

意志決定理論における自己位置づけ

— 眠り姫問題を通じて —

Self Locating Belief in Decision Theory

— Sleeping Beauty Problem —

榛葉 豊*

Yutaka SHINBA

Abstract: When we evaluate the credence or the degree of conviction, the subjective probability may differ from credence by the character of the proposition to the credence by outside world observer. Credence also differs by the ensemble which if evaluating character belonged. We consider mainly on this problem, what is called 'self locating problem', through the 'Sleeping Beauty' illustration. This illustration contains many other hard questions. For example, the followings are involved, what is information in Bayesian reasoning, the principle of indifference, solipsism and self-consciousness, the Anthropic principle and Copernican principle, interpretation of subjective and objective probability, Doomsday logic and the Fine Tuning of our world, and eventually the interpretation of quantum probability. We discuss these subjects and at a same time several theories in controversy up to now are compared under some independent viewpoints. When we alter the set up in the Sleeping Beauty, more difficulty which show us new world, shall emerge. We discuss that "Information" can be generated by itself within the thinker by his/her location. Three Prisoners and Drunk Home-liner problem also discussed.

1. 序論：眠り姫問題

信念の度合い或いは主観確率を、ある事象に付与しようとするときに、その事象を記述する文に、信念の度合いを考える主体が参加しているかどうか、そしてどういう経緯でそこに登場し認識主体となったのか。また主体はどういう参照集団に属していると考えて評価しているのか、そのとき注目している参照集団を規定する特徴のカテゴリーは何なのかで信念の度合いが違って来る事が有り得る。

その様な事態を考えるための典型として「眠り姫問題」(the Sleeping Beauty Problem)^{1-24, 31)} がある。この問題は、意志決定理論やゲームの理論分野では 1997 年頃から議論されているそうであるが、科学哲学では自己位置づけ問題に関して最近注目されている主題である。宇宙のファイン・チューニング問題における人間原理や、地球文明

の存続に関する終末論法、主観確率の Bayes 改訂と情報取得の関係などの分野に関係するのであるが、哲学の分野では、2000 年の Elga の論文¹⁾ からであろう。

まず、その問題を Elga のオリジナル形で見てみる事にしよう。

眠り姫問題の設定：

あなたは日曜日に睡眠薬で眠らされる。その後フェアなコインが投げられる。

表 → 月曜

裏 → 月曜 火曜

その結果、裏表に従って、上に示したように、表なら月曜

2009 年 3 月 6 日受理

* 総合情報学部 人間情報デザイン学科

に起こされて、質問をされてまたすぐ眠らされる。もし裏の場合は、月曜に起こされるのは表の場合と同様であるが、また眠らされた後、火曜にまた起こされ、月曜同様に質問されまた眠らされる。

このとき月曜に起こされたという記憶は完全に消去される。火曜に起こされたときには、前の記はなく、今日が何曜か分からない。

あなたは実験の設定は完全に分かっているが、とにかく目覚めたときに何曜か分からないし、コインが表だったのか裏だったのか分からない。水曜日に起こされて実験終了となる。

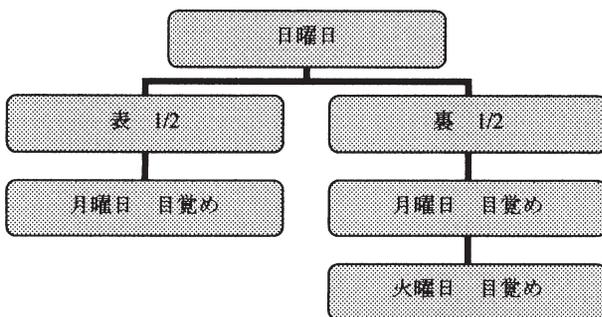


図1 眠り姫問題

さて、あなたは起こされたとき「コインは表だったと思いますか」とコインが表だった確率を聞かれる。ここで言う「表だった確率」とは、現在の目覚めがコインの表の結果として起こっているのか、そうでなく月、火と起こされるべき系列の一つとして起こされているのかの見込みが問われている。

この答えに $p(H)=1/2$ (H は表, T は裏とする)

と答える考えと、 $p(H)=1/3$ と答える考えが先ずはあるのである。

目覚めた時の質問は、2段階になされる。

質問0 まず、日曜日に、「コインをあなたが眠った後投げるときに、表の確率はどれだけだと思いますか」

質問1 目覚めたとき先ず「コインは表だったと思いますか、その確率は」

質問2 次に「今は、月曜日です」という情報を与えられて、「改めて聞きます。コインは表だったと思いますか、その確率は」

質問0に対する答えは、誰でも一致して $1/2$ であろう。フェアなコインとは、表が出る確率が $1/2$ のコイン

である。これは必ずしも物理的に対称で、頻度確率が $1/2$ であると言うことでなくてもよい。ここで聞いているのは、信念の度合いを尋ねているのである。主観確率が何らかの理由で、たとえばそのサイコロや実験者に対する情報が無いなどでもよい、表とか裏が出やすいとは思えないという不充足理由律でよいのである。

2. 質問1に対する $p=1/2$ 説と $1/3$ 説：眠り姫問題

意見が分かれるのは、質問1と質問2に対する答えである。質問2 $1/2$ 説と、 $1/3$ 説である。

その根拠はそれぞれの説で各人色々である。例えば Elga は $1/2$ 説を次のように説明する。

日曜日に眠る前に $1/2$ であった。今起こされてみて、一度はどのような場合でも必ず起こされることははじめから分かっていたのだから、そして周りの状況も区別は付かないのであるから、情報は何も入ってきていない。したがって $1/2$ から Bayes 改訂する理由はない。Lewis⁸⁾ はこの派である。

一方 $1/3$ 説に対しては次のような説明を与える。Elga 自身はこの派である。

H_1 : Head (表) の結果, 月曜日に起こされている

T_1 : Tail (裏) の結果, 月曜日に起こされている

T_2 : Tail (裏) の結果, 火曜日に起こされている

としよう。まず $\{H_1, T_1, T_2\}$ は互いに背反で、事象の完全系をなす。不充足理由律より

$$p(T_1|T_1 \vee T_2) = p(T_2|T_1 \vee T_2)$$

である。したがって

$$p(T_1) = p(T_2)$$

である。一方次のことは自明であろう(と、Elga は言うが、筆者は、それは間違いだと思う。後述する。)

「眠り姫問題の設定と次の設定は同じ確率を与える。すなわち、日曜に眠らせ、そのまま月曜に起こす。質問の後また眠らせ、その後でコイントスを行う。表ならそれで終わり。裏なら火曜日にもう一度起こす」図2参照

この方法の実験で置き換えて考えよう。

もし、今が月曜日であるという情報もたらされたとき、即ち質問2であるが、完全系は $\{H_1, T_1\}$ となる。このとき

H_1 とはこの設定の場合、この後自分は眠らされた後、まもなくフェアなコインが投げられて、それが表になると言

う事についての信念の度合いである。もしそうならそれで終わり。裏なら翌日もう一度起こされるということになる。

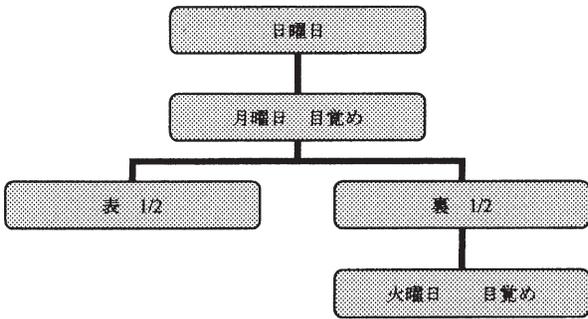


図2 コイントスを事後にした眠り姫問題

言い換えれば、質問2の時点での確率を p_+ と表記することにして、

$$p_+(H_1|H_1 \vee T_1) = 1/2$$

である。したがって

$$p_+(H_1|H_1 \vee T_1) = p_+(T_1|H_1 \vee T_1)$$

となり

$$p_+(H_1) = p_+(T_1)$$

前半の結果と、さらに p と p_+ は同じであるという Elga の考えと合わせて

$$p(H_1) = p(T_1) = p(T_2) = 1/3$$

が得られる。

この導出は、(図2が図1と同じかどうかという点を除いても)、質問2の情報が得られた場合の評価を、未だ情報が得られない時点での評価のために使って良いのかどうかと言う問題点を含んでいると言えよう。

この Elga の導出よりも次のようにするのが分かりやすいであろう。

単一事象設定ではなく、眠り姫実験自体を多数回繰り返すというスキームで、何回 H_1 が起こったかという頻度解釈を採用するとき妥当性を持つと言われる³¹⁾。いわば、実験者の立場である。

もし、頻度確率で無い単一事象に対する主観確率とすると、参加者が、裏が出ていると考え推論しているとき、既に表が出ている系列は過去の潜在でしかあり得ないのである。だが、実験者からすればそうではない。

1万回眠り姫実験をしたとしよう。コイントスが表にな

った実験も、裏になった実験も1万回のうち約5000回実際に存在するという解釈である。5000回のコインが表であつたうえでの月曜の目覚めが存在し、また一方コインが裏で月曜の5000回の目覚めが存在し、そしてそのそれぞれに対応している、引き続いて起こる5000回の火曜の目覚めが実在する。あわせて目覚めは実際に1万5000回起こるのである。

それらの15000回の目覚めに不充足理由律を適用すると言っても、今ひとつ同じ重みを持っているとは考えにくい人もいるであろう。それは同じ世界の出来事の違いだからではないだろうか。しかし、後述するように、設定を目覚めではなく、人間にしたら同じ重みという感じが自然であると思えるであろう。「神はサイコロを振って、表なら1人を創造する。裏なら1億人を創造する」という設定を考えるのである。これなら、あなたの見るサイコロが表である可能性はほとんど無いと判断するであろう。

さてその上で、目覚めと言うことにたいして、不充足理由律を使う。すなわち、目覚めに関して眠り姫が無差別という事を根拠にするのである。

統計力学では、巨視的変数では区別出来ない微視的状態の数を勘定した。微視的変数達の値は異なるのであるが、エルゴード仮説により微視的状態は互いに可能性が同じであるとしたわけである。等重率である。

あなたは実験者によって起こされた。実際には、裏の部屋にあるコインの裏表は違うし、また月曜日か火曜日かと言うこと、それに外の世界の事件なども異なっているかも知れない。しかし、あなたには、3つの目覚め $\{H_1,$

$T_1, T_2\}$ は、互いに区別出来ない。不充足理由律により、

この3つに同じ重みを与えると、それぞれ確率 $1/3$ である。

統計力学との対比で言うところなる。微視的状態は莫大な変数により記述されるが、その位相空間の全ての変数(ある制約条件はあるのであるが)一つ一つの Lebesgue 測度的違いに対し等重である。眠り姫問題の場合これになぞらえて言うと、場合は4通りになる。即ち(表, 月), (表, 火), (裏, 月), (裏, 火)である。しかし(表, 火)という目覚めは、眠り姫問題の設定から起こりえない。これは統計力学で言えば、エネルギー超平面に乗っていない状態、と言う制約に対応する。その結果残りの3通りが互いに区別出来ない、等重な目覚めとなる。

目覚めたと言う事は、確率1で起こるので、情報ではないと思えるが、時間軸の可能性に着目するとそうではない。月曜に起きると言うことは、2通りの可能性なのに、火曜日に起きると言うことは1通りの場合しかないと考え得る。これは目覚めを特徴付ける2変数の間の絡み合い具合、制約の付き具合によっているのだとも言えるであろう。

日曜に眠らされ、コインが投げられ、起こされる。「今が月曜である確率は」という質問だったらどうなるか考えてもらいたい。時間に対する位置づけを顕わに聞くのである。1/3派は、月曜であろう、と答えねばならない。

3. 何に対して無差別か

この事情は、頻度解釈だけではなく主観解釈でも同じように起こりうることであるが、不充足理由律をどの様な変数に対して適用するのが自然かという問題に戻ってくる。

確率論では、何について一様無差別とすればよいのかと言う事に関して、例えば「ベルトランのパラドックス」であるとか、「ワインと水」という問題が知られている^{28,37)}。前者は幾何学的問題で、円の半径上の長さに対して一様か、円周上の一点から見込む確度に対して一様か、円盤上の面積に対して一様かで、ある確率が $1/2$ $1/3$ $1/4$ と算出される。また後者はワインと水を混合したときの濃度に関する命題で、ワインの濃度を問題にするか、その逆数である「薄さ」に関して一様かで、ある命題の確率が異なって計算される。

これと同様で、眠り姫問題もどの変数に着目するのかである。

曜日とコインで記述された位相空間にたいして不充足理由律を発動して、問題設定という制約条件を入れて得られる、「実在」としての目覚めに自己を同定して位置づけるのか（このばあい自己は目覚めた事後の認定である）。

それとも、実験者の視点で、実験が行われると想定される時間順に、逐次不充足理由律を発動するのか。即ちフェアなコインは不充足理由律より確率 $1/2$ である。その次に、可能性別々に片方には（表）、実験設定より火曜日の目覚めは存在し得ないという制約条件を入れ、別の可能性である裏には、更に2つの可能性、月曜日と火曜日があつて不充足理由律により $1/2$ のさらに $1/2$ と計算する。

不充足理由律の逐次適用か（ $1/2$ 派）、それとも2変数の直積への一括適用（ $1/3$ 派）であると言えるかも知れない。このことは更に考察していきたい。

しかしコインの裏表と曜日で記述された目覚めに自己を位置づける事までは出来ても、そしてそれらが無差別であっても、当銃とは認定出来にくいと思う。それはやはり、実験者の視点がどうしても入ってきてしまうためではないだろうか。すなわち、その無差別な目覚めに至った「経緯」と言うことに思いが言ってしまうのではと思われる。

そこで、宇宙論の例題を考えてみよう。実験者は神である。神はサイコロを振り、表が出たら宇宙を1回だけ創世する。1回の宇宙とはビッグバン（目覚め）に始まりビッグクランチ（睡眠薬と記憶消去剤）で終わる。もし裏なら、

複数回宇宙を創生する（ここで1つの宇宙の中の時間と神が複数回創世すると言ったときの神の時間とは何かが問題になるが）。我々の宇宙は裏の結果の宇宙であろうか、それとも表の結果の宇宙であろうか。

各宇宙はある準拠集団の一員として無差別であるが、全く同じ問うわけではないという。人類などのその中の知的生命体には分からないが、極極微妙に、物理法則の詳細や、物理定数が異なっている。もちろん、この様な問題を考えようとする程度の知的生命体が生じうる程度には共通になっているとする。ビッグクランチごとに全ては（姫の記憶は）消去される。この設定なら、その宇宙の中の知的生命体にとって他の宇宙（あったとしたら）と無差別であろう。そしてその宇宙での、ビッグクランチがどの様に破滅的なかなどの可能性の議論が出来るであろう。これはすなわち、終末論法に対し、議論の根拠を与えるものである。

急いで説明すると、終末論法とは次のような型の論法である。論者がどの様な世界に属しているかと言うことと、論者はどの様な性質を持っているかから、その世界の運命、確率を導く種々の論法である。ごく簡単な例を示す。「私は特別な人ではない。アダムとイヴのような人とか、地球滅亡に立ち会っているという特別な人であるはずはない。人類を並べたとき私は順番で言って中央付近の人であろう。私より前に生まれた人と私より後に生まれる人の数は同じである。しかるに人類は指数関数的に増えている。時間の尺度で言えば人類の滅亡は近い」。ここに現れる「私は特別ではない」、私は平凡であるという原理は、地球は特別ではないというコペルニクスの原理の拡張形である。また、人間原理とは、論者である人間がまさに存在しうる条件としてこの宇宙の奇跡的な物理定数等の特別の値（ファイン・チューニングという）があるという、「観測選択効果」を基礎とする原理である。

質問2の議論に入る前に、ここでしばらく、眠り姫問題全体の位置づけを見ておこう。

4. 眠り姫問題の論点

眠り姫問題は、意志決定理論の自己位置づけ問題として出発したが、それ以外の科学哲学の様々な問題の試合場でもある。人間原理、終末論法、コペルニクスの原理、自意識の問題、独我論、確率の解釈問題、ファイン・チューニング、情報とは何か、ベイズ改訂とはなど多岐にわたる。

それらの問題はそれぞれ幾つかの要素的論点に分けられるが、その幾つをも、眠り姫問題は含んでいる。

そのことをはっきりさせるために、設定のもう少し詳細な互いに直行するヴァリエーションを列挙しておこう。

(設定軸 1) 単一事象設定か繰り返し実験設定か

問題で議論するのが、主観確率であるのか、客観確率であるのかの違い

(設定軸 2) いつコインを投げるのか

日曜に姫を眠らせた後でコインを投げ世界を分岐させるのか。それとともにかく共通に月曜に目覚めさせて、質問をして事後眠らせ、その後でコインを投げ世界を分岐させるのか。この場合、コインが表ならそれでお終い。裏なら実験は続くということになる。

この違いは、確率の評価に影響しないと Elga は言い、またその用に見る論者は多いが、筆者にはそう考えることは出来ない。これは確率の収縮問題、或いは量子力学の言葉で言えば、波束の収縮(崩壊)である。

未だ決定されていない潜在と、決定されていて実在となっているが、認識主体が知らないだけと言うのは違うのである。

(設定軸 3) 実験者の視点か参加者(姫)の視点か

この違いが(設定軸 1)の主観と客観の違いと一致するのかどうか。

以上の組み合わせで 8 通りの設定があるわけだが、1/2 派 1/3 派 はどう答えうるかを考えてみなくてはならない。

更に

(設定軸 4) 眠り姫の目覚めはどの時点での特定か

一つの世界(この場合目覚め)が議論の中でフォーカスされたとき、そこに参加登場する主体が、「わたくし」「今考えている眠り姫」であるとして、結果的に事後に特定されるのか(Self Sampling Assumption という)。終末論法の立場である。そうであったかも知れない自分(眠り姫)を勘定するのである。

それとも、実験が始まる前から特定の価値を持った存在であるのか(Self Identification Assumption という)。このばあい、「わたくし」場合によっては実在化しないかも知れない。

設定軸 1 での区別とは微妙に異なると思われる。も関連した重要な視座の違いである。

5. 質問 2 に対する答え: 情報の取得と発生

ここで情報が得られたとき、すなわち質問 2 がなされた段階に戻ろう。「今日は月曜です」と言う情報が与えられた。質問 1 に対する 1/2 派 と 1/3 派は、それぞれどう答えるのが、それなりに首尾一貫して合理的であろうか。

1/2 派:

設定軸 1 でいうと、この考えは繰り返し実験設定を採用していると言われる。質問 2 に際してはこう考えなければならない。(表, 月) が 1/2, (裏, 月) が 1/4, (裏, 火) が 1/4 と考えているところに、火曜日の可能性はなくなったと言われたわけである。単純に考えると 1/2 と 1/4 に再規格化すると 2/3 が答えとなる。

3 囚人問題^{28, 38)}で似た状況があり(眠り姫との比較を後述する)、この単純な考えは誤りに導いたので、ベイズ改訂の計算を試みる。

$$p(H|Mon) = \frac{p(Mon|H)p(H)}{p(Mon|H)p(H) + p(Mon|T)p(T)}$$

$$= \frac{1 \cdot 1/2}{1 \cdot 1/2 + 1/2 \cdot 1/2} = 2/3$$

こうして 1/2 派 は、質問 0 → 1 → 2 に際して、

$$1/2 \text{ 説: } 1/2 \rightarrow 1/2 \rightarrow 2/3$$

と変化することになる。

ここで 1/2 説に対するここでパラドックスを述べておこう。Elga はじめ多くの論者は、設定軸 2 について、月曜の目覚めの前後で、どちらでコイン投げても同じとの意見である。仮にそうだとすると、図 2 の状態で、月曜という情報が入り、表の確率が 2/3 である。しかし眠り姫が眠った後、「フェアな(即ち表が出る確率が 1/2 である)コインを投げると、確率 2/3 で表が出ます」と言う矛盾に逢着する事になる。どこからこの矛盾が発生したのだろうか。

1/3 派:

1/3 派は、3 つの目覚めを等確率と見ている。そこに月曜日であるという情報が入り、(裏, 火) という 1/3 の確率であった可能性が消える。先と同じベイズ改訂により、表である確率は 1/2 である。こうして

$$1/3 \text{ 説: } 1/2 \rightarrow 1/3 \rightarrow 1/2$$

と確率が変化する。

2 段目の 1/2 は、質問 1 の時にも 1/3 派は、頻度解釈となじみが良いと述べたように、設定軸 1 において、質問 1 の時と同様に 1 万回の眠り姫実験をしたと考えればいい。約 5000 回ずつの 3 通りの実在の目覚めのうち、月曜のラベルが付いている目覚めは 1 万回、うち 5000 回でコインは表である。

Elga は自身 1/3 派であるが、この変化に対して、質問 0 から質問 1 の間になんの情報取得もないのに 1/2 → 1/3 と確率が変化して van Fraassen の原理^{39,40)}に

反している。これは追求されるべき事態であるとなると論じている。

目覚めると言うことは、設定から必ず1度は起こされるのであり、完全に予測されていたことでなんの情報でもないと一見思われる。しかし目覚めたと言うだけで、ある種の情報と考えることも出来る。時間軸を先に考えてみると、「今」は月曜日か火曜日である。今が月曜日であるとしよう。するとコインが表だったときと裏だったときと2通りの可能性がある。一方今が火曜日であるとしよう。するとコインが表だったら目覚めることはあり得ず、必ずコインは裏である。すなわち、今は月曜日である可能性が高い。これは1/3派の考えである。

同様にコインが裏か表かと言うことに対しても、目覚めたとするだけから議論出来る。頻度説を採った途端次の議論が出来る、表だと火曜日の目覚めはなく、2回の目覚めを含む裏である可能性が高い。これは将に1/3説である。

以上の話は、目覚めではなく、神による人間創造の話だったとすれば、分かりやすいだろう。私が存在すると言うからには、「今」は月曜日である可能性が高い。この場合の私は、事後の認定である。

この様に何も聞かなくても、確率が変化するのである。もちろんある設定の下では、であるが、確率を変化させるのが「情報」であるなら、何も情報が入らなくても情報は取得される。と言うよりも、

「情報は主体の中で、主体と世界の関係、すなわち位置づけによりにより発生する」

のである。情報は物理的担体によって運ばれるのではなく、主体の世界の中での位置づけによって、自ら変化するのである。この観点はこれから考究されなければならないと考えている。量子力学の観測理論でも、Reningerの否定結果実験では、対象と観測装置が相互作用せず何も起こらないことで波動関数は収縮する。

そうだったかも知れない世界について注意を喚起しておこう。実験者がコインではなくサイコロで、6通りのシナリオを用意してある場合、それらのシナリオ中に登場する参加者である姫達は、世界の設定について間主観的考察で、確率を評価せねばならず、この事も情報が担体によって保存量の様にして運ばれるのではないことが分かる。

折衷派：

Bostrom は¹⁶⁻²¹⁾、折衷派である。単一事象設定をとる。そして質問0→1→2に対して、

折衷案： 1/2 → 1/2 → 1/2

と言う変化を主張する。これは、真ん中の1/2は「1/2説」を採用し、最後の1/2では「1/3説」を採用しているのである

6. 3 囚人問題との対比

1/2 説, 1/3 説共に、月曜であるという情報が入ること、(表, 月)の確率が增大している。Bayes 改訂で他の可能性消えたので上がったと解釈したい状況であるが、心理学で有名な3囚人問題^{28, 38)}では、次のような例が知られている。

先ずオリジナル版3囚人問題を見てみよう。

「3人の囚人A, B, Cがいる。1人だけ皇帝により恩赦になって後の2人は死刑である。看守は誰が死刑か知っている。Aが看守に対して「少なくともBかCのどちらかは必ず死刑である。2人の内どちらが死刑か教えてくれない私に何も情報を提供していない事になるから、どちらが死刑か教えてくれ」。そこで看守は「Bは死刑だ」と教えた。但し看守は嘘をつかず、またBとCの両方とも死刑の時には確率1/2でどちらかの名前を答えることを前提とする。」

この問題のBayes解は次の通り。先ず事前確率は、A, B, Cそれぞれに1/3 1/3 1/3である。看守が、情報であるとも情報でもないとも付かない「Bは処刑」と伝えたとき、これは「今日は火曜ではない」と言ったことに当たるのであり質問2に相当するのであるが、Aが処刑されるBayes解は1/3である。逆に言うとAが助かる確率は2/3である。

1/3 → 2/3

ライバルBが生存競争から脱落したために、Aの助かる確率が増加したと取れそうな状況である。しかし、次の変形3囚人問題では全く状況が違う。

事前確率として1/4 1/4 1/2の恩赦確率を設定しよう。この場合のBayes解は、Aが助かる確率が1/5に減少すると教える。

1/4 → 1/5

ライバルが減ったが、減ったのは弱いライバルで、強敵が残ってしまったためである。弱いライバルでももしかしたら偶然に強いライバルを倒してくれていたかも知れないのである。ライバルの消滅が直ちに自分の助かる確率を増すのではない。

眠り姫問題では、ライバルになる目覚めは、1/3派では等確率であるためこの限りではない。しかし1/2派では、1/2 1/4 1/4の事前確率で、等確率ではない。だが、「月曜」という情報は最後の1/4を消すのであるから、

変形 3 囚人と同じではなくて $1/4$ $1/4$ という同じ条件の中の片方を消す。変形 3 囚人とはここが違う。

しかし、眠り姫問題と 3 囚人問題は、一見よく似ている。選択肢が 3 つあるのは共通である。そして (眠り姫問題の方はその一部でしかないが)、問題で問われるのは、情報が入ったときにどう確率が変化するかである。ここで情報とは、3 囚人問題の時には、「囚人 B は処刑される」という看守の言葉である。囚人 B または C のうちどちらかは (排他的 OR でなく OR として) 必ず処刑されるのであるから、新しい情報は何もないと考えていいのかわ。それとも看守が B と言った経緯を考えるのかと言う事である。A が恩赦になる場合、看守は B と言っても C と言っても良かったのである。オリジナル問題では 確率 $1/2$ で B か C を指名するとなっている (これは囚人 A にも周知されている)。

(仮に、A が恩赦になり、B と C は処刑される時、看守は確率 $1/100$ で B は処刑されるというのだったら、これは尤度であるが、そしてその条件で C と現に言ったのだったとする。この場合 Bayes 解は、A が助かる確率は $1/2$ と教える。

「もし自分が助かる場合 C と言うのは当然」という推論からも A が助かる場合と A が助からない場合は等確率と推論される、などである。等尤度でないから、どれかが実現した場合には、当然ながら情報がある.)

しかし、眠り姫問題と 3 囚人問題には本質的な大きな違いがある。それは次のことである。

眠り姫問題ではコインが裏の目覚めの中から片方を消すのに、確率的な選択をしているのではなくて、最初から「月曜 (火曜は消える)」と言う事に決まっているのである。3 囚人問題では、確率変化を問題にする発言は尤度により選択される。

3 つのドア :

3 囚人問題よりも直感的に分かりやすい「3 つのドア」を見ておこう。これは「モンティ・ホール・ジレンマ」とも呼ばれ、アメリカの TV ショウが元になっている。

「ドアが 3 つ有り、そのどれかに新車が入っている。残りの 2 つは外れである。一般性を失わずに、自分はドア A を選んでいるとする。司会者が「では関係ない 2 つを見てください。ここには入っていませんでした」とドア C を開ける。あなたが当たる確率は $1/3$ でしょうか、 $1/2$ でしょうか。B に選択を変えても良いですよ。」

2 つのドア A と B が残ったときの A の当たる確率は、Bayes 解では $1/3$ である。ここでは、ライバルが消えると、自分の成功する確率は下がっている。自分が当たる確率 $1/3$ は変わらない、といえ、オリジナル 3 囚人問題と全く同型である。

ここで、納得しにくいのは、目の前に残った 2 つのドアのどちらかに車が入っているのだが 確率 $1/2$ ではないと言うことである。しかし、3 囚人問題で学んだように、目の前の条件だけ考えて、それらに不充足理由律を適用すればいいのではない。経緯を知り、尤度も知っているならば、主観確率は $1/3$ になるべきであると言うことは納得されるであろう。

眠り姫問題と 3 つのドアを比べるということは、すなわち 3 囚人問題との比較と同じである。情報をもたらされたとき、3 つのドアでは、ある尤度でドア C が開けられる。眠り姫問題では、月曜であるという情報をもたらされて、(裏, 月) と (裏, 火) のうち消えるのは、(裏, 火) の可能性である。

7. 極端な設定での考察

次に、極端な眠り姫実験を取り上げよう。図 3 の様に、コインが裏の場合に、1 億日間起こすという設定である。本質的違いはないのではあるが、こうすると、コインが表である確率の質問 1 から質問 2 への変化は、

1/2 説 : $1/2$ → ほぼ 1
1/3 説 : ほぼゼロ → $1/2$

となる。

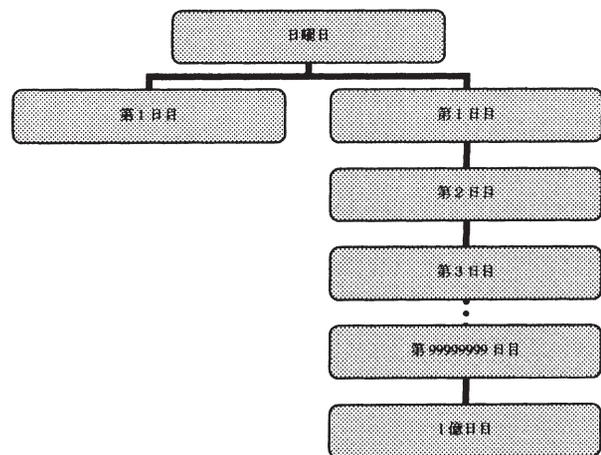


図 3 極端な眠り姫実験

仮にコインを第 1 日目以後で投げて同じであるとする、 $1/2$ 説では、第 1 日目であると聞いたとき、運命を決めるべきフェアなコインをこれから投げるとほぼ確実に裏が出ると予測される事になる¹⁷⁾。こうなると受け入れにくさが際だつてあろう。そのコインが自分の運命を決めるコインであると、ほぼ確実に裏が出て、隣にある、物理的には全く同じコ

インは裏の確率が $1/2$ であるというのも異様である。

しかしこの異様な状態は無前提に起こるのではない。第1日目と判明した時点であって、その前の時点では、1億1回の目覚めのうち眠り姫はどの目覚めなのか判断出来ないのであるから ($1/2$ に極僅か足りない確率で、コインは既に投げられている) この事態は起こっていない。

この事態を避けるために Bostrom¹⁶⁾ は、質問2の場合には(選択的に) $1/3$ 説を採用したのである。また一方質問1に対応する場合の困難も指摘して、その場合は $1/2$ 説を採用するという、折衷説を採用した。

しかしここで、「フェアなコイン」という設定に着目してみよう。問題設定より、眠り姫はフェアなコインと知らされるのは、日曜日である。ここで意味されるのは、「偏っているとは言えない」という意味であろう。例えば物理的に対称であって、とか実験的或いは統計的に保障されていて等と言うことでなくても、偏っているという根拠がないと言う事で充分である。つまり、実際には偏っているかも知れないが、そのことに関する情報がないと言う事である。

すると、この情報は事態の進行に従って Bayes 改訂されるべきものである。フェアなのは日曜日のみのことである。目覚めという事態に進行したときには、特に月曜日(第1日目)と知らされたときには改訂されなくてはならない。第1日目と知らされたとき、目覚めに自己を同定したとすると、表は奇跡である。一方裏なら1億の目覚めの一つに過ぎない。とすると Copernicus の原理により奇跡を避けるなら、裏が出やすいコインだったと見込みが変わる。

そうするとパラドキシカルな状況はこう解釈される。裏が出やすい(Bayes 改訂によれば、ほぼ確実に裏が出ると言うことになる) コインを投げて、裏がでそうというのは当然のことである。

設定軸2 の、コインを目覚めの先に投げるのか、後投げなのかは、確率の収縮問題²⁵⁾ だけではなく、自己位置づけ問題でも、やはり定説に反し大きな違いがあるようである³¹⁾。

コインがフェアと言うことは Bayes 改訂されるべきである、と言う視点をとる以前に戻ろう。あくまでどの時点でもフェアなコインと考えるべきであるとしよう。 $1/2$ 説は後投げ設定では困難に陥った。しかし $1/3$ 説は妥当する。先投げ設定では, Elga や Lewis の議論で充分である。単一事象設定と切り返し設定を区別しながらも混在させた謝った証明だったと言うだけで、彼らの議論は単一か、繰り返ししかを区別すれば正しい。

ここで、設定軸1の、単一事象設定か、繰り返し実験設定での頻度解釈かという軸と組み合わせて、クロス表的に整理すると表1のようになる。マスに書き込まれた説が、その設定の組み合わせで正しい説である。と言うよりこの表は逆に読むべきで、それぞれの説が主張する確率値が、どの設定の組み合わせにおいて妥当であるかと読まねばならない。

表1

設定軸1 \ 設定軸2	先投げ 図1	後投げ 図2
単一事象	$1/2$ 説	$1/3$ 説
繰り返し設定	$1/3$ 説	$1/3$ 説

先投げ後投げで事態が全く違うのであるなら、論者の確率値の導出の都合で設定軸2に関してぶれることは許されない。オリジナルの眠り姫問題は、先投げ設定なのである。また、問題文から単一事象設定であるから、 $1/2$ 説が正しい、というのが一応の正解であろう。しかし、眠り姫問題は、オリジナル設定での正解を求めるのが目的では決していない。局所実在論での EPR パラドックスのようにそこから議論が発展し、思いもよらない種々の根本的知見が得られるのが価値なのである。

8. 結語

Leslie¹⁵⁾ の提出する例を見てみよう。ある特定の3人称的に特定される(一人称としての考察に移行していくのではあるが) 人間に位置づける問題と、空間的な位置づけ問題である。

・ブラック博士とグリーン博士:

神がコイン投げで人類を創造する。表なら1人で裏なら9000万人である(表の方はもっと多くしなくてはならない。なぜなら自分以外に人がいるかどうかで確定的な答えが分かってしまうから) 表の時の人はグリーン博士、裏の時の9000万人のうちの一がブラック博士である。「あなたはブラック博士かグリーン博士である」という情報もたらされた。あなたがグリーン博士である見込みは。

全く極端な眠り姫問題と同型である。ブラックとグリーンは眠り姫の時の月曜(第1日目)という特定に当たる。Leslieはグリーンである可能性が高いという。

・ロンドンとリトルパドル:

あなたは窓のない部屋で記憶喪失になった。何所にいるのだろう、人口50人のリトルパドルだろうか、

それとも人口 1000 万人のロンドンであろうか、自分はロンドンにいると思うべきであろう。

この様に自己位置づけは、空間、時間、人格、目覚めなどに関して等多岐にわたる。眠り姫問題に限れば、一応、表 1 が解答であろう。しかし、何とかけりをつけるという段階で、設定軸 1～4 をどう扱うか、設定軸の間の関係はどうなっているのかなどは、それぞれは古い問題であると共に、将に今これから始まった問題である。その中では、目覚めのような「状況」ではない、人間に位置づける場合の方がわかりやすいであろう。ただし「わたくし」とはどういうことかと言う問題を抜きにすればではあるが。

そうであったかも知れない世界、私であったかも知れない人などが準拠集団になるのである。しかしそれで「わたくし」言えるか、感じられるかである。

やり始めなくてはならない問題は、眠り姫問題に限定しても沢山ある。

- (1) 情報とは何か。情報は、世界自体の変化/変化しないことにより、内部的に発生しうる。
- (2) 定説に反し、先投げと後投げは全く違うこと。特に量子眠り姫問題では顕わな大問題であるが、古典確率であっては、ここにこそ相違の重大性が認識される。
- (3) 主観確率と客観確率の関係と単一事象と繰り返し設定の関係が並行的であるか
- (4) 主体が特定されるとはどういう事か
- (5) 時間軸上の位置づけでは、未だ起こっていない未来の主体と、過去現在の主体を同列に扱って良いのか
- (6) 量子確率が、頻度確率とも主観確率とも違う本質的非決定であること。傾向性解釈の再検討。

等について、精力的に考察していく所存である。

最後に、ほしくずスターダスト⁴³⁾のアイデアによる、1つの同型問題を見ておこう。

酔っぱらいの郊外電車帰宅⁴³⁾：

あなたは酔っぱらってターミナル駅から自宅のある A 駅に帰ろうと電車に乗る。ターミナル駅では、電車発車前にコインを投げ、表なら列車は A 駅止まり。裏なら A 駅の先の B 駅まで行く。あなたは酔っぱらっているので A 駅と B 駅の区別は付かない。各駅の前には橋がありかなり揺れるので目が覚める。各駅では「コインは表だったと思いますか(列車は A 駅止まりでしょうか)」と聞かれる。しかし(もし B 駅止まりとしても)駅を出ると眠ってしまって、ターミナル駅でのことしか覚えていない。

A 駅であなたは、先の質問をされた。次に「ここは A

駅です。電車は、この先の B 駅まで行くと思いますか」と聞かれた。

眠り姫問題と同型であるが、話は現実的であろう。

後投げ設定に対応するのは、「電車の行く先はターミナル駅を出る時点では決まっていない。A 駅であなたは起き、質問 1, 2 をされて、その後コインが投げられ、電車がそこ止まりなのか、更に先まで行くのか決定される。」となる。A 駅で再度眠った後、A 駅で起きた記憶は無くなり何度目の目覚めか分からないと言うのも、酔っぱらいならば自然である。

この話に、前節までの考察を翻訳してみれば、納得がいくのではないだろうか。

謝辞：1年半ほど前、「情報」とは何かに乗って運ばれてくるものではない」という事について、注意を喚起してもらいました、筑波大学の緑川信之教授に感謝します。自己位置づけ問題を考えている中で、役に立たないと言われるこの様な問題が、現実の問題に分かりやすい形で役立つという喜びを味わえました。

参考文献

- 1) A. Elga, "Self-locating Belief and the Sleeping Beauty Problem", *Analysis* 60(2000) 143
- 2) A. Elga, "Defeating Dr. Evil with Self-Locating Belief", *Philosophy and Phenomenological Research*, 69 (2004) 383
- 3) R. Weintraub, "Sleeping Beauty: a simple solution", *Analysis* 64 (2004) 8
- 4) T. Hogan, "Sleeping Beauty awakened: new odds at the dawn of the new day", *Analysis* 64 (2004) 10-21
- 5) B. Monton, "Sleeping Beauty and the forgetful Bayesian", *Analysis* 62 (2002) 47
- 6) F. Arntzenius, "Reflections on Sleeping Beauty", *Analysis* 62 (2002) 53
- 7) C. Dorr, "Sleeping Beauty: in defence of Elga", *Analysis* 62 (2002) 292
- 8) D. Lewis, "Sleeping Beauty: reply to Elga", *Analysis* 61(2001)171
- 9) D. Lewis, "Human Supervenience Debugged", *Mind* 103 (1994) 473
- 10) D. Lewis, "A Subjectivist Guide to Objective Chance", *Studies in Inductive Logic and Probability*, (1980) Univ. California
- 11) C. Hitchcock, "Beauty and the Bets", *Synthese* 139 (2004) 405
- 12) J. Leslie, "Observer-relative Chances and the Doomsday Argument", *Inquiry*, 40,(1997),427-36

- 1 3) J. Leslie, *Universe*, Routledge (1989)
- 1 4) J. Leslie: *Infinite Minds - A Philosophical Cosmology*, Oxford (2001)
- 1 5) J. Leslie, 『世界の終焉』, 青土社 (1998 年)
原書: *End of the World: The Science and Ethics of Human Extinction*, Routledge (1996)
- 1 6) N. Bostrom, "Sleeping Beauty and Self-Location: A Hybrid Model", *Synthese* 157(2007)59-78
- 1 7) N. Bostrom, "The Mysteries of Self-Locating Belief and Anthropic Reasoning", *The Harvard Review of Philosophy* 11(2003)59
- 1 8) N. Bostrom, "Self-Locating Belief in Big Worlds: Cosmology's Missing Link to Observation", *The Journal of Philosophy* 99(2002)607*-
- 1 9) N. Bostrom, *Anthropic Bias: Observation Selection Effects in Science and Philosophy*, Routledge (2002)
- 2 0) N. Bostrom, "The doomsday Argument, Adam & Eve, UN++, and Quantum Joe", *Synthese* 127(2001)359
- 2 1) P.J. Lewis, "Credence and Self-Location" (2009)
http://philsci-archive.pitt.edu/view/subject/decision_theory
- 2 2) P.J. Lewis, "Probability, Self-Location and Quantum Branching", (2008)
<http://philsci-archive.pitt.edu/archive/00004309/>
- 2 3) L. Delabre, "Sleeping Beauty :Debate on a Paradox", (2008) http://philsci-archive.pitt.edu/view/subject/decision_theory
- 2 4) 榛葉豊, 「平凡の原理と主観確率 —多世界・意識・参照集団—」, 静岡理科大学紀要 16 卷 (2008 年) 37
- 2 5) 榛葉豊, 「遅延選択と確率的遡及因果 —確率はどの段階で崩壊するか—」, 静岡理科大学紀要 15 卷, (2007 年) 47
- 2 6) 榛葉豊, 「多世界論と「私」の謎」, 静岡理科大学紀要 15 卷 (2007 年) 77
- 2 7) 榛葉豊, 「統計的確率に対する信念の度合い —確率的言明の反証可能性—」, 静岡理科大学紀要 14 卷 (2006 年) 73
- 2 8) 榛葉豊, 「信念の度合いと不充足理由律 —3 囚人問題の Bayes 解をめぐる—」, 静岡理科大学紀要 13 卷 (2005 年) 53
- 2 9) 榛葉豊, 「人間中心原理と確率の本性 —確率過程量子化の解釈のために—」, 静岡理科大学紀要 12 卷 (2004 年) 177
- 3 0) 三浦俊彦, 「人間原理的推論における「この宇宙」
『科学基礎論研究』 103 (2005 年)
『ゼロからの論証』, 青土社 (2006 年) に収録
- 3 1) 三浦俊彦, 「人間原理のパラドクス 1 ~ 1 2」,
『現代思想』2006 年 11 月~2007 年 10 月 (青土社)
『多宇宙と輪廻転生 —人間原理のパラドクス』,
青土社 (2007 年) に収録
- 3 2) 三浦俊彦, 「人間原理と独我論 —「私」の位置の再解釈」, 『和洋女子大学紀要』 第 40 号 (2000 年) 17
- 3 3) 三浦俊彦, 「観測選択効果と多宇宙説 —伊東邦武『偶然の宇宙』について」, 『科学哲学』36-1 (2003 年) 121
- 3 4) 三浦俊彦, 『論理学入門』, NHK 出版 (2000 年)
- 3 5) 三浦俊彦, 『可能世界の哲学』, NHK 出版 (1997 年)
- 3 6) 三浦俊彦, 『論理パラドクス』二見書房 (2002 年)
『論理サバイバル』二見書房 (2003 年)
- 3 7) D. Gillice, 『確率の哲学理論』, 日本経済評論 (2004 年) 原書 *Philosophical Theories of Probability* (2000)
- 3 8) 市川伸一, 『確率の理解を探る』共立出版 (1998 年)
- 3 9) B.C. van Fraassen. "Belief and will", *Journal of Philosophy* 81(1984)235
- 4 0) B.C. van Fraassen. "Belief and the problem of Ulysses and the Sirens", *Philosophical Studies* 77(1995)7
- 4 1) B.C. van Fraassen, 『科学的世界像』, 紀伊國屋書店, (1986 年),
原書: "*The Scientific Image*", Oxford(1980)
- 4 2) 緑川信之, 『「情報」概念の再考』, *Library and Information Science*, 56(2006)23
- 4 3) hoshikuzu | star_dust,
<http://d.hatena.ne.jp/hoshikuzu/20080509>