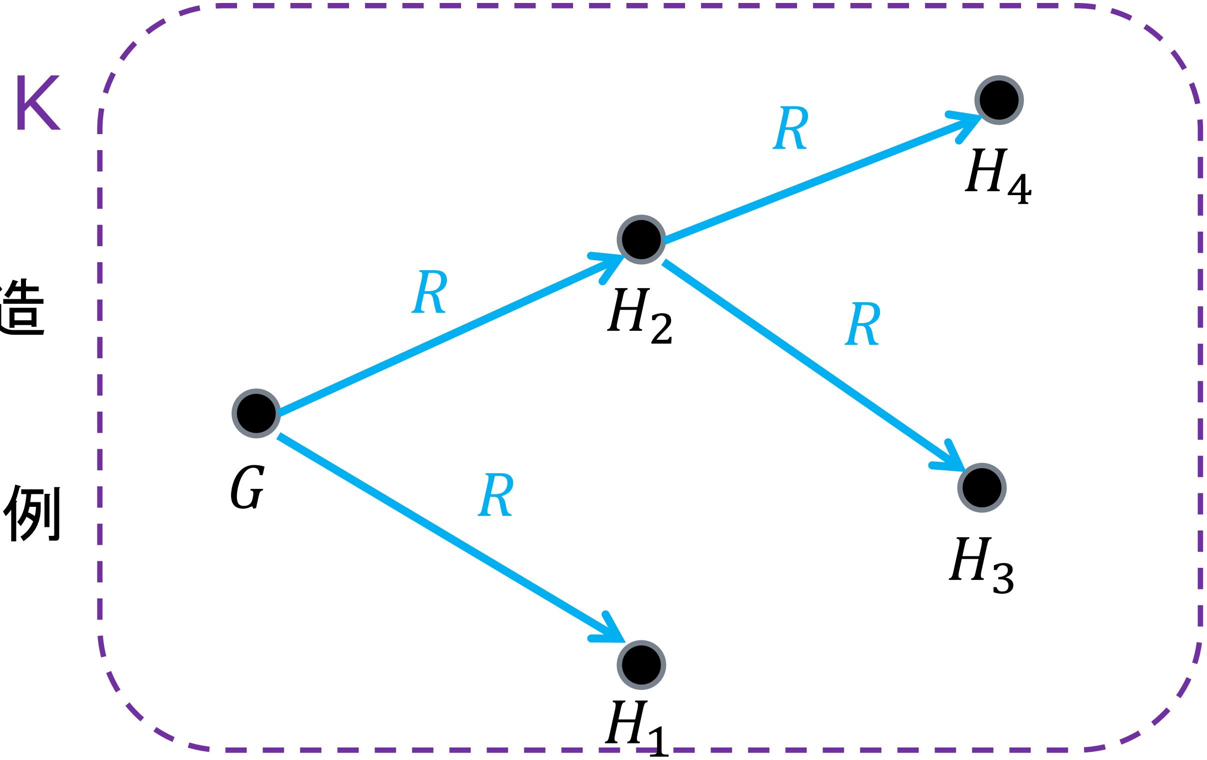
モデル構造
(G,K,R)
の一つの例



世界のあいだの可能的関係 R

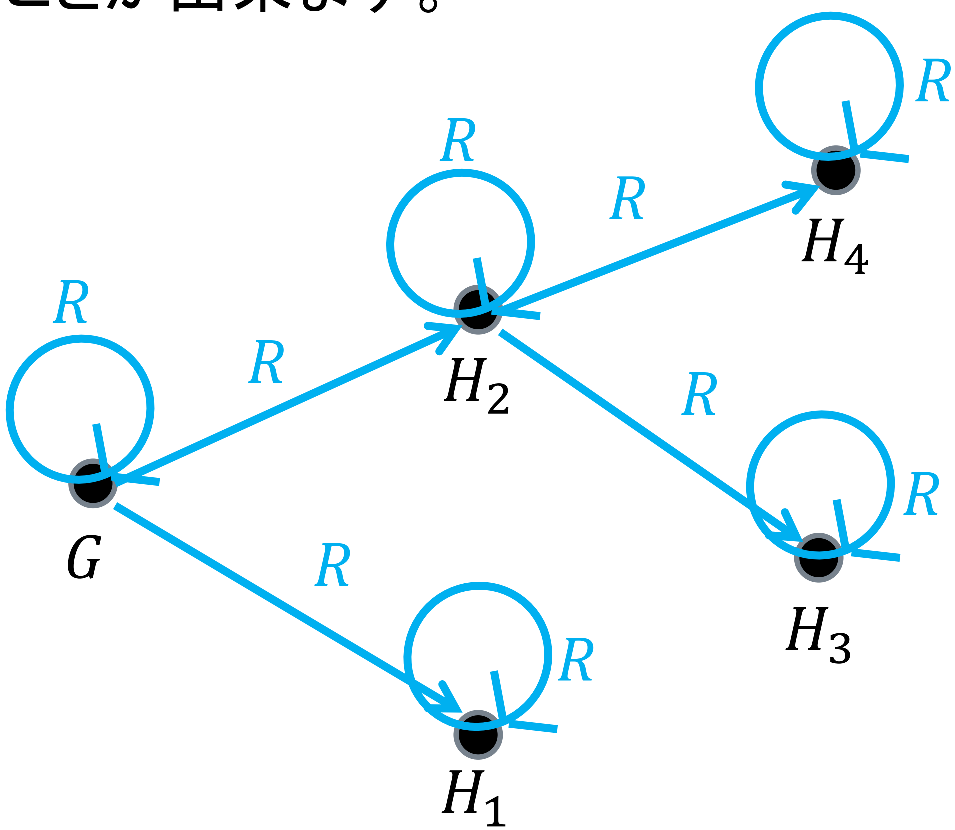
世界Hがどんな状況であっても、我々が世界Hから世界H′に移動する可能性を持つなら、HRH′ であるといいいます。

モデル構造
(G,K,R)
の一つの例



可能的関係 R の性質
反射性: HRH

世界Hで我々の持つ情報は，どんなに長い間そこにとどまっていても、我々がうるすべての情報であるかも知れないので，HRHと考えることが出来ます。

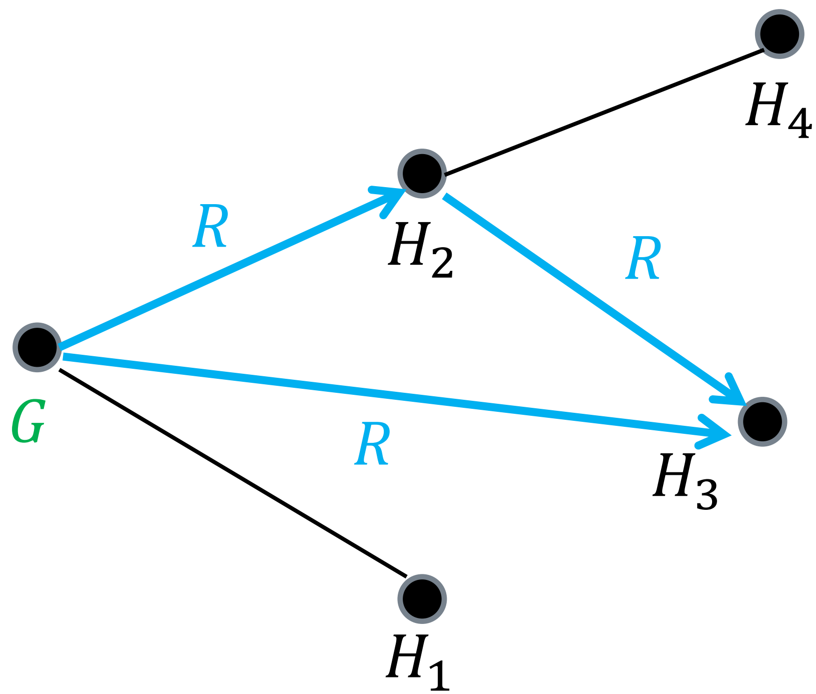


可能的関係 R の性質
推移性： HRH′でH′RH″ならHRH″

関係Rの推移性は、直観的に明らかです。

例えば、我々がGからH_2に移動し、その後H_2からH_3に移動したのなら、我々はGからH_3に移動したことになります。

ですから、下のモデルでは、GRH_3が成り立っています。



モデル (G,K,R)で認識の特徴を表現する

モデル (G,K,R)が興味深いのは、このモデルにある性質を持つことを要請することで、認識あるいは論理の、いくつかの特徴を記述・表現することができることです。

ここでは、次の二つの例を紹介します。

- 知識の「累積性」の表現
- 直観主義論理の表現

知識の「累積性」の表現

このモデルに、次の性質を持つことを要請したとします。

「すべてのAに対して$\phi(A, H) = T$で、HRH′ なら$\phi(A, H') = T$である。」

この要請は，もし，我々が，HでAの証明をすでに得ているなら，いかなる後であってもAを証明されたものとみなしてよいことを意味しています。

別の言い方をするなら、この要請は、このモデルでは、「我々は、一度知ったことは忘れない」という特徴、すなわち、知識が「累積性」を持つことを表現しています。

モデル (G,K,R)で
直観主義論理の特徴を表現する

このモデルが、次の性質を持つことを要請したとします。

$$\phi(\sim A, H) = T \leftrightarrow (\forall H' \in K)(HRH' \dashrightarrow \phi(A, H') = F)$$

$\phi(\sim A,H)=T$ すなわち、世界Hで論理式～Aを証明する十分な情報があると主張するためには、Aが Hでは検証されないことを示すだけでは十分ではなく、Kの中のHRH'であるすべての世界H'で、いかに多くの情報がえられても，いかに時がたっても，Aが検証されえないこと、すなわち $\phi(A, H') = F$であることを，我々はHで知らねばならないということです。

この否定の解釈は、直観主義論理のものです。

モデル (G,K,R)で論理の特徴を表現する 直観主義論理と古典論理

こうしたKripkeの直観主義論理の解釈は、以前に見た、Grzegorczykの解釈とほとんど同一のものです。

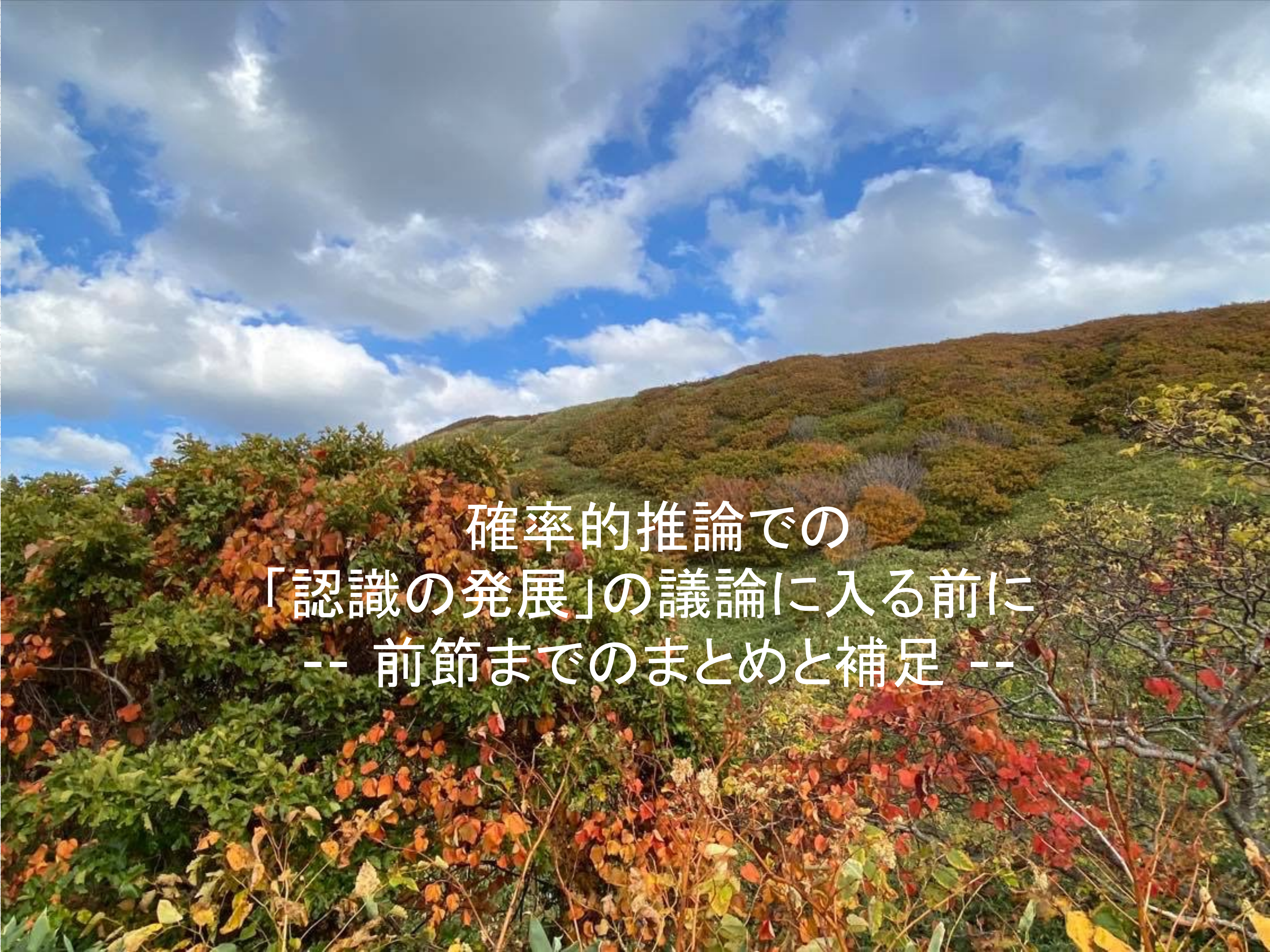
もちろん、モデル(G,K,R)では、古典論理の解釈も与えることができます。次の二つの条件を比較してみて下さい。

直観主義論理

$$\phi(\sim A, H) = T \leftrightarrow (\forall H' \in K)(HRH' \dashrightarrow \phi(A, H') = F)$$

古典論理

$$\phi(\sim A, H) = T \leftrightarrow (\forall H' \in K)(HRH' \dashrightarrow \phi(\sim A, H') = T)$$


確率的推論での
「認識の発展」の議論に入る前に
-- 前節までのまとめと補足 --

Grzegorczyk と Kripkeのモデルの背景
Cohenの「連続体仮説」の独立性証明

Grzegorczyk と Kripkeのモデルには、共通の背景があります。それは、1963年のJ.P.Cohenの仕事です。

Cohenは、Cantor以来の難問であった「連続体仮説」を、それが集合論ZFから独立であることを示すことによって、否定的に解決します。Hilbertが20世紀の数学が解くべき問題の筆頭に挙げた問題がついに解かれたのです。

Grzegorczyk と Kripkeのモデル構成の試みは、この20世紀の数学史上の大きな出来事を、どのように理解するのかという問題意識と、直接的・間接的に結びついています。

Grzegorczyk と Kripkeのモデルと「直観主義論理」

Grzegorczyk と Kripkeのモデルは、彼らの論理式の否定の解釈によれば、ともに、「認識の発展」の論理が「直観主義論理」に従うことを主張しています。次の二つの式は、同じことを言っています。

Grzegorczyk：

$$\alpha \rhd \lceil \sim\phi \rceil \leftrightarrow \forall\beta\big(\beta \succ_R \alpha \rightarrow \sim(\beta \rhd \phi)\big)$$

Kripke：

$$\phi(\sim A, H) = T \leftrightarrow (\forall H' \in K)(HRH' \rightarrow \phi(A, H') = F)$$

これは、とても興味深いことです。もちろんCohenのForcing methodも直観主義論理に従います。

Grzegorczyk と Kripkeのモデル
情報は命題として表現される

Grzegorczyk と Kripkeのモデルは、もう一つ大事な共通点があります。それは、両者のモデルにおいて、情報は「命題」という形で表現されていることです。

Grzegorczykの場合は、直接的です。「情報 αがアトミックな論理式ϕを成り立たせる」のは、$\alpha \rhd \phi \leftrightarrow \phi \in \alpha$ の時です。

Kripkeの場合は、すこし間接的です。「世界Xで、ある論理式Aを証明する十分な情報が存在する」は、$\phi(A, X) = T$ で表現されます。しかし、ここでも「十分な情報が存在する」という「情報についての情報」は、命題として表現されます。

Grzegorczyk と Kripkeのモデル 情報は命題として表現される？

情報が論理的な命題として形式的に表現されるというのは、ある意味自然なことのように見えるのですが、実は、認識論的には自明なことではありません。

情報は、まず、自然言語で表現され、その後、形式的な表現に変換されているはずです。（この問題は、今回は取り上げません。「言語の情報理論」が必要です。）

しかしながら、この点では、Grzegorczyk と Kripkeのモデルでは違いも存在します。

Grzegorczykの場合には、その命題は、仮説・実験を通じて「経験的」に獲得されたものです。

Kripkeの場合には、その命題の数学的証明可能性が問題となる、「数学的な命題」です。

判断としての認識と確率的認識

Thomas Bayes
1701~1761

Edwin Thompson Jaynes
1922~1998

Part III

Bayesian推論と相対エントロピー

Part III

Bayesian推論と相対エントロピー

1. 相対エントロピーとは何か？

2. PriorとPosteriorで「認識の発展」を記述する

3. Deep Learning と相対エントロピー


Bayesian推論と相対エントロピー 1
相対エントロピーとは何か？

確率分布とエントロピー

ある確率分布 p_iが与えられた時、そのエントロピー Sは次の式で与えられます。

$$S = -\sum_i p_i \log p_i$$

確率分布とエントロピー

ただ、どんな確率分布についても、アプリオリに一つのエントロピーが先の公式で天下り的に定まるということに、すこし違和感を持つ人がいるかもしれません。

そもそも、確率分布がアプリオリに与えられるものかは、自明ではありません。

相対エントロピー

そういう人には、次の「相対的なエントロピー」という考え方の方が、納得が行きやすいと思います。

「エントロピー」は、絶対的な確定したものではなく、事前に知っていたこととの関係で決まる、相対的なものだと考えるのです。

事前に知っていた確率分布をpとし、実際に、観測して得られた新しい確率分布 qとします。

この時、次の式で確率分布pに対する確率分布qの「相対エントロピー」 $\mathrm{H}_{\mathrm{rel}}(\mathrm{q}, \mathrm{p})$を定義します。

相対エントロピー

確率分布pに対する確率分布qの「相対エントロピー」
$\mathrm{H}_{\mathrm{rel}}(\mathrm{q},\mathrm{p})$を、次の式で定義します。

$$\boxed{H_{rel}(q,p) = \sum_x q(x) log \frac{q(x)}{p(x)}}$$

確率分布pに対する確率分布qの「相対エントロピー」
$H_{rel}(q,p)$を $H(q(x)||p(x))$ と表すことがあります。

$$H(q||p) = H_{rel}(q,p) = \sum_x q(x) log \frac{q(x)}{p(x)}$$

‘Prior’ と ‘Posterior’

確率分布pに対する確率分布qの相対エントロピー$H(q||p)$で、

　p を、「事前確率」＝ ‘Prior’

　q を、「事後確率」＝ ‘Posterior’

と呼びます。

$$H(q||p)$$

Posterior probability　　Prior probability

認識の発展のBayesian的解釈

こうした考え方は、Bayesianのものです。

「相対エントロピー」というのは、アプリオリな「シャノンのエントロピー」を、Bayesianの考え方で、相対化したエントロピーと考えることができます。

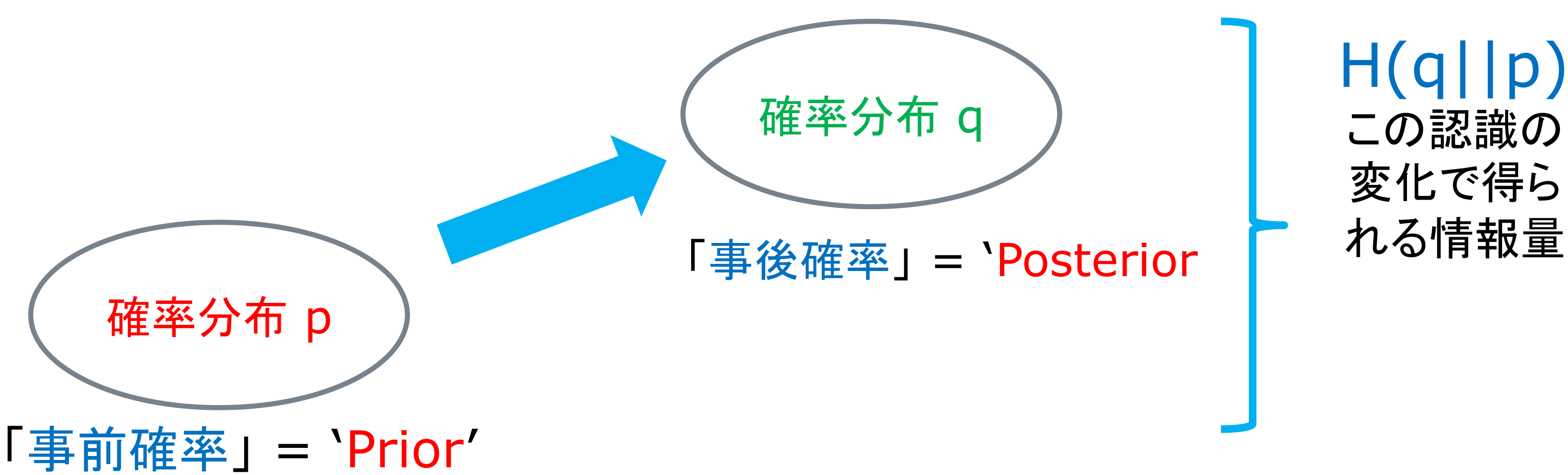
エントロピー＝情報量のこのBayesian的な解釈は、人間の認識で得られる情報量の解釈には、とても向いています。

Prior の pという仮説的な確率の認識は、Posterior のqという確率の認識に「発展」したと考えることができるからです。

Bayesian inferenceと相対エントロピー 2
PriorとPosteriorで「認識の発展」を記述する

相対エントロピーの直観的意味

直観的に言えば、相対エントロピー H(q||p) は、あるシステムが確率分布 p に従っているという仮説的認識から出発して（これが‘Prior’ です）、その後、そのシステムの「正しい」あるいは「実際」は、確率分布 qに従っていることを学んだ（これが‘Posterior’）時に、得られる情報量です。



コイントスを例に相対エントロピーを計算する

Case 1:例えば、コイントスに使われるコインが、かたよりがなく公正なものだという仮定から出発して、実際に、コインの表がでたとすれば、その相対エントロピーは、log 2 となって、我々は1bitの情報を得たことになります。

Case 2:しかし、コインは常に表が出るという仮説から出発すれば、表が出たとしても、我々の得る情報、すなわち相対エントロピーはゼロになります。

$$H(q||p) = \sum_x q(x) log \frac{q(x)}{p(x)}$$

$$H(q||p) = q(表) log \frac{q(表)}{p(表)} + q(裏) log \frac{q(裏)}{p(裏)}$$

コイントスを例に相対エントロピーを計算する

$$H(q||p) = \sum_x q(x) log \frac{q(x)}{p(x)}$$

$$H(q||p) = q(表) log \frac{q(表)}{p(表)} + q(裏) log \frac{q(裏)}{p(裏)}$$

Case 1:

仮定: コインは公正 p(表)=p(裏)=1/2; 結果: q(表)=1,q(裏)=0

$$H(q||p) = 1\ log \frac{1}{1/2} + 0 log \frac{0}{1/2} = \mathrm{log2} = 1\ \mathrm{bit}$$

Case 2:

仮定: コインはインチキ p(表)=1,p(裏)=0; 結果: q(表)=1,q(裏)=0

$$H(q||p) = 1\ log \frac{1}{1} + 0 log \frac{0}{0} = 0\ \mathrm{bit}$$

認識の発展： Step t = 0

Step	Posterior Prior
$t = 0$	$H(p_1 \| \| p_0)$

認識のある段階で、我々はあるシステムの p0で与えられると考えていた確率分布が、本当は確率分布p1に従うことを発見したとする。
この発見によって得られる情報量は、相対エントロピー H(p1||p0) によって与えられる。

認識の発展:　Step t = 1

Step	Posterior　Prior
$t = 0$	$H(p_1 \| \| p_0)$
$t = 1$	$H(p_2 \| \| p_1)$

認識の次の段階で、あるシステムの p1で与えられると考えていた
確率分布が、本当は確率分布p2に従うことを発見するかもしれない。
この発見によって得られる情報量は、相対エントロピー H(p2||p1)
によって与えられる。

認識の発展： Step t = 2

Step	Posterior Prior
$t = 0$	$H(p_1 \| \| p_0)$
$t = 1$	$H(p_2 \| \| p_1)$
$t = 2$	$H(p_3 \| \| p_2)$

$\vdots$

認識の次の段階で、あるシステムの p2で与えられると考えていた
確率分布が、本当は確率分布p3に従うことを発見するかもしれない。
この発見によって得られる情報量は、相対エントロピー H(p3||p2)
によって与えられる。

PriorとPosteriorで「認識の発展」を記述する

Step　　Posterior　Prior

$t = 0$　　$H(p_1||p_0)$

$t = 1$　　$H(p_2||p_1)$

$t = 2$　　$H(p_3||p_2)$

$t = 3$　　$H(p_4||p_3)$

$\vdots$　　$\vdots$

$\vdots$　　$\vdots$

$t = i - 1$　　$H(p_i||p_{i-1})$

$t = i$　　$H(p_{i+1}||p_i)$

$\vdots$　　$\vdots$

Bayesian inferenceと相対エントロピー 3
Deep Learning と相対エントロピー

この「認識の発展」が終わるのは？

Step	Posterior Prior
$t=0$	$H(p_1 \| p_0)$
$t=1$	$H(p_2 \| p_1)$
$t=2$	$H(p_3 \| p_2)$
$t=3$	$H(p_4 \| p_3)$
⋮	⋮
$t=i-1$	$\boxed{\boldsymbol{H(p_i \| p_{i-1}) = 0}}$

p_{i-1}に対するp_iの相対エントロピー

$$\boldsymbol{H(p_i||p_{i-1}) = 0}$$

がゼロになる時。

すなわち

p_{i-1}から新しい情報が、もはや得られない時。

「学習」のBayesian的解釈

逆に、もしも、最初から正しい分布q を、何らかの方法で我々が知っていて、実験 $p(t)$を繰り返すのなら、$\mathrm{H}\big(q||p(t)\big)$ は、実測値$p(t)$から、正しい答えq に至るために「学習しなければいけない情報量」を表すことになります。

q は、「常に正しい」と仮定しているので、それは時間には依存しません。qは、tを含まないことに注意してください。

ここでは、$\mathrm{H}\big(q||p(t)\big) = 0$ は、「もはや、学習すべき情報が残されていない」ことを意味して、その状態で、学習は終わります。

「学習」のモデルと相対エントロピー

Step	Posterior Prior
$t = 0$	$H(\boldsymbol{q} \| p_0)$
$t = 1$	$H(\boldsymbol{q} \| p_1)$
$t = 2$	$H(\boldsymbol{q} \| p_2)$
$t = 3$	$H(\boldsymbol{q} \| p_3)$
$\vdots$	$\vdots$
$t = i - 1$	$H(\boldsymbol{q} \| p_{i-1})$

学習すべき
正しい目標 **q**
時間の中で
変わらない

q に近づくように
p の値を変えて
繰り返し「学習」
する。

この「学習」が終わるのは？

Step	Posterior Prior
$t=0$	$H(\boldsymbol{q}\|\|p_0)$
$t=1$	$H(\boldsymbol{q}\|\|p_1)$
$t=2$	$H(\boldsymbol{q}\|\|p_2)$
$t=3$	$H(\boldsymbol{q}\|\|p_3)$
$\vdots$	$\vdots$
$t=i-1$	$\boxed{H(\boldsymbol{q}\|\|p_{i-1})=0}$

p_{i-1}に対するqの相対エントロピー

$$\boldsymbol{H(q\|\|p_{i-1})=0}$$

がゼロになる時。

すなわち

もはや、学習すべき情報が残されていない時

クロス・エントロピー

ディープラーニングでコスト関数として利用される「クロス・エントロピー」は、こうした「相対エントロピー」の一種です。

「正しい」分布をq、実測値を pとしたとき、クロス・エントロピー $H_{cross}(q,p)$は、次の式で定義されます。

$$H_{cross}(q,p) = \sum q_i \log p_i$$

クロス・エントロピーと相対エントロピー

先の相対エントロピー $\mathrm{H}(q||p)$ の定義から、次のことがわかる。

$$\mathrm{H}(q||p) = \sum q_i \log \frac{q_i}{p_i} = \sum q_i \log q_i - \sum q_i \log p_i$$

$$= -H(q) - H_{cross}(q,p)$$

クロス・エントロピー

個人的には、シャノンのエントロピーの定義も、ぶっきらぼうだと思いますが、クロス・エントロピーの定義の意味は、式の形からはわかりにくいように感じています。

ディープラーニングでの「学習」の直接の目的は、クロス・エントロピーを最小にすることなのですが、相対エントロピーの言葉で言えば、それは、「正しい」認識に至るために、「残された学習すべき情報を最小のものにすること」となります。

こちらの方がわかりやすいように思います。

Part IV

Jaynesの「最大エントロピー 原理」

Part IV

Jaynesの「最大エントロピー 原理」

1. Gibbsの論理の不思議なパワー

2. 認識の発展と「最大エントロピー原理」

3. 「最大エントロピー原理」の応用
 としての「アルゴリズム論的熱力学」


JaynesのMAXENT
Gibbsの論理の不思議なパワー

JaynesのMAXENT

現代のBayesian理論の最大の貢献者は、E.T.Jaynes です。相対エントロピー の重要性を最初に指摘したのも彼です。

このセクションでは、Jaynes の「MAXENT ＝ 最大エントロピー原理」を紹介します。

それは、統計理論の枠を超えて、科学の方法論、さらには、認識の理論としも、大きな影響力を持っています。

このスライドでは、JaynesのMAXENTのアイデアの出発点となった、Gibbsの方法とその不思議なパワーについて述べてみようと思います。

現代のBayesian理論の開拓者 E.T.Jaynes

Thomas Bayes
1701~1761

Edwin Thompson Jaynes
1922~1998

Gibbsの方法の不思議なパワー

Gibbsの方法は、不思議なパワーを持っています。

たとえば、ある系のエネルギーについて考えてみましょう。

我々は、ある系がどのようなエネルギーをもつかを、観測によって知ることができます。

しかし、それだけでは、明らかに情報が不足していて、その系がどのようなエネルギー分布に従っているかを知ることは出来ません。観測された平均的なエネルギーだけから、系のエネルギー分布を知ることは不可能です。

ところがGibbsは、この不可能を可能にしました！

我々が知っていること

我々が知っていることは、系の平均的エネルギーが E であるということだけです。系の確率分布をp_iとすると、この条件は次の式で表すことができます。

$$\sum_i p_i E_i = E$$

もう一つ、我々が知っていることがあります。それは、

$$\sum_i p_i = 1$$

だということです。

ただ、あきらかに、情報が不足していて、この二つの式だけからp_iを求めることは不可能です。

系のエントロピーSを考える

Gibbsは、欠けている情報を、利用できる他の情報で補おうとします。

彼は、系のエントロピーSを考えます。

$$S = -\sum_i p_i \log\, p_i$$

これと、先の式を合わせて、次のような式をたてます。

系のエントロピーSを考える

Gibbsは、欠けている情報を、利用できる他の情報で補おうとします。

彼は、系のエントロピーSを考えます。

$$S = -\sum_i p_i \log\, p_i$$

これと、先の式を合わせて、次のような式をたてます。

$$-S = \sum_i p_i \log\, p_i + \alpha \left[\sum_i p_i - 1 \right] + \beta \left[\sum_i p_i E_i - E \right]$$

知っている情報
ただしこの項は
ゼロである

知っている情報
ただしこの項は
ゼロである

系のエントロピーが最大になるという条件を追加する

Gibbsが追加した条件は、「系のエントロピーは最大になる」というものでした、

この条件は、次の式で表せます。

$$\frac{\partial(-S)}{\partial p_i} = 0$$

この式から、彼は系の確率分布p_iを求めることに成功します。

計算省略していますので、詳細については、次の資料を参照ください。

https://www.marulabo.net/docs/info-entropy2/#エネルギーとエントロピー

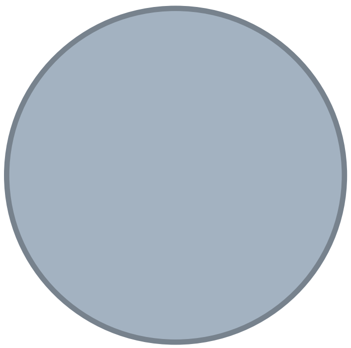

JaynesのMAXENT
「認識の発展」とMAXENT

Gibbsの方法

知っていたこと

平均的 エネルギー <E>

確率　$\sum_i p_i = 1$

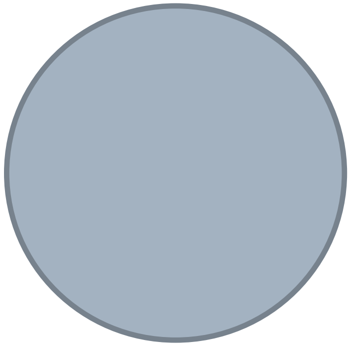


Gibbsの方法

知っていたこと

平均的 エネルギー <E>

確率　$\sum_i p_i = 1$



システムのエントロピー

$$S = -\sum_i p_i \log p_i$$

Gibbsの方法

知っていたこと

平均的 エネルギー <E>

確率 $\sum_i p_i = 1$

システムのエントロピー

$$S = -\sum_i p_i \log p_i$$

$$-S = \sum_i p_i \log p_i + \alpha\left[\sum_i p_i - 1\right] + \beta\left[\sum_i p_i E_i - E\right]$$

Gibbsの方法

知っていたこと

平均的 エネルギー <E>

確率　$\sum_i p_i = 1$

$$-S = \sum_i p_i \log\, p_i + \alpha\left[\sum_i p_i - 1\right] + \beta\left[\sum_i p_i E_i - E\right]$$

システムのエントロピー

$$S = -\sum_i p_i \log\, p_i$$

$$\frac{\partial(-S)}{\partial p_i} = 0$$

Gibbsの方法

知っていたこと

平均的 エネルギー <E>

確率　$\sum_i p_i = 1$

システムのエントロピー

$$S = -\sum_i p_i \log p_i$$

$$\frac{\partial(-S)}{\partial p_i} = 0$$

新たに知ったこと

$$p_i = \frac{1}{Z} e^{-\beta E_i}$$

$$Z(\beta) = \sum_i e^{-\beta E_i}$$

JaynesのMAXENT

事前に知っていたこと

平均的 エネルギー <E>

確率 $\sum_i p_i = 1$

システムのエントロピー

$$S = -\sum_i p_i \log p_i$$

認識の発展

$$\frac{\partial(-S)}{\partial p_i} = 0$$

MAXENT

事前に知っていたこと

平均的 エネルギー <E>

確率 $\sum_i p_i = 1$

事後に知ったこと

$$p_i = \frac{1}{Z} e^{-\beta E_i}$$

$$Z(\beta) = \sum_i e^{-\beta E_i}$$

「認識の発展」の原理として MAXENTを解釈する

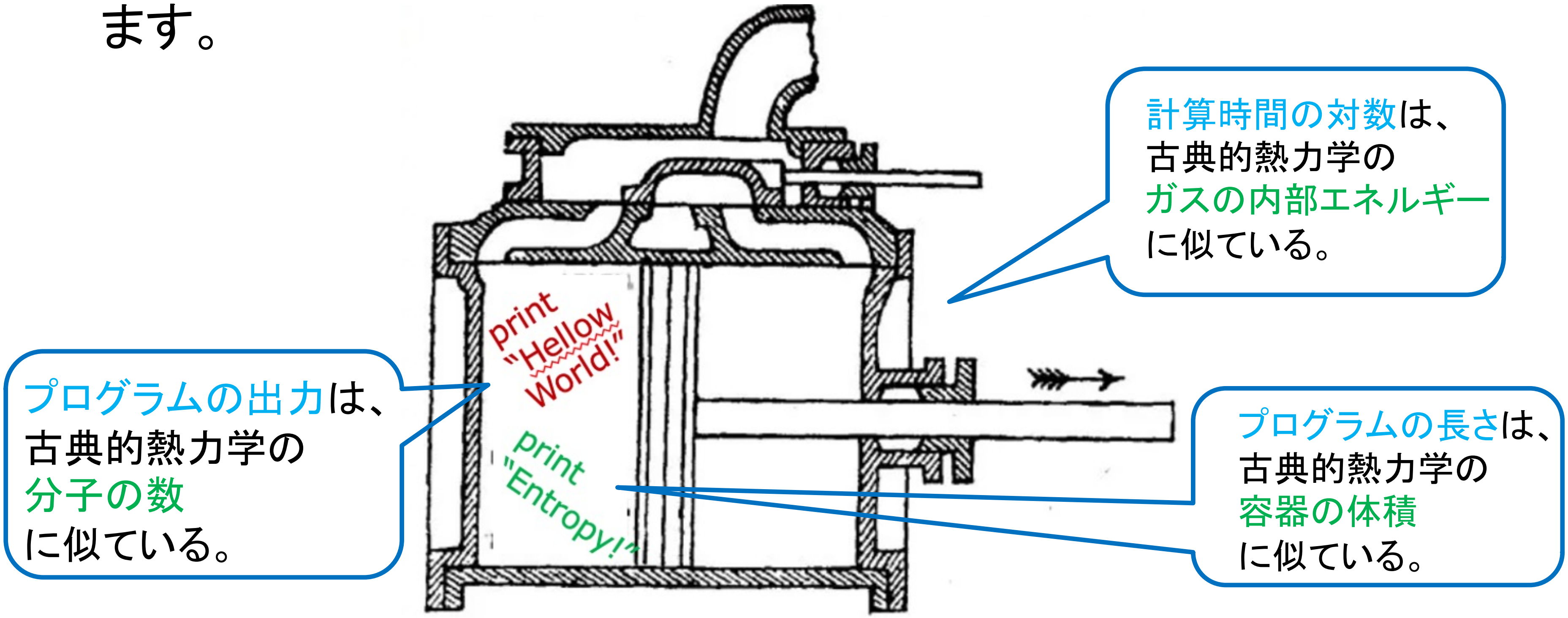
MAXENT – 最大エントロピー原理は、ある系に対する現在の認識の状態を最もよく表現する確率分布は、もっともエントロピー の大きいもの、すなわち、もっともその系の情報が少ない確率分布だと考えようと言うことを表しています。

ある系に対する認識が発展するのなら、その認識の発展は、現在の系が含んでいる最少の情報(それは最大のエントロピーを持ちます)からも得られるはずだと言うことです。

JaynesのMAXENT － 最大エントロピー原理
アルゴリズム論的熱力学

「最大エントロピー原理」の応用としてのアルゴリズム論的熱力学

ここでは、先日のマルゼミ(「コロモゴロフ複雑性とアルゴリズム論的情報理論」 https://www.marulabo.net/docs/info-entropy4/)で紹介した、少し奇妙な「アルゴリズム論的熱力学」も、「最大エントロピー原理」の応用として導出されるという話をします。



振り返り
エントロピーの最大化と分配関数Z

系のエネルギーの平均値Eが与えられたとします。エネルギーEが、各部分系にE_iとして分配されているとします。

エントロピーが最大になるという条件を用いれば、エネルギー分布p_iを求めることができます。

確率p_iは、次の式で与えられます。

$$p_i = \frac{1}{Z} e^{-\beta E_i}$$

この時、分配関数Zは

$$Z = \sum_i e^{-\beta E_i}$$

になります。

古典的な熱力学
エントロピーの最大化とGibbs Ensemble

ガスの分子からなるある系について、エネルギーEだけでなく、体積V、圧力P、分子数Nについても観測量の期待値が与えられているとします。この時、エントロピーが最大になるという条件用いれば、(マルゼミ「情報とエントロピー」を参照ください。
https://www.marulabo.net/docs/info-entropy2/ ）

次のように確率分布p_iを求めることができます。

$$p_x = \frac{1}{Z} e^{-\frac{1}{T}(E(x)+PV(x)-\mu N(x))}$$

この時、Zは、次のような形をしています。

$$Z = \sum_{x \in X} e^{-\frac{1}{T}(E(x)+PV(x)-\mu N(x))}$$

これらから、観測量 E(x), V(x), N(x) の共役変数 β, γ, δを使って、期待値 E,V,Nを定義できます。

$$\beta = \frac{1}{T}, \qquad \gamma = \frac{P}{T}, \qquad \delta = -\frac{\mu}{T}$$

とすると、

$$Z = \sum_{x \in X} e^{-\beta E(x) - \gamma V(x) - \delta N(x)}$$

ですので、

$$E = \sum_{x \in X} p(x) E(x) = -\frac{\partial}{\partial \beta} \ln Z$$

$$V = \sum_{x \in X} p(x) V(x) = -\frac{\partial}{\partial \gamma} \ln Z$$

$$N = \sum_{x \in X} p(x) N(x) = -\frac{\partial}{\partial \gamma} \ln Z$$

最大エントロピー原理

maximize entropy
subject to the constraints imposed by what you believe

あなたが信ずる拘束条件たちに
従うエントロピーを
最大のものにせよ

John Baezが考えた拘束条件

John Baez は、先に見た古典的な熱力学の観測量を、アルゴリズム論的情報理論の観測量で置き換えた、これらを拘束条件とするモデルを考えました。

- E(x)を、プログラムxの実行時間の対数
- V(x)を、プログラムxの長さ
- N(x)を、プログラムxの出力

とします。

最大エントロピー原理が、古典的な熱力学のGibbs Ensembleの式を導くのなら ......

そうすると面白いことが起きます。

最大エントロピー原理が、古典的な熱力学のGibbs Ensembleの式を導くのなら、拘束条件とみなしたものの解釈を全く別なものに置き換えたとしても、最大エントロピー原理は、熱力学のGibbs Ensemble が満たすのと同じ式の導出を可能にします。

Baezが選んだ拘束条件 P,V,N は、熱力学のGibbs Ensemble の圧力、体積、分子数とは、全く異なるものなのですが、最大エントロピー原理は、この新しい拘束条件の解釈のもとでも、全く同一の式が成り立つことを教えてくれます。

Déjà vu！
アルゴリズム論的熱力学

次のような分配関数を考えます。

$$Z = \sum_{x \in X} e^{-\frac{1}{T}(E(x)+PV(x)-\mu N(x))}$$

この和が収束する時、次のようなX上の確率分布p(x)を考えることができます。

$$p_x = \frac{1}{Z} e^{-\frac{1}{T}(E(x)+PV(x)-\mu N(x))}$$

これらから、観測量 E(x), V(x), N(x) の共役変数 β, γ, δを使って、期待値 E,V,Nを定義できます。

$$\beta = \frac{1}{T}, \qquad \gamma = \frac{P}{T}, \qquad \delta = -\frac{\mu}{T}$$

とすると、

$$Z = \sum_{x \in X} e^{-\beta E(x) - \gamma V(x) - \delta N(x)}$$

ですので、

$$E = \sum_{x \in X} p(x) E(x) = -\frac{\partial}{\partial \beta} \ln Z$$

$$V = \sum_{x \in X} p(x) V(x) = -\frac{\partial}{\partial \gamma} \ln Z$$

$$N = \sum_{x \in X} p(x) N(x) = -\frac{\partial}{\partial \gamma} \ln Z$$

二つの理論に、アナロジーが成り立つ！

Baezの選んだ拘束条件の名前はまぎらわしいものでしたが、次のように考えることもできます。

- アルゴリズム論的熱力学のE、すなわち、計算時間の対数は、古典的熱力学のガスの内部エネルギーに似ている。
- アルゴリズム論的熱力学のV、すなわち、プログラムの長さは、古典的熱力学の容器の体積に似ている。
- アルゴリズム論的熱力学のN、すなわち、プログラムの出力は、古典的熱力学の分子の数に似ている。

$dE = TdS - PdV + \mu dN$ もそのまま成り立ちます。

二つの理論には、アナロジーが成り立つのです。