

博士論文

自己制御と情動

西堤 優

目次

はじめに	1
第一章 二重システム理論と自己制御	6
1.1 二重システム理論とは何か	6
1.1.1 心理学における二重システム理論	7
1.1.2 二重システムの神経基礎	9
1.2 二重システム理論に対する批判とその検討	10
1.3 二重システム理論における自己制御	19
第二章 自己制御に情動はどう関係するか	22
2.1 情動とは何か	22
2.1.1 情動とは	22
2.1.2 ジェームズ＝ランゲ説	23
2.1.3 認知説	24
2.1.3.1 認知説とは	24
2.1.3.2 認知説の問題点—情動における認知的侵入不可能性—	27
2.1.4 Damasio や Prinz の説	28
2.2 自己制御における情動の必要性	29
第三章 ソマティック・マーカー仮説について	33
3.1 アイオワ・ギャンブル課題	33
3.1.1 実験目的と実験デザイン	34
3.1.2 実験結果	34
3.1.3 Damasio の考察	35
3.2 問題点	35
3.2.1 Colombetti の批判	36
3.2.2 情動による行動の説明はどこまで可能か	37
3.2.2.1 ポジティブな情動による説明	37
3.2.2.2 目先の報酬への傾向性による説明	38
3.2.3 事前の SCRs と事後の SCRs	39
3.2.4 「長期的な利害の考慮」の多義性	40
3.3 まとめ	42

第四章 知的評価と情動	43
4.1 身体状態・脳状態・意思決定	43
4.2 あたかも身体的情動と身体的情動	45
4.3 情動と知的評価	48
4.4 意思決定への影響の違い—身体状態の独自の役割—	50
4.4.1 身体的情動と行動の準備	51
4.4.2 あたかも身体的情動と紋切り型	52
4.4.3 知的価値評価と動機付けの欠如	54
4.5 まとめ	56
第五章 情動の合理性	57
5.1 情動と知的評価はどう関係するか	57
5.2 情動について	58
5.2.1 情動の焦点の重要性	60
5.2.2 情動の一貫性について	61
5.3 欲求について	62
5.3.1 欲求と目標志向的状态の区別	63
5.3.2 欲求と情動の合理的相互結合	65
5.4 判断について	65
第六章 自己制御の喪失のあり方	69
6.1 意思決定過程における自己制御とその喪失	69
6.2 執行過程における自己制御とその喪失	72
6.2.1 システム1の過活動	73
6.2.2 システム2による事後的正当化	74
6.2.3 自己資源の消耗	74
6.2.4 時間割引	76
第七章 信念形成における自己制御とその喪失	78
7.1 はじめに	78
7.2 認知的アクラシアとは何か	78
7.3 認知的アクラシア不可能論	81
7.4 認知的アクラシア可能論	82
7.5 欲求の影響	85
7.6 信念の本性	87
7.7 信念形成における自己制御の喪失	88

第八章 意志についての二つのモデル	91
8.1 価値の双曲的時間割引と意志の弱さ	91
8.2 意志のモデル	96
8.2.1 意志の弱さの非限定性：器官理論	97
8.2.2 Ainslie の器官理論批判	98
8.2.3 異時点間交渉モデルによる意志の弱さの限定性	98
8.2.3.1 異時点間交渉モデルとは何か	99
8.2.3.2 意志の弱さの限定性	102
8.2.4 異時点間交渉モデルと器官理論の調停	102
第九章 意志力の枯渇	104
9.1 自己制御（意志力の行使）の神経基盤	105
9.2 リソースモデル	106
9.3 プロセスモデル	109
9.4 どちらのモデルが正しいか	111
第十章 責任の二つの面	114
10.1 責任の二つの側面とその関係	115
10.2 正当化と弁解	116
10.3 なぜ弁解は可能か	117
10.4 まとめ	119
むすび	121
参考文献	125

はじめに

日常生活の中で、私たちは自己制御が必要な場面なのに、それがうまくできないことがよくある。たとえば、今月の出費は控えるべきだと思っているにもかかわらず、旬の食べ物を見かけると今しか食べられないからと買ってしまったり、翌週月曜日から朝5時に起きてジョギングをすると決意したとしても、実際に起きてみたら眠くてついまた寝てしまったりする。これらはすべて、あることをするのが最善だと判断しているにもかかわらず、その通りに実行できなかった諸事例である。このような自分で自分を制御できないという事態は、現代の私たちにとってあまりにも身近なことであると同時に、古くから「意志の弱さ」として論じられてきた問題でもある。もし、誰もが皆、特別な事情がない限り、常に最善の判断（何をするのが最善かの判断）に従って行為することができるとしたなら、つまり常に自分で自分を制御できるとしたなら（もしくは現段階で完全に自己制御できるための術が判明しているのなら）、このようなことは話題にすらされないに違いない。だが、楽しそうな誘いの誘惑や食べたいという衝動、または眠りたいという欲求に抗うのは、非常に困難なことなのである。

自己制御とは最善の判断に従って、実際に行為するように自分を律することである。とくに最善の判断に反する欲求¹や衝動がある場合に、そのような欲求や衝動に屈せず、最善の判断に従って行為するようにすることである。従来、このような自己制御は理性と情動の対立として捉えられてきた。欲求や衝動は情動と結び付けられ、それが合理的な選択・行動を可能にする理性的な働きをかき乱すと考えられ、それゆえ自己制御のためには情動をできるだけ排除し、理性に従わねばならないとされたのである。もしそうであるなら、私たちが情動的な部分を持たないとしたら、それが理性を邪魔することもないだろうから、私たちは理性だけに基づいて、あらゆることを最善の判断通りに実行することができるようになるように思われる。そのような人は理想的な合理的存在者であり、意志の弱さとは無縁であるように思われるのである。

だが、昨今の情動に関する科学的成果によって明らかになったことは、私たちが理性的に振る舞うためには、ある程度の情動が必要不可欠だということである。情動なしでは私たちは理性的に振る舞うことができない。極端な場合には、自らの生存さえも脅かすような非常に不合理な行動をしてしまうという。このことが示していることは、私たちにとって自己制御の喪失と呼ばれる状態は、情動によってしばしば引き起こされるが、一方で、情動が欠如することによってもときに引き起こされるということである。このような事態を私たちはどのように理解したらいいのであろうか。

¹ 本論文で用いる「欲求」は、Davidson のような広い意味 (Davidson 1980, pp. 3-4, 27, 邦訳, pp. 3, 37-8)、すなわち美的判断や経済的な価値判断などを含む賛成的態度 (pro attitude) のような意味ではなく、通常の狭い意味で用いる。すなわち知的な評価を含まないような欲望、衝動、誘惑などを意味するものとする。

この論文の主題は、自己制御とは何かを特に情動の役割に注目しながら明らかにすることである。情動が自己制御を妨げることからすれば、自己制御には、情動が邪魔であるように見えるが、実はそうではなく、適切な情動が必要なのである。もし情動が過剰であったり不足していたりすると、私たちは為すべきことを為すのに意志の力に頼らなければならなくなる。しかし、意志の力は消耗しやすいので、意志の力に頼るだけでは、自己制御の喪失に陥りやすい。概ね自己制御ができていくというようにするためには、概ね適切な情動を持つことが必要なのである。本論文ではこのようなことを明らかにしていきたい。前半では適切な情動によって自己制御が為される場合の自己制御を考察し、後半では適切な情動がないために意志の力によって自己制御がなされる場合の自己制御を考察する。そして、意志の力が消耗しやすいために、結局、自己制御には適切な情動が必要であることを明らかにしたい。ここで本論文の概要を章ごとに簡単に述べておこう。

まず第一章では、私たちが誘惑に抗して自己制御を行おうとするときに感じる二人の自分（制御する自分と制御される自分）について、それを理論的に捉えた二重システム理論を詳細に検討する。二重システム理論は心理学的なモデルとして提唱された認知に関する理論であり、私たちの認知はシステム1とシステム2という二つの異なるシステムによって行われるとする。システム1は情動的なシステムであり、システム2は理性的なシステムである。昨今ではこの二つのシステムに対応する二つの神経システムが見出されたと主張する神経科学者もあり、二重システム理論は神経科学的に基礎付けることが可能であるように思われる。本章では、とくに二重システム理論を脅かす恐れのあるいくつかの重要な批判を取り上げて検討し、二重システム理論を維持することが可能であることを確認する。そしてその上で、二重システム理論のもとでは、自己制御はシステム2がシステム1を制御することとして捉えられることを論じる。

続く第二章では、自己制御における情動の必要性について論じる。そのためにまず、そもそも情動とは何かについて、情動論の歴史的な背景を踏まえながら、情動の本質が情動に含まれる身体状態にあるとする説がもっとも説得的であることを確かめる。A. R. Damasio による「ソマティック・マーカー仮説」(SMH) はそのような説の代表的なものである。この仮説によれば、情動に含まれる身体状態は対象の価値に応じて誘発される状態であり、それゆえ対象の価値を反映している。そしてそのような身体状態が、私たちの意思決定場面において、常に無自覚的に影響を与えている。このようなことから、情動が自己制御に不可欠であることを見ていく。

第三章では、前章で取り上げた Damasio の SMH についてさらに詳しく掘り下げて考察する。特に SMH の根拠として示されている実験「アイオワ・ギャンブル課題」に注目し、SMH がその実験結果によってどの程度確証されているのか、さらに Damasio によるその実験結果の解釈について問題がないのかについて吟味を行う。それを通じて、SMH における情動概念を洗練し、SMH の内容そのものの明確化を図る。こうして自

己制御における情動の必要性をさらに明確化していく。

第四章では、意思決定における情動の役割について、その最も本質的な要素と考えられる身体状態が実際にどのような役割を果たすのかについて詳細に考察する。そのためにまず、情動の本質は身体状態にあると考えられるが、それにもかかわらず、Damasio は身体状態を含まない情動、すなわち「あたかも身体的情動」を認めているという点に注目する。そして身体状態を実際に含む「身体的情動」とそうでない「あたかも身体的情動」について、それらがそれぞれどのように外界の価値を表すのかについて考察する。さらに、身体的情動であれ、あたかも身体的情動であれ、ともかく情動とは異なる純粋に知的な価値評価、つまり全く身体性を帯びない価値評価が存在することを確認し、そのような知的価値評価が意思決定においてどのような役割を果たすのかについても考察する。その上で、これら三つの価値評価、すなわち身体的情動、あたかも身体的情動、知的価値評価がそれぞれ意思決定にどのような影響を及ぼすのかを比較し、それを通じて身体的情動がその身体性ゆえに価値評価と動機付けの点で意思決定に対して独自の影響を及ぼすことを論じる。このような考察により、情動の身体性が自己制御において独自の重要な役割を果たすことを見ていく。

第五章では、前章でその違いを強調した情動的評価と知的評価について、今度はそれらがどのように関係し合うのかを考察する。そのために、情動に関する Damasio や J. Prinz らの新ジェームズ主義的見解を批判し、その不備を補おうとする B. W. Helm の情動論を援用することにする。ただし、援用するといっても、彼の情動論はかなり独特なものであり、かつ、詳細な検討に値するものであるので、ここでは Helm が情動をどのように捉えているのか、そして情動が欲求とどのような関係にあり、また情動や欲求が価値判断（つまり知的評価）とどのような関係にあると彼が捉えているのかについて詳しく見ていく。それを通じて情動的評価と知的評価の関係を明らかにするとともに、情動にそれなりの合理性を認めることができることを明らかにしていく。そしてこの観点からも、情動が自己制御にとって重要であることを確認する。

第六章からは、自己制御の喪失（意志の弱さ）へと焦点を移し、その考察を通じて、結局、適切な情動が自己制御においていかに重要であることを示すことを試みる。そのためにまず第六章では、自己制御が失われた状態、すなわち自己制御の喪失のあり方に注目し、それを考察する。特に二重システム理論の枠組みの中でその考察を試みる。自己制御とは何を為すべきかについての最善の判断（知的評価に基づく判断）に従って意思決定を行い、その意思通りの行為を行うことである。これを二重システム理論の枠組みで言えば、システム2が最善の判断に従って意思決定を行い、その決定がシステム1の決定（情動に基づく決定）と一致するときは、そのままその行為を行うが、食い違ふときは、システム2がシステム1を制御して、システム2の決定した行為を実行することである。自己制御の喪失は、システム1とシステム2の決定が食い違ふときの自己制御に関して生じる。この種の自己制御が失われる場合、その喪失が一連のプロセス

のどの段階で生じるかによって大きく分けて二つのタイプがありうる。すなわち、システム2が最善の判断に従って意思決定を行おうとする段階で生じる自己制御の喪失（最善の判断とは異なる意思決定を行ってしまうこと）と、システム2が最善の判断に従って意思決定したにも関わらず、その意思通りの行為を行おうとする段階で生じる自己制御の喪失（決定した意思とは異なる行為をしてしまうこと）である。自己制御の喪失にはこのような二つのタイプがあるということは、意志の弱さにも二つのタイプがあることを示している。このような自己制御の喪失のあり方を明らかにすることにより、システム1とシステム2の決定が食い違うときの自己制御（情動が適切でないときの自己制御）がどのようなものであるかを逆照射する。

第七章では、これまで見てきた意思決定や行為についての自己制御とその喪失に関する考察からいったん離れて、信念形成における自己制御とその喪失を見ていく。信念形成における意志の弱さに相当する事態は「認知的アクラシア」と呼ばれるが、認知的アクラシアが可能かどうかについては両論がある。それらを吟味しつつ、認知的アクラシアが不可能であることを論じる。しかし、認知的アクラシアが不可能だとしても、信念形成においても自己制御の喪失と呼びうるような事態は確かに存在し、そこでは欲求や情動の影響が鍵となることを明らかにする。

第八章では、自己制御ができないことは意志が弱いことであり、逆に自己制御ができるということは意志が強いことだと言われるときの「意志」について考察することにより、システム1とシステム2の決定が食い違うときの自己制御の実質にさらに迫る。意志とは何かについて、これまで様々なモデルが提示されてきたが、ここではそのうちの二つのもっとも有力だと思われるモデルを取り上げて検討する。一つは意志とは器官のようなものであると主張する「器官理論 (Organ Theory)」であり、もう一つは意志とは異なる時点での自己の交渉状態であると主張する「異時点間交渉モデル (Intertemporal Bargaining Theory)」である。この二つのモデルを比較検討することにより、どちらが意志のあり方を説明するモデルとしてより相応しいのかを考察し、それらが両立しうることを論じる。

第九章では、前章で確認した意志とは何かを受けて、ここではシステム1とシステム2の決定が食い違うときに自己制御を行うために意志が発揮する力、すなわち意志力とはそもそもどのようなものであるのかを考察し、それに基づいて自己制御の喪失（意志の弱さ）がどのようにして起こるのかを明らかにする。そのためにまず、自己制御およびその喪失に関与する脳部位を探究した科学的成果を参照し、それによって意志力には確固たる脳基盤がありそうであることを確認する。続いて、意志力とは「自己資源 (ego resource)」であり、自己制御の喪失（意志の弱さ）は自己資源の消耗 (ego depletion) によって起こるという「リソースモデル (Resource Model)」と、そもそも自己資源というような何か実体的な資源があるわけではなく、ある種の認知的な過程によって自己制御の喪失が起こるという「プロセスモデル (Process Model)」を比較検討し、それに

より意志力の実態を明らかにすることを試みる。

最後の第十章では、私たちの現実の実践において情動が果たす役割と意義を責任という場面に即して考察する。責任には悪い行為の帰属がなされるという意味での「帰属責任」と悪い行為に対する非難がなされるという意味での「非難責任」の二つの面がある。帰属責任が成立するとふつう非難責任が成立するが、弁解により非難責任が免除されることがありうる。弁解の場合、正当化と違って、悪い行為をなしたことの帰属責任を免れないが、それでも同情という情動の働きにより、非難責任は免れうるということを論じる。但し、非難責任を免れうるといっても、それは現実的な非難責任であり、原理的な非難責任は免れえないことを明らかにする。そして、二重システム理論の枠組みからは、帰属責任と原理的な非難責任はシステム2の働きによるものであり、現実的な非難責任はシステム1の働きによるものであることを示す。

以上のように、本論文では情動の役割に注目しながら、自己制御とは何かを明らかにすることを試みる。

第一章 二重システム理論と自己制御

どうすれば誘惑に耐えることができるのかは、私たちの日常生活において重要なことである。誘惑に負けると、多かれ少なかれ自分の利益を損なうようなことを行ってしまふ。たとえば、今書かなければならないメールがあるにもかかわらず、ついつい面白そうなネットニュースを次から次へと読んでしまい、メールを書く時間が無くなってしまふ。また、ダイエットをしているにもかかわらず、目の前のドーナツを食べてしまい、痩せるどころか太ってしまう。なぜ私たちは誘惑に晒されると、しばしばそれに負けてしまうのだろうか。そもそも誘惑とは何なのだろうか。また、誘惑に打ち克つということ、すなわち自己制御とは一体どのようなことなのだろうか。

誘惑に抗して自己制御を行おうとするとき、いわば二つの自分がある。誘惑する自分とそれに打ち克とうとする自分である。先ほどのダイエットの例でいえば、ダイエットするよりもドーナツを食べたいという自分と、その欲求に打ち克ってダイエットをしようとする自分である。ダイエットしようとする自分がドーナツを食べたいという自分に打ち克つことができれば自己制御に成功し、そうでなければ自己制御に失敗して誘惑に負けたことになる。心理学や神経科学において提唱された二重システム理論は、この二つの自分という考えを洗練したものと考えることができる。この理論は、私たちの意思決定 (decision-making) や行動を二つの認知システムないし神経システムの産物として捉えようとする理論である。本章では、二重システム理論をかなり詳しく検討するとともに、そのもとで自己制御がどのように捉えられるかを考察しよう。

1.1 二重システム理論とは何か

私たちの行為を生み出すのに、二つの認知システムが関わるという考えは、哲学においても古くから見られる。理性と感情の対立ないし葛藤という見方は、そのような考えに基づくと見えよう。心理学では特に 20 世紀後半になって、行為を生み出す二つのシステムという考えが盛んになってきた。さらに 21 世紀になると神経科学において、二つの認知システムの神経基盤となるような二つの神経システムがあるということが言われるようになってきた。本節では、これまで心理学における二重システム理論²がどのようにして誕生してきたのかを概観しつつ、それがどのような理論かを見る。それとともに、二つの認知システムが脳において確かに神経基盤を持つことを示した神経科学の二重システム理論を概観する。

² 心理学では、「二重システム理論 (Dual System Theory)」のほかに「二重プロセス理論 (Dual Process Theory)」や「二重タイプ理論 (Dual Type Theory)」など、いくつかの異なる名称がある (Evans and Stanovich 2013) が、ここでは「二重システム理論」という名称を採用する。

1.1.1 心理学における二重システム理論

心理学における二重システム理論に直接かかわる成果が出てきたのは、1960年代から1970年代にかけて行われた学習・記憶、演繹的推論、社会認知、意思決定の各領域における実験研究によってである (Evans and Frankish 2009, pp. 10-15)。これらの研究から二重システム理論の萌芽となるような考えが、それぞれの領域でどのように出てきたのかを見ていこう。年代順に学習・記憶、演繹的推論、社会認知、意思決定の順番で見ていくことにする。

まず、学習・記憶についてである。1960年代、心理学者 A. S. Reber は、意識にのぼらない暗黙的学習 (implicit learning) があるということに注目し、それを実験によって検証した。Reber は人工的な言語を用いて、その言語の文法規則が暗黙的学習によって獲得されるかどうかを調べた (Reber 1967)。この人工言語は五つの単語 (P, S, T, V, X) とそれらを組み合わせる文法規則から成る簡単な言語である (この言語では一文字が一単語となっている)。被験者らは文法規則に従って構成された単語列 (文) を提示される群と、そうではないランダムな単語列を提示される群に分けられ、次々と単語列を提示された。その結果、文法規則に適った単語列を提示された人々は、やがて文法規則に適った単語列とそうでない単語列を直観的に選別できるようになった。つまり、彼らは文法規則を明示的に理解することなく、それを習得したのである。こうして、明示的な学習・記憶のプロセスとは異なる、暗黙的な学習・記憶プロセスがあるということが明らかにされた。この成果はこの後、二重システム理論を発展・展開させる重要な成果となった。

次に演繹的推論の領域を見ていこう。ここでは1970年代の J. St. B. T. Evans の研究が二重システムの考えの最も重要な起源として挙げられる。Evans らは被験者が様々な演繹的推論の妥当性を判断する際に、被験者の心の中で論理的な過程と信念バイアスに影響される非論理的な過程という二つの過程が競合しているということを発見した (Evans et al. 1983)。この実験では妥当である推論で結論が信じやすい三段論法と、妥当であるにもかかわらず結論が信じがたい三段論法が用いられており、被験者には与えられた推論が妥当かどうかの判断が求められた。但しその際、被験者が「妥当」という言葉の正確な意味を知らないかもしれないので、その言葉を用いずに、前提から結論が「論理的に導き出される」かどうかという表現を用いて、それに答えてもらうという仕方で妥当性の判断が求められた。しかも前提が実際に真だと仮定した上でこの問いに答えるようにとの指示が出された (Evans et al. 1983, p. 298)。この実験の結果、一つには、私たちは信念バイアスによって、妥当である推論で結論が信じやすい三段論法を妥当でないと判断する傾向よりも、妥当であるが結論が信じ難い三段論法の方を妥当でないと判断する傾向の方が強いということが示された。たとえば、「どんなタバコも安価ではない。一部の依存性のあるものは安価である。それゆえ、一部の依存性のあるものはタバコではない」という推論は妥当であり信じやすい結論を持ち、また「どんな依

存性のあるものも安価ではない。一部のタバコは安価である。それゆえ、一部のタバコは依存性がない」という推論は妥当であるが信じ難い結論を持つ。このとき、前者の推論を妥当でないと判断する人よりも、後者の推論を妥当でないと判断する人の方が多かったのである。

もう一つの結果として、私たちは、妥当でない推論で信じやすい結論を持つ三段論法を妥当であると判断する傾向の方が、妥当ではない推論で信じ難い結論をもつ三段論法を妥当であると判断する傾向よりも強いということが示された。たとえば、「どんな依存性のあるものも安価でない。一部のタバコは安価である。それゆえ、一部の依存性のあるものはタバコではない。」という推論は妥当でないが信じやすい結論を持ち、また「いかなるタバコも安価ではない。一部の依存性のあるものは安価である。それゆえ、一部のタバコは依存性がない」という推論は妥当でなく信じ難い結論を持つ。このとき、前者の推論を妥当であると判断する人の方が、後者の推論を妥当であると判断する人より多かったのである。

要するに、推論は妥当であるが結論が事実に反する場合、私たちは信念バイアスにより推論を妥当でないと判断してしまうことがある。また、推論は妥当でないが結論が事実に一致する場合、私たちは信念バイアスによって推論を妥当だと判断してしまうことがある。このように、推論の妥当性を判断するとき、論理的な過程と信念バイアスに影響される非論理的な過程という二つの過程が競合していることが伺われる。この種の研究は、推論における分析的な過程とヒューリスティックな過程という二重の過程を措定する理論へと発展していった (Evans and Frankish 2009, p. 13)。

第三に、社会認知の領域において二重システムの考えがどのように芽生えてきたのかを見てみよう。社会心理学では、明示的に述べられた態度と実際の社会的行動の間に乖離があるという事実が 1960 年代から 1970 年代に確立された (Evans and Frankish 2009, ch. 9)。たとえば、ある人は自分はいつも平等に人に接するとはっきり述べながら、実際の振舞いとしてはある種の人たちを避けるような差別的な行動をとる。このような事実をどのように説明するかということが社会心理学において重要な問題となっていたが、認知心理学における発展が解決の鍵となった。認知心理学では一つの重要な区別として、自動的なプロセスと制御されたプロセスの区別が立てられた。この区別を援用して、社会心理学では明示的に述べられた態度は制御されたプロセスによって形成されるものであり、実際の社会的行動は自動的なプロセスによって形成されるものであるという理解がなされるようになった。こうして、二つの異なるプロセスという観点から、明示的に述べられた態度と実際の社会的行動の乖離が説明されることとなった。

最後に、意思決定の領域における二重システム理論がどのように発展・展開したのかを見ていこう。心理学では 1970 年代位から、不確実な状況下における人々の意思決定に関わるバイアス研究が発展してきた。中でも、D. Kahneman と A. Tversky は、リスクを伴う選択に関して、リスクを回避するバイアスや、逆にリスクを冒そうとするバ

バイアスが働いていることを示した (Kahneman and Tversky 1984)。たとえば、私たちは 85% の確率で 1000 ドルの利益が得られる選択肢と 100% の確率で 800 ドルの利益が得られる選択肢のうちどちらを選ぶかという場合、前者の期待値は 850 ドルで後者の期待値は 800 ドルなので、期待値に基づくと前者を選ぶことになる。しかし実際には、私たちは後者を選ぶ傾向がある。これは、私たちが利益に関してリスクを回避しようとする傾向があるからだと考えられる。

また、利益ではなく、損失になると逆の傾向が見られる。すなわち、85% の確率で 1000 ドルの損失になる選択肢と 100% の確率で 800 ドルの損失になる選択肢のうちどちらを選ぶのかという場合、前者の期待値は 850 ドルの損失で後者は 800 ドルの損失であるので、期待値に基づくと後者を選ぶことになる。しかし実際には、私たちは前者を選ぶ傾向がある。これは、私たちにリスクを冒そうとする傾向があるからだと考えられる。

以上は、期待値に基づいて選択するプロセスとは別に、リスクの評価を行うもう一つ別の選択プロセスがあることを示している。このように、意思決定に関しても、期待値に基づく合理的なプロセスと、バイアスの影響を受けるプロセスの二つがあると考えられる。

これまで、学習・記憶、演繹的推論、社会認知、意思決定という個別の認知領域において二重システムの考えが発生してきたことを見てきた。ここで挙げたような諸成果がやがて洗練され、現代の二重システム理論へと発展していったのである。現代の理論では、私たちの認知や行為を形成する二つの一般的なシステムがあり、それらはシステム 1 とシステム 2 と呼ばれる³。システム 1 は、進化論的にも古いシステムであり、無意識的で自動的、迅速で、直観的に働くシステムである。それに対して、システム 2 は、進化的に新しく、人間において特に発達しているシステムであり、意識的で制御され、処理が遅く、熟慮的に働くシステムである。この二つのシステムが相互作用することによって私たちの複雑な認知や行為が形成されるとみなされている。

但し、システム 1 とシステム 2 はいずれも、それぞれ単一のシステムであるというよりも、多くのサブシステムの集まりである。システム 1 の各サブシステムは無意識的、自動的、迅速、直観的といった処理特徴を共有しており、処理タイプが同じであるとみなせる。それゆえ、それらは同じシステム 1 に属するとされるのである。システム 2 の各サブシステムは意識的、制御され、遅く、熟慮的といった処理特徴を共有し、処理タイプが同じであるとみなせるために、同じシステム 2 に属するとされるのである。

1.1.2 二重システムの神経基礎

これまで見てきたように、二重システム理論は心理学的なモデルとして発展してきた

³ システム 1 とシステム 2 という用語を最初に用いたのは、心理学者の K. E. Stanovich と R. F. West だと言われている (Kahneman 2012, p. 20、邦訳, p. 32)。

が、近年では、この二つのシステムに対応する二つの神経システムがあることが分かってきた。たとえば、Burns and Bechara (2007) は、「衝動システム (Impulsive System)」と「熟慮システム (Reflective System)」という二つの神経システムを提唱している。衝動システムはシステム 1 に、熟慮システムはシステム 2 に相当する。

システム 1 に相当する衝動システムは、扁桃体という情動に深く関係する脳部位を中心とした神経システムであり、迅速で自動的で、意識的な努力なしに暗黙のうちに働く。そして、何らかの選択を行う際、それぞれの選択肢によって直ちにもたらされる快苦に基づいて、どの選択肢を選ぶかを決定する (Burns and Bechara 2007, pp. 264-5)。たとえば、今、ドーナツとレタスがあり、A 子がそのうちのどちらかを選んで食べるとしよう。ドーナツは食べるとすぐ美味しいという快感が生じるが、後々の健康には良くない。レタスは食べてもあまり美味しくないが、後々の健康には良い。つまり、ドーナツは短期的な利益が大きいのに対し、レタスは長期的な利益が大きい。A 子の衝動システムは短期的な利益のみを考慮するのでドーナツを選ぶ。

一方、システム 2 に相当する熟慮システムは、前頭前野を中心とする神経システムであり、緩慢で随意的で、意識的な努力を要し、明示的に働く。このシステムは、それぞれの選択肢の長期的な快苦をも考慮に入れて、全体的な利益を最大化するような選択を行う (Burns and Bechara 2007, pp. 264-5)。先の例でいえば、A 子の熟慮システムは、レタスが長期的には大きな利益を持つので、それを考慮に入れて全体的にレタスの方がドーナツよりも利益が大きいと判断してレタスを選ぶ。

Burns と Bechara によると、最終的な行為決定はこの衝動システムと熟慮システムという二つのシステムが相互に作用し合うことによって導かれる。この二つのシステムの行為決定は一致している場合が多いが、それが食い違う場合は二つのシステムのうち相対的に強い方のシステムが決定した行為が実現されることになる。先の例で言うと、A 子の衝動システムがドーナツを選び、熟慮システムがレタスを選ぶ場合、熟慮システムの方が強ければ、レタスを食べるという行為が実現される。だが、ドーナツを食べたいという欲求が非常に強ければ、衝動システムの方が熟慮システムより優位となり、ドーナツを食べるといふ行為が実現される。このように、衝動システムと熟慮システムが相互に作用し合い、相対的に優位なシステムに基づいた行為が実現されるのである (Burns and Bechara 2007, p. 265)。

以上のように、システム 1 とシステム 2 に相当する二つの神経システムがあるということが神経科学において提唱されており、心理学で提唱された二重システム理論に対して、神経科学的な基礎がある程度与えられていると言うことができよう。

1.2 二重システム理論に対する批判とその検討

人間の認知現象をシステム 1 とシステム 2 という二つのシステムで捉えようとする試みは、今日では上記で述べられたさまざまな認知領域だけではなく、それらにまたが

る横断的研究も盛んに行われており、この理論を支持する新たな成果が日々生み出されている。だが、二重システム理論が広範囲の領域を巻き込んで展開され新たな成果が生み出されるにつれ、この理論に対する批判も見られるようになってきた。その批判の中には、二重システム理論の成立そのものを脅かすと考えられるものが幾つかある。ここでは二重システム理論の批判として特に重要であると考えられる三つの批判を取り上げ、その妥当性を検討しよう。まず、システム1とシステム2は人間の認知処理過程として程度の違いはあるが連続的なものであり、それゆえ質的にタイプの異なる二つのシステムとするのは不当ではないかという批判を取り上げ検討する。次に、そもそも人間の認知を説明するには単一プロセスの方が尤もらしいのではないかという批判を取り上げ考察する。そして最後に、二重システム理論における二つのシステムの関係が不明瞭ではないかという批判を取り上げ吟味する。

批判1：システム1とシステム2は、それぞれ別個の異なるタイプのシステムではなく一つの同じタイプのシステムであり、ただこのシステムには処理の仕方に連続的な違いがあり、その連続性の両極に位置するのがシステム1およびシステム2とされるものではないか (Newstead 2000; Osman 2004)。

二重システム理論に対する特に重要な批判として、私たちの認知活動を処理するシステムは、実は二重システム論者が主張するようなシステム1とシステム2という二つの異なるタイプのシステムではなく、一つの同じタイプのシステムであり、この同じタイプのシステムが人によってその処理の仕方に連続的な違いがあり、この違いの一方の極に位置するのがシステム1とされるものであり、他方の極に位置するのがシステム2とされるものなのではないかという批判がある。批判者がこの主張の拠り所としているのは、二重システム理論のもとでは思考の傾向性 (**thinking disposition**) に関する個人差が説明できないと考えられる点である (Newstead 2000; Osman 2004)。

たとえば、迷信の信じやすさについては、かなり個人差がある。「夜に笛を吹くとヘビが寄ってくる」という迷信について、それをそのまま鵜呑みにしていかなる証拠も考慮せずに信じてしまう人から、ヘビの生態や行動習性について多くの証拠を考慮することでそのような迷信を信じない人まで様々である。迷信の信じやすさに関する人々の分布は、その迷信に反する証拠の考慮の度合いや、個人的な思い込みの強さによって、簡単に信じるから容易に信じないまで、一つの連続体をなして分布すると考えられる。

だが、二重システム理論のもとでは、迷信はいかなる証拠も考慮せずに信じるか、十分な証拠を考慮して信じないかのどちらかになってしまい、その中間のさまざまな程度の信じやすさというのは認められなくなるように思われる。上記の例で言えば、システム1はいかなる証拠も考慮せずに、言われるがままに「夜に笛を吹くとヘビが寄ってくる」と信じ、システム2はいろいろ証拠を考慮して、それを全く信じない。そしてシ

システム1が勝利すれば、言われるがままに信じ、システム2が勝利すれば、全く信じないことになる。それゆえ、言われるがままに信じるか、それとも全く信じないかの両極に分かれ、その中間の様々な程度の信じやすさというものはないことになる。

しかし、実際に私たちの迷信の信じやすさについては、まったく証拠を考慮せずに信じるか、あるいは十分な証拠を考慮して信じないかのどちらかではなく、個人によって証拠を考慮する程度は様々である。迷信を信じる人の中にも、言われるがままに信じる人から多少の証拠を考慮して信じる人まで様々であり、迷信を信じない人の場合も同様にそうである。二重システム理論の批判者らは、このような現に存在する個人差が二重システム理論では上手く説明できず、それゆえ二重システム理論は私たちの認知活動を説明するには相応しくないとみなすのである。

Evans and Stanovich (2013) は、この批判に関して、確かに二重システム理論にとっては重要な批判であることを認めつつ、このような批判は、二重システム理論におけるタイプとモードという二つの概念に関する混乱した理解と、システム2の認知処理に対する誤解によって生じると主張している。そして、確かにこのような混乱や誤解は二重システム理論を理解する上で生じやすいことを認めつつ、タイプ概念とモード概念を正しく区別し、システム2の認知処理を的確に捉えることにより容易にそのような混乱や誤解を解消することができるのであり、それゆえこの批判は二重システム理論に対する決定的な批判にはならないと述べている。以下では、二重システム理論におけるタイプ概念とモード概念の区別、およびシステム2の認知処理に関する彼らの説明を見てみよう。

まず、二重システム理論におけるタイプ概念とモード概念の区別である。二重システム理論のもとでは、私たちのあらゆる認知活動はシステム1とシステム2の相互作用によって説明されうるが、この二つのシステムを区別するのがタイプの概念である。システム1とシステム2は、それぞれ以下のような様々な質的属性の集まりによって特徴付けられる。たとえば、先に述べたとおり、システム1は、直観的で、迅速で、無意識的で努力を要しない等といった質的属性の集まりによって特徴付けられ、対照的にシステム2は熟慮的で、緩慢で、意識的で努力を要するといった質的属性の集まりによって特徴付けられる。このような質的諸属性の集まりがそれぞれ処理の一つのタイプを形成し、それゆえシステム1とシステム2は認知システムとしてタイプが異なることになり、別のシステムとして区別されるのである。

ところで、もう少し詳しく言うと、このようなシステム1とシステム2の諸特徴には、それぞれのシステムの認知処理に必ず伴う必然的なものと、そうではなく典型的に伴うに過ぎないものがある。Evans らによると、それぞれのシステムによる認知処理の必然的な特徴とは、システム1にとっては自律性 (autonomy) とワーキングメモリへの負担の無さである。自律性とは、処理を行うのに必要な情報はすべてシステムの中に蓄えられており、システム外の情報を参照せずに行われるような処理の特性である。たと

例えば、知覚では、処理に必要な情報は知覚システムの中にあり、その外にある信念は参照されない。また、ワーキングメモリ（作動記憶）とは、すぐに行う作業のために一時的に必要とされる記憶のことである。記憶を持続期間によって分類する場合、以前は短期記憶と長期記憶の二つに分類するのが一般的であったが、近年ではそれらにワーキングメモリを加えることで、三つの分類となっている（加藤 2011, 5 章）。システム 1 では、このようなワーキングメモリを用いずに処理が行われる。

これに対して、システム 2 にとっての必然的特徴とは、認知的分断（*cognitive decoupling*）とワーキングメモリへの負担である。認知的分断とは、環境との知覚的な関わり及び行為的な関わりを遮断して、認知が行われることである。たとえば、あることをしたと仮定して、その結果どうなるかを色々と熟慮するような場合である。また、ワーキングメモリへの負担とは、この熟慮にもみられるように、既に考えたことを覚えておいて、それを適宜参照するような仕方で行われることである。

これら必然的な属性は、それぞれの処理タイプの定義的特徴とみなされ、それによってある認知処理ないし認知システムがどのタイプのものかが決まるのである。一方、典型的な特徴とは、システム 1 にとっては迅速で無意識的、文脈依存的、自動的、などであり、システム 2 にとっては緩慢で、意識的、抽象的、制御された、などである。Evans によると、これら典型的特徴は、あるシステムにおいて常に一緒に生じているわけではなく、ふつう一緒に生じるというに過ぎない（Evans and Stanovich 2013, p. 224-6）。

このように、二重システム理論におけるタイプとは、自律性とワーキングメモリへの負荷の無さという定義的特徴および様々な典型的特徴を含む属性の集まりによって特徴付けられるシステム 1 の処理タイプと、認知的分断とワーキングメモリへの負荷という定義的特徴および様々な典型的特徴を含む属性の集まりによって特徴付けられるシステム 2 の処理タイプのことを指す。二重システム理論は、このような二つのタイプの認知処理によって人間の認知がなされると説くのである。

では二重システム理論を理解する上で重要なもう一つ概念であるモードとはどのようなものなのだろうか。モードとはタイプと対比される概念である。タイプは、非連続的な認知処理様式で、二つの異なるタイプの間には中間的な処理様式はありえない。それとは対照的に、モードは、一見すると非連続的に見えても、中間的な処理様式が実際にはあるような認知処理様式である。Evans によると、二重システム理論においては、まさにシステム 2 の認知処理のスタイルが連続的な個人差を許容するのであり、それゆえ批判者が二重システム理論で説明不可能と指摘した思考の傾向性に関する個人差も、思考の傾向性をシステム 2 の認知処理の一つのモードと考えれば、容易に説明可能となる。このことは K. E. Stanovich による心の三部モデルにおいて詳しく示されている（Stanovich 2009a, 2009b, 2011）。それによると、私たちの認知処理はこれまで見てきたように大きく分けてシステム 1 とシステム 2 に区別されるのに加え、システム 2 はさらに「反省的な心（*Reflective Mind*）」と「アルゴリズム的な心（*Algorithmic*）

Mind)」に分かれる（合計三つの部分）。迷信の信じやすさのような思考の傾向性の連続的な個人差は、迷信を信じるかどうかの認知処理がシステム2の反省的な心においてなされることによる。すなわち、ほとんど証拠を比較考量しないで信じるのと、証拠を考慮したけれど信じてしまうのと、証拠を十分に考慮して信じないというような個人的な程度差は、このシステム2の反省的な心において証拠をどれほど慎重に考慮するかが人によって連続的に変化しうることによるのであり、それゆえモードの概念によって説明できるのである。こうして、迷信の信じやすさのような思考の傾向性の個人差も、二重システム理論によって説明できるのである。

このように、思考の傾向性の連続的な差異について説明する際にも、システム1とシステム2を二つの極とみなすような単一のタイプのシステムを想定する必要はなく、異なるタイプの二つのシステムであるシステム1とシステム2の相互作用によって認知処理がなされるという二重システム理論の基本枠組みを維持したまま、システム2の反省的な心の認知スタイルをモードの概念で捉えてやれば、十分説明可能なのである。

以上のように、Evansらは思考の傾向性に基づく二重システム理論批判に対して、タイプとモードの区別を導入することによって応答を試みるが、この応答はおおむね妥当だと考えられる。システム1とシステム2を二つの異なるシステムとして区別するタイプの違いは確かに連続性を許容しないカテゴリー的な違いであり、またシステム2の中に思考の傾向性のような連続的な差異を許容するモードを認めることも決して不自然ではない。Evansらの応答は成功しているとみなしてよいであろう。しかし、システム1とシステム2をその処理タイプによって異なるシステムとして区別することに対しては、そのような区別はそもそも成り立たないのではないかという根源的な批判がある。次に、その批判を考察しよう。

批判2：私たちの認知現象は二重システム理論でなくとも単一システム理論で十分説明可能ではないのか（Osman 2004; Kruglanski and Gigerenzer 2011）。

Kruglanski and Gigerenzer (2011) は、私たちの認知現象を明確に区別された二つのタイプの処理様式で説明しようと試みる二重システム理論に対して、その二つのタイプの処理様式が一つのタイプのそれへと収束可能であるとして、その論拠や経験的証拠を提示している。特に二重システム理論で想定されている直観的判断と熟慮的判断の区別に注目し、これらが実はどちらも根本的には規則に従う過程によって形成されるものであり、それゆえそれらは一つのタイプに収束可能であり、統一的な単一理論によって説明可能であると主張している。それとともに、彼らは一般に規則に基づく処理と、連合処理のような規則に基づかない処理を区別するのは誤りであり、いずれも規則に基づく処理に他ならないのだと主張するのである。

二重システム論者らは、このような Kruglanski らの主張をどのように捉えているのだ

ろうか。たとえば **Evans** らのような二重システム論者は、二重システム理論によってシステム1に帰属させられる行動も、またシステム2に帰属させられる行動も、規則を用いることで記述されうるし、コンピュータープログラムによってモデル化されうるという点でどちらも規則に従うものであることを認めている (**Evans and Stanovich 2013, p. 231**)。特に、システム1による連合処理は、規則を用いて実装されるニューラルネットワークでモデル化可能であり、この点では規則に基づくと言うことができる。つまり、システム1とシステム2はどちらも規則に基づいているといえる側面をそれぞれ持つのである。**Evans** らはこのように規則に基づいた処理と連合処理の両方が規則的な処理でありうることを認めた上で、このことが直ちに二重システム理論による直観的判断と熟慮的判断の区別を脅かすことはなく、ましてや単一理論で説明可能だということなどなりえないと主張するのである。

Evans らによると、**Kruglaski** らによって指摘されているシステム1の連合処理における規則性は、連合処理を実現する一つ下のレベル、すなわちニューラルネットワークのレベルでの話である。ニューラルネットワークにおいては、確かに神経細胞の活動は一定の規則に従って行われている。従って、システム1の連合処理が規則的であるといっても、それは連合処理そのもののレベルでの規則性ではなく、連合処理を実現する下位のレベル、すなわちニューラルネットワークのレベルでの規則性であり、システム2の処理が規則的だと言われるのはレベルが異なるのである。システム1とシステム2が異なるレベルで規則的であるに過ぎないとすると、このことは同じレベルでのタイプの異なる二つの処理方式が存在することを否定することにならない。あるレベルで二つのシステムの属性がタイプの異なるのであれば、たとえ異なるレベルで属性が同一であったとしても、そのことによって二つのシステムが実は同じタイプのシステムであることにはならないのである。このように **Evans** らは、**Kruglaski** らが指摘した直観的判断と熟慮的判断がいずれも規則に基づいているということは、それらの判断が同じタイプのシステムによるかどうかということと何の関係もなく、従ってそれらの判断が異なるタイプのシステムによるという二重システム理論を否定するわけではないと主張するのである。

しかし、このように二重システム論者らが単一システムで十分だという **Kruglaski** らの批判を退けることができるとしても、逆に、彼らは二重のシステムが必要だという積極的な証拠を示すことができるのだろうか。**Evans** らは私たちの認知現象を説明するものとして、二重システム理論を積極的に擁護するために、三つの論拠をあげている。第一に、二つのタイプの処理過程の一方にだけ影響するようデザインされた実験によって示された結果である。この種の実験では、通常、システム2の認知処理に必要なワーキングメモリの量を指示や動機付けによって増加させたり、あるいは同時に別の課題を課してワーキングメモリの負荷を大きくしたり、さらに熟慮的に考える時間がほとんどないスピード課題を課してワーキングメモリの働きを抑制したりといったことが行わ

れる（たとえば、De Neys 2006 や Evans and Curtis-Holmes 2005）。このような実験操作によって、システム1とシステム2が別個の異なるシステムであるということを実証する成果が蓄積されてきたのである。

第二に、昨今よく用いられるニューロイメージングによる成果が二重システム理論の論拠として挙げられている。この成果によれば、システム1の認知処理やシステム2の認知処理が観察されるとき、異なる脳領域が活性化されていることが推論や社会的認知、意思決定などの領域で確認されている。たとえば意思決定においては、McClure et al.

(2004) で示されたところによると、直近の報酬に基づいてなされる決定と、あとでもらえる報酬に基づいてなされる決定とでは、別個の異なる神経システムが関係していることがわかる。すなわち、遅れた報酬に基づいた場合、外側前頭前皮質や後側頭葉が活性化され、直近の報酬に基づいた場合は辺縁系システムが活性化することが示されているのである。二重システム理論では、遅れた報酬にもとづく処理は、ワーキングメモリを用いて未来の可能性の心的シミュレーションを行うシステム2によってなされ、直近の報酬に基づく処理はワーキングメモリを用いない自動的なシステム1によってなされると考えられるが、このような別個のシステムの存在が、関連する脳部位の違いによって証拠立てられているとみなせるのである。

最後に二重システム理論の論拠として挙げられるのは、精神測定アプローチによる成果である。このアプローチでは、推論や社会的認知などの精神能力が測定される。その結果、推論や社会的認知がヒューリスティックを用いて行われる場合にはあまり個人差がないが、熟慮によって行われる場合には個人差があることが見出された。この結果は、ヒューリスティックスを用いた認知はシステム1で行われ、熟慮による認知はシステム2で行われると解釈できる。というのも、ワーキングメモリの容量や知性の洗練度に個人差があるため、システム2の能力には個人差があるが、システム1の能力には、そのような個人差があまりないと考えられるからである (Stanovich 2011, pp. 234-5)。

これまで見てきたように、Evansらは、Kruglaskiらが二重システム理論を掘り崩すものとして挙げる論拠がまったく正当性がないことを示すとともに、二重システム理論がなぜ支持されうるのかについて三つの積極的な論拠を提示した。Evansらの応答は十分納得のいくものだといえる。つまり、二重システム理論が主張するように、二つの異なるシステムがあるということを認めてよいように思われる。しかし、そうすると、その二つのシステムの関係はどうなっているのだろうか。この点が不明瞭なままだと、二重システム理論は大きな欠陥を抱えることになるだろう。そこで次にその点が不明瞭ではないかという批判を検討していくことにしよう。

批判3：二つのシステムの関係が不明瞭ではないか。

これまでの二つの批判は認知現象を説明するために二つのシステムを立てることに

対する批判であったが、第三の批判は二つのシステムを立てるのはよいとしても、その二つの関係が明瞭になっていないのではないかという批判である。この批判に答えるために、ここではシステム1とシステム2が具体的にどのように関係し合っているのかについて考察していこう。

システム1とシステム2はそもそもどのような関係にあるのだろうか。それぞれが互いに独立に処理を行い、最終段階で照合が行われ、食い違っている場合は、強いほうが勝つのだろうか。それとも、システム1とシステム2の処理は独立に進行することなく、最初から相互作用しながら進行し、強い方の処理が最終結果に反映されるのだろうか。もしくは、そもそもいずれかのシステムに優位性がある、優位システム主導で処理がなされるのだろうか。この点に関して、たとえば、S. A. Slomanらは、システム1とシステム2の処理がそれぞれ並行に進行し、最終結果に関してこの二つのシステムに葛藤が生じた場合、強い方が勝つという形で葛藤が解消されるという並行競合形式（parallel-competitive form）を提唱している（Sloman 1996; Smith and DeCoster 2000）。たとえば、システム1により煙草を吸おうという決定がなされ、システム2により禁煙をしなければならないという決定がなされた場合、システム2がシステム1よりも強ければ、禁煙するという事で葛藤が解消される（もしくは、システム1が相対的に強い場合、後で後悔するかもしれないが、煙草を吸うことで葛藤が解消される）。この説は、それぞれのシステムが独立に処理を実行し、途中のどの段階でも相互作用することはなく、最終段階で葛藤が生じた場合、競合が起こり、常に相対的に強い方のシステムによって行動が出力されると考える説である。

しかし Evansらは二重システム理論の並行競合形式に伴う困難を指摘している（Evans and Stanovich 2013）。それはシステム1の処理がシステム2の処理よりも極めて早いということと大いに関係がある困難である。先の煙草の例でみると、システム1のタバコを吸おうという決定は、システム2の禁煙しようという決定よりも早く行われる。その場合、システム1がシステム2の決定を待たなければ、この二つのシステムは競合することなく、常にシステム1が勝って、それが決定した通りの行為が引き起こされることになる。つまり、もし二重システムが並行競合形式なら、ある葛藤状態を解消するのに、早いシステムは遅いシステムの決定を待つ必要があるということになるのである。Evansらは二重システム理論における二つのシステムの関係として、この考えは尤もらしくないと判断して、並行競合形式を退けるのである。

では、Evansらは二つのシステムの関係をどのように捉えるべきだと考えているのだろうか。彼らは、迅速なシステム1の処理が、直観的な応答をデフォルトとして生み出すと考える。この直観的なデフォルト応答とは、それに対して熟慮的なシステム2が場合によっては干渉するような応答のことである。この見方はデフォルト干渉主義（default-interventionism）と呼ばれ、私たちの行動がふつうシステム1によってなされ、何か問題があるときのみ、システム2が干渉して別の行動がなされるとみなすので

ある。これは Kahneman からも支持している考えであり、鍵となるのは、システム 1 によるデフォルト直観をシステム 2 が、反省的推理によりときに干渉するという見方である (Kahneman, 2011; Kahneman and Frederick 2002)。たとえば、通常、人々は新奇な問題に直面した場合、直観的かつ最小限の努力で素早く答えを出す。この直観に誤りがありそうな場合、システム 2 はシステム 1 の答えに抑制をかけ、その通りの行動を起こさせないようにする。たとえば、慣れない山での登山の途中で茂みがガサガサと揺れると、システム 1 により直観的にクマだという判断が下され、逃げようという決定がなされるかもしれない。だが、この山にクマは生息していないはずだというシステム 2 の推理がなされ、クマではないから逃げなくてもよいという決定がなされると、それによる干渉が行われ、逃げるという行動は行われなくなるのである。Evans らはこのようなデフォルト干渉主義が、二重システム理論における二つのシステムの関係として尤もらしいと考えるのである。

また、デフォルト干渉主義は並行競合説と比べて、システム 2 がワーキングメモリ資源を利用していることについても、より整合的な説明を与えることができる (Evans and Stanovich 2013)。ワーキングメモリ資源は非常に限られた貴重な資源なので、困難な課題や新奇な課題、意志的な動機付けを必要とするような課題に選択的に割り当てられる必要がある。デフォルト干渉主義では、ほとんどの課題はシステム 1 によってデフォルトとして解決され、システム 2 が働くことはない。システム 2 が働くのは、困難な課題や新奇な課題、意志的な動機付けが必要な課題の場合だけである。従って、システム 2 が貴重なワーキングメモリ資源を利用しているとしても、特に問題はない。しかし、並行競合説では、どんな課題でも、システム 2 はシステム 1 と並行して課題の解決に当たっており、従ってどんな課題にもワーキングメモリ資源が利用されることになってしまう。限られた貴重な資源であるワーキングメモリをそのように利用することは不可能である。このように、Evans らによると、システム 2 におけるワーキングメモリの利用についても、並行競合仮説よりもデフォルト干渉主義の方が整合的なのである。

Evans らのデフォルト干渉主義は、システム 1 とシステム 2 の関係として、並行競合説よりも尤もらしいものと言えよう。但し、システム 2 がシステム 1 にいつどのようにして介入するのかについて、デフォルト干渉主義では明らかにされていない。その点でデフォルト干渉主義にも大きな問題があるといえる。しかし、これは非常に厄介な問題である。システム 2 はシステム 1 の決定に何か問題があるとき、システム 1 に介入するが、システム 2 は自らどうするかを決定して、それがシステム 1 の決定と食い違っていれば、システム 1 の決定に問題があると判断するわけではない。そのように判定するならば、デフォルト干渉主義ではなく、並行競合説になってしまう。そうではなく、システム 2 は自らどうするかを決定する前に、システム 1 の決定に問題があるかどうかを判定するのである。しかし、そのようなことが一体如何にして可能なのであろうか。デフォルト干渉主義が抱えるこの問題は確かに難問である。だが、この問題についてはデフォ

ルト干渉主義の今後の課題と捉え、二つのシステムの間を説明する現時点における最も有力な理論として、デフォルト干渉主義を受け入れることができよう。

ここまで、二重システム理論にとって決定的に重要であると考えられる三つの批判について検討してきた。それにより、二重システム理論におけるシステム1とシステム2はタイプの異なる別個のシステムであるということが確保され、さらに私たちの認知現象を説明する理論として、単一システム理論よりも二重システム理論の方が尤もらしいことが明らかになった。その上で、二つのシステムの間については、システム1がデフォルト処理を行い、それに対してシステム2が必要に応じて干渉するというデフォルト干渉主義の見方は十分支持できると思われる。

1.3 二重システム理論における自己制御

これまで二重システム理論について見てきたが、この理論のもとでは、自己制御はどのようなものとして捉えられるのかを明らかにしておこう。この点を明らかにするうえで鍵となるのは、二つのシステムが物事の価値を評価する仕方の違いである。この点は、これまで心理学や神経科学で提唱されてきた二重システム理論においてあまり明確化されてこなかったが、ここでは以下のように捉えることにしたい。

まずシステム1は、欲求や情動⁴に基づいて物事の価値を評価すると考えられる。たとえば、A子のシステム1は、レタスよりもドーナツを選ぶが、それはレタスよりもドーナツを強く欲しているか、あるいはドーナツを食べることに至福の喜びを覚えるからである。A子はドーナツよりレタスのほうが健康によいと思っているだろうが、それでもドーナツを食べたいという欲求（あるいはドーナツを食べることへの喜び）が強いので、システム1はドーナツを選ぶ。このようにシステム1は物事の価値を欲求ないし情動によって評価し、それに基づいてどの行為を選ぶかを決定すると考えられる。一方、システム2は、物事の価値を知的に評価すると考えられる。A子のシステム2はドーナツよりレタスを選ぶが、それは知的な観点からドーナツとレタスの価値を評価して、健康の価値の方が美味しさの価値より勝ると評価するからである。システム2は物事の価値を知的に評価することによって、どの行為を選ぶかを決定する。

このようにシステム1とシステム2の行為決定が食い違う場合は、システム2がシステム1を制御して、システム2が決定した行為を実現することが、自己制御だと言える。A子は、ドーナツを食べたいと思いながら、それを我慢してレタスを食べたとすれ

⁴ 私たちの欲求はふつう何らかの情動、たとえば渴望、憧憬、怒り、恐怖などの形を取ると考えられる。ことによると、いかなる情動でもないような欲求もあるかもしれないが、そうだとしても非常に稀であろう。従って、本稿では、欲求も情動の一種と捉えて、欲求を情動から特に区別しない。また、選好という言葉も用いるが、それも欲求や情動と特に区別せず、文脈に応じて適宜それらを用いることにする。欲求と情動と選好はいずれも物事の価値を表しており、またそれと同時に、そのような価値を実現する（あるいは負の価値ならそれを回避する）ような行動を促す動機付けの力を含んでいる。

ば、自己制御ができていくことになるが、これは彼女のシステム2がシステム1を制御して、システム2が決定した行為を実現したということに他ならない。逆に、レタスを食べようと決意しながら、ドーナツを食べたいという誘惑に負けて、ドーナツを食べてしまえば、自己制御ができていないことになるが、これはシステム2がシステム1を制御できず、システム1が決定した行為が実現したということに他ならない。

しかし、システム1とシステム2の行為決定がシステム1のそれと一致することももちろんある。たとえば、A子がドーナツよりレタスが好きであれば、A子のシステム2だけではなく、システム1もレタスを食べるという決定を行うだろう。そうだとすれば、システム2がシステム1を制御する必要もなく、レタスを食べるという行為が実現されるだろう。このような場合も、システム2が決定した行為が実現されているので、自己制御ができていくと言ってよいだろう。この場合は確かに、自己制御が実質的に行われているわけではないが、それでも自己制御がなされていることには変わりがない。しかも、私たちが概ね自己制御ができるというようになるためには、このような自己制御が極めて重要である。というのも、システム1の行為決定とシステム2の行為決定が一致しない場合、システム2がシステム1を制御して自己制御を行わなければならないが、それは必ずしも容易なことではなく、自己制御の失敗が起りやすいからである。私たちが概ね自己制御ができていくというようになるためには、システム1の行為決定ができるだけシステム2の行為決定と一致するようにして、おのずと自己制御ができていくというようにする必要があるのである。

このように自己制御には、システム1とシステム2の行為決定が食い違う場合になされる自己制御と、システム1とシステム2が一致する場合におのずとなされる自己制御の二種類がある。

ここで一つ注意すべきことがある。システム2の決定がシステム1の決定に一致するのは、実際にシステム2が行為決定を行い、その決定がシステム1の決定に一致する場合だけではなく、デフォルト干渉主義が言うように、システム2がシステム1のデフォルト決定を容認してその決定に干渉しない場合も含むということである。このような場合、システム2は自ら行為決定を行うことなく、システム1のデフォルト決定をそのまま容認する。これに対して、システム2が行為決定を実際に行って、それがシステム1の決定と一致する場合というのは、システム1の決定に問題がありそうなときに、システム2が自ら行為決定を行うが、それでもその決定が結局、システム1の決定と一致してしまう場合である（ふつうは、システム2が行為決定を行うのは、システム1の決定に問題があるとシステム2が判定した場合であるから、システム2の決定はシステム1の決定と異なるものになる）。システム2がシステム1の決定を容認するにせよ、自ら行為決定を行ってそれがシステム1の決定と結局一致するにせよ、いずれの場合も、自己制御が為されていることに変わりはない。これらの場合は、自己制御がおのずと為されているのである。

システム2がシステム1を実質的に制御する必要があるのは、システム2の行為決定がシステム1のそれと食い違う場合である。この場合は、システム2はシステム1を抑制して、自分が決定した行為を実現するようにしなければならない。意志 (will) という概念が登場するのも、このような場合である。A子のシステム2がレタスを食べるといふ行為を実現するために、ドーナツを食べようとするシステム1を抑制する場合、A子はレタスを食べるために、意志の力でドーナツを食べたいという誘惑に打ち克とうとしていると言えよう。意志はシステム2とシステム1の行為決定が食い違うとき、システム2がシステム1を抑制する働きをすると言えよう。しかし、このような意志は消耗しやすく、意志の力で自己制御を行おうとすると、しばしば失敗してしまうのである。

また、意思決定という場合の「意思 (decision)」、すなわち自分の「決意 (resolution)」や「意図 (intention)」と言われるものは、システム2が決定した行為⁵ (すでに為された行為ではなく為すことに決めた行為) だと言えよう。A子がレタスを食べようと思いつながら、ドーナツを食べたいという誘惑に負けて、ドーナツを食べてしまうとき、確かにA子はドーナツを食べたのだが、ドーナツを食べることがA子の意思 (決意、意図) であったわけではない。A子の意思はあくまでレタスを食べることである。そして、レタスを食べることはA子のシステム2が決定した行為である。また、仮にもし、A子がドーナツを食べたいという誘惑に負けて、レタスを食べようという意思を翻して、ドーナツを食べようという意思に変えたとすれば、その場合はもちろん、A子の意思はドーナツを食べることになる。つまり、A子はドーナツを食べようという新たな意思を持って、ドーナツを食べることになるのである。そしてこの場合、A子のシステム2の行為決定はレタスを食べるといふ決定からドーナツを食べるといふ決定へと変化していると考えられる。従って、やはりA子のシステム2の決定した行為がA子の意思なのである。このように、意思というものはシステム2が決定した行為だと言える。

以上をまとめれば、二重システム理論のもとでの自己制御とは、システム2とシステム1の行為決定が一致するにせよ、一致しないにせよ、ともかくシステム2が決定した行為が実現されることである。システム1とシステム2の行為決定が一致する場合 (システム2がシステム1の決定をそのまま容認する場合も含む) は、システム2がシステム1を制御する必要がなく、おのずと自己制御が行われる。それに対して、システム1とシステム2の行為決定が食い違う場合は、システム2がシステム1を制御して、自己制御が行われる。この実質的な制御を行うのが意志の働きである。また、私の意思とは私のシステム2が決定した行為である (システム2が容認した行為も含む)。このように言うことができよう。

⁵ システム2は、デフォルト干渉主義の場合、自ら行為決定せずに、システム1が決定した行為を容認する場合もあるので、システム2が決定した行為には、システム2がそのように容認した行為も含めることにする。

第二章 自己制御に情動はどう関係するか

親に口答えしてはいけないと思いつつも、あまりにも腹が立つので、ついつい口答えしてしまったり、油断してはいけないと思いつつも、望外の結果に有頂天になって、ついつい油断してしまったりする。このように情動は私たちの自己制御を妨げることがある。しかし逆に、情動が自己制御に役立つこともあるように思われる。たとえば、ダイエットすべきだと思っけていても、ついつい甘いものを目にすると、食べてしまいそうになるが、そこで太ることへの恐怖が沸き起こることで、かろうじて食べずにいることができる。このような場合、システム1では甘いものを食べたいという欲求と、太ることへの恐怖が生じ、恐怖の方が強いため、システム1がダイエットするという行為決定を行い、システム2はこのシステム1の決定を容認して、ダイエットをするという行為が行われる。ここでは、このような仕方システム2がシステム1を制御することなくおのずと自己制御が行われ、情動が重要な役割を果たしている。

この章では、自己制御における情動の役割について考察しよう。そのためにまず、そもそも情動とは何かについて、情動経験に含まれる身体状態こそ情動の本質であるということを示唆した情動論を、歴史的な背景を踏まえながら確認する。その上で、自己制御と情動の関係について、A. R. Damasioによる「ソマティック・マーカー仮説」(SMH)に基づいて考察する。それにより、情動が対象の価値に応じて誘発された身体的変化を含む状態であるということだけではなく、そのような情動が、私たちの意思決定の場面において、常に無自覚的な影響を及ぼしており、それが意思決定を左右する重要な役割を果たすことを確認する。

2.1 情動とは何か

2.1.1 情動とは

古くから情動とはどのようなものなのかについて様々な考察がなされてきた。情動を経験している当事者はどのような状態にあるのか、また、情動はどのような性質をもち、どのような働きをするのかなどである。その際、情動とは、理性を脅かす存在でありそれに屈せず理性的に振る舞うにはどうすればよいのかや、情動を理性に従わせるためにはどうすればよいのかなどという仕方しばしば理性と対比されてきた。歴史的にはヒュームのように、理性が情動に従属するとみなす論者もいたが、多くの論者はそうではなかった⁶。いずれにせよ、ある一個人において情動とは何か問題となるとき、多くの場合は、理性と対比的に捉えられることが多かったと言えよう。

⁶ 古代ギリシア時代、プラトンはソクラテスの教えに倣い、理性を「賢馬」や「主人」に、熱情を「悍馬」や「奴隷」に例えた(遠藤 2013, p. 8; プラトン 2014, p. 69-70)。18世紀には、ヒュームが情動は理性の奴隷ではなく、むしろ理性が情動の奴隷であるとし独自の見解を展開した(ヒューム 2010)。

今日では、情動について考察する際、単に理性と対比的に捉えるだけではなく、情動が様々な側面を含むものであり、そのような多様な側面から情動を明らかにしていこうという試みが盛んである (LeDoux 1996; ピネル 2005)。たとえば、情動を分析する際に、情動を一人の個人内の現象と捉えて分析することもできるし、個人と個人の間の現象と捉えて分析することもできる。また、個人内の現象として捉えた場合でも、情動には信念や欲求、身体状態、特有の行動傾向など、さまざまな側面があり、それらの多様な側面から情動を明らかにしていこうという試みがなされている。但し、そもそも情動の本質とは何かということを問題にし、その解明を試みようとするそれが容易ではないことに否が応でも気づくことになる。哲学者 De Sousa (2011) によると、現在様々な論者により定義された「情動」概念は、身体状態や欲求、特有の行動傾向などそれらのどの側面に注目するかによって、少なくとも 29 の定義があると言う。情動とは何かということは、私たちにとっては日常生活の中で主観的に経験するものとしては明々白々なことである。だが、それ自体がそもそもどのようなものなのかということに関して、暫定的でもよいかから何らかの共通理解を得たいとなると、研究分野や研究者によって様々な解釈や定義が数多く存在していて收拾を付けがたいという事態に直面してしまうことになる。

本論文では、情動に関する多くの見解がひしめき合う中で、個人内における情動の機能に焦点を絞り、一個人の情動経験に含まれる身体状態こそが情動にとって本質的なものとみなす説を支持したいと考えるが、そのために、それがどのような説であるのかを情動論の歴史的な流れも踏まえながら考察することにする。まずは個人内の情動生起のメカニズムを考察することによって、情動における身体状態の重要性を初めて強調したジェームズ＝ランゲ説から見ていこう。

2.1.2 ジェームズ＝ランゲ説

近代科学の成立以降、科学的に人間を理解する様々な試みがなされ、急速に人間についての科学的知識が蓄積された。19 世紀の末に W. James や C. Lange が人間の解剖学的な知識に依拠し、情動の発生メカニズムについて一定の見解を示したのは、今から見るとまさにそのような時期の真っ只中であつたと言えよう。情動経験とそれに対応する身体反応の関係について当時一般的に考えられていたのは、情動経験(たとえば「恐怖」や「悲しみ」)が、何らかの身体反応(たとえば「身震い」や「涙が出る」)を引き起こすという因果関係であつた。だが James らはこの常識的な見方を否定し、身体反応の方が情動経験に先立つと考えたのである⁷。この説は現在では「ジェームズ＝ランゲ説」

⁷ 当時、James (1884) は内臓や骨格筋に、Lange (1885) は循環器や血管運動に注目しているものの、両者とも外部環境の認知によって引き起こされた身体反応が脳にフィードバックされ、それゆえ主観的な情動経験が形成されるという点で共通している。

として広く知られている⁸。この説を具体的な事例でみると以下のようなになる。たとえば私が北海道の山林で山登りをしている最中にヒグマに遭遇したとしよう。その場合、私は恐怖で縮み上がるに違いない。その際、私が経験する恐怖の情動は、James によると、ヒグマを視覚的に捉え、それにより引き起こされた心拍数の増加や血管の収縮、発汗などといった身体部位の反応が、脳にフィードバックされたゆえに引き起こされるのである。つまり、ジェームズ＝ランゲ説によると、情動経験（ここでは恐怖の情動経験）とは、生理学的な身体反応（ここでは心拍数の増加など）とその感受されたもの（ここでは心拍数の増加などを感じ取ること（*feeling*））の二つの側面を持つということになる⁹。このように、ジェームズ＝ランゲ説は情動を、外界の刺激により引き起こされた身体状態の変化とその感受という二つの側面からなるものであり、その因果関係は前者が後者を引き起こすものとみなすのである。

2.1.3 認知説

2.1.3.1 認知説とは

ジェームズ＝ランゲ説に端を発する一連の情動発現のメカニズムに関する研究は、一個人における主観的な情動経験と身体状態の変化の関係を明らかにすることを主な目的としたため、情動発現に関わるとみなされていた中枢神経や末梢神経、またそれらを含む生理的なメカニズムの解明に大きく寄与し著しくその領域を進展させた（遠藤 2013, p. 174）。このような情動研究の流れの中では、先に見た通り、情動経験とは、身体状態の変化とその感受だと言える。たとえば私がヒグマと遭遇したなら、それによって引き起こされる身体状態の変化とその感受がまさに恐怖の情動経験である。しかし、

⁸ James らのこの説は、情動の発現メカニズムにおいて、中枢（脳）ではなく、末梢（身体部位）がその起源になることから「情動の末梢起源説」ともよばれている。

⁹ 当時、ジェームズ＝ランゲ説の考え方は、生理学者 W. B. Cannon と P. Bard によって真っ向から批判された（Cannon 1927; Bard 1928）。彼らの説は「キャノン＝バード説」として広く知られる。Cannon らは、James らの説では上手く説明できない具体的な事例（たとえば四肢麻痺の人による豊かな情動経験など）を提示し、情動経験の起源は末梢（身体部位）ではなく、むしろ中枢（脳）であるとみなした。彼らの説は、情動の末梢起源説と対比的に「情動の中枢起源説」とも呼ばれる。Cannon らの説によると、外界の知覚（ヒグマの知覚）によって得られる情報が脳で処理され、そこから独立並行的に枝分かれした情報が、主観的な情動経験（恐怖の情動）と身体反応（心拍数の増加など）をそれぞれ生み出すのである。このようにジェームズ＝ランゲ説とキャノン＝バード説は、前者は情動経験における生理学的身体反応とその感受の因果関係を認めるのに対し、後者は認めないという点でも大きく異なる。その後、Cannon らの説が James らの説に取って代わったかという点ではなく、むしろ James らの説の復権を掲げて心理学者の S. Schachter と J. E. Singer により身体反応とその原因の推察の二つの要因から情動が構成されるとする「情動二要因説」が提唱され（Schachter and Singer 1962）、また後述する A. R. Damasio により「ソマティック・マーカー仮説」が提唱されている（Damasio 1994）。とくに、ソマティック・マーカー仮説は、James らの説の焼き直しとも言われる説である（遠藤 2013, pp. 174-7）。

そうだとすると、私の内に生じる恐怖の感じ、つまり主観的に怖いと感じる心の状態は、身体状態の変化を感受したものであるということになる。つまり、それは外界のあり方に関わるものではなく、単に身体のある方に関わるものに過ぎないということになる。だが、直観的に言えば、情動は外界のあり方にも本質的に関わるように思われる。たとえば、私の恐怖の情動経験は、ヒグマが危険であるというヒグマに対する評価を含むように思われる。つまり、恐怖の情動経験は、James らのいうように外界の刺激により私の内に引き起こされた身体状態とその感受という非志向的なものではなく、外界のあり方に向けられた志向的なものであるのではないかということである。このような情動の役割、すなわち情動が外界のあり方について一定の評価を下すという役割に注目し、その点で James らの説を批判したのが心理学者 M. B. Arnold である。以下で彼女の主張を確認していこう。

1960年代、Arnold (1960) はジェームズ＝ランゲ説に対する直接的批判として、情動における身体反応とその感受ではないような側面、すなわち外界についての認知的側面の役割の重要性を強調した。これは外界の知覚による身体反応とその感受でもって情動の説明を試みた James らの枠組みにおいては、完全に見落とされていたものである。James らの説によると、私たちが外界の刺激に対して何らかの行動的反応をするのは、身体の生理的反応とその感受を通して、行動に関する一定の考えが生じるからである。たとえば、ヒグマに遭遇し私が逃避という行動をとるのは、ヒグマの知覚により引き起こされた一連の生理的な身体反応およびその感受を通して、私が逃げようという考えを持つようになるがゆえである。この考え方のもとでは、私がヒグマをどう評価するかという側面は無視されている。その評価なしにヒグマから逃げようという考えが生じ、そして逃げるとされるのである。

だが、私がヒグマから逃走するのか、それともヒグマと闘争するのかは、ヒグマが私にそのような考えを付与したからではなく、私がヒグマをいかなるものとして評価するのかに関わっている。そのことをまさに指摘したのが Arnold (1960) である。彼女によると、私がヒグマに対してどのような行動を行うのかは、私自身がヒグマを価値的にどう認識するかによって決まるのである。Arnold が想定していたこととは、遠藤のまとめを借りれば、「事象そのものではなく、事象に対する評価、すなわち事象のその個人にとっての利害、あるいは幸・不幸の意味などに対する直接的で即自的な、そして大概は自動的に生じる判断こそが、個人が経験する情動の質を分ける」(遠藤 2013, p. 178) ということである¹⁰。このように Arnold が情動における認知的評価の重要性を指摘して以降、情動の発現メカニズムにおける認知的評価の役割やその有無について大きな論

¹⁰ ジェームズ＝ランゲ説において、たとえ身体の生理的反応が同じであっても、なぜ主観的に異なる情動経験が生じるのかについては一つの大きな問題であった。これに対して、Arnold は、認知的評価の重要性を説き、それによって整合的な説明を与えた。

争が繰り広げられることとなる¹¹。

情動に認知の役割が導入されることによって、ジェームズ＝ランゲ説の場合と大きく変わった点は、ジェームズ＝ランゲ説において情動は外界の価値的なあり方を反映していないもの、すなわち情動はその点で非志向的なものであったが、それが外界の価値的なあり方を反映するもの、すなわち志向的なものとなったことである。James らの説によると、たとえば、襲いかかってくるヒグマに対して恐怖を抱くというのは、ヒグマの知覚により私の内にある身体的な変化が起こり、その感受として怖いという感覚が生じるということにしかない。一方、認知を導入してこの状況を説明すると、襲いかかってくるヒグマに対して恐怖を抱くというのは、まさにそのヒグマが危険であるという外界の対象物であるヒグマについての価値的な判断もしくは評価がなされることだということになる。このように、情動とは価値的なあり方についての志向的状态（つまり認知的な状態）であり、世界のあり方について何らかの評価を行うことだという見方は、主に哲学において、認知説（cognitivism）と呼ばれる（Nussbaum 2001）。

では、情動経験において、身体反応とその感受とは別のあり方であるような、認知（志向的状态）とはどのような心の状態なのだろうか。そもそも「認知（志向的状态）」とは、何を指しているのだろうか。ふつう認知とは、外界を捉えることであり、それには私たちの感覚器官（五官）による知覚的認知と、知覚でないような知的な認知、すなわち信念的認知がある。知覚的認知は知覚を通じて世界のあり方を認識し、信念的認知は信じるというあり方で世界のあり方を認識する。そして、そのような認知を通じて知ることができる外界のあり方そのものも、事実的なあり方なのか、それとも価値的なあり方なのかという二通りのあり方がある。つまり、世界の事実的なあり方についての知覚的認知と信念的認知があり、また世界の価値的なあり方についての知覚的認知と信念的認知があるのである。たとえば、目の前にコップがあるという事実的なあり方についての知覚的認知とは「コップがあるように見える」ということであり、その信念的認知とは「コップがあると信じる」ということである。また、目の前に美しいコップがあるという価値的なあり方についての知覚的認知とは「美しいコップがあるように見える」と

¹¹ 心理学においては、サイアンス＝ラザルス論争が当該の論争として広く知られている。R. S. Lazarus は情動生起における認知的評価は不可欠なものであり、かつ、認知的評価は情動経験に先立つとみなした（Lazarus 1984）。一方、R. B. Zajonc は情動生起において必ずしも認知的評価が先行する必要はないとし、両者は真っ向から対立したのである（Zajonc 1984）。この論争を遠藤（2013）は、両者が用いる「認知」や「感情」といった術語の意味が大きく乖離していることにその原因があり、現在ではこの論争は殆ど収束したといっても過言ではないと総括している。それぞれの術語の内包を Zajonc と Lazarus がどのように捉えていたかということ、「認知」については、Zajonc は「反省的で意識の介在する情報処理のみ」としていたのに対し、Lazarus は「半ば自動化された無意識レベルでの情報処理も含む」としていた。また「感情」について行われていた議論は、Zajonc は「未分化な感情、すなわち低次の無意識的で自動化された情報処理プロセスに焦点化」していたのに対し、Lazarus は「主観的な意識経験を伴う完全なる情動状態に特化」していたのである（遠藤 2013, p. 187-2）。

いうことであり、またその信念的認知とは「美しいコップがあると信じる」ということである。通常、知覚的認知と信念的認知は一致しているので、心の状態として信念的な状態と知覚的な状態の二つのタイプの心の状態があるようにはみえないかもしれないが、知覚と信念にズレがあると、この二つが別であることがよくわかる。そのようなズレはふつう錯覚を修正しその上で形成される信念的認知でみられる。たとえば、事実的なあり方でいえば、実際には円である図形が楕円に見えるとき、信念的認知としてはその知覚的認知を修正してこの図形は円であると信じるような場合である。また価値的なあり方でいえば、実際にはまずいケーキが美味しそうに見えるとき、信念的認知としてはその知覚的認知を修正してこのケーキはまずいと信じるような場合である。

では、認知説において情動が外界の価値を「認知する」という場合、どのような認知（志向的状态）が意味されているのだろうか。それは外界における価値的なあり方についての信念的認知（信念というタイプの志向的状态）である。たとえば、ヒグマに遭遇し恐怖の情動が引き起こされる場合、私たちは知覚的にヒグマを認知すると即座にそのヒグマは危険であると判断する。これは言い換えると「このヒグマは危険である」という、ヒグマの価値的なあり方についての信念的認知を行っているのである。認知説における認知とは、価値的なあり方についての信念的認知なのであり、知覚でもなく身体状態でもなく、まさに信念という形で、外界のあり方を評価するものなのである¹²。

また、認知説によると、様々な情動に含まれるそれぞれに固有の身体反応は、価値の信念的認知のゆえに生じるものであり、それ自体が世界のあり方を反映しているわけではない。たとえば、ヒグマとの遭遇による恐怖の情動には身震いや冷や汗などのそれ特有の身体反応が伴うが、それらはあくまでも「目の前のヒグマは危険である」という信念的認知の結果生じたものであり、身震いや冷や汗といった身体状態自体が、ヒグマの危険さを評価しているわけではないのである。外界についての評価は情動に含まれる信念によるのであり、たとえそこに身体状態が伴うことがあってもその身体状態は外界のあり方について評価しているのではなく、それは信念の結果として生じたに過ぎない。世界を評価するのは信念なのであり、その信念が情動なのである。つまり、認知説によると、情動は様々なものを含みうるが、情動を情動たらしめるその本質的なものとは外界の価値的なあり方についての信念なのであり、情動に含まれる信念こそ情動の本質的なものなのである。

2.1.3.2 認知説の問題点—情動における認知的侵入不可能性—

認知説のいうように、情動は世界のあり方についての信念的な評価であり、そのような評価をする信念こそ情動の本質だとすると、世界についての評価的な信念、すなわち

¹² 外界の価値的なあり方の知覚が可能なので、Nussbaum が言うような価値判断としての情動というのは、実は価値の知覚に基づく判断だと考えられるかもしれないが、ここでは特にこの点を論じないことにする。

世界についての価値判断が変われば、その判断であるような情動も当然変わるはずである。だが、必ずしもそうではない。たとえば、山中でイノシシに追いかけてまわされるといふ恐怖体験をした人は、そのあとそのイノシシが捕えられ檻の中に入れて自分に危害を加えることは不可能であるような状態になっても、しばしば「怖い」という恐怖を抱き続けるだろう。情動の本性が信念であり、恐怖の情動が「このイノシシは危険である」という判断そのものであるなら、「このイノシシは危険である」という判断から「このイノシシは危険ではない」という判断に変化したなら、もはや恐怖の情動を抱かないはずである。だが、檻の中にいるイノシシを見て「このイノシシは危険ではない」と判断できたとしても、依然として「危険ではない」イノシシを目にして恐怖の情動が続くことがあるのだ。P. Goldie はこの事態を説明するために、情動が知覚と同様「認知的に侵入不可能 (cognitive impenetrability)」であるという理由を与えた (Goldie 2000, pp. 74-8)。知覚は認知 (ここでは信念的認知) によって影響を受けない。たとえば、まっすぐな棒を水に浸した場合、その棒について「棒はまっすぐだ」という信念を形成できたとしても、それを知覚に反映してまっすぐに見えるようにすることはできない。目の前にある水に浸った棒は、そのような信念の形成如何に関わらず曲がって見えるのだ。つまり、知覚システムに対して認知は侵入不可能なのである。知覚と同じように情動も認知的に侵入不可能なのだとすれば、たとえ判断が変化しても情動が変化することはないのである。

だが、このように情動と判断が食い違うことがあり、情動が認知的に侵入不可能だということは、情動がもはや単なる判断ではない、すなわち情動の本質は信念的認知ではないということを示している。では、情動の本質とはどのようなものなのだろうか。この問いに対して、それはまさに情動に含まれる身体状態であり、但しジェームズ＝ランゲ説とは違って、その身体状態そのものが世界の価値的なあり方を反映しているのだと主張する説がある。本論文が数ある情動に関する見解の中で支持するのはこの説である。

2.1.4 Damasio や Prinz の説

これまで見てきたように、外界の価値的なあり方についての判断 (信念) が情動であるという認知説においては、判断と情動の食い違いを説明できないままであり、そしてこのことが示しているのは、情動が単なる判断ではないということであった。そうだとすれば、情動に本質的なものとは何だろうか。この問いに対し、情動に含まれる身体状態こそ情動の本質であるとし、ジェームズ＝ランゲ説の難点を乗り越えて情動について整合的な説明を与えているのが A. R. Damasio や J. Prinz である。Damasio は「ソマティック・マーカー仮説」(Damasio 1994) を唱え、Prinz はこの Damasio の説を踏まえて「身体的評価説」(Prinz 2004) を唱える。

彼らの説によると、たとえば、イノシシを目にして生じる恐怖の身震いは「このイノシシが危険である」ことを示すものであり、目の前のイノシシが危険であるという外界

の対象の評価を行うものである。つまり、情動に含まれる身体状態は、世界について身体的な評価を下しているのであり、身体状態そのものが志向性を持つのである。このように考えると、「このイノシシは危険だ」という判断から「このイノシシは危険ではない」という判断へと判断が変化したとしても、それは恐怖の身体状態が続くことと整合的である。というのも、このイノシシは危険だという判断が信念的な評価であるのに対し、そのイノシシの危険さを表す身体状態は信念的ではなく、それとは別種の評価、つまり知覚的評価に類した評価であり、それゆえ認知的に侵入不可能だからなのである。また、「このイノシシは危険である」という判断によって恐怖の身体状態が生じたとしても、その身体状態はそれ自体としてイノシシの危険さを表している。なぜなら、それは危険なものを知覚したときに生じるような身体状態だからである。このように、**Damasio** らの説によると、情動に含まれる身体状態が判断と別種の仕方で外界のあり方を評価するのであり、そのような身体状態こそ情動にとって本質的なのである。

情動に関する見解として **Damasio** らの説が画期的な点は、身体状態そのものが外界のあり方の評価であるとみなすという点である。先に述べたように、私たちが外界を捉えるさまは、大別して感覚器官による知覚的認知か、そうではないような信念的認知の二通りである。こうした分類において、身体状態によって外界を評価する身体的認知は、知覚的認知の一種であると言えるだろう。だが、いわゆる五官による知覚的認知ではなく、身体による認知は、勘や第六感とみなされ、これまで詳細に説明されてこなかった。そのまさに勘や第六感を科学的な成果を踏まえ整合的に説明しようとする試みが **Damasio** らの「ソマティック・マーカー仮説」による情動論なのである。

本論文では、情動に関するこのような **Damasio** らの見方を支持し、それに依拠して論を進めていくことにする。

2.2 自己制御における情動の必要性¹³

情動が外界の価値的なあり方の評価であるならば、情動が自己制御にとって重要な役割を果たすであろうことが予想される。実際、**Damasio** はそのように議論を進めるのである。情動は自己制御の妨げになるとしばしば主張されるが、脳の特定の部位を損傷した人々は、知的な能力は損なわれないにもかかわらず、情動的な反応を示さなくなり、また同時に自己制御を欠いた振舞いを示すようになる。これは情動の低下が自己制御の喪失を招いたのではないかと考えられる。では、情動は自己制御においてどのような役割を果たすのだろうか。本節では、情動が自己制御にとって実は不可欠であることを考察する。

まず、ここでは **Damasio** が情動研究の中で詳細に描き出したエリオットという患者について少し詳しく見ていこう。エリオットは、社会的に恵まれた職業に付き、私生活においてもよき夫・よき父として豊かな生活を送っていたが、ある日、脳の前頭葉に腫

¹³ 本節は西堤（2014）第二節に多少の変更を加えたものである。

瘍が見つかり、その除去の為に手術を受けた。幸い手術も成功し、手術後も、彼の知性は損なわれることなく、記憶や認知の障害は見られなかったため、彼はまもなく元の職場へ復帰することができた。だが、彼は一見すると手術前と変わらず正常であるように見えたのだが、明らかに前の彼とは変わってしまっていたのである。彼は、事物の価値に関する知的な評価（つまり信念的認知ないし判断）を適切に行うことができ、それに基づいて複数の選択肢の中から何が最善かを知的に判断することはできるが、実際にその判断に基づいて実行することができず、何もできずにグズグズするか、目先の利益にとらわれた衝動的な行動をしてしまうかになってしまったのである (Damasio 1994, ch. 3, 邦訳, 三章)。

このようなエリオットの行動傾向は、前頭葉の手術により、前頭前野の腹内側部 (ventromedial prefrontal cortex: VMPFC) を手術によって損傷させられた結果引き起こされたと考えられる。この部位の損傷は、将来の計画を立ててそれを実現するために必要な能力や、社会規範に従う能力、そして生存のために最も有利な行動をする能力に異常を引き起こす (Damasio 1994, pp. 32-3, 邦訳, p. 78)。つまり、自分で自分を制御する能力が損なわれてしまうのである。また、VMPFC は情動に関わる部位であり、同じく情動に関わる扁桃体を中心とした辺縁系に信号を送ったり、視床下部や脳幹に信号を送ったりすることで、情動的な身体反応を引き起こす役割を担っている部位であると考えられている (村上 2010, pp. 198-213; Damasio 1994, ch. 7, 邦訳, 7章)。従って、この部位の損傷は情動の低下を引き起こす。このように、VMPFC の損傷は、自己制御に必要な能力を損なわせるとともに、情動の低下を引き起こすのである¹⁴。

では、自己制御の喪失と情動の低下はどう関係しているのだろうか。Damasio によれば、情動の低下によって自己制御の喪失が起こるのである。しかし、なぜ情動の低下は、自己制御の喪失を招くのだろうか。通常、私たちは事物を知覚したり想起したり、あるいは予期したりすると、それに対する反応としてある情動的な身体状態が生じる。たとえば、マムシを見かけると鳥肌が立つような恐怖の身体状態が引き起こされたり、大事な会議に遅刻したことを思い出すと恥ずかしくて身体が熱くなるような身体状態が引き起こされたりする。Damasio は、このような身体状態を「ソマティック反応」と呼んでおり、それはその反応を引き起こした世界の事物の価値を反映したマーカーとして働くと考えている。マムシを見て引き起こされる鳥肌が立つような恐怖の身体状態であるソマティック反応は、まさに回避すべき負の価値を持つものとしてマムシを評価しているというわけである。つまり、先に述べたように、このソマティック反応という身体的反応がそれ自体としてマムシの危険さを表しているのである。そして、そのソマ

¹⁴ 厳密に言えば、エリオットは皮質下にある扁桃体は損傷していなかった (Damasio 1994, p. 39, 邦訳, p. 88)。それゆえ、扁桃体によって誘発される一次情動は持ちうる。以下の本文の叙述では、情動を VMPFC が関係する二次情動に限定しているが、そのように限定しても議論の本筋に影響はない。

ティック反応は、VMPFC を介して引き起こされ、私たちが意思決定を行う際、それが評価する価値に合致する選択肢を選ぶようにバイアス¹⁵として働く。Damasio はこの身体反応を「ソマティック・マーカー (somatic marker)」と呼び、これがまさに私たちの情動の核心だとみなすのである (Damasio 1994, p. 173、邦訳, pp. 270-2)。

ここで注意すべきなのは、Damasio のいう情動とは、身体状態 (somatic state) の変化¹⁶を含むような状態だという点である。彼は、ある状況に直面したときに自動的に生じる身体状態の変化と、脳に伝えられるその変化した身体状態の情報を合わせて「情動 (emotion)」と呼ぶ¹⁷。ここで情動に含まれる身体状態の変化とは、たとえば体内環境と体腔内諸器官の変化 (内分泌ホルモンの遊離、心拍数の上昇、筋収縮など) や骨格筋系の変化 (姿勢や表情など) 等々である。

ソマティック・マーカーについてももう少し詳しく言えば、それは各選択肢の有利・不利に対する予測因子として働き、選択に無意識的なバイアスを掛ける。つまり、各選択肢の有利・不利を自覚的に考慮するのに先立って、好ましい選択肢と避けたい選択肢の評価 (選好の値の決定) とそれに基づく篩い分けが、あらかじめソマティック・マーカーの働きによって無自覚レベルで行われているのである。無自覚レベルでの評価と篩い分けというアイデア自体は目新しいものではないが、Damasio の仮説では、そのような無自覚レベルでの評価と篩い分けが身体状態を含む情動の働き、すなわち身体状態およびそれと情報をやり取りする大脳皮質や多数の皮質下構造体を巻き込んだ大規模な神経システムの活動によって行われるものとされており、その点で新たな提案となっているのである (Damasio 1994, pp. 180-4、邦訳, pp. 270-4)。

Damasio は、VMPFC を損傷した患者らは、その損傷の為、このソマティック・マーカーが形成されず、それゆえ物事の価値を適切に反映した意思決定が行われないと考えた。彼は情動的な身体反応が世界の価値的なあり方を示すソマティック・マーカーとして働き、そのソマティック・マーカーとしての身体状態が意思決定および行動を決定的に左右するという考えを「ソマティック・マーカー仮説 (Somatic Marker Hypothesis: SMH)」(Damasio 1994, pp. 173-5、邦訳, pp. 270-4) と呼んだ。この仮説によれば、情動の低下が自己制御の喪失をもたらすのは、情動の低下により、物事の価値を表すソマティック・マーカーとしての身体的反応が形成されないために、物事の価値を適切に反映した意思決定が行われず、それゆえ不合理な行動が為されてしまうか

¹⁵ 心理学や脳科学で用いられる「バイアス」というのは、たんに意思決定において何らかの影響を及ぼす働きを意味し、必ずしも合理的な意思決定をゆがめる悪い働きを意味するわけではない。

¹⁶ 現実の状況によって引き起こされる身体状態の変化だけではなく、想起、想像、予期によって引き起こされる身体状態の変化もある (Damasio 1994, pp. 134-9、邦訳, pp. 219-25)。

¹⁷ Damasio は、脳に伝えられる身体状態の情報が意識的に感受されるとき、その情報をとくに「感情 (feeling)」とよぶ。このようにして Damasio は「情動」と「感情」を区別している (Damasio 1994, pp. 143-8、邦訳, pp. 230-36)。

らである。しかし、情動が低下しても、システム2がちゃんと機能していれば、システム2によってシステム2の決定通りに行為が行われて自己制御がちゃんとできるということにならないのか、という問題が残る。この問題については章を改めて詳しく考察することにしよう。但し、次の章では、ソマティック・マーカー仮説について、もっと詳細に論じておくべきことがあるので、まずそれを論じ、その後でこの問題を論じる。

日常生活の中で、ひとびとはかならずしも合理的とは思えない行動をとる。しかも、それらの行動の中には、単なる情報の不足に帰因させることができない事例も多数含まれている。たとえば、移動手段の選択などがそうである。合理的なリスク計算の結果、都市の間を自動車で移動するよりも、飛行機を利用したほうが、はるかに生存率が高いことを示されても、飛行機より自動車による移動を選択してしまうひとのほうが圧倒的に多いのである (Damasio 1994, pp. 191-192、邦訳, p.295-6)。このような合理的とは言い難い行動傾向を理解するためには、従来の経済学が前提している合理的行為者としての人間というモデル (von Neumann and Morgenstern 1944) では不十分であるように思われる。

従来の経済学が前提するモデルは、人間の行動を信念と欲求によって合理的に説明しようとするモデルであり、本来的に人間は合理的な意思決定者であるとみなされ、不合理な行動は、結局は情報の不足や信念の誤りに帰着させられる。しかし、私たちの行動には、単に情報の不足や信念の誤りでは説明できないような不合理な行動が多数ある。また、必ずしも不合理ではない行動でも、直感による行動のように、信念と欲求だけでは説明しがたいと思われる行動もある。

A. R. Damasio は、そのような人間の行動には身体的な情動が深く関与していると考え、前章で述べたように、ソマティック・マーカとして働く情動によってすべての行動が影響されると主張する「ソマティック・マーカ仮説」(SMH)を提案した (Damasio 1994, pp. 173、邦訳, pp. 270-2)。この仮説は、不合理な行動だけではなく合理的な行動もすべて、情動の働きが関与すると主張する点に大きな特色がある。ところで、情動や意思決定と深い関わりを持つと考えられている前頭前野の研究はまだ始まったばかりで、まだまだ多くの謎を解決できずにいる。従って現段階では SMH は、証拠となる実験データやその解釈についてはもちろんのこと、仮説の内容そのものの明瞭さについても不十分さを免れない。そこで本章では、SMH とその証拠として示されている実験を概観した後、その実験結果によって SMH がどの程度確証されているのか、また Damasio による実験結果の解釈にどのような問題があるのかを明瞭にし、それとともに SMH の内容そのものの明確化をはかり、情動が自己制御にどう関係するかをさらに明らかにしたい。

3.1 アイオワ・ギャンブル課題¹⁹

Damasio とその同僚たちは SMH を実験的に検証するために、「アイオワ・ギャンブル課題 (Iowa Gambling Task)」とよばれる実験を考案している (Damasio 1994, p.

¹⁸ この章は西堤優 (2010a) に多少の変更を加えたものである。

¹⁹ 本節では、主として Bechara et al. (1997) に基づく。

212-2、邦訳, p. 322-34)。そこで、このギャンブル課題の内容と結果を少し詳しく見ておこう。これは SMH にとって唯一と言ってもよい実験的証拠を提供した課題であり、VMPFC の損傷に起因する情動の欠如によってもたらされる行動異常を劇的な形で示している。

3.1.1 実験目的と実験デザイン

この実験の目的は、熟慮（意識的な考慮）に先立つ段階ですでに、被験者が長期的な利害に基づく選択を行うかどうかを判定することである。つまり、熟慮によらないでも合理的な選択が可能かどうかを判定することである。その判定のために次のような課題が被験者に課せられた。

被験者はテーブル上の 4 組のカードの山（A 組、B 組、C 組、D 組）から、実験者によって止められるまで、自由にカードをめくることができる。カードの裏にはプラスあるいはマイナスの金額が記入されており、被験者はめくったカードに記入された額に応じてお金をもらうか、あるいは支払う。被験者は最終的により多くの利益を得るように求められる。4 組のカードのうち、A 組と B 組はハイリスク・ハイリターン、C 組と D 組はローリスク・ローリターンの組となっている。つまり、具体的に言うと、A 組と B 組の場合、大きな額のプラスのカードとマイナスのカードからなり、10 回ごとの総利益が 1000 ドルであるが、損失額は 1250 ドルとなっており、長期的には損失をもたらす。他方、C 組と D 組では、小さな額のプラスのカードとマイナスのカードからなり、10 回ごとの総利益は 500 ドルであるが、損失額は 250 ドルで、長期的には利益をもたらす。被験者にはこのようなそれぞれの山の特徴を教えず好きなようにカードを引かせるが、A 組と B 組、C 組と D 組のように特徴の似た山を用意したのは、カードを引くことですぐに山の特徴が分かってしまうのを防ぐためである。被験者にはあらかじめ 2000 ドルの貸付金が与えられている。被験者は何枚のカードをめくったらゲームが終了するのかは知らされていない。

この実験において記録されたデータは、被験者がめくった各組のカードの枚数と、課題中に生じた被験者の情動的な身体状態の変化の指標となる SCRs (skin-conductance responses : 皮膚伝導反応) の強度変化²⁰、そしてゲーム途中でのゲームについての理解および採用している戦略に関する被験者自身の口頭説明（実験途中で実験者の質問に答えて行われる報告）の三つである。

3.1.2 実験結果

記録されたデータを詳しくみてみると、健常者と VMPFC 損傷患者の間で大きな違いがみられた。健常者の場合、A 組と B 組に含まれる大きな額の損失カードを数回めくったあとでは、それ以降、どの組からカードをめくろうとしても、めくる直前に著し

²⁰ SCRs とは、汗の分泌量によって影響を受ける皮膚の電気伝導率の変化である。

く強い（ほぼ3倍強度の）予期的なSCRsが出るようになった。つまり、カードをめく
ることを考えただけで、強いSCRsが出たのである。そして、さらに課題が進むと、健
常者は、「A、B組は不利な気がする」という予感を口にする段階（「予感期」）に達し、
この段階では、予期的なSCRsは不利な組であるA、B組からカードを引くことを考え
るときのほうがそうでないC、D組からカードを引くことを考えるときよりも明らかに
大きくなった。そして実際の選択も、A、B組よりもC、D組からカードを多く引くよ
うになった。さらに課題が進むと、健常者はたんなる予感ではなく、なぜ、A、B組が
不利であり、C、D組が有利かを意識的に述べることができる段階（「概念期」）に到達
し、やはり一貫して有利なC、D組からのカードをめくり続け、不利なA、B組からカ
ードをめくろうとすると、強い予期的なSCRsが生じた。

一方、VMPFC 損傷患者の場合、A組とB組の大きな額の損失カードをめくった後
でも、予期的なSCRsの強度に顕著な変化が見られず、それは最後まで常に低いままで
あった。また、彼らには予感期が見られなかったが、彼らも、それぞれの組がなぜ有利
ないし不利かを意識的に述べるができる概念期に到達することはできた。しかし、
驚くべきことに、彼らの多くは概念期に到達した後でさえ、依然として不利な組から
の選択に固執し続け、有利な組からの選択と同程度か、あるいはそれを上回るくらい、不
利な組からの選択を行い続けたのである。

3.1.3 Damasio の考察

このギャンブル課題の実験により、健常者は、課題の特徴（各組がなぜ有利ないし不
利なのか）を明確に意識的に理解している概念期に達する前に、すでに無意識的に、長
期的には有利な選択を行っていることが示された。つまり、彼らには概念期に先だつて
予感期があることが見出されたのである。しかし、損傷患者のほうには予感期がなく、
しかも彼らは、概念期に達した後でさえ、長期的には不利になる選択をし続けることが
見出された。Damasioはこの実験結果から、健常者たちの場合は、意識的な評価に先
立って、無自覚的に形成される情動的な身体反応（SCRsによって示される反応）によっ
て、長期的には利益となるような選択が予感期および概念期になされるのに対し、患者
たちの場合は、そのような情動的な身体反応が形成されないために、そうした選択が概念
期でさえ行えないとみなした。

こうしてDamasioは、アイオワ・ギャンブル課題を用いた実験の結果が、身体的な
情動反応が意思決定と行動に大きな影響を及ぼすと唱えるSMHの非常に重要な実験
的証拠になると解したのである。

3.2 問題点

前節では、SMHを支えるギャンブル課題のデータやそれに対するDamasioの解釈
を見てきたが、このDamasioの解釈についてはいくつか問題がある。以下では詳しく

そのことを論じ、それを通じて SMH の内容をより明瞭化していこう。そのために、まず最初に、すでに Colombetti によって SMH の批判がなされているので、それを検討してみよう。

3.2.1 Colombetti の批判

Colombetti (2008) は、SMH において情動が演じる役割には曖昧さがあり、それをなくすためには、SMH を一般性の異なる二つの仮説に分離したほうがよいと主張する。そしてその上で、彼女はどちらの仮説もギャンブル課題の実験によっては十分確証されているとは言えないと批判する。

まず、Colombetti は、SMH を「SMH-G」という一般的な仮説と「SMH-S」という限定的な仮説に区別する。SMH-G では、情動は短期的な利害か長期的な利害かを問わず、一般に任意の利害を反映するものとされ、そのような情動によって複数の選択肢のそれぞれに選好の順位が付けられるとする。それに対して、SMH-S では、情動は長期的な利害のみを反映して、選択にバイアスを掛けるとされる。

このように Colombetti は、SMH を SMH-G と SMH-S に明確に分離してその明確化を図った上で、そのどちらの仮説も、ギャンブル課題の実験によっては十分な裏付けがなされていないと主張する。まず、一般的な SMH-G については、次のように批判する。すなわち、ギャンブル課題についての Damasio の解釈では、VMPFC 損傷患者は、予期的な情動によるバイアスが掛からない状態で選択を行っており、その時の患者の振る舞いはいかなる選好の結果でもないと言われることになるが、このことは実験結果のデータによって明確に示されているわけではない。実験結果のデータをみる限りでは、VMPFC 損傷患者は、選好が欠如しているわけではなく、むしろ不利な組をかなり選択している。それゆえ、Colombetti に言わせれば、彼らは不利な組を好むという極めて特異な選好をもつと解釈することも可能である。つまり、Damasio の解釈はデータの可能な解釈の一つに過ぎず、ギャンブル課題は、患者に特異な選好を帰属させる解釈を排除するにはデザインされていないのである。

また、Colombetti は、情動の役割を限定した SMH-S のほうも、ギャンブル課題の実験によっては必ずしも支持されないと主張する。この仮説では、情動は、長期的な利害のみを反映して選択にバイアスを掛けるという役割を果たすが、Colombetti は Tomb et al. (2002) らが行った実験結果を援用して必ずしもそうは言えないと指摘する。Damasio らが設定したギャンブル課題では、カードを一回引くごとの利益と損失は、長期的には不利な結果をもたらすカードの組の方が、長期的には有利な結果をもたらす組よりも大きな額となっているが、Tomb らの課題では、これを逆転させて、有利な組における一回ごとの利益と損失を、不利な組のそれよりも大きくした。その結果、情動の指標として用いられた SCRs は、不利な組からカードを引く直前ではなく、有利な組からカードを引く直前に検出されたのである。Tomb らは、この結果から、SCRs は長

期的な利害を反映するのではなく、目先の一回ごとの利害の大きさを反映するのだと考えた。Colombetti はこれを受けて、Damasio らのギャンブル課題だけでは、情動が長期的な利害ではなく、むしろ目先の利害を反映すると解釈することも可能だと指摘する。

このように、Colombetti は、SMH をより明確化するためにそれを二つの仮説に分離し、その上でそれぞれの仮説を検討することによって、どちらの仮説も、ギャンブル課題だけでは、Damasio らが主張するほど強くは確証されていないと論じる。このColombetti の批判は確かにおおむね妥当だと思われる。しかし、SMH-G の批判の際に出された解釈、すなわち VMPFC 損傷患者は不利な組を好むという特異な選好を持つという解釈は自然なものとは言い難い。VMPFC 損傷患者のカード選択を説明するもっとも自然な解釈はないだろうか。この点については、次項で詳しく検討したい。また、SMH-S の批判においては、Tomb らの実験を援用して、情動が目先の利害を反映している可能性があるという解釈が出されたが、たしかにその解釈も可能である。しかし、情動が長期的な利害を反映するという解釈もなお排除されていない。そしてそのようにDamasio に好意的に解釈したとしても、ギャンブル課題が SMH を支持するとする彼の解釈には、まだまだいろいろな問題がある。以下では、それらを検討しながら、SMH の更なる明確化を図っていこう。

3.2.2 情動による行動の説明はどこまで可能か

ギャンブル課題において VMPFC 損傷患者は、健常者と違って、長期的には有利な状況をもたらす組 (C、D 組) からではなく、不利な状況をもたらす組 (A、B 組) から選択的にカードをめくる。Damasio によると、損傷患者は VMPFC の損傷により、長期的には不利なカードの組に対して予期的な情動をほとんど欠くために、それらの組を回避せずに選択してしまうのである。しかし、VMPFC 損傷患者は、A、B 組だけではなく、C、D 組に対しても、同じようにほとんど予期的な情動を示さない。そうだとすれば、SMH に従うかぎり、彼らはどの組に対しても特別な選好をもたないはずである。従って、彼らは結局、A、B、C、D すべての組からランダムにカードを引くことになるのではなかろうか。ところが、実際には、損傷患者は A、B 組から好んでカードを引く。予期的な情動が欠如しているにもかかわらず、なぜどの組に対してもとくに選好を持っていないような (つまり、ランダムにすべての組からカードを引くという) 振る舞いをするのではなく、とくに A、B 組に対して選好を持つような振る舞いをするのだろうか。この点が、Damasio の解釈では十分に説明されていない。

3.2.2.1 ポジティブな情動による説明

Damasio によると、この実験において、予期的な SCRs は、長期的には不利な状況を回避するように働く予期的な情動を反映しており、しかも予期的な SCRs によっては反映されないような予期的な情動は存在しない (Damasio 1994, p. 209, 邦訳, p. 316)。

VMPFC 損傷患者は、この予期的な SCR の値が著しく低く、不利な状況を回避するためのバイアスが働いておらず、それゆえ、不利な状況を選択するとみなされる。しかし、SMH によれば、意思決定において予期的な情動が不可欠であり、ある選択肢に対して選好を持つには、長期的には不利な選択肢であれ有利な選択肢であれ、なんらかの予期的な情動が必要不可欠なはずである。従って、損傷患者のように A、B 組に対して選好を持つのであれば、SCR では検出できないような、A、B 組に対して選好をもたせる予期的な情動が存在しなければならないのではなかろうか。

課題中の VMPFC 損傷患者の特異な行動を説明するために、損傷患者は罰への感受性（予期される罰に対してネガティブな情動を引き起こす傾向性）を欠くが、報酬への感受性（予期される報酬に対してポジティブな情動を引き起こす傾向性）は保持すると仮定してみよう。そうすると、彼らは、罰への感受性を欠くために、A、B 組に対してとくにネガティブな予期的情動をもつことはないが、報酬への感受性を保持しているために、A、B 組に対して C、D 組よりも大きな値のポジティブな予期的情動をもつことになる（報酬だけ見ると、A、B 組は C、D 組よりもはるかに大きな報酬が与えられる）。そのため、より多くのカードを A、B 組から引くのだと解釈される。

この解釈を立証するためには、情動には、SCR で検出されるような、罰を回避させるネガティブな情動と、SCR では検出されないような、報酬の獲得を促すポジティブな情動の二種類があることを示さなくてはならない。しかし、SCR の測定によっては、ポジティブな情動を検出できないということについては、きわめて疑わしいように思われる。実際、Damasio さんもポジティブな情動が SCR の測定によって検出されることを認めている（Damasio et al. 2002）。従って、VMPFC 損傷患者がなぜ長期的には不利な状況をもたらすようなカード選択を好むのかということ、SCR では検出できないようなポジティブな情動の存在によって説明しようすることは、あまり望みがないということになる。むしろ、情動によらない説明が必要であるように思われる。

3.2.2.2 目先の報酬への傾向性による説明

これまで述べてきたように、VMPFC 損傷患者が A、B 組からカードを好んで引くことは、ただランダムにカードを引くのと違い、A、B 組から引くようにバイアスを掛ける何らかの要因があると考えられる。しかし、それはどうやら情動ではなさそうである。むしろ、人間には本来的に目先の報酬を求める傾向性²¹があり、それが損傷患者に A、B 組から好んでカードを引くという行動を起こさせているのではなかろうか。つまり、損傷患者は、罰へのネガティブな情動を欠くだけでなく、報酬へのポジティブな情動も欠いているが、目先の報酬を求めるという本来的に人間に備わった傾向性をもつ

²¹ ここで言う目先の報酬とは、目先の正の報酬のことであり、目先の報酬を求める傾向性とは目先の正の報酬を求める傾向のことであって、目先の負の報酬（罰）を避ける傾向は含まない。この目先の報酬を求める傾向性は、システム 1 に属すると考えられる。

ており、それによって A、B 組から好んでカードを引くのではなかろうか。ここで想定される目先の報酬への傾向性とは、生得的もしくは幼児期に獲得された神経回路によって固定化されているような傾向性であり、そのつどの情動とは無関係に働くものである。健常者であれば、いくらそのような目先の報酬への傾向性をもっているとしても、ネガティブな情動によるバイアスが適切に働くので、大きな罰カードを含む A、B 組を回避して、C、D 組を選択するようになり、目先の利益に惑わされずにすむ。しかし、損傷患者の場合は、そのようなバイアスが適切に働かないために、A、B 組を回避できず、ただ本来の目先の報酬への傾向性に従って A、B 組からカードを引いてしまうのである。

このような目先の報酬への傾向性が人間にあることを示すためには、それを支える神経回路の発見が必要であろう。しかし、現在のところ、そのような回路の存在を示す証拠はまだ見出されていない。それは、今後の脳科学の発展を待たなければならないだろう。

なお、この目先の報酬への傾向性による説明を受け入れる場合、ギャンブル課題の実験において、SCRs は情動全体を反映しているものとみなされることになり、VMPFC 損傷患者の特異な行動選択はまったく情動によるものではないことになる。従って、この説明が正しければ、SMH は部分的に否定されることになる。そしてこのような部分的な否定は妥当だと思われる。なぜなら、確かに人間には情動とは無関係な選択や行動の傾向性が生得的に存在したり、あるいは習慣的に形成されたりすると考えられるからである。

また、情動とは無関係な傾向性の導入に加えて、SMH をより明瞭にするためには、ポジティブな情動とネガティブな情動を区別し、それぞれがどのような働きをするのかを SMH に組み込む必要があるということもいえるだろう。

3.2.3 事前の SCRs と事後の SCRs

Damasio によると、VMPFC は長期的には不利な状況を避けるための予期的な情動を形成するのに決定的に重要な役割を果たす。VMPFC が損傷すると、意思決定にバイアスを掛ける予期的な情動が形成されなくなる。ところが、VMPFC 損傷患者は、予期的な SCRs が検出されないにもかかわらず、A、B 組から大きな損失カードを引いた後、しばらく A、B 組を避けてもつぱら C、D 組から引くという行動を示す (Damasio 1994, p. 217, 邦訳, p. 328)。確かに、その後はまた A、B 組から好んでカードを引くようになるのだが、直後はしばらく C、D 組から引くのである。VMPFC の損傷により、長期的には不利な状況を避けるための予期的な情動が生じないのであれば、いくら A、B 組から大きな損失のカードを引いたとしても、その直後も迷わず A、B 組から引くのではないか。このことは SMH の反証となりかねない事実であるが、Damasio はそれに説明を与えていない。

Damasio によると、ギャンブル課題の中で検出される SCRs は、あるカードを引こ

うと考えたときに、それによって生じる予期的な情動の指標であり、そのカードによってもたらされると予想される利益ないし損失を反映している。しかし、実際には、事情はもう少し複雑である。あるカードを引こうとするとき、そうしようという思考によって生じる予期的な SCR_s だけではなく、前のカードを引いたときに生じた SCR_s もいくらか残存している。あるカードを引いたとき、その結果（利益もしくは損失）に応じて事後的な SCR_s が形成される（Bechara and Damasio 2005, p. 346, Fig. 4, p. 347, Fig. 5）。従って、あるカードを引こうとするときの事前の SCR_s は、そのカードを引こうという思考によって引き起こされる予期的な SCR_s と、前のカードを引いたときに生じる事後的な SCR_s が残存したものとを足し合わせたものである。

健常者の場合、あるカードに対する事前の SCR_s は、確かにそのカードに対する予期的な SCR_s と前のカードに対する事後的な SCR_s の残存したものの両方からなる。しかし、VMPFC 損傷患者の場合、予期的な SCR_s が形成されないため、事前の SCR_s は前のカードに対する事後的な SCR_s が残存したものだけである。予期的な SCR_s は VMPFC によって形成されるが、事後的な SCR_s は扁桃体によって形成される。それゆえ、VMPFC 損傷患者は、VMPFC を損傷しているだけで、扁桃体を損傷しているわけではないから、事後的な SCR_s を形成することは可能である。VMPFC 損傷患者において、A、B 組からカードを引こうとするときに検出される事前の SCR_s は、前に A、B 組からカードを引いたときの事後的な SCR_s の余韻であり、予期的な SCR_s をまったく含んでいない。しかし、それでも、事後的な SCR_s が残存するために、損傷患者は A、B 組からカードを引くことを避けて、C、D 組からカードを引くのである。そしてその残存する SCR_s が次第に弱まってくると、再び A、B 組からカードを引くようになるのである。

このようにして、VMPFC 損傷患者が予期的 SCR_s を欠くにもかかわらず、A、B 組から大きな損失カードを引いた後しばらくは、それらの組からカードを引かないという事実は、残存する事後的 SCR_s によって説明することが可能であるように思われる。そうだとすれば、その事実は SMH への反論とはならず、むしろ SMH のもとで説明できることとなる。

3.2.4 「長期的な利害の考慮」の多義性

Damasio の解釈では、ギャンブル課題において、カードを引く前に検出される予期的な SCR_s は、カードの選択を長期的な見通しに基づくものにするような情動の指標とみなされる。健常者の場合、予感期と概念期に達すると、そのような予期的な SCR_s が検出され、そのため長期的な考慮に基づく選択がなされるとされる。しかし、VMPFC 損傷患者の場合は、そのような予期的な SCR_s が検出されないため、長期的な利害に基づく選択がなされないのだとされることになる。

しかし、このギャンブル実験では、明らかに、損傷患者たちもやがて概念期に達する。

つまり、彼らも、どの組からカードを多く引くのがなぜ長期的には有利かを意識的に理解するようになる。そうであれば、情動によって長期的な利害に基づく選択を行えないとしても、意識的な理解に基づいてそのような選択を行うことができるのではないだろうか。彼らはなぜ概念期に達してもなお、不利な組からカードを引き続けてしまうのだろうか。その理由として、Damasio によっては必ずしも明示されていない二つの可能性が考えられる。

まず、VMPFC 損傷患者はそもそも、情動の欠如のゆえに、長期的な利害の意識的な考慮を意思決定のプロセスに活かすことができず、それゆえ長期的には有利な意思決定を行うことができないのかもしれない。彼らはたしかにどの組が長期的には有利かを考慮することができるが、その考慮は単なる理論的なものとどまってしまう、意思決定にバイアスをかける実践的な考慮とはならない。従って、彼らは長期的には有利な意思決定を行うことができず、目先の利益に基づいた衝動的な意思決定を行ってしまうのである。

しかし、もう一つ別の可能性も考えられる。すなわち、VMPFC 損傷患者は、どの組が長期的には有利かを考慮して、それに基づいて長期的には有利な組からカードを引こうという意思決定を行うことができるが、それにもかかわらず情動の欠如のために、その意思決定を行動に移すことができないのかもしれない。彼らは長期的な利害の考慮を意思決定に活かすことはできる。つまり、彼らの考慮は意思決定にいたる実践的な考慮の一部となっており、意思決定に何の影響も与えない単なる理論的な考慮ではない。しかし、せつかく長期的な利害を考慮に入れて意思決定を行っても、情動の欠如のゆえに、その意思決定を行動に移すことができないのである。つまり、意思決定を行動に移す「執行プロセス」²²には情動が必要であり、それゆえ、情動の欠如は執行プロセスの正常な働きを妨げるのである。これは、意思どおりに行為できないという意味での意志の弱さともいえる。情動の欠如は、執行プロセスの不全、ないしその意味での意志の弱さをもたらすのである。

VMPFC 損傷患者は、長期的な利害を考慮できるが、それでも長期的には有利な行動を行うことができない。それは、彼らがそもそも長期的には有利な意思決定を行うことができないからだろうか、それとも長期的には有利な意思決定を行うことができるものの、その意思決定を行動に移すことができないからだろうか。言い換えれば、彼らの長期的な利害の考慮は意思決定のプロセスにバイアスを掛ける実践的な考慮なのだろうか、それとも、単なる理論的な考慮に過ぎないのだろうか。さらに言い換えれば、情動は意思決定のプロセスに関与するのだろうか、それとも意思決定から行動にいたる執行プロセスに関与するのだろうか。もし、情動が意思決定から行動へ至る執行プロセスに

²² Strack and Deutsch (2004) は、行動を実行するには、行動の決定とそれを実行に移す行動スキーマの働きが必要であるとして、行動の決定プロセスと決定された行動を実行に移すプロセスを区別している。

働きかけていたとしたら、意思決定に情動が影響を及ぼすと主張している SMH の枠組みそのものが危うくなる。Damasio は情動が意思決定に影響を及ぼすと主張しているので、第一の可能性を擁護すると考えられるが、この二つの可能性を明示的に考察して、第一の可能性を支持しているわけではない²³。

情動が意思決定に影響を及ぼすという SMH の根本的な枠組みの妥当性を明らかにするために、この二つの可能性のいずれが正しいのかを明らかにする必要がある。但し、もちろん、第三の可能性も考えられる。すなわち、情動は意思決定のプロセスと執行のプロセスの両方に影響を及ぼすという可能性である。後の章で詳しく考察するように、意志の弱さのあり方を考えると、この第三の可能性が妥当であるように思われる。そうだとすれば、SMH は情動が意思決定に影響を及ぼすとするだけでなく、意思決定から行為の実行に至る過程にも影響するとして、その拡張をはかったほうがよいだろう。

3.3 まとめ

これまで、SMH の内容そのものの明瞭化を行うために、SMH が、その証拠として挙げているギャンブル課題の結果によって、どの程度確証されているのか、また Damasio によるその実験結果の解釈にはどのような問題があるのかを指摘してきた。その結果、情動が意思決定に影響を及ぼすといっても、情動にはネガティブなものもポジティブなものがあり、それぞれがどのように意思決定に影響を及ぼすかを明確にする必要があることを明らかにした。また、情動がなくても目先の報酬への傾向性によって意思決定が行われる可能性があることを示して、SMH を部分的に修正する必要があるかもしれないことを示唆した。さらに、ある種の行動を SMH のもとで説明するためには、予期的な情動だけではなく、以前の行動の結果として生じた事後的な情動の残存したものが意思決定に影響を及ぼすことを考慮に入れる必要があることを論じた。これらの問題提起は、情動が意思決定に影響を及ぼすという SMH の根本的な枠組み内部での問題提起であった。それに対して、SMH の根本的な枠組み自体を問う問題提起も行った。すなわち、情動は意思決定のプロセスに影響を及ぼすのではなく、決定した意思を行動へともたらず執行プロセスに影響を及ぼすのかもしれないという可能性、あるいはその両方に影響を及ぼす可能性を指摘した。

これらの問題を最終的に解決するためには、たんに思弁的な考察だけではなく、認知科学や脳科学などによる更なる経験的な探求が必要である。しかし、そのような探求をどのように進めるべきかに関して、一つの方向性を示すという意味でも、このような問題提起は決して無意味なものではないように思われる。

²³ Damasio は VMPFC 損傷患者のエリオットを詳細に観察・研究しているが、その報告をみると、エリオットはそもそも意思決定ができないのか、それとも決定した意思を行動へ移すことができないのかが判然としない (Damasio 1994, ch. 3、邦訳, 3 章)。

第四章 知的評価と情動²⁴

情動が意思決定に影響を及ぼすとき、情動に含まれる身体状態は意思決定においてどのような役割を果たすのだろうか。身体状態は情動をまさに情動たらしめる最も本質的な要素であり、従って情動が意思決定に影響を及ぼすのなら、身体状態が何か重要な役割を果たしているはずである。それはどんな役割であろうか。本章では、この問題を考察していきたい。そのために、まず 4.1 では、Damasio の SMH に依拠しつつ、情動に含まれる身体状態が外界の価値を反映することにより、その身体状態を感受した脳状態もその同じ外界の価値を反映し、直接的にはこの脳状態が意思決定に影響を及ぼすことを示す。続いて 4.2 では、Damasio において情動は本来身体状態を含むものであるが、それを含まない「あたかも身体的情動」の存在を Damasio は認めており、このあたかも身体的情動は本来の情動とどのような関係があるのか、また、あたかも身体的情動がどのように外界の対象の価値を表すのかを考察する。4.3 では、情動による価値評価とは別に、身体状態から独立したまったく身体性を帯びない純粋に知的な価値評価が存在すると考えられるが、それがどのようなものかを考察する。以上を準備として、身体的情動、あたかも身体的情動、知的価値評価という三種類の価値評価がそれぞれ意思決定においてどのような役割を果たすのかを考察し、それを通じて意思決定における身体状態の役割を明らかにしたい。その過程で、価値評価と動機付け、意思決定の三つの関係の見直しを行うことにもなる。

4.1 身体状態・脳状態・意思決定

すでに述べたように、Damasio によると、情動が生じるとき、まず外界の刺激によって身体状態の一連の変化が引き起こされ、続いてその身体状態の変化を感受する脳状態が形成される。身体状態の変化には、たとえば頬の紅潮、姿勢、発汗など外部の観察者が見てとれるものから、動悸や筋肉の硬直といった本人にしか直接的には分からない変化まで含まれる。Damasio においては、情動はこのような身体変化とそれを感受する脳状態の両方からなる (Damasio 1994, pp. 131-9、邦訳, pp. 215-22)。そして身体変化を特に意識的に感受する脳状態を「感情 (feeling)」と Damasio はよぶ (Damasio 1994, pp. 143-6、邦訳, pp. 230-33)。私たちがふつう恐怖や悲しみなどといっているのはこの「感情」に相当すると言えよう。

さて、情動がこのようなものだとする、それはどのようにして意思決定に影響を及ぼすのだろうか。Damasio によれば、情動は単に意思決定に影響を及ぼすだけではなく、意思決定を合理的なものにするのに貢献するはずである (Damasio 1994, pp. 173-4、邦訳, pp. 270-1)。では、いったいどのようにしてそのような貢献を行うのだろうか。情動における身体状態の変化は、それを引き起こした外界の対象の価値を反映している。

²⁴ この章は西堤 (2011) を加筆修正したものである。

しかし、情動が意思決定に関与するとき、それに直接関与するのは、身体状態の変化ではなく、それを感受した脳状態である。脳状態も外界の対象の価値を反映しているのだろうか。

Prinzによれば、身体状態の変化が脳に伝えられてそれを感受する脳状態が形成されるとき、この脳状態は身体状態の変化を反映している (Prinz 2004, pp. 68-9, pp. 77-8)。従って、この脳状態は、それが反映する身体状態が外界の対象の価値を反映することによって、間接的に外界の対象の価値を反映することになる。こうして Prinz は、身体状態によって引き起こされた脳状態は外界の対象の価値を反映していると考えるのである。

そうすると、結局、情動は、それに含まれる身体状態と脳状態がともに外界の対象の価値を反映することになる。情動が合理的な意思決定に貢献しうるのは、情動に含まれる脳状態（身体状態の感受）が外界の対象の価値を反映することから説明できる。たとえば、ヘビを目にすることで恐怖の情動が生じるとき、まずヘビを目にすることで恐怖に特徴的な身体状態が生じ、それが脳に伝えられる。この脳状態は、身体状態が反映している対象の価値、すなわちヘビが危険であるという価値を反映する。このような脳状態が意思決定にバイアスを掛けることにより、ヘビの危険性を考慮に入れた意思決定が可能になる。つまり、ヘビの危険性とその他の関連することがら（たとえば、ヘビから逃げると笑い者になるだろうといったこと）を比較考量して、ヘビから逃げる（あるいは逃げない）という意思決定を行うという仕方で、ヘビの危険性が考慮に入れられるのである。こうして、情動が外界の対象の価値を表すことにより、そのような価値のあり方に応じた意思決定が可能になる。情動は、外界の対象の価値に適応した意思決定を可能にするという意味で、合理的な意思決定に貢献するのである。

情動が合理的な意思決定に貢献するとき、身体状態が中心的な役割を果たしていることを強調しておきたい。意思決定に直接関与するのは、情動に含まれる脳状態のほうである。身体状態は脳状態に反映されることで、間接的に意思決定に関わるに過ぎない。しかしながら、外界の対象の価値を直接反映するのは身体状態である²⁵。脳状態は身体状態の反映を通じて間接的に外界の対象の価値を反映するに過ぎない。ヘビを目にして恐怖が生じたときに、その場から立ち去るという意思決定を行う場合、その意思決定に直接関与するのは、ヘビが危険であるという価値を反映する脳状態であるが、その脳状態がそのような価値を反映しうるのは、身震いなどの身体状態がヘビの危険性を反映しているためである。もし、身体状態が外界の対象の価値を反映しなければ、脳状態は外

²⁵ 身体を介さないで価値が知覚できるということも考えられるが、それはあたかも身体的情動だと解釈できるかもしれない。人間における原初的な価値の把握というのは、身体的情動という形をとり、それが繰り返されることで、確かに身体性を失ったあたかも身体的情動になる。身体性抜きで価値の知覚と言われるものは、このあたかも身体的情動に他ならないと考えることができるかもしれない。だがここでは、この問題を特に論じることはしないことにする。

界の対象の価値を反映しえず、それゆえ外界のあり方に適した合理的な意思決定は不可能になるだろう。このように、身体状態は、外界の事物の価値を直接反映するという点で、情動が合理的な意思決定に貢献するのに中心的な役割を担っているのである。

情動が合理的な意思決定に貢献するとき、情動に含まれる身体状態はこのように無くてはならない重要な役割を担っているように思われる。しかしながら、Damasio によると、情動には、身体状態を含むものだけではなく、それを含まない「あたかも身体的情動」が存在する。次に、あたかも身体的情動が外界の対象の価値を反映して合理的な意思決定に貢献しうるのかどうかを考察しよう。

4.2 あたかも身体的情動と身体的情動

これまで見てきたように、身体状態を含む情動（以下では「身体的情動」と呼ぶ）は、外界の価値を反映しており、それゆえ外界のあり方に適した合理的な意思決定を可能にする。Damasio はこのような外部の事物を直接的な原因として引き起こされる身体的情動が、熟慮の評価に比べて迅速で、場当たりの決定よりの確な意思決定を可能にする要因であることを認めている (Damasio 1994, pp. 173-4, 邦訳, pp. 270-2)。しかし、彼は同時に、SMH の枠組みの中で、身体状態を含まないが、あたかも含んでいるかのような情動（以下では「あたかも身体的情動」と呼ぶ）を認めており、それが意思決定において一定の影響を及ぼすことにも言及している (Damasio 1994, p. 155, 邦訳, p. 245)。もしあたかも身体的情動が意思決定に影響を及ぼすとすれば、それはどのような影響を及ぼすのだろうか。それは合理的な意思決定に貢献しうるのだろうか。そして、もしあたかも身体的情動が身体的情動と同様に合理的な意思決定に貢献できるとしたら、身体的情動に含まれる身体状態は、情動が合理的な意思決定に貢献するうえで、実は不要だということにならないだろうか。本節では、この二つの問題に焦点をあてて考察する。そのために、まず身体的情動とあたかも身体的情動がそれぞれどのように生じ、どう異なるのかを見ていこう。

Damasio によると、すでに述べたように、身体的情動は、現前する刺激により身体状態が誘発され、その身体状態から脳状態が形成されることによって生じる情動である²⁶。たとえば、ヘビを目にすることで身震いが引き起こされ、その身震いからそれに応じた脳状態が形成されることにより、恐怖の身体的情動が生じる。この過程をもう少し

²⁶ 身体的情動における身体状態は、詳しく言えば、非常に複雑な事象の連鎖によって生じている。まず、情動を誘発しうる対象の刺激が、感覚連合皮質と高次の大脳皮質に伝えられ、そこで対象の知覚が形成されて、その知覚信号が脳の別の場所にある扁桃体や VMPFC といった情動誘発部位で利用可能になる。情動誘発部位で処理された信号は、さらに神経回路を介して、情動実行部位に伝えられ、この部位の活動によって身体反応が引き起こされる。情動実行部位は、視床下部、前脳基底部、脳幹の被蓋にある核などから成る。これらはホルモン物質あるいは末梢神経系を通じて、身体の内部環境、内臓、骨格筋システムを、ある時間、特定のパターンに変化させ、特定の身体状態をもたらすのである (Damasio 2003, pp. 57-64, 邦訳, pp. 87-95)。

詳しく述べれば、まずヘビを目にすると、その視覚情報が、網膜から視覚皮質を介して扁桃体や VMPFC といった、恐怖に関係する脳部位へ送られる。それからその情報は、視床下部や自律神経核を介して身体へと送られ、それによって身震いという身体状態が誘発される。このようにして引き起こされた身震いの情報は、今度は、脳幹へ送られ、さらに島領域にある体性感覚皮質へと伝えられる。こうして身震いという身体状態を反映する脳状態が形成されるのである。Damasio はこのような身体的情動に関係する一連のプロセスを、身体を介するという意味を込めて「身体ループ (body loop)」(Bechara and Damasio 2005, pp. 342-3)

と呼んでいる²⁷。

一方、あたかも身体的情動とは、身体状態から脳状態があたかも引き起こされているように見えながら、実際にはそうではなく、単に脳状態だけからなる情動である。あたかも身体的情動は、現前する刺激により、実際には身体状態が生じないにもかかわらず、過去において身体状態が生じたときと同様の脳状態が形成される²⁸。たとえば、ヘビに遭遇すると、その視覚情報は、扁桃体や VMPFC に伝えられる。それから、身体的情動と違って、その情報は身体へは送られず、直接脳幹に送られ、そこから島領域へ伝達される。こうしてヘビを見て身震いしたときと同じ脳状態が形成される。このように、あたかも身体的情動においては、あたかも身体状態によって引き起こされたかのような脳状態が身体状態抜きに形成されるのである。あたかも身体的情動において身体状態抜きにそのような脳状態が形成されるのは、以前に何度も身体状態を介してそうした脳状態が形成されたからである。身体的情動が繰り返されると、身体状態をバイパスして、直接脳状態を形成するショートカットができあがってくる (Damasio 1994, p. 156, 邦訳, p. 245)。あたかも身体的情動はこのようなショートカットを通して形成される情動である。Damasio はこのように身体を完全にバイパスした一連のプロセスを「あたかも身体ループ (as if body loop)」(Bechara and Damasio 2005, pp. 343-4) と呼び、身

²⁷ 情動には、身体状態の誘発過程において、扁桃体などの辺縁系回路に依存する一次情動 (primary emotion) と、VMPFC (つまり前頭前野の腹内側部) に依存する二次情動 (secondary emotion) が存在する。一次情動は生得的に引き起こされるような情動であり、早い時期から経験するのに対し、二次情動はそのような初期の情動経験を基礎として引き起こされうる情動である (Damasio 1994, p. 131, 邦訳, p. 214)。ただし、同じ情動が一次情動として経験される場合と、二次情動として経験される場合がありうる。たとえば、ヘビに対する恐怖は、ヘビの知覚によって引き起こされる一次情動の場合と、そのような一次情動の経験の想起によって引き起こされる二次情動の場合がある。

²⁸ あたかも身体的情動においても、身体的情動が誘発される場合と同じように、情動を誘発しうる対象の刺激によって、感覚連合皮質と高次の大脳皮質、そしてさらに情動誘発部位が活性化され、その活動が情動実行部位である視床下部や前脳基底部などの活動を上昇させる。身体的情動とあたかも身体的情動の違いは、身体的情動では、この情動実行部位の活動によりある身体状態が生じ、それによって一定の脳状態が形成されるのに対し、あたかも身体的情動では、情動実行部位によって身体を介さずに直接、その脳状態が形成される点にある (Damasio 1999, pp. 68-9, 邦訳, pp. 96-7)。

体を介する身体ループとは異なることを強調している (Damasio 1994, pp. 157-8、邦訳, p. 248)。

では、あたかも身体的情動における脳状態は、身体的情動における脳状態のように、外界の対象の価値を反映しているといえるのだろうか。身体的情動では、現前する刺激によって引き起こされた身体状態がその刺激の価値を反映し、それゆえその身体状態によって引き起こされた脳状態も間接的に刺激の価値を反映していた。しかし、あたかも身体的情動では、現前する刺激から扁桃体や VMPFC、そして脳幹を経由するだけで、身体を介さずに直接、島領域に関連する脳状態が形成される。あたかも身体的情動における脳状態は、身体を介さずに引き起こされたものであるため、一見すると外界の対象の価値を反映していないように思われる。しかし、あたかも身体的情動における脳状態は、身体を介して形成される身体的情動における脳状態と同じである (Bechara and Damasio 2005, p. 344)。従って、身体的情動における脳状態が外界の対象の価値を反映しているなら、あたかも身体的情動における脳状態もそうなのである。ヘビを目にすることによって形成される恐怖の情動における脳状態は、身体を介さない場合でも、身体を介する場合と同じ状態であり、それゆえヘビが危険であるという外界の価値を反映しているのである。こうしてあたかも身体的情動も、外界の対象の価値を反映し、それによって外界のあり方にふさわしい合理的な意思決定に貢献しうるのである。

しかし、合理的な意思決定において、あたかも身体的情動が身体的情動と同じ貢献をなしうるとすると、情動が合理的な意思決定に貢献するうえで、身体状態は不要ではないか、という疑問が生じてくる。身体的情動であっても、合理的な意思決定に貢献するのは脳状態であって、身体はその脳状態を引き起こす付随的な役割を果たすに過ぎず、本質的には脳状態さえあればよいということになるのではないだろうか。しかし、決してそうではない。Damasio も強調しているように、あたかも身体的情動は、あくまでも身体的情動の反復によって生じるのであり、その点でそれは身体的情動に依存している (Damasio 1994, p. 156、邦訳, pp. 245-6)。過去に身体を介した身体的情動が繰り返し生じていなければ、あたかも身体的情動は生じないだろう。あたかも身体的情動に含まれる脳状態が身体的情動に含まれる脳状態と同じでありうるのは、過去においてその脳状態が身体状態を介して形成されたからである。あたかも身体的情動が外界の対象の価値を反映しうるのは、過去の身体的情動においてその価値を反映する身体状態が形成されたからに他ならない。従って、あたかも身体的情動が合理的な意思決定に貢献できるのも、それが現在の身体状態には依存していないとしても、過去における身体状態には依存しているからである。

これまで対象の価値評価として、身体状態を伴う身体的情動と身体状態を伴わないあたかも身体的情動を見てきた。次の節では、このような情動という形態の価値評価ではなく、身体状態とは全く独立に形成される純粋に知的な価値評価を検討してみよう。

4.3 情動と知的評価

これまで見てきた身体的情動とあたかも身体的情動は、どちらもなんらかの仕方で身体を介して外界の価値を反映することにより、外界のあり方に適した合理的な意思決定を可能にするものであった。しかし、そのような情動とは別に、いかなる身体性も帯びていない純粋に知的な価値評価²⁹が存在するように思われる。もしそのような知的な価値評価が存在するならば、身体状態が合理的な意思決定に実質的な貢献をなしうるのかどうかやはり怪しくなってしまう。身体状態に全く依存しない純粋に知的な価値評価は存在するのだろうか。

純粋に知的な価値評価とは、現前する刺激によって、いかなる身体状態も生じず、従って身体状態を反映する脳状態も生じずに（あるいは、身体状態をあたかも反映するような脳状態すら形成されずに）、全く知的に対象の価値を評価するものである。たとえば、ヘビに遭遇して、いかなる情動反応もなしに、冷静な推論の結果として「このヘビは危険だ」という価値評価が行われるとすれば、それは知的な価値評価である。この知的価値評価は、外界の対象の価値に関する評価であるものの、身体状態と本質的な関係が無く、この点で身体的情動はむしろ、あたかも身体的情動とも異なる。このような価値評価は本当に存在するのだろうか。

知的価値評価の存在に関しては、Damasio が詳細に記述した VMPFC 損傷患者エリオットの事例が参考になる (Damasio 1994, ch. 3, 邦訳, 3 章)。すでに触れたように、エリオットは純粋に知的な判断に関しては何の問題もないにもかかわらず、実生活では簡単な行動選択の場面でさえ、しばしばまったく意思決定ができないという特異な症状を示した。彼は VMPFC を損傷する以前は、商社で働き社会的にも恵まれた立場につき、同僚や後輩の鑑であるような人物であり、その上、私生活の面でもよき夫、よき父として順風満帆な人生を送っていた。だが、両半球の前頭葉を圧迫する髄膜腫を患い、その腫瘍の除去のための手術を受けてからは、術後の経過は極めて良好で腫瘍が再度成長する兆候も見られなかったにもかかわらず、彼の人格はまるで別人のように変化したのである。

手術によるエリオットの脳の損傷は前頭前野に限定されており、その損傷の中心は VMPFC であった (Damasio 1994, p. 39, 邦訳, p. 88)。情動を誘発するために決定的に重要な VMPFC を損傷すると、情動の平板化が生じるが、エリオットはまさにそのような状態であった。彼は一見すると、多くの理性的な人々がそうであるように、情動

²⁹ 知的な価値評価は信念的な評価に相当する。ただし、信念的な評価であっても、情動に完全に支配され、情動による価値評価をただ受け継ぐだけのものもあり、そのような信念的な評価は知的な評価とはいえない。また、情動ではない知覚的な価値評価も、身体性を帯びないという点で、知的な価値評価の一種とみなせそうであるが、この価値評価は情動から身体性が失われたもの、つまり、あたかも身体的情動と考えられるかもしれない。そうだとすれば、情動でない知覚的な評価というのは、実は情動でないというよりも、むしろ情動だとみなした方がよいことになるように思われる。

的に控えめな人間のように見えたが、実は控えめというよりも、きわめて情動が希薄だったのである。

治療の初期の段階で Damasio がエリオットのこのような無情動とでもいうべき状態に気が付かなかったのは、彼の知覚能力、過去の記憶、短期記憶、ワーキングメモリは正常で、その上、学習、推論、言語、計算能力についても少しも損なわれていなかったためである (Damasio 1994, pp. 39-43、邦訳, pp. 89-94)。たとえば、彼は知的能力を計るテストで、社会的状況における適切な反応や、社会的目標を達成する手段やその帰結の予測を正しく導き出したり、高い発達レベルの道徳的推論を行ったりすることができた (Damasio 1994, pp. 48-9、邦訳, p. 102)。彼の出来事や物事についての価値評価は、VMPFC を損傷する以前と同等の水準であったのである。しかし、エリオットは、VMPFC を損傷しているために、身体的情動もあたかも身体的情動も持ちえない。従って、このことを考えれば、彼の価値評価は、身体的情動でも、あたかも身体的情動でもなく、身体状態とは全く無関係な知的な価値評価だといってよいであろう。

VMPFC 損傷患者の存在は、身体状態と関係しない純粋に知的な価値評価が存在することを示している。しかし、もちろん、純粋に知的な価値評価は健常者においても見られる。VMPFC 損傷患者は、情動が欠如しているため、外界の価値評価を専ら知的な評価によって行うが、健常者は、情動的な価値評価と知的な価値評価の両方で外界の評価を行うことができる。健常者の場合、通常、情動的な価値評価と知的価値評価は一致しているが、ときに外界のあり方に適した知的な価値評価を行いながら、それと一致しない情動的な価値評価を行うことがある (Damasio 1994, pp. 179-80, pp. 191-2、邦訳, p. 280, pp. 295-6)。たとえば、檻に入れられたヘビを目にして「安全だ」という知的価値評価を行いながら、同時に、ヘビに恐怖を感じることもある。つまり、情動的には、ヘビは「危険だ」という価値評価を行うのである。また、友人が成功したのは自分にとっても良いことだという知的価値評価をおこないながら、その一方で、妬ましいという情動が起こることがある。知的には自分にとって良いことだと評価していても、情動的には自分にとって悪いことだと評価しているのである。このように、知的価値評価と情動的価値評価が一致しないことがあることからわかるように、健常者の場合でも、情動的価値評価とは別に、まさに頭では分かっているという知的価値評価が外界のあり方を評価するために日常的に用いられているのである。

では、実際にこのような知的価値評価が行われるとき、脳の中ではどのようなことが生じているのだろうか。エリオットの場合、正常な知識、記憶、言語、推論の能力を有しているにもかかわらず極端な情動反応の衰退が生じたが、それは彼が前頭前野、特に VMPFC に損傷を負っていたからであった。前頭前野は、認知・実行機能と情動・動機付け機能の両方に関わっており、認知・実行機能には外側部、情動・動機付けには内側部と眼窩部が強く関わっていると考えられている (渡邊 2008, pp. 208-9)。内側部や眼窩部を損傷すると、エリオットや前頭葉ロボトミー患者に見られるように、情動が平板

化し、積極的に行動しようとする意欲を示さなくなる。一方、外側部を損傷すると、選択的な注意や状況の理解、推論の能力などに障害が見られるのである³⁰。知的価値評価の脳内メカニズムについては、まだ不明瞭な点が多いが、一応、前頭前野の内側部や眼窩部が情動に関係しているのに対し、外側部が知的な価値評価に関係していると言えようである。

情動とは異なる知的な価値評価は、外界の価値についての信念であるような価値評価であるが、そのような価値評価があることは、価値評価が形成される仕方からもうかがえる。たとえば、私たちはサソリが危険であるということ、自分でサソリに襲われたことがなくても、他人がサソリに襲われて大変なことになっているのを見たり、本から学んだりすることによって知ることができる。つまり、身体状態を伴わない知覚経験や伝聞経験によって、私たちはサソリが危険だという価値的な信念を形成しうるのである。このような信念は身体状態には関係しない純粋に知的な価値評価だと言えよう。

この節では、身体状態に全く依存しない純粋に知的な価値評価が実際に存在することを見てきた。それゆえ、意思決定に影響を及ぼす外界の価値評価として、現在の身体状態に依存する身体的情動と過去の身体状態に依存するあたかも身体的情動、そしていかなる身体状態にも依存しない知的価値評価という三つの異なる種類の価値評価が存在することになる。しかし、知的価値評価のように、身体状態に全く依存しない価値評価が可能であるなら、身体状態そのものは合理的な意思決定において何か重要な役割を果たすことができるのだろうか。次節では、これまで見てきた身体的情動とあたかも身体的情動、および知的価値評価がそれぞれ意思決定に対してどのような影響を及ぼすのかを考察し、その影響の違いから合理的な意思決定における身体状態の独自の役割を明らかにしたい。

4.4 意思決定への影響の違い—身体状態の独自の役割

意思決定における身体状態の役割という問題の鍵を握るのは、価値評価と動機付けの関係である。価値評価が意思決定に影響を及ぼすためには動機付けの力が必要である。意思決定は、たとえその決定通りの行為がなされないこともあるとはいえ、それなりの動機付けの力を内包した決定であり、そのような意思決定をもたらすためには、価値評価それ自体がそもそも動機付けの力を内包していなければならないのである。分別してゴミを出すことはよいことだという価値評価を行っても、その評価が分別してゴミを出すという行動を促す動機付けの力を伴わないなら、その行動をしようという動機付けの

³⁰ もっぱら理性的な推論能力を発揮して行う「合理的推論」と、好き嫌い、脅威や情動に左右される「情動的推論」の二種類の推論を調べた研究によると、情動内容を含まない合理的推論には前頭前野背外側部 (dorsolateral prefrontal cortex、略して DLPFC) の活動上昇と腹外側部 (ventrolateral prefrontal cortex、略して VLPFC) の活動抑制が見られ、情動内容を含む情動的推論には VLPFC の活動上昇と DLPFC の活動抑制が見られた (Goel and Dolan 2003)。

力をもった意思決定を生み出すことができない。ただ価値評価を行ったというだけで、意思決定には至らないのである。価値評価が合理的な意思決定に貢献するためには、その評価に見合うだけの動機付けの力が必要である。もし、外界の価値を反映する身体状態がその価値の大きさに見合うだけの動機付けの力をもたらすのであれば、身体状態は合理的な意思決定において独自の重要な役割を果たすことになるだろう。本節では、身体的情動とあたかも身体的情動、および知的価値評価がそれぞれどのような動機付けの力を伴うのかを考察しながら、身体状態が意思決定において果たす独自の役割を明らかにしていきたい。このことは、同時に、価値評価と動機付けの関係の再考を図ることにもなるだろう。

4.4.1 身体的情動と行動の準備

まず、身体的情動がどのような動機付けの力を伴うのかを考察しよう。身体的情動では、外界の対象の価値を反映した身体状態が形成される。この身体状態は具体的にはどのような状態なのだろうか。それは何らかの動機付けの力をもたらすのだろうか。たとえば、ヘビに遭遇して恐怖の身体的情動が生じるとき、ヘビが「危険だ」という価値評価に応じた身体状態が形成されるが、そのとき実際に身体では、どのようなことが生じるのだろうか。そのときの身体状態は、ヘビから逃げることを促す動機付けの力をもたらすのだろうか。

ヘビを目にして恐怖の情動が起こるとき、心臓の高鳴りや身震い、発汗といった興奮した身体状態が起こる。このような身体状態は、逃げるという行動が起ころうとするときに生じる状態であり、その行動の準備状態だと言えるかもしれない³¹。つまり、恐怖の身体的情動が生じるとき、身体状態は既に逃げる態勢に入っていると言えるかもしれない。そうだとすれば、この身体状態は実際に逃げる行動へと繋がっていくのが自然であり、逃げないという行動に対しては大きな抵抗として働くことになるだろう。たとえば、臆病だと思われたくないので逃げないようにしようという動機が働いたとしても、恐怖の身体状態はそのような動機に抗して逃げるという行動を促すだろう。こうしてヘビへの恐怖に含まれる身体状態は、ヘビから逃げるという行動を促す動機付けの力をもつと考えられよう。

しかし、いかなる身体的情動においても、それに含まれる身体状態が必ず行動の準備となっているとは言えないように思われる。むしろそうでないことのほうが多いのではなかろうか。たとえば、大好物のケーキが目の前にあるのを見ると、ついつい嬉しくなっていて、実際に手にとって食べてしまうことがある。このようなとき、嬉しいという情動

³¹ 時には恐怖が生じるときに、逃げるできないような身体状態（たとえば、足がすくむとか、腰が抜けるとか）が起こることがあるが、通常は逃げることをむしろ可能にするような準備状態の身体状態が起こる。また、悲しみの場合は、一見いかなる行為も動機付けないように見えるが、その悲しみの状態から逃れさせようとする動機付けの力があると考えることもできる。

は、ケーキのおいしさの価値を反映するとともに、ケーキを食べることを促す動機付けの力を備えていると言えよう。しかし、この嬉しいという情動に含まれる身体状態は、おそらく、ケーキを食べるという行動の準備状態であるとは言えないだろう。そのわくわくとした身体状態は、むしろケーキを食べて得られる快感に伴う身体状態であり、ケーキを食べる行動の準備というよりも、むしろその行動の結果の先取りである。しかし、そうだとすると、それはケーキを食べるという行動を促す動機付けの力を持っているだろう。たとえば、太ることを気にして食べないようにしようという動機が働いたとしても、そのわくわくとした身体状態は、そのような動機に抗して、ケーキを食べる行動を促すだろう。そのような身体状態があるかぎり、食べることへと向かわせる力はなくなるのでない。

結局、身体的情動に含まれる身体状態は、行動の準備であるかどうかにかかわらず、外界の対象の価値を反映し、その価値に見合うだけの動機付けの力を持っていると考えられる。身体的情動が意思決定に影響を及ぼすとき、直接、意思決定に影響を及ぼすのは、身体状態ではなく、それによって引き起こされる脳状態である (Bechara and Damasio 2005, pp. 342-3)。この身体的情動に含まれる脳状態は、脳のメカニズム上、一定の動機付けの力を持って意思決定に働きかけるようになっているのであろう。しかし、脳状態が持つ動機付けの力を支えているのは、身体状態である。へびを見て恐怖を感じたときの身体状態は、それが存続するかぎり、それを反映した脳状態を維持し続け、へびから逃げようという動機を与え続ける。従って、たとえ臆病だと思われたくないという思いが逃げようという動機を抑え込もうとしても、恐怖の身体状態が存続する限り、逃げようという動機はなくなる。もちろん、その動機よりも逃げないという動機のほうが強くて、結局、逃げないという意思決定が行われるかもしれないが、それでも恐怖の身体状態が続く限り、逃げようという動機は存在し続けるのである。ケーキを見て喜びを感じたときの身体状態についても同様である。

このように、身体的情動に含まれる身体状態は、評価した外界の対象の価値に見合うだけの動機付けの力を提供する。こうして身体状態は脳状態の持つ動機付けの力を下支えすることにより、合理的な意思決定において独自の重要な役割を果たすのである。

4.4.2 あたかも身体的情動と紋切り型

あたかも身体的情動においては、実際には身体状態は生じないが、それが生じる身体的情動と同じような脳状態が形成される。従って、あたかも身体的情動も、身体的情動と同じような対象の価値評価を行っていると考えられる。へびを目にして生じる恐怖のあたかも身体的情動は、同じへびを目にして生じる恐怖の身体的情動と同じように、へびは「危険だ」と評価しているのである。しかし、身体的情動の場合、実際に身体状態が生じ、その身体状態が脳状態のもつ動機付けの力を支えている。それに対して、あたかも身体的情動の場合、そのような身体状態による支えがない。それゆえ、身体的情

動に含まれる身体状態が対象の価値評価に見合うだけの動機付けの力をもたらすとすれば、身体状態を含まないあたかも身体的情動は、それほどの動機付けの力をもたらさないことになる。あたかも身体的情動は、動機付けの力をもたらすとしても、身体状態を欠くために、その価値評価に見合うだけの動機付けの力をもたらさないと考えられるのである。

また、あたかも身体的情動では、過去の身体的情動の記憶が呼び起こされているだけであり、Damasio の表現を借りれば、身体的情動が「実演 (live performance)」であるのに対し、あたかも身体的情動はいわば過去の身体的情動の「再放送 (rebroadcasts)」である (Damasio 1994, p. 158、邦訳, p. 248)。つまり、身体的情動においては、対象が与えられたときに、その現場で身体状態が形成され、それによって脳状態が形成される。このようにその都度の現場で形成される身体状態は、その時の個別的な状況のあり方によって異なりうる。たとえば、普段ヘビを目にして恐怖を感じる人でも、テレビを通してヘビを見たときは恐怖の身体状態は生じないかもしれない。また、恋人がペットのヘビを可愛がるのを見たときも、そのヘビに対して恐怖の身体状態は生じないかもしれない。こうして、身体状態は、同じ対象に対してもその時の個別的な状況のあり方によって異なりうる。そしてそのような身体状態を介して形成される脳状態も、状況のあり方に応じて異なりうるのである。

これに対して、あたかも身体的情動においては、過去においてこのような身体的情動を繰り返し経験することによって、次第に身体をバイパスした「あたかも身体ループ」が形成され、このループによって身体を介さずに脳状態が形成される。従って、あたかも身体的情動で形成される脳状態は、その時の個別的な状況のあり方に応じて形成される身体状態を介さないで、過去の身体的情動によって引き起こされた脳状態と同じである。それは過去の身体的情動で形成された脳状態の再現なのである。たとえば、ヘビを目にして恐怖のあたかも身体的情動を感じる際の脳状態は、過去においてヘビを目にして恐怖の身体的情動を感じたときの脳状態を再現したものである。このような意味で、あたかも身体的情動は過去の身体的情動の再放送なのである。

このようにあたかも身体的情動は、過去の身体的情動の再放送であるため、その都度の個別的な状況に合わせて変化するということはない。ヘビを目にしてあたかも身体的情動が引き起こされる時は、過去の恐怖の身体的情動が再現されるため、常に恐怖を感じるようになる。従って、同じヘビに対して身体的情動とあたかも身体的情動がともに生じることがあるとすれば、両者は状況によっては一致しないことがありうることになる。たとえば、テレビを通してヘビを見たとき、身体的情動としては、ヘビを見ても恐怖を感じず、むしろ可愛いと感じるにもかかわらず、あたかも身体的情動としては、過去の恐怖の情動が再現されて、相変わらず恐怖を感じるということが起こりうるのである³²。

³² もっとも、あたかも身体的情動が変わることもありうる。これは元になる身体的情動が

以上のように、身体的情動がその都度の状況に応じて変化し、それゆえ柔軟性があるのに対して、あたかも身体的情動は紋切り型で硬直的である。従って、価値評価に関しても、あたかも身体的情動は対象の価値を固定的に評価するのに対して、身体的情動はそのときの個別的な状況のあり方に応じて柔軟に対象の価値を表す。そのため、身体的情動はあたかも身体的情動よりも、外界のあり方に適した柔軟な価値評価を可能にするのである。価値評価に対応する動機付けの力についても、同じように、身体的情動の方が外界のあり方に適した動機付けの力を与えることができる。従って、身体的情動が意思決定に対してその場の状況に応じた融通のある貢献をするのに対し、あたかも身体的情動は硬直的な影響しか及ぼせないのである。

たしかに、このようなあたかも身体的情動にも有益な面があると言えよう。Damasioによると、あたかも身体ループは身体をバイパスすることによって、時間とエネルギーを節約し、それゆえ状況によっては非常に有用な働きをする (Damasio 1994, pp. 155-6, 邦訳, p. 245; Damasio 1999, p. 283, 邦訳, pp. 336-7)。しかし、そうだとした場合、あたかも身体的情動の場合、定型化されたプロセスが脳の中で再現されるだけで、その時の状況に合わせて柔軟な変化が生じたりはしないのである (Damasio 1994, p. 158, 邦訳, p. 284)。

以上をまとめれば、あたかも身体的情動は身体的情動よりも身体状態を含まないという点で動機付けの力が弱く、また、意思決定に対して固定的な影響しか及ぼすことができないため、外界のあり方に適した柔軟な意思決定を妨げる恐れがあると言えよう。

4.4.3 知的価値評価と動機付けの欠如

最後に、知的価値評価がどのような動機付けの力をもたらし、意思決定に対してどのような影響を及ぼすのかを考察してみよう。知的価値評価はいかなる身体状態にも依存しない価値評価である。たとえば、ヘビを目にしても、いかなる身体反応もなしに、冷静な推論の結果として、ヘビは「危険だ」という価値評価が行われる。このような知的価値評価がどのような動機付けの力をもたらしかについては、VMPFC 損傷患者のエリオットの事例が参考になる。彼は、知的価値評価によって次にどうすべきかを判断することができたとしても、実際にそれを行動に移すことができなかった。つまり、ある行動をするのが最善だと判断することはできても、そうしようという意思決定を行うことができなかったのである。VMPFC 損傷患者の事例を見る限り、知的価値評価は、身体的情動やあたかも身体的情動とは違い、意思決定に対してほとんど影響を及ぼさず、従って、それ自体としては動機付けの力を持たないのではないかと考えられる。

一般に、価値判断それ自体に動機が含まれるかという問題は、哲学において長年議論されている問題である (Roskies 2005, p. 22, 邦訳, p. 35)。この問題については、価値

変化し、その変化した身体的情動が新たにあたかも身体的情動の元になることがありうるからであり、そうでない限り、あたかも身体的情動は変わらない。

判断それ自体に動機が内包されているとする「動機内在主義 (motive internalism)」と、価値判断それ自体には動機は含まれないとする「動機外在主義 (motive externalism)」の二つの立場がある。VMPFC 損傷患者のあり方をみるかぎり、価値判断には必ずしも動機付けの力が伴わないようである。つまり、動機は価値判断にとって外在的なようである。

実際、Roskies (2005) は、価値判断の一種である道徳的判断について、VMPFC 損傷患者の存在を論拠にして、動機外在主義を主張している。VMPFC 損傷患者は、健常者とほとんど同じ道徳的判断を行う³³。健常者と同様に、「正直であるべきだ」とか「約束を守るべきだ」とか判断する。しかし、彼らはそのように判断しながら、平気で嘘をついたり、約束を守らなかったりする。つまり、正直にしようという意思決定や、約束を守ろうという意思決定を行わないのである。これは、彼らが「正直であるべきだ」とか「約束を守るべきだ」と判断しても、そうしようという動機を欠いていることを示していると考えられる。この場合、動機付けなしの道徳的判断は、実は道徳的判断ではないのではないかという反論が考えられるが、Roskies はこの反論を詳しく吟味してそれらを斥け、その上でこのような判断は確かに道徳的判断だとみなせると結論づけている (Roskies 2005, pp. 24-5, 邦訳, pp. 40-2)。VMPFC 損傷患者はまさに動機内在主義に対する生きた反例であり、それに基づいて Roskies は道徳的判断には動機が内在しないと主張するのである (Roskies 2005, p. 28, 邦訳, p. 47)。この Roskies による道徳的判断についての考察も、やはり知的な価値評価がそれ自体としては動機付けの力を持たないことを示唆していると考えられる。

知的価値評価が通常、動機付けの力を持つように思われるのは、知的価値評価にふつう情動が伴っているからであろう³⁴。たとえば、ヘビは危険だという知的な価値評価が行われるとき、ふつうヘビに対する恐怖の情動が生じている。この情動は、自分でも明らかほどの心臓の高鳴りや身震いなどを伴っていて、はっきりと意識される場合もあるが、内臓や骨格筋などのわずかな変化しか伴わないため、意識されない場合もある (Bechara and Damasio 2005, p. 341)。また、身体的情動ではなく、単にあたかも身体的情動である場合もあろう。このような無意識的な情動やあたかも身体的情動も含めれば、知的価値評価が行われるとき、ふつうその価値評価と合致する情動がそれに伴っていると考えられる。そうだとすれば、知的価値評価が動機付けの力を持つように見え

³³ 但し、VMPFC 損傷患者は、道徳的なジレンマ状況下では、健常者とはしばしば異なる判断を下す。5人を助けるために1人を橋から突き落とすかどうかというタイプのトロリー問題 (trolley problem) では、VMPFC 損傷患者は1人を犠牲にしてでも5人を助けるという判断を行う者が多いが、健常者の場合は1人を橋から突き落とすことをためらって、そうしないと判断する者が多い (Roskies 2005, p. 21, 邦訳, pp. 33-5)。

³⁴ 知的価値評価に情動が伴っていないにもかかわらず、その価値評価が動機付けの力を持つように思われる場合がある。それは知的価値評価に意志の力が伴う場合である。意志の力については、後の章で詳しく扱うことにして、ここではとくに取り上げないことにする。

でも、それは知的価値評価そのものに動機が内在するからではなく、知的価値評価に伴う情動に動機が内在しているからである。

一般に、価値評価が意思決定に影響を及ぼすには、動機付けの力が必要である。へビは危険だと評価しても、へビから逃げようという動機付けの力がなければ、へビから逃げ方へと意思決定を向かわせることができない。つまり、ただ評価が行われるだけで、意思決定には何の影響も及ばない。従って、知的価値評価それ自体として動機付けの力を持たないとすれば、それはそれ自体としては意思決定を左右する力を持たない。知的価値評価が意思決定に影響を与えるためには、身体的情動にせよ、あたかも身体的情動にせよ、ともかく何らかの情動を伴う必要がある。知的価値評価は、情動から動機付けの力を借りることによってはじめて、意思決定に影響を及ぼすことができるのである。

知的価値評価が意思決定に貢献しうるのは、それに伴う情動が動機付けの力を提供するからである。そうだとすれば、情動はやはり意思決定において不可欠の役割を果たすと言える。しかも、あたかも身体的情動は本質的には身体的情動に依存しており、身体的情動のほうがあたかも身体的情動よりも柔軟で適切な価値評価と動機付けの力をもたらす。従って、身体的情動は、三つの価値評価の中で合理的な意思決定に対して最も重要で根本的な役割を果たしていると言える。そして身体的情動のこのような重要な役割を可能にしているのが、まさにそれに含まれる身体状態にほかならない。こうして身体状態は、柔軟で適切な価値評価と動機付けの力を提供するという点で、意思決定において極めて重要な役割を果たすのである。

4.5 まとめ

これまで、意思決定における身体状態の役割を明らかにするために、まず、三つの価値評価、すなわち身体的情動、あたかも身体的情動、知的価値評価が、それぞれ意思決定にどのような影響を及ぼすのかを相互に比較した。その結果、身体的情動は、外界の対象の価値評価に見合うだけの十分な動機付けの力をもたらすのに対し、あたかも身体的情動は、身体状態を含まないため、対象の価値評価に見合うほどの動機付けの力を持たないことを示した。また、あたかも身体的情動が固定的な価値評価と動機付けの力しかもたらしえないのに対して、身体的情動はその都度の状況に応じた柔軟な価値評価と動機付けの力をもたらすことを明らかにした。知的価値評価については、それ自体としては動機付けの力を持たず、情動の力を借りて初めて意思決定に影響を及ぼしうることを示した。

以上のように身体的情動に含まれる身体状態は、状況に応じた柔軟な価値評価とその評価に見合う動機付けの力をもたらす、それゆえ外界のあり方に応じた適切な意思決定を可能にするという点で、意思決定において極めて重要かつ独自の役割を果たすのである。

第五章 情動の合理性

外界の事物の価値評価には、動機を内在する情動的な評価と、そうでない知的な評価があると考えられるが、これらはどのような関係にあるのだろうか。それらは互いに独立に存在し、別々に機能するのだろうか。それとも、それらの間には、互いに制約し合うような相互的な関係があるのだろうか。情動的評価と知的評価はふつう一致する。目の前のヘビに恐怖を感じる時、ふつうそのヘビを危険だと判断する（つまり、知的に評価する）。しかし、一致しない場合もときにある。ヘビに恐怖を感じているのに、そのヘビが檻に入れられているために、危険ではないと判断することがある。そしてこのような場合、ヘビへの恐怖は不合理な情動だとみなされ、生じるべきでないといわれる。そうだとすれば、情動は本来、知的評価に合致すべきものであろうか。しかし、逆に、不合理なのは情動ではなく、知的評価であり、知的評価が情動に合致すべきだということもあるのではなかろうか。本章では、情動的評価と知的評価がどのような関係にあるかを考察するとともに、それを通じて情動にもそれなりの合理性があることを明らかにする。

5.1 情動と知的評価はどう関係するか

これまで見てきたように、Damasio や Prinz は、外界からの刺激によって引き起こされる身体状態を伴う情動は、外界についての一定の価値評価を下しているものであり、情動とは外界についての身体的評価であるとみなす。そしてこの身体的評価は、私たちが危険から遠ざけたり、報酬に接近させたりすることで生存にとって有利に働くと考えられる。情動をこのように考える彼らの立場は、情動の哲学的理解の流れの中で、身体状態をとくに重視する James らの立場を継承するものであり、それゆえ新ジェームズ主義 (neo-Jamesian) と呼ばれる (Helm 2010)。

新ジェームズ主義の主張をここで確認しておこう。この見方における情動とは、私たちを取り巻く状況の身体的な価値評価である。たとえば、赤色と黒色からなる縞模様のヘビに遭遇すると、そのヘビを危険であると評価する恐怖の身体状態が生じる。このような価値評価を行う身体状態のことを、これまで見てきたように、Damasio はソマティック・マーカー (Damasio 1994, ch. 8, 邦訳, 8章) と呼ぶ。また、Prinz はヴェイランス・マーカー (valence marker) (Prinz 2004, p. 173) と呼ぶ。新ジェームズ主義の情動に関する見解は、このように、情動は置かれた環境に合わせて対象に関する身体的な価値評価を行うものだという見方だと考えられるのである。

一方で、Damasio や Prinz は、前にも述べてきたように、身体状態を含む情動的価値評価とは別に、身体状態を含まないような価値評価、すなわち知的価値評価も想定している。たとえば、赤と黒の縞模様のヘビと遭遇したときに、一方でそのヘビは危険であることを示す情動的評価（身震いなどの身体状態による評価）を行い、他方で、その

ヘビが図鑑で見たとおり毒ヘビであり、それゆえ危険であるといった知的評価も下す。もしくは、この二つの価値評価が相反することもある。一方では満員電車で押しつぶされそうな経験をしたために満員の急行電車に乗るのは嫌だという情動的評価を下すのに、他方では遅刻しないためには満員の急行電車に乗らなければならないと知的評価を下す場合がそうである。この場合、情動的評価は満員の急行電車に乗るのは良くないという評価を行い、知的評価は満員の急行電車に乗るのは良いという評価を行っている。このように、先に述べたとおり、確かに、私たちは情動的評価とは別の知的評価も行っている。

ではこの情動的評価と知的評価は一体どのような関係にあるのだろうか。Damasio や Prinz はこの点に関してあまり明らかにしていない。ある対象に対して、私たちは知的にも情動的にも評価し、その二つが一致することも一致しないこともあると述べるに留まっているのである。情動的評価と知的評価の関係をもっと明らかにするために、この点に関して興味深い考察を詳細に行っている Helm (2010) を参照しよう。まず、Helm が情動をどのように捉えているのか、次に、彼の情動概念が欲求とどのような関係にあるのか、そして最後に判断が欲求や情動とどんな関係にあるのかを見ていく。それを通して知的評価と情動的評価の関係を明らかにするとともに、情動がどのような合理性を有するかを明らかにする。

5.2 情動について

新ジェームズ主義によると、これまで見てきたように、情動とは外界の対象によって誘発される身体状態を含むものであり、その身体状態はその対象の価値を表している。たとえば、赤と黒の縞模様のヘビと遭遇した場合に引き起こされる身体反応（心拍数が上がったり、脂汗をかいたりするなど）は、まさに恐怖という情動に含まれる身体反応であり、そのヘビが危険であるという価値的なあり方を表しているのである。言い換えると、「恐怖」という情動は、「ヘビが危険である」というヘビの価値的なあり方に対する反応なのである。このように、新ジェームズ主義の中では、情動を対象の価値への反応として捉えてきたのである。

だが、Helm は、情動をより適切に理解するためには、情動を単なる価値への反応とみなさず、「意味」へのある種の関与とみなす必要があると主張する (Helm 2010, p. 310)。これは一体どういうことなのだろうか。ここで言われている意味とは何か、また、意味への関与とは具体的にどのようなことなのだろうか。そもそも Helm は情動をどのように捉えているのだろうか。

Helm は私たち人間が日常生活の中で抱く情動について、それは各々の人が持つ各自の価値体系の基盤があってはじめて一貫性のある仕方で表出されるものであるとみなしている。情動表出の根底にある価値体系こそ、私たちの情動表出の一貫性を可能にするものであり、それがあからこそ人々の情動は、理解可能なものとなるのである。

Helm が主張するような一貫性のあるものとして情動を理解するためには、まず情動が三つの対象を持つことに注意しなければならない。

Helm は情動の対象を以下の三つに分ける。まず一つ目は、情動の「標的 (target)」である (Helm 2010, p. 310)。これは世界において情動が向けられる対象のことである。たとえば、私が友人 A と待ち合わせをしたとしよう。もしその友人 A が待ち合わせ時間になっても来ず、約束をすっぽかしたとすると、私はその友人 A に対して怒りを抱くに違いない。その場合、私の怒りはまさにその友人 A へ向けられるのであり、その友人 A は私の怒りの標的である。

情動の二つ目の対象は、情動の「形式的対象 (formal object)」であり、これは標的が持つ価値的特徴のことである (Helm 2010, p. 310)。情動は標的を、ある一定の価値的特徴を持つものとして評価する。たとえば、私が友人 A に対して怒りを抱くのは、友人 A が約束をすっぽかしたという事実のゆえであり、もし約束通り現れていたなら怒りを抱くことはなかつただろう。怒りを抱く場合、確かに私は情動の標的である友人 A に対して怒りを抱いていると言えるが、より詳しく言うなら、友人 A が約束をすっぽかしたということに対して怒りを抱いているのである。ではどうして約束をすっぽかしたということに対して、私は怒りを抱くのだろうか。それはもちろん約束をすっぽかすことが悪いことだからである。つまり、友人 A は悪いことをしたのであり、悪いことをしたことに対して私は怒っているのである。この悪いことをしたというのが友人 A の価値的特徴であり、私の怒りの形式的対象である。

情動の三つ目の対象は、情動の「焦点 (focus)」である (Helm 2010, p. 310)。これは情動の背景的对象とも呼ばれる。先の事例では、私が友人 A に対して怒りを抱くのは、情動の標的である A が約束をすっぽかすことにより、私に対して悪いことをしたという形式的対象 (価値的特徴) を持つゆえである。だが、よくよく考えてみると、なぜ私は A をそのような形式的対象を持つものとみなすのだろうか。確かに友人 A は約束をすっぽかしたのだが、だからといって友人 A が悪いことをしたという形式的対象を持つことには必ずしもはならない。私が A をそのような形式的特徴を持つものと見なすのは、まさに私が「約束を守ることは重要で、すっぽかすことなどあつてはならないことだ」と考えているからである。つまり、約束は守られるべきだということを含むような価値体系を持っているからなのである。そのような価値体系を持っているがゆえに、私は約束をすっぽかした A を悪いことをしたとみなすのであり、それゆえ A に対して怒りを抱くのである。もし、私が約束を守ることをそれ程重要だと考えていないなら、私は A に対して怒りを抱くことはなかつただろう。この事例における私の怒りの焦点とは、約束は守るべきだという、私が自分の価値体系の一部として持っていることがらであり、これは情動を抱く際にその背景となる対象なのである。

5.2.1 情動の焦点の重要性

私たちが抱く情動には、標的、形式的対象、焦点の三つの対象があり、一つの情動がこの三つの異なる種類の対象を同時に持つことを Helm に即して確認した。ところで、私たちは情動について、それが正当かどうかをしばしば問題にする。約束をすっぽかした A に対して私が怒りを覚えるとき、この私の怒りは正当であろうか。Helm によれば、ある情動が正当かどうかは、その情動の焦点によって決まる。この点で、焦点は情動の三つの対象の中でもとりわけ重要である。Helm はある情動が正当かどうかを知るための条件として二つの条件を提示しているが、それらを見ることにより、焦点の重要性を確認しておこう。

第一に情動の焦点は、主体にとって実際に意味を持つのでなければならない (Helm 2010, p. 310)。たとえば、もし私が約束を守るべきだと口先では言いながら、実際には常に約束をすっぽかすような人であったなら、つまり約束を守るべきだということが私の常日頃の行動からは実現されておらず、私自身も約束を守るべきだと実は思っていない場合、私が約束をすっぽかした友人 A に対して怒りの情動を抱いたとしても正当化されないだろう。この場合、約束を守るべきだというのは、私にとって本当は意味を持つことではなく、それゆえ実際は私の情動の焦点ではないのである。情動の焦点の一つひとつがまさに私たちの価値観（価値体系）の構成要素なのであり、それゆえ焦点は情動を抱く当人にとって実際に「意味」のあることでなければならないのである。

第二の条件は、情動の標的が情動の焦点と適切に関連付けられており、それゆえ標的が実際に焦点と関連する形式的対象を持っていないなければならないということである (Helm 2010, p. 310)。もし私が友人 A との待ち合わせの日を間違えていたなら、私の怒りは正当化されないだろう。その場合、友人 A は実際には約束を破っていないのであり、それゆえ約束を守るべきだという怒りの焦点と適切な関係をもっていない。従って、友人 A は実際には悪いことをしたという怒りの形式的対象を持っていないのであり、私が友人 A に対して怒りの情動を抱くことが正当化されないのは、そのためである。情動の標的と情動の焦点の関連は、情動の正当性にとって極めて重要なのである。

このように情動の焦点は、実際にそれが情動を抱く当人にとって意味を持つことかどうか、また、標的が焦点と適切に関連付けられているかどうかという点で、情動の正当性を決定する重要なものなのである。但し以下のことに注意しておく必要がある。それは、Helm のこのような情動の正当性の捉え方は、一定の価値観のもとでの主観的な正当性を捉えたものに過ぎないのであり、価値観の是非をも問題にするような客観的な正当性を捉えたものではないということである。私が約束を守るべきだという価値観を持つ場合、その焦点に照らして、私のある特定の情動が正当だとされる（あるいは不当だとされる）が、その根拠である「約束は守られるべきだ」という価値観自体の正当性は問われていない。価値観自体の正当性をも問題にする場合は、各人の主観的な価値観を超えた何らかの客観的な価値の尺度を持ち出す必要が出てくるだろう。

5.2.2 情動の一貫性について

これまで見てきたように、情動は三つの対象を持ち、その内の一つである焦点に照らして、情動が正当かどうかが決まる。だが、焦点の重要性はそれだけにとどまらない。Helm は私たちの情動には一貫性があることやそこから情動に一定のパターンが出現してくるということにも情動の焦点が大きく関わっていると指摘する。私たちの情動の一貫性について、それが共通の焦点によって可能になるという Helm の主張を次に見ていこう。

情動の焦点は、これまで見てきたとおり、情動を抱く当人にとって意味のあることであり、私たちはそれぞれそのような意味に自らを関与させながら日々生活している。そのような意味の一つひとつが私たちの価値観を構成するのである。私たちはそのような自らの価値観に基づいて行動し、ある状況では憤りを感じたりそれと違う状況では喜びを感じたりする (Helm 2010, pp. 310-1)。たとえば、先の事例の友人 A の場合、友人 A が約束をすっぽかしたなら怒りを感じるし、約束通り現れたらなら嬉しく感じるのである。これは約束を守るべきだという共通の焦点を基礎にして、置かれた状況に応じた情動が形成されるからだと考えられる。これは約束を守るべきだという焦点が、私にとって意味がある焦点だからこそ、友人 A が約束をすっぽかしたなら私は怒りを感じるし、約束通り現れたならうれしく感じるのである。

そうだとすると、共通の焦点を持つ様々な状況で生じた情動は、合理的な相互関係を持つ一定のパターンを為すことになるだろう。たとえば、私は友人 A が約束をすっぽかしたことに対して怒るなら、その日の夕食で友人 A を紹介すると約束していた友人 B に対しては、A を紹介できなくなったことを申し訳なく思うだろう。また、友人 A が私との約束をすっぽかしたものの、次は必ず友人 B との夕食には行くと再度約束したなら、友人 A を友人 B に紹介できることについて安心するだろう。私が抱くこれらすべての情動は、約束を守るべきだという共通焦点のもとに、互いに整合的であり一定の合理的なパターンを為している。逆に、友人 A が約束をすっぽかしたことを怒りながら、友人 A がそのことを私に詫びてきたときに私が狼狽えたり、友人 A が次の機会に友人 B との食事会に来ることを私が恐れたりするのは、互いに整合的ではなく一貫した合理的なパターンを欠いた状態だと言えよう。情動の間にこのような一貫した合理的なパターンがあるかどうかは、共通の焦点に照らすと明らかである。つまり、私たちが状況に応じて様々な情動を抱いたとしても、ふつうなお一貫性がありうるのは、共通の焦点を基にして一定の合理的な情動パターンが形成されるからなのである (Helm 2010, p. 311)。

このことは、ある情動が合理的なのは、それが合理的なパターンに属する場合であり、それゆえ共通の焦点のもとで形成された場合であり、そうでなければ不合理であるということである。たとえば、私が借金を踏み倒す人や借りたものを紛失しても平気で居直るような人に対しては憤りを感じ、借りたものを返す人に対しては当然のことだと感じ

るとする。これらの情動が、互いに整合的であり一定のパターンをなしているとみなすことが可能であるのは、私が借りたものは返すべきだという焦点を持ち、それを基にしてこれらの情動が形成されている場合である。そしてその場合、このような情動は合理的なパターンを形成しているのであり、それゆえそれぞれ合理的な情動なのである。もし、そのような焦点を持ちながら、借金を踏み倒した人に賞賛の念を感じたり、借りたものを返す人を軽蔑したりしたとすると、それらの情動は合理的なパターンに属さないもので、不合理なのだと言えよう。なぜなら、それらの情動は借りたものは返すべきだという焦点を持つ限り、正当でない情動であり、その焦点のもとで形成される他の正当な情動と合理的な関係をもたないからである。このように、ある情動の合理性は、共通の焦点のもとに形成された情動による合理的なパターンに属するかどうかによって決まるのである。

5.3 欲求について

これまで Helm により提示された情動概念を見てきたが、ここからは彼が欲求について述べていることについて見ていこう。彼は情動と欲求を注意深く区別しており³⁵、かつ、情動は欲求とも合理的な相互関係にあり、情動はそれ自体として直接的に行動を動機付けるだけでなく、欲求との合理的関係を通じて間接的にも行動を動機付けるという見解を示している (Helm 2010, p. 312-3)。この点を確認するために、まず欲求について Helm が自身の見解を新ジェームズ主義の見解からどう区別しているかを見ておこう。

Helm によると、Damasio や Prinz の新ジェームズ主義は、欲求に関して伝統的なヒューム主義的見解をとっている (Helm 2010, p. 307)。ヒューム主義的見解では、欲求のような行為を引き起こす動機付け状態と信念のような認識的な状態とは明確に区別され、その区別は世界への適合方向によってなされる。信念やその他の認識的な状態は、心から世界への適合方向を持つ。たとえば、「外は雪だ」という私の信念と、「外は雨が降っている」という実際の世界のあり方との間に不一致があることが判明した場合、私は世界のあり方に一致するように私の信念を「外は雨だ」という信念に変更すべきであり、そうすることが合理的だと考えられる。一方、欲求のような動機付け状態は、世界から心への適合方向を持つ。たとえば私が「P店のアボカドアイスクリームが食べたい」

³⁵ Helm は情動と欲求を区別するが、それらをどのようにして区別するかについては特に明示していない。従って、恐らくごく常識的な区別をしているのだろうと思われる。すなわち、情動は状況に対する評価的な反応であるのに対し、欲求は一定の望ましい事態を実現させようとする働きだと考えられているのであろう。しかし情動には多くの場合、一定の望ましい事態を実現させようとする動機付けの力が含まれるし、また一定の望ましい事態を実現させようとする欲求の働きは多くの場合、何らかの情動の形をとるとと思われる。従って、この常識的な情動と欲求の区別は、それほど厳密には成立しないと思われる。しかし、ここでは、Helm に従って、一応、情動と欲求を区別して扱うことにする。

という欲求と P 店のアボカドアイスクリームを食べていないという現実との間に不一致があることが判明した場合、私は自分の欲求に一致するように世界を変えるべきであり、そうすることが欲求にとって合理的だと考えられる。つまりこの場合、P 店のアボカドアイスクリームを食べるようにするというこそ合理的なのである。このようにヒューム主義的見解によると、欲求と信念は、世界への適合方向によって区別され、欲求とは世界から心への適合方向を持つ状態であるとみなされるのである。

欲求をこのように世界から心への適合方向を持つ状態であると解釈するなら、欲求は目標志向的状态に他ならないということになる。目標志向的状态というのは、目標を達成しようとする状態のことである。たとえば、P 店のアボカドアイスクリームを食べたいという欲求は、そのアイスクリームを食べていないという世界のあり方をそのアイスクリームを食べるというあり方に変えようとする状態である。それゆえ、その欲求はそのアイスクリームを食べるという目標を達成しようとする状態である。つまり、それは目標志向的状态なのである。

この伝統的見解は、目標を達成するための行為の動機であるという欲求の役割をよく捉えている。だがこの見解では、欲求それ自体の合理性・不合理性はまったく問題にされない。どんな目標であれ、ともかく目標を達成しようとする状態なら、すべて欲求であり、そのような欲求について合理的かどうかということは一切問われない。だが、欲求にも実は合理的なものとしてでないものがあるように思われる。たとえば、「夕食に水銀 200 g を食べたい」という欲求は、「自動車の運転免許を取得したい」という欲求と違って、不合理であるように感じられる。ヘルムはこのような違いから、欲求にも合理性の基準が適用可能であることを示している (Helm 2010, pp. 307-8)。この Helm の欲求に関する反ヒューム主義的な見解をもう少し詳しく見ていこう。

5.3.1 欲求と目標志向的状态の区別

私たちはある目標の達成を欲する場合、通常その目標は私たちにとって意味があるものであり追求に値するものであるからこそその目標達成を欲すると考える。たとえば、「夕食に水銀 200 g を食べたい」という欲求と「自動車の運転免許を取得したい」という欲求がどちらも欲求であるとしても、合理性に関して大きな違いがあると感じるのは、前者は自分にとって追求するに値しない欲求であるが、後者は追求するに値する欲求であるとみなすからだろう。このことは、欲求が世界から心への適合方向を持つ目標志向的状态であるというヒューム主義的見解では説明できない。「夕食に水銀 200 g を食べたい」という欲求は、この欲求を実現する手段に関わる信念とともに「夕食に水銀 200 g を食べる」という状態を引き起こし、「自動車運転免許を取得したい」という欲求は、この欲求を実現する手段に関わる信念とともに、「自動車運転免許を取得する」という状態を引き起こすが、ヒューム主義的見解では欲求についてそれ以上述べるべきことはない。なぜなら、欲求が単なる目標志向的状态なら、目標の実現を目指す状態だという

ことが述べられればそれで十分だからである。従って、そこからは、なぜその目標が追求に値するのかという目標の価値に関わる説明は出てこない。Helm はこの点を問題にし、欲求については目標が追求するに値するのかどうか、自分にとって目標の実現が意味のあることなのかどうかを考慮する必要があると主張するのである。

Helm によると、欲求がある目標の実現を目指す目標志向の状態であるだけでなく、その目標が追求するに値するものでなければならない (Helm 2010, pp. 307-8)。たとえば、自動車免許を取得したいという欲求は、自動車免許の取得という目標を達成しようとする状態であるのと同時に、その目標は私にとって追求するに値するものであるような状態である。欲求とは、目標志向の状態であることと目標が追求するに値するということの両方から成り立つ状態なのである。

さらに Helm は欲求について目標が追求するに値するということは、まさにその目標が自分にとって意味があるということであり、それゆえにこそ自分はその目標を追求するのだと考える (Helm 2010, p. 307-8)。このことは、欲求のヒューム主義的見解では、実は欲求が持っている行為の動機の側面が理解できないということを示している。動機の理解には、自分がどのようなことに意味を見出しているのか、つまりどのような価値体系に基づいて目標を追求するに値するものとしているのかという自分の意味や価値体系の理解が必要なのである。この見解からすると、動機の本性を理解する上で中心となるのは、欲求の強さや欲求を実現するための手段に関わる信念ではなく、意味による目標の価値付けである。つまり、動機は単にある目標を達成させようとするだけのものではなく、価値のある目標を達成させようとするものなのである。

また、欲求の合理性も意味による目標の価値付けの観点から理解できる。「自動車運転免許を取得したい」という欲求が合理的なのは、その目標が自分の意味に照らして追求するに値するものだからである。自分は自動車運転免許の取得を意味あるものとするような価値体系を持っており、それゆえそれは追求するに値するものになっている。それに対して、「夕食に水銀 200g を食べたい」という欲求が不合理なのは、その目標が自分にとっての意味に照らして追求するに値するものではないからである。自分は水銀 200g を食べることを意味のあるものとするような価値体系を持っておらず、それゆえそれは追求するに値するものになっていない。このように欲求の合理性も意味の観点から理解できることになる。

但し、「夕食に水銀 200g を食べる」ということを意味のあるものとするような奇妙な価値体系を持った人もいるかもしれない。そのような人にとっては、「夕食に水銀 200g を食べたい」という欲求は合理的だということになるだろう。しかし、この場合、やはり価値体系の是非が問題にされるべきであろう。Helm は欲求の合理性を当人の主観的な価値体系に照らして問題にするが、価値体系そのものの客観的な是非を問わない。欲求について単にその主観的な合理性を問題にするのではなく、客観的な合理性を問題にしようと思えば、価値体系の客観的な是非を問う必要がある。しかし、ここではこの

点には深入りしないことにする。

5.3.2 欲求と情動の合理的相互結合

これまで見てきたように、Helm が欲求に関して示した見解は、欲求には目標を追求するという目標志向的状态の側面と、目標が追求するに値するという評価的側面の二つがあり、追求する価値のあるものかどうかは自分にとって実際に意味のあることかどうかによって決まるということであった。また、欲求の合理性も意味による目標の価値付けの観点から理解されたが、この欲求の合理性は先の情動の合理性とはどう関係するのであろうか。次にこの点をみてみよう。

情動に関して先述したとおり、情動は何らかの焦点を背景にして形成される。そしてそのような情動は、焦点を共有するその他の情動と合理的な相互関係を形作る (Helm 2010, pp. 312-3)。Helm がここでさらに指摘する重要な点は、この情動の合理的な相互関係には情動を超えてさらに欲求も含まれるということである。たとえば、友人 A が約束をすっぽかしたり遅れたりすることを私が恐れるなら、私はこれを防ぐことを欲すべきである。私の恐れ焦点、すなわち約束を守るべきだという焦点は私にとって意味のあるものであり、この意味が、A 氏が約束を破りそうな時は約束を守らせることを追求するに値するものとし、そうである限りにおいて、A 氏に約束を守らせようという欲求を持つべきものにする。もしそのような欲求を持たないとしたら、それは合理的ではないのであり、さらに一貫してそのような欲求を持たないのであれば、私は約束を守るべきだという意味には実は関与していない(約束を守るべきだということをも自分にとって意味のあることだとしていない)ということになるだろう。つまり約束を守るべきだという焦点を持たないことになるだろう。要するに、ある情動を持つとき、その情動の焦点を意味のあるものにするような意味があるが、その意味によって追求に値するとされる欲求があることになり、それゆえそのような欲求を持つべきだということになるのである。このように、共通の焦点を持つ情動の合理的なパターンには情動を超えてさらに関連する欲求も含まれるのである。

これまで見てきたように、Helm によれば、あることを欲求することは、単にそれを一つの目標として追求する傾向性があるというだけではなく、それを追求する価値のあるものとして見出すということでもある (Helm 2010, pp. 312-3)。そしてそうである以上、欲求それ自体が意味への関与(目標を追求することを自分にとって意味のあることだとすること)であり、それゆえ欲求は意味(焦点)を共有する情動の合理的なパターンに含まれることになるのである。このように共通の意味を持つ合理的な相互結合には情動と欲求の両方が含まれることになるのである。

5.4 判断について

これまでである共通の意味(つまり共通の焦点)のもとで情動と欲求が合理的な相互結

合を形成するという Helm の見解を見てきたが、ここではそのような相互結合と物事の価値に関する判断(知的評価)がどのような関係にあるのかについて見ていこう。通常、ある状況において何をすべきかに関して下される知的判断とは、参照可能な諸事情を比較考量して、時には情動や欲求に反して、何が最善の行為かを決定するものだと考えられる。このような見方のもとでは、判断は欲求や情動とは全く異なる役割を果たすものであり、それらとは独立しているように思われる。だが Helm は、判断も情動と欲求の合理的な相互結合のなかに組み入れられるものであり、それゆえそれらと合理的な相互結合を形成すると説くのである (Helm 2010, p. 318)。ここでは Helm によるこのような判断についての見解を見ていこう。

通常、判断は情動や欲求と整合的である。たとえば、私が赤と黒の縞模様のへびに遭遇し、恐怖の為に震え上がったとしよう。このとき、私は恐怖の情動によって、へびを危険だと評価するとともに、ふつう知的にもへびを危険だと判断する。そしてさらにへびから逃げたいという欲求を形成する。これらの情動と判断と欲求は互いに整合的である。このような場合、判断は情動や欲求と合理的な相互結合を形成しており、そのような結合のもとで判断それ自体も合理的だと言われる。

だが、判断と情動や欲求の間に葛藤が生じることもある。赤と黒の縞模様のへびに遭遇して震え上がり、へびから逃げたいと思うにもかかわらず、知的にはそのへびは毒のない大人しいへびであるから、危険ではないと判断するような場合である。この場合、このへびは危険ではないという知的判断と、そのへびへの恐怖の情動およびそのへびから逃げたいという欲求の間には、葛藤が生じている。このような状況を Damasio 的に分析するなら、いったん恐怖の情動によって危険だというソマティック・マーカ―が付けられたへびが私の知的な推論過程によって危険ではないと判断されたということになる。こうした状況では、判断は情動や欲求と整合的ではなく、合理的な相互結合を形成していない。従って、判断か情動・欲求の少なくともどちらかが不合理とされることになる。

このような判断と情動や欲求の間で葛藤が生じた場合、合理的であるのはふつう判断の方であるとされる。そして葛藤は、通常、情動や欲求が変わることによって解消される。だが、Helm は判断と情動や欲求の間で葛藤が生じた場合、判断のほうが不合理であるとしてその変更が求められる場合もあると主張する (Helm 2010, p. 316)。これはかなり独特の見方であるといえる。では、このような見方がどうして出てくるのだろうか。

Helm によると、判断と欲求や情動の関係とは、それらが葛藤状態にある場合、単に判断が一方向的に情動や欲求に影響を与えてそれらを変えることによってその葛藤を合理的に解消すべきであるような関係ではない。それらは双方向に影響を及ぼしあうのである。それが可能であるのは、そもそもそれらは一緒になってある評価的展望(つまり、主体にとって意味あるもの、主体の価値体系)のもとで形成されるからである。その評

価的展望のもとでは、通常、情動や欲求と判断は調和的に統合され整合的であるが、ときに葛藤を生じることもある。しかし、その葛藤は対等なものどうしの葛藤なのである (Helm 2010, p. 317)。

単一の評価的展望のもと、判断と情動が食い違った場合、どのように解消されうるのだろうか。たとえば、誰もが約束を守るべきだという評価的展望のもと、友人 A に約束をすっぽかされたという状況において、友人 A は悪いことをしたという判断をしながらも、憤りを感じないとしよう。このような場合、私の情動が間違っているのであり、判断に合わせて然るべき情動を持つことが必要となる。これは情動に関する合理性の欠如である (Helm 2010, pp. 320-1)。

しかしながら、次のような判断と情動の食い違いも想定可能である。たとえば、私は最初、誰もがみな約束を守るべきであるということに意味を見出し、その意味に確かに関与していたとしよう。だが、時には約束を破ることが自己利益を最大化させるのであり、誰もがみなそのように自己利益を最大化するように振る舞うべきだという方向へ評価的展望 (私にとっての意味) を変えたとしよう。そのような評価的展望を持つ私は友人 A に約束をすっぽかされても、友人 A のしたことは悪くないと判断する。しかし、その一方で、私は友人 A に対して怒りを感じたとしよう。このとき、私の判断と情動の間には葛藤が生じている。私は、A がしたことは悪くないと判断しながら、情動的には悪いと評価している。この場合、自己利益を最大化すべきだという私の評価的展望に合致しているのはやはり判断であり、それゆえ情動が不合理なものとして変更されるべきであるということになる。しかし、私が評価的展望を変えたというのは、私が誤ってそう思っただけで、実は私は依然として約束を守るべきだということに意味を見出していたのだとしよう。そうだとすれば、実は評価的展望に合致しているのは、判断でなく、情動だということになるのであり、情動に合わせて判断を変える必要がある。これは判断に関する合理性の欠如の例であり、この場合、情動が判断に対して影響を及ぼすのが合理的だということになる。

このように、世界に対する単一の評価的展望のもと、判断と情動や欲求は合理的な相互結合を構成しており、それらは双方向的に影響を及ぼし合うべき対等な関係にあるのである (Helm 2010, pp. 319-21)。

これまで見てきたように、Helm は、情動が判断 (知的評価) や欲求とともに合理的な相互結合を形成し、それぞれが双方向に影響を及ぼし合うような関係であることを指摘し、その点に情動の合理性を見出した。但し、そのような情動は常に合理的なものではないということではない。つまり、合理的でなければ、もはや情動でないというわけではない。たとえば、情動と判断が食い違った場合、先の事例のように判断が誤っていることもあるが、情動のほうが誤っていることもありうるのであり、その場合、情動が変更されるべきものとなる。このように情動が時に不合理になることもある。Helm のいう情動の合理性とは、情動がときに不合理であることを許容するような合理

性なのである。情動の合理性に関する Helm のこのような捉え方は、基本的に正しいと言えよう。それは当人の価値観に相対的な合理性を捉えるにとどまるとはいえ、情動の合理性を的確に捉えていると考えられる。情動がふつう自己制御に貢献しながら、ときに自己制御を妨げることがあるということも、情動がおおむね合理的でありながら、時に不合理になることがあるということから容易に説明することができる。Helm の捉え方は情動の合理性に関する非常に優れた見方だと言える。

第六章 自己制御の喪失のあり方³⁶

自己制御とは最善の判断（何を為すのが最善かに関する知的な判断）に従って意思決定を行い、その意思どおりの行為を行うということである。二重システム理論の枠組みでいえば、システム2が最善の判断に従って意思決定を行い、その決定がシステム1の決定（欲求や情動に基づく決定）と一致するときには、そのままその行為が行われるが、食い違うときは、システム2がシステム1を制御して、システム1の決定した行為を実行するということになる。自己制御の喪失は、システム1とシステム2の決定が食い違うときの自己制御に関して生じる。この種の自己制御が失われるとき、その喪失には大きく分けて、二つのタイプがある。一つは、意思を形成する過程（意思決定のプロセス）で生じるものであり、もう一つは、決定した意思を実行する過程で生じるものである。本章では、自己制御の喪失のこの二つのタイプについて、それぞれどのようなものか考察していこう。自己制御の喪失はこれまで「意志の弱さ」と呼ばれてきた現象にほかならないので、本章の考察は意志の弱さのタイプとその内実についての考察ということもできる。

6.1 意思決定過程における自己制御とその喪失

意思決定の過程で自己制御の喪失が起こることがある。意思決定はシステム2で行われるが、システム2は意思決定を行うとき、まず物事の価値を知的に評価して、何をするのが最善かを判断し、その判断に基づいてどうするかを意思を決定する。たとえば、満員の急行電車に乗るか、それとも空いた各駅停車に乗るかを決めるとき、それぞれの選択肢の価値を知的に評価し、満員電車に乗るほうがよいと判断すれば、その判断に基づいてそうしようという意思を形成する。

ここで、最善の判断とそれに基づいて形成される意思は決して同じものではない。最善の判断はそれぞれの選択肢の知的評価に基づく判断であり、必ずしも動機付けの力を伴わないが、最善の判断に基づく意思は必ず動機付けの力を伴う。満員の電車に乗るのが最善だと判断しても、そうすることへの動機付けの力があるとは限らないが、満員の電車に乗ろうという意思を形成すれば、そうすることへの動機付けの力が必ずある。動機付けの力を含まないような意思は実は意思ではない。

最善の判断と意思が異なるとすれば、ふつう最善の判断に基づいて意思が形成されるとしても、ときには最善の判断とは合致しない意思が形成されることもありうる。満員の急行電車に乗るのが最善だと判断しても、それでも何らかの原因で満員の急行電車ではなく、空いた各駅停車に乗ろうという意思決定をしてしまうこともありうる。VMPFC 損傷患者は、その病的な事例だと考えられる。彼らは情動が平板化しているものの、知的な能力には殆ど異常がなく、知的な評価を十分行うことができる。従って、

³⁶ この章は西堤（2010b）の一部に基づくものである。

アイオワ・ギャンブル課題でも、どのカードの組が長期的に有利であり、どの組が不利かを知的に評価でき、それに基づいてどの組からカードを引くのが最善かを正しく判断することができる。それにもかかわらず、不利な組からカードを引こうという意思決定を行ってしまうのである。しかし、このような最善の判断と意思の不一致は VMPFC 損傷患者だけではなく、健常者にも見られる。満員の急行電車に乗るのが最善だと判断しながら、それでもやっぱり空いた各駅停車に乗ろうと意思決定してしまうというようなことは、健常者でもそれほど珍しいことではない。

意思決定の過程で生じる自己制御の喪失というのは、最善の判断に基づいてそれに一致するような意思を形成すべきであるにもかかわらず、そうすることができずに、最善の判断にそむく意思決定をしてしまうことである。最善の判断に一致するように意思を形成せよという原理は自制の原理 (Principle of Continenence) とよばれるが、意思決定の過程で生じる自己制御の喪失とはようするに自制の原理に背くことであり、それゆえ自制の欠如なのである³⁷。どうしてこのような自己制御の喪失が起こるのであろうか。

健常者であれば、システム2における知的評価であっても、常にシステム1における情動による評価に基づいており、それゆえその評価には情動のもつ動機付けの力が伴っている。しかも、通常、知的評価とそれに伴う動機付けの力はちょうど対応している。知的評価が大きいのに、動機付けの力は小さいとか、逆に知的評価は小さいのに動機付けの力が大きいということは、通常ない。従って、システム2において、知的評価に基づいてどの選択肢が最善かを定める最善の判断が、動機付けの力に基づいてどの選択肢を選ぶかを定める意思決定とずれることはない。たとえば、知的な価値評価に基づいて、満員の急行電車に乗るのが最善だと判断するなら、動機付けの力に基づいて、やはり満員の電車に乗ろうという意思を形成することになる。

しかし、VMPFC 損傷患者の場合、一部の情動³⁸に低下が見られるために、システム2における知的評価に必ずしもそれに対応する動機付けの力が伴わない。つまり、知的評価と動機付けの力が乖離するのである。そのため、システム2において知的評価に基づく最善の判断と動機付けの力に基づく意思決定がずれることがありうる。VMPFC 損傷患者がアイオワ・ギャンブル課題において、ある戦略が長期的には不利であると判断しながら、それでもその戦略を選択しようという意思決定を行って、それを実行するのは、彼らがある種の情動を十分持たないために、最善の判断の際に働く知的評価と意思

³⁷ 自制の原理については Davidson 1970, p. 41, 邦訳, p. 59 を参照。但し、Davidson は自制の原理を、最善の判断に一致するように行為せよという原理として述べているが、彼の場合、意思通りに行為がなされるということを前提しているため、これは結局、最善の判断に一致するよう意思を形成せよという原理となる。

³⁸ Damasio によると、情動は扁桃体を介して引き起こされる一次情動と VMPFC を介して引き起こされる二次情動の二種類がある (Damasio 1994, pp. 131-9, 邦訳, pp. 215-25)。VMPFC 損傷患者の場合、VMPFC を介して引き起こされる二次情動反応は低下するが、扁桃体を介して引き起こされる一次情動の反応は低下しない。

決定の際に働く動機付けの力がうまく対応しないからである。

システム2の知的評価を裏打ちする情動があるかどうかということと並んで考慮すべきことは、システム1の働きである。VMPFC 損傷患者は情動が低下しているが、いかなる情動も低下しているとすれば、損傷患者のシステム1は全く機能しないはずである。なぜなら、これまで見てきた通り、システム1は情動的な評価によって行為を決定し実行するシステムだと考えられるからである。従って、VMPFC 損傷患者がシステム1に基づく自制を欠いた行為をする以上、損傷者は何らかの情動をそれなりの強さで持つはずである。実際、彼らは VMPFC を損傷しているものの、前述したとおり、扁桃体など、他の情動に関係する部位を健常に保っている³⁹。従って、VMPFC 損傷患者は、アイオワ・ギャンブル課題において、不利な山から引くことへの否定的な情動をほとんど失っているものの、不利な山に含まれる大きな正の金額のカードを引くことへの賭博的なスリル感を強く持っている (Damasio 1994, pp. 214-217, 邦訳, pp. 325-8)。アイオワ・ギャンブル課題において、健常者も VMPFC 損傷者と同じくそのようなスリル感を持っているだろうが、健常者の場合、不利な山への否定的な情動も持っているため、いわばそれらが相殺されて、システム1において不利な山から引こうという動機付けは強くないか、あるいは逆の方が強い。それゆえ、健常者はシステム2がシステム1を制御して有利な山から引くことができる。それに対して、VMPFC 損傷者の場合、不利な山への否定的な情動が低下しているため、システム1において不利な山から引こうという動機付けが強い。従って、システム2がシステム1を制御するのが難しいのである。

知的評価と動機付けの力の乖離は、VMPFC 損傷患者の場合だけではなく、健常者においても起こりうる。たしかに健常者においては、通常、両者はちょうど対応している。しかし、健常者においても、知的評価が自分自身の十分な経験に裏打ちされておらず、他人から得た単なる知識に基づく場合には、それにふさわしい動機付けの力が伴わないことがありうる (Damasio 1994, pp. 179-80, 邦訳, p. 280)。遅刻しないように満員の急行電車に乗るのが最善だと判断するとき働く知的評価が、遅刻して上司から厳しく注意されて落ち込んだことがあるとか、得意先を失って落胆したことがあるというような情動経験に裏打ちされておらず、単に遅刻してはならないと教えられているだけであ

³⁹ 健常被験者、VMPFC 損傷患者並びに扁桃体損傷患者を対象にしたギャンブル課題において、扁桃体損傷患者は情動反応の指標である皮膚伝導反応をほとんど示さなかった。VMPFC 損傷患者は不利な山からカードを引こうとするとそれを否定的に評価する情動反応は健常被験者に比べてかなり弱かったものの、カードを引いた後はその報酬や損失に応じてそれなりの情動反応が見られた。だが、扁桃体損傷者の場合、いかなるカードを引いても情動反応がほとんど見られなかったのである。扁桃体は VMPFC と密接に関係しており、この部位を損傷すると、情動全般が損なわれると考えられる。それに対し、VMPFC 損傷患者はシステム2の知的評価に合致する情動のみが損なわれるのである。この点が VMPFC 損傷患者と扁桃体損傷患者の情動不全の異なる点であると考えられている (Bechara and Damasio 2005)。

れば、その評価に見合うだけの動機付けの力を伴わないだろう。従って、満員の急行電車に乗るのが最善だと判断しつつも、そうしようという意思決定を行わず、むしろ空いた各駅停車に乗ろうという意思決定を行うことになるだろう。そのような意思を形成する方が動機付けの力に適っているからである。

また、逆に、知的評価がそれに見合う以上の大きな動機付けの力を伴ってしまうこともありうる (Damasio 1994, pp. 191-2, 邦訳, pp. 295-6)。疲れたときに空いた電車に乗ってとても気持ちがよかったという経験をしたために、空いた電車に乗ることに不当に強い喜びを抱く場合がそうであろう。このような場合、遅刻しないように満員の急行電車に乗るのが最善だと判断しても、どうしても空いた各駅停車に乗ってしまうだろうが、それは、空いた各駅停車に乗るという選択肢の知的評価がそれに見合う以上の大きな動機付けの力を伴ってしまい、それゆえ空いた各駅停車に乗ろうという意思を形成してしまうからである。

このように、システム2における最善の判断に一致しない意思の形成、つまり意思決定過程における自己制御の喪失は、システム1における情動の異常などのせいで、知的評価と動機付けの力が乖離することによって生じると考えられる。但し、知的評価と動機付けの力の乖離があまり大きくなければ、システム2がシステム1を制御して、最善の判断に従う意思決定を行うことができる。システム2が最善の判断に従って意思決定を行うことができないのは、知的評価と動機付けの乖離がかなり大きい場合である。

6.2 執行過程における自己制御とその喪失

前節では、システム2において最善の判断が下されるが、それと一致する意思決定が行われないという意思決定過程における自己制御の喪失がどのようにして起こるのかを考察した。本節では、システム2が決定した意思が実現されず、システム1の決定した行為がなされるという執行過程における自己制御の喪失がいかんして生じるのかを考察していく。執行過程における自己制御の喪失は、四つの仕方で生じると考えられる。第一は、システム1が過剰な活動をすることによって、システム2がシステム1を制御できなくなる場合である (Burns and Bechara 2007, p. 271)。第二は、システム2が不本意にもシステム1に支配され、システム1の行為選択を事後的に正当化させられる場合である (Bechara 2002, p. 1702)。第三は、システム2の認知資源が何らかの原因で消耗することによって、システム2がシステム1を制御する力を失い、システム1の主導で行為が遂行される場合である (Levy 2007, 2011)。第四は、システム1で起こる価値の時間割引により、システム2で意思決定が行われた時点では、システム2はシステム1を制御できていたが、実際に行為する時間が近づくにつれて、システム1で決定された行為の価値がシステム1において非常に大きく見積もられるようになり、その結果、システム1が過剰に活動して、システム2による制御が効かなくなる場合である (Ainslie 2001)。以下では、これら四つのケースを順に詳しく考察していく。

6.2.1 システム1の過活動

システム2は、通常、システム1を制御している。しかし、ときにシステム1が何らかの原因で過剰な活動を起こすと、システム2がシステム1を制御できなくなる。システム2は通常の制御力を保持しているが、それにもかかわらず、システム1がシステム2を「圧倒」したり「ハイジャック」したりして、システム1主導で行為が行われるようになるのである (Burns and Bechara 2007, p. 267)。従って、システム2の意思決定がシステム1の行為決定と食い違う場合、システム2の決定した意思は実現されないことになる。このように、システム1の過活動により、執行過程における自己制御の喪失が生じる。

たとえば、薬物依存者の例を見てみよう (Burns and Bechara 2007, p. 271)。彼らは、しばらく薬物を摂取しないと、薬物に対する渴望がきわめて強くなる。そのため、システム1が薬物を摂取するという行為決定を行い、その実現に向けて過剰に大きく活動する。従って、システム2において薬物を摂取しないという意思決定を行っても、システム2はシステム1を抑制することができず、システム1が決定した行為が実現されてしまう。むしろ、薬物依存者は、薬物を摂取しないのが最善だと判断しても、システム1の過剰な活動に影響されて摂取しようという意思決定を行ってしまうこともある。しかし、そうではなく、摂取しないという意思決定を行ったとしても、システム1が決定した行為が実現されてしまうのである。こうして薬物依存者は、薬物を摂取しないという意思を形成しながらも、システム1の過活動によって、薬物を摂取することになってしまうのである。

また、薬物依存者を対象にしたアイオワ・ギャンブル課題の結果も、システム1の過活動による自己制御の喪失を示している (Bechara et al. 2002; Bechara 2003)。この結果によると、薬物依存者らは VMPFC 損傷患者と同じく、不利な山からばかりカードを引き、最終的な利益を最大化するという課題の目的を達成できない。なぜなら、彼らの扁桃体 (システム1の一部) が一時的に大きな報酬 (大きな正の金額のカード) に対して過剰に活動する (つまりそのような報酬の価値を不当に大きく評価する情動が生じる) からである。薬物依存者のシステム2が最終的な利益を最大化するために、有利な山からカードを引くという意思決定を行っても、過剰に活動するシステム1を抑止することができないため、システム1が決定した行為、すなわち一時的に大きな報酬を含む不利な山からカードを引くという行為が行われてしまうのである。

このように、システム1の過活動によって、自己制御の喪失が生じることがある。システム2が過剰に活動するシステム1を制御できなくなり、システム2が決定した意思ではなく、システム1が決定した行為が実現されてしまうのである。次の第二のケースは、システム1の過活動による自己制御の喪失の一つの特殊なケースであるともいえるが、非常に興味深いケースなので、別建てにして考察してみよう。

6.2.2 システム2による事後的正当化

禁煙している人が、喫煙の誘惑に負けて煙草を吸うことがある。しかも単に衝動的に吸ってしまうというわけではなく、一本の煙草にはほとんど健康への害はないとか、我慢する方が体に悪いとかいった理由を付けて、禁煙の意思を覆して新たに煙草を吸おうという意思を形成して、そのうえで煙草を吸うことがある。この場合、システム2によって新たに形成された意思に従って行為がなされており、そのかぎりでは意のままに行為がなされているといえる。確かに禁煙するというもともとの意思からすれば、それに反する行為がなされているが、もともとの意思を変更して新たな意思が形成されており、この新たな意思からすれば、意思通りのことが実現している。従って、そのかぎりでは、自己制御の喪失は起きていないように見える。しかし実際には、このようなケースは、システム1の過活動によって生じる自己制御の喪失だと考えられる。

Bechara et al. (2002, p. 1702)によると、システム1の過剰な活動のせいで、システム2はシステム1が決定した行為を後から正当化させられる場合がある。この場合、システム2はシステム1から独立に自律的に活動しているように見えてもそうではなく、システム1に支配され、その行為決定を事後的に正当化させられているのである。喫煙の誘惑に負けて、一本の煙草には健康への害はないなどの理由を考えて、煙草を吸おうという新たな意思決定を行うのは、システム2が自律的に行っていることではなく、システム1がシステム2に行わせていることである。つまり、システム1が煙草を吸おうという行為決定を行い、その後でそうするのが良いという理由をシステム2に考え出させているのである。この意味でシステム2はシステム1の決定を事後的に正当化させられているのである。そうだとすれば、新たな意思と合致した行為がなされたとしても、その新たな意思そのものがシステム2の自律的な活動によって形成されたものではない。従って、この新たな意思は一見、意思のように見えながら、実は本当の意思ではなく、意思もどきに過ぎず、行為はそのような意思もどきに合致した行為に過ぎない。それゆえ、ここにも自己制御の喪失が起きているのである。

6.2.3 自己資源の消耗⁴⁰

Levy (2007, pp. 209-15)によると、自己制御の喪失は自己資源の消耗(ego depletion)によって生じる。自己資源の消耗とは、自己資源、すなわちシステム2の認知資源(たとえば、忍耐力、集中力、比較考慮能力など)が何らかの原因によって消耗することである。自己資源の消耗が生じると、もはやシステム2がほとんど働かなくなり、システム2からシステム1へ主導権の移行が起こる。その結果、システム1主導の行為が遂行されることになる。このようにして、自己資源の消耗は、執行過程における自己制御の喪失を招くのである。

では、自己資源の消耗はどのようにして生じるのだろうか。自己資源の消耗は、大き

⁴⁰ 自己資源の消耗については第九章で主題的に扱う。

く分けると、三つの仕方（必ずしも排他的ではない）で起こると考えられる。第一に、誘惑に耐えることによって、自己資源の消耗が起こる（Mischel 1989）。誘惑に耐えるには、忍耐力のようなシステム2の認知資源が必要であり、それゆえ誘惑に耐えていると、システム2の認知資源が次第に消耗してくる。第二に、困難な課題に従事することによって、自己資源の消耗が起こる（Baumeister and Vohs 2007）。たとえば、何か選択を行ったり、計算を長時間行ったり、注意力を要する作業を行ったりすると、システム2の認知資源が消耗する。第三に、利用可能なグルコースが減少することによって、自己資源の消耗が起こる（Gailliot et al. 2007）。たとえば、糖尿病や低血糖症の人、あるいはダイエット中の人は、利用可能なグルコースが減少しており、誘惑に負けやすい傾向がある。グルコースはシステム2の認知資源を支えるエネルギーだと考えられる⁴¹。

自己資源の消耗によって、自己制御の喪失がどのように起こるのかを一つの例に即して具体的に見ておこう。Mischel (1989) とその同僚たちは、子供を被験者として、遅延満足課題（delay gratification task）の実験を行い、子供たちの遅延能力を測ることによって、誘惑に耐えることが自己資源の消耗を引き起こすことを明らかにした。この実験では、子供たちは実験者から、目の前に置かれたマシュマロを食べないように言われる。実験者は子供たちにそう言ってから部屋を離れ、しばらくしてから戻ってくる。そして子供たちがマシュマロを食べないでいられたかどうか、またその間、彼らが何を考えていたかを調査した。その結果、マシュマロを食べた子供たちは、マシュマロを目の前にして、甘いマシュマロのことを考え続けており、一方、マシュマロを食べなかった子供たちは、マシュマロを目の前にして、白い雲のことを考えていたり、マシュマロの魅力的でない側面について考えていたりした。

この実験結果は次のように解釈できよう。マシュマロを食べた子供たちは、甘いマシュマロのことを考え続けていたために、マシュマロを食べたいという誘惑にずっとさらされ、それゆえその誘惑にずっと耐えていた。そのせいで、彼らは忍耐力というシステム2の認知資源を消耗させ、その結果、システム2からシステム1への主導権の移行が生じた。こうして子供たちは、システム1の主導による行為、つまりマシュマロを食べるという行為を行ったのである。一方、マシュマロを食べなかった子供たちは、マシュマロとは別のことを考えたり、マシュマロの魅力的でない特徴を考えたりしていたので、マシュマロを食べたいという誘惑にさらされず、それゆえその誘惑に耐える必要がなかった。そのため、システム2の認知資源が消耗することもなく、システム2からシステム1への主導権の移行が生じなかった。こうして彼らはマシュマロを食べないでいることができたのである。

⁴¹ 自己資源の消耗を引き起こす以上の三つの要因は、必ずしも独立ではないかもしれない。Gailliot らは、血中グルコースの減少を自己制御的な認知プロセスの喪失の指標に用いており、結局は血中グルコースの減少によって自己資源の消耗が起こるとみなしていると考えられる（Gailliot et al. 2007）。

このように忍耐力というシステム2の認知資源が保持されていれば、自己制御の喪失が起こらないが、それが消耗すると、システム2からシステム1への主導権の移行が起こり、自己制御の喪失が生じるのである。

6.2.4 時間割引

私たちは時間の経過によってものごとの評価を変える場合がある。たとえば、昼食を終えた時点では、太ることを考慮して、3時のおやつケーキを食べるのは控えようと思いつき決定したとしよう。しかし、3時になって実際にケーキを目の前にすると、どうしても食べたくなくなってケーキを食べてしまう。この場合、お昼の時点でも、ケーキを食べると「おいしい」というプラスの評価をしてあったが、食べると「太る」というマイナスの評価の方が上回っていたために、食べないと決意した。だが、3時にケーキを目の前にした途端、太るというマイナスの評価はお昼と同程度のままであったが、おいしいというプラスの評価が急激に上昇したため、結局、食べてしまうのである。このような時間の経過による価値の逆転は私たちの日常においてときどき生じるが、どうしてこのようなことが生じるのだろうか。

それは価値の時間割引によって起こると考えられる。人はもともと、将来手に入る報酬と、現在ただちに手に入る報酬とでは、同じ報酬でも、その価値の受け止め方が異なる。ただちに得られる報酬はその価値が大きく評価され、将来得られる報酬はその価値が小さく評価される。従って、報酬が得られる時点が現在から遠いほど、その価値は小さく見積もられ、報酬が得られる時点が近づいてくると、その価値は大きく見積もられるようになる。つまり、報酬の価値は報酬が得られるまでの時間的な遅延と反比例して評価されるのである (Ainslie 2001, ch. 3, 邦訳, 3章; 佐伯 2011, 2章)。

このような価値の時間割引によって、異なる選択肢のあいだの選好の逆転が生じると考えられる (詳しくは 8.1 を参照)。先ほどの、ケーキを食べる例では、お昼の時点ではケーキのおいしさというプラスの価値は小さく見積もられたが、3時になりケーキを目の前にすると、おいしさのプラスの価値は急激に大きく見積もられるようになる。しかし、太ることのマイナスの価値は、お昼でも、3時でも、太るまでの時間がまだかなりあるために、ほとんど同じ大きさで評価される。従って、お昼の時点では、太ることのマイナスの価値がおいしさのプラスの価値を上回っていても、3時になると、逆に、おいしさのプラスの価値が太ることのマイナスの価値を上回るようになる。こうして、お昼には、ケーキを食べないという選択肢の方がケーキを食べるという選択肢よりも選好されたが、3時になると、その選好が逆転するのである。こうして、お昼には、ケーキを食べないと決意したにもかかわらず、3時になると、ケーキを食べてしまうのである。

但し、この場合、お昼に決定した、ケーキを食べないという意思は、3時になって、選好が逆転しても、必ずしもケーキを食べるという意思に変わるわけではない。ケーキ

を食べないという意思がなお存続する場合があります。3時になって、ケーキのおいしさが大きな価値をもつようになり、そのためケーキを食べてしまうが、相変わらず、ケーキを食べないという意思は持ち続けており、従って、不本意ながら、ケーキを食べてしまうのである。

では、このような意のままにならない行為は、二重システム理論のもとでは、どのように捉えられるのであろうか。まず、システム2では、ケーキのおいしさと太ることのそれぞれの価値の見積もりがいずれも、お昼と3時で、ほとんど変わらない。つまり、システム2においては、ものごとの価値が知的に見積もられるため、価値の時間割引があまり生じないと考えられるのである。これに対して、システム1においては、ものごとの価値が情動的に見積もられるため、価値の時間割引が大きく生じると考えられる。従って、ケーキのおいしさの価値は、3時になると、お昼よりも急激に大きく見積もられるようになるが、太ることの価値は、3時になっても、太るまでの時間がまだかなりあるため、お昼とほとんど同じくらいに見積もられる。こうして、システム1においては、お昼の時点で、ケーキを食べないという行為決定を行っても、3時になると、ケーキを食べるという行為決定に変わることになる。そして、このとき、システム1の活動が強いと、システム2はシステム1を抑止することができず、システム2がケーキを食べないという意思を持ち続けても、システム1の主導でケーキを食べるという行為が行われてしまうことになる。このように、システム1において生じる価値の大きな時間割引によって、執行過程における自己制御の喪失が生じることがあるのである。

以上、見てきたように、自己制御の喪失（意志の弱さ）には、大きく分けて、意思決定の過程で起こるものと執行過程で起こるものの二つのタイプがある。意思決定の過程で起こる自己制御の喪失とは、最善の判断に合致しない意思を形成してしまうことであり、執行過程で起こる自己制御の喪失とは意思に合致しない行為をしてしまうことである。この後者の自己制御の喪失には、システム1の過活動、システム2による事後的正当化、自己資源の消耗、価値の時間割引という四つの原因が考えられる。

第七章 信念形成における自己制御とその喪失

これまで意思決定および行為における自己制御の問題を考察してきた。本章では信念形成における自己制御の問題を論じよう。行為における自己制御の喪失はアクラシアと呼ばれるが、信念形成においてそれに対応する「認知的アクラシア」はありうるだろうか。この問題についてはその可能性を巡って論争がある。本章ではその論争を検討し、認知的アクラシアは不可能だという結論を導き出す⁴²。しかしながら、認知的アクラシアは不可能だとしても、信念形成において自己制御の喪失（システム2によるシステム1の制御の失敗）と呼びうるものはありうる。最後にこの点について論じる。

7.1 はじめに

手持ちの証拠からすると、あることを信じることに不思議はないどころか、むしろそう信じるべきだとさえ思うのに、実際はそう信じることができず、その逆のことを信じてしまうということがあるように思われる。たとえば、恋人が浮気をしているのかどうかということは、彼の携帯電話の履歴や最近の動向から明らかであるのに、彼のことが好きなあまり浮気をしているとは信じることができず、浮気をしていないと信じてしまう。また、証拠からすれば祖父が死んだことは明らかであり、それゆえ死んでいると信じるべきだとわかっているのに、そうだと信じることができず、死んでいないと信じてしまう。これらは一見すると、証拠からすればあることを信じるべきなのに、それとは反対のことを信じてしまう事例であるように思われる。本章では、このような認知的アクラシア（epistemic akrasia）が本当に成立可能かどうかを探究する。そのために、まず7.2では認知的アクラシアが厳密にはどのようなものかを明らかにし、7.3ではそのような認知的アクラシアの不可能性を主張する議論を検討し、そして7.4では逆に認知的アクラシアの可能性を探る議論を検討する。以上の検討を通じて、認知的アクラシアが不可能であるという結論を導く。さらに7.5では欲求の影響を考慮に入れた新たなタイプの認知的アクラシアが可能かどうかを検討し、それも不可能であるという結論を導く。そして7.6では、そのようなアクラシアが不可能な理由を信念の本性に照らして明らかにする。最後に、7.7では、認知的アクラシアが不可能だとしても、信念形成において自己制御の喪失とよびうる事態はありうることを論じる。

7.2 認知的アクラシアとは何か

認知的アクラシアが生じていると思われる場面を想像することは容易である。たとえば、ある殺人事件が起こって、友人のA氏が容疑者として浮かんだとしよう。犯行現場に落ちていた被害者の血の付いたナイフにはA氏の指紋が付いており、その時間にA氏のアリバイはなかった。これらの証拠からするとA氏が犯人であるように思える。

⁴² 本章はAdler（2002）の論文から大きな示唆を得ている。

だがその一方で、日ごろの A 氏の温厚な人柄や態度からすると、彼が犯人であるとは到底思えない。A 氏が犯人であることを示す諸々の証拠と、A 氏が犯人でないことを示す諸々の証拠を比較考量した結果、私は A 氏が犯人であると信じるべきだという結論に至る。だが、私は証拠に基づくと A 氏が犯人であると信ずべきだと思いながら、それに反して、A 氏は犯人でないと思ってしまう。このように、証拠からすると P だと信じるべきだと判断しながら P ではないと信じるのが認知的アクラシアである⁴³。また、そのような信念はアクラシア的信念 (akratic belief) と呼ばれる (Adler 2002, p. 2)。

アクラシアはもともと行為の場面で考えられたものである。認知的アクラシアは、この行為の実践場面で想定された実践的アクラシア (practical akrasia) を認識の場面に類比的に適用したものである (Adler 2002, p. 2; Levy 2004, p. 150)。そこで、認知的アクラシアがどのようなものかをより詳細に明らかにするために、まず実践的アクラシアがどのようなものとして特徴づけられているかを見ておこう。実践的アクラシアとは、あることを行うのが最善だと判断しながら、それとは異なる行為を実際には行ってしまふことであり、いわゆる意志の弱さ (weakness of will) と呼ばれるものである (Adler 2002, pp. 2-3; Levy 2004, p. 150)。たとえば、ハンバーガーとポテトを一緒に食べることで得られる満足感とそれによってもたらされるカロリーの過剰摂取を比較考量し、ハンバーガーを買うときにポテトは一緒に注文しないのが最善だと判断したとする。つまり、諸々の事情を比較考量してハンバーガーだけを注文するべきだと判断する。しかしそうであるにもかかわらず、結局はハンバーガーとポテトを一緒に食べることの満足感に屈して、ポテトも一緒に注文してしまう。このように A するのが最善であると判断したにもかかわらず、実際は A でない行為をするのが実践的アクラシアである。ただし、ここでは、行為は意思通りになされるものと想定されている。従って、この実践的アクラシアは、第六章で分類した自己制御の喪失タイプでいえば、意思決定過程における自己制御の喪失に相当する。つまり、それは執行過程における自己制御の喪失を含まない。

Adler (2002, pp. 2-3) によると、実践的アクラシアには特に二つの重要な特徴がみられる。ひとつは一人称的不合理性 (first-person irrationality) である。たとえば、靴かジャケットのどちらを買おうか考えて、諸々の事情を比較考量した結果、ジャケットを買うのが最善だと判断したとしよう。だが、靴やジャケットの魅力、値段、必要性などすべてを考慮してジャケットを買うべきだと判断したにもかかわらず、結局靴の魅力に負けて靴を買ってしまう。この靴の魅力は、靴とジャケットのどちらを買うのがよいかを比較考量する際、すでに考慮に入れられていたものであり、それを考慮に入れた

⁴³ 認知的アクラシアは信念形成が証拠だけにもとづく場合と、それに加えて欲求の影響も受ける場合に分けられるが、ここではまず、証拠だけに基づく場合を考察し、欲求の影響を受ける場合は、7.5 と 7.6 で考察する。なお、欲求の影響を受ける場合の認知的アクラシアは、Adler (2002, p. 19) が指摘するように、欲求の影響で自らの信念を歪めようとする自己欺瞞 (self-deception) の一種と言える。

うえでジャケットを買う方が良いと判断したのである。それにもかかわらず、靴の魅力ゆえに実際は靴を買ってしまう。この不合理性は本人自身も気づいているものである。つまり、本人の一人称的な視点からも不合理だと思われるものである。このように、実践的アクラシアには一人称的不合理性がみられる。

もう一つの特徴は、動機上の理解可能性 (motivational intelligibility) である。実践的アクラシアにおいては、最善の判断を下す際に、必ずしも動機の大きさが正しく反映されていない。ジャケットを買うのが最善だという判断は、靴の方がジャケットよりも魅力的だが、ジャケットの方が靴よりも安く、安さの方が魅力よりも重要だという比較考量によるものである。しかし、動機については、魅力の方が安さよりも動機付けの力が大きい。最も動機付けの力の強いものが最善の判断の際に反映されないために、ジャケットを買うのが最善だと判断しながら、実際には靴を買うということが起こってしまうのである。このように実践的アクラシアがなぜ生じるかは、動機の観点から十分理解可能である。以上のように、実践的アクラシアは、一人称的不合理性と動機上の理解可能性という二つの重要な特徴をもつ。

認識的アクラシアはこのような実践的アクラシアと類比的に規定される。ただし、それは行為ではなく認識に関わるために、欲求ではなく証拠が主要な役割を果たすことになる。すなわち、考慮した証拠の全体からすると P だと信じるべきだと判断しながら、その考慮した証拠の一部に基づいて P ではないと信じるのが認識的アクラシアである (Adler 2002, pp. 2-3; Levy 2004, p. 150)。認識的アクラシアにおいても実践的アクラシアと同様に、一人称的不合理性と動機上の理解可能性が成り立つとされる。ナイフに付いた A 氏の指紋や A 氏の温厚な性格などの証拠を比較考量して A 氏が犯人だと信じるべきだと思いながら、すでに考慮した A 氏の温厚な性格によって A 氏が犯人でないと思えるとき、私は自分がそのような不合理な信念を形成していることに気づいているように思われる。このような一人称的不合理性が成り立っているとされる。また、何を信じるべきかを比較考量しているときには、証拠がもっている信念形成を促す力は必ずしも正確に反映されない。従って、ナイフに付いた A 氏の指紋の方が A 氏の温厚な性格よりも重視されたために、A 氏が犯人だと信じるべきだと思つたとしても、ナイフに付いた A 氏の指紋という証拠がもっている「A 氏が犯人だ」という信念を形成する力は、A 氏の温厚な性格という証拠がもっている「A 氏が犯人でない」と信じさせる力より小さいかもしれない。そうだとすれば、実際には A 氏が犯人ではないと信じてしまうということが起こりうるように思われる。ここで、証拠がもっている信念形成を促す力は、行為における動機に対応している。それゆえ、動機上の理解可能性も成り立っていると考えられる。

このように認識的アクラシアは、実践的アクラシアと同様に、一人称的不合理性と動機上の理解可能性が成り立つものとして規定される。しかし、このような認識的アクラシアは本当に成立しうるのだろうか。次の節からは認識的アクラシアの不可能性を主張

する議論と、逆にその可能性を主張する議論を順に検討し、それを通じて認識的アクラシアが可能かどうかを吟味していく。次の節では、まず認識的アクラシアの不可能性を唱える論者の主張を検討してみよう。

7.3 認識的アクラシア不可能論

認識的アクラシアとは、考慮した証拠の全体からすると P だと信じるべきだと判断するにもかかわらず、その証拠の一部に基づいて P ではないと信じることである。認識的アクラシアの不可能論を主張する論者が依拠しているのは、P であることを示す証拠とその反対の証拠を比較考量して P だと信じるべきだと判断すれば、そのときもはや反対の証拠の力は無化され、それゆえ P ではないことを信じることはありえないという点である (Adler 2002; Owens 2002)。

たとえば、ナイフに付いた A 氏の指紋、彼のアリバイのなさ、彼の温厚な性格など、支持する証拠と反対の証拠を比較考量することによって、最終的に A 氏が犯人だと信じるべきだと判断すれば、その時点で反対の証拠は私にその証拠に基づいて信念を形成させる力をもはや持っていないように思われる。A 氏は確かに温厚な性格であるが、だからといって殺人を行わないとは限らない。何かやむにやまれぬ事情によって殺人を犯してしまったのであろう。そう考えて、A 氏が犯人だと信じるべきだと結論するとき、もはや A 氏の温厚な性格は、A 氏が犯人であることに反対する力（言い換えれば、A 氏が犯人でないことを支持する力）を持っていないように思われる。

もし反対の証拠がその力を保持しているとすれば、私は心から A 氏が犯人だと信じるべきだとは思えないであろう。確かに証拠は A 氏が犯人である方を強く支持しているように思われるが、そうでない可能性も払拭できない。そうだとすれば、A 氏が犯人だと心から信じるべきだということではなく、せいぜいかなりの強さでそう信じるべきだということにしかならないであろう。つまり、十全な信念 (full belief) ではなく、部分的な信念 (partial belief) ⁴⁴を形成すべきだということにしかならないだろう

(Adler 2002, pp. 12-3)。A 氏が犯人だという判断を十全な信念として形成するためには、反対の証拠の力を無化することができなければならない。反対の証拠の力が無化されてしまえば、もはや反対の証拠に基づいて A 氏が犯人でないと信じることはありえないからである。A 氏の温厚な性格がもはや A 氏が犯人であることに反対する力を持っていないければ、それに基づいて A 氏が犯人ではないという信念を形成させることはできないし、それゆえ A 氏が犯人ではないと信じることもありえない。その際形成さ

⁴⁴ 部分的な信念の場合、それに反する証拠がその力をまだ保持しているので、一見、アクラシア的信念でありうるようにみえるかもしれない。だが、この信念は理論的推論における信念形成へと向かう一段階に過ぎず、本来、まだ信念とは言えないものである (Adler 2002, pp.12-3)。なお、部分的な信念は確率的な信念（あることがある確からしきを持つという信念）と似ているが、確率的な信念が確率的な内容に対する十全な信念であるのに対し、部分的な信念は十全な信念ではない。

れる信念は、A氏が犯人であることを示す証拠のみに基づいたA氏が犯人であるという（十全な）信念でしかありえないのである。

認識的アクラシアの不可能論者によると、認識的アクラシアと実践的アクラシアの根本的な違いはここにある（cf. Adler 2002, pp. 12-5）。実践的アクラシアの場合、諸々の事情を比較考量して、ジャケットよりも靴を買う方が良いと判断しても、靴の魅力は相変わらず靴を買わせる力（動機付けの力）を保持したままである。ジャケットを買うのが最善だと判断したからといって、靴の魅力に備わる動機の力が必ずしも無化されるわけではない。従って、靴の魅力に屈して、靴を買ってしまうということも起こりうるのである。しかし、認識的アクラシアの場合には、A氏が犯人だと信じるべきだと判断すれば、反対の証拠はもはや力を失い、十全な信念が形成される。それゆえ、A氏が犯人ではないと信じることはありえないのである。

これを別の観点からいえば、認識的アクラシアにおいては、動機上の理解可能性が原理的に成立しえず、それゆえ認識的アクラシアは不可能だということになる。認識的アクラシアにおいて動機上の理解可能性が成立するためには、諸々の証拠からPを信じるべきだと判断するとき、その反対の証拠は依然としてその力を失わず、しかもそう判断する過程で想定されていた力よりも実際は大きな力を持つということが可能でなければならない。しかし、Pだと信じるべきだと判断したとき、それに反対する証拠がその力を無化されるとすれば、そのようなことは原理的に不可能である。従って、認識的アクラシアは成立不可能なのである。

7.4 認識的アクラシア可能論

認識的アクラシア不可能論の核心をなすのは、Pだと信じるべきだと判断すれば、Pに反対する証拠の力は必ず無化されるという見方である。Levyはこの証拠に関する見方を取り込み見解（subsumption view）と名付け、これが成り立たない可能性を論じている（Levy 2004, p.149）。つまり、Pだと信じるべきだと判断しつつもPに反する証拠の力が保持されることがありうるということである。もしそうであれば、認識的アクラシアが生じる可能性が開かれるということになる。以下では、認識的アクラシアの可能性を開こうとするLevyの主張を見ていこう。

Levy（2004, pp.153-6）は、信念に反する証拠があるにもかかわらずそれが無化されないものとして、哲学的信念に関する証拠を挙げている。哲学的信念は、一般的にある哲学的論争において、どの立場がもっとも正しいかを見極めることによって形成される。だが、哲学的論争の場合、たとえどの立場をとったとしても、自分の立場よりも他の立場のほうがある種の事実をより良く説明するように思われるということがしばしば起こる。たとえば決定論と自由意志論（決定論と両立しない自由な意志があるとする説）の対立の場合、決定論を心から信奉しているとしても、決定論よりも自由意志論の方が私たちの自由な行為をより良く説明するように思われる。つまり、私たちの自由な行為

は自由意志論を支える力を失わない。哲学的信念を形成する場合、決定的な証拠 (knockdown evidence) を手に入れることはほとんど不可能であり、対立する見解を支える論拠が哲学者を悩ませ続けることになる。しかしそうではあっても、哲学者は徹底した考察によってある立場が正しいと信ずるに至るのであり、その時形成される信念は部分的な信念ではなく十全な信念なのである。このように Levy は哲学的信念の場合、反対の証拠が力を保持したまま十全な信念を抱くことがありうることを主張するのである。

Levy が主張するように、哲学的信念の場合、十全な信念が形成されてもなお、反対の証拠が力を失わないことがあるとすれば、それはある信念を形成すべきだと思ってもなお、その信念に反対する証拠が力を失わないことがありうることを意味する。従って、認識的アクラシア不可能論者が拠り所としている取り込み見解は必ずしも成り立たないのである。Levy の主張するような反対証拠が無化されない事例は、哲学という特異な場面だけではなく、もっと通常の場合でも想定可能かもしれない。たとえば裁判の場合、裁判官は有罪か無罪かについて断定的な判決を下さなければならない。その場合、無罪であることを示す証拠の力を感じつつも、有罪だという判決を下さなければならないことがあるかもしれない。そのとき、裁判官は有罪だという十全な信念を抱きつつも、なおそれに反対する証拠の力を無化することはできないのではなかろうか。そうだとすれば、通常の場合でも、反対の証拠に力がありながら十全な信念が形成されることは可能であるように思われる。

しかし、本当に十全な信念が形成されながら、反対の証拠が力を失わないということはあるのだろうか。十全な信念が形成される以上、やはり反対の証拠はその力を失うように思われる。もし力を失わないなら、形成された信念は実は部分的な信念に過ぎないように見える (Adler 2002, pp. 12-3)。十全な信念が形成される時、通常、反対の証拠の無化が試みられる。たとえば、A 氏が犯人であるという十全な信念を形成しようとするとき、A 氏が実は凶暴な一面も持つという新たな証拠を見出すことにより、A 氏の温厚な性格は A 氏が犯人であることに反対する力を喪失させられる。このように具体的に反対証拠の力を無化することができない場合は、反対証拠が何らかの仕方で無化されうるだろうと信じることで、その力が無化される。これは、反対証拠の無化は今のところできていないが、いつかある仕方で無化することができるだろうという展望を持つことができ、それによって無化がなされるということである。たとえば、温厚な性格の A 氏がなぜ殺人という凶悪な犯罪を行ったのかは依然として謎であるが、それでもたとえば、A 氏のような温厚な性格の人でもかっとなって殺人を犯すことがあるといった類のことがやがて見出されるだろうという展望を持つことができれば、それによって A 氏の温厚な性格は反対の証拠の力を失うだろう。

もしこのような仕方で反対の証拠の力を無化することができなければ、やはり形成される信念は部分的な信念にとどまるであろう。A 氏の温厚な性格が反対の力を保持している限り、私は心から A 氏が犯人だと信じることはできないだろう。哲学的な信念の

場合も同様であろう。心からある立場を信奉するのであれば、その立場に反対する論拠は具体的に無化されるか、あるいは何らかの仕方では無化されると信じていることができなければならないだろう。そうでなければ、その立場を心から信じていることはできないように思われる⁴⁵。裁判官の場合も、同様であろう。おそらく裁判官は、たとえ断定的な判決を下したとしても、反対の証拠の力を無化できないときは、心の中でせいぜい部分的な信念を形成しているに過ぎないであろう。そして、そのような部分的な信念しか形成できなくても、とにかく断定的な判決を下さざるを得ないので、そのような判決を下すのであろう。

以上のように、十全な信念が形成されるときは、やはり反対証拠が無化されるとすれば、認識的アクラシアは不可能であろう。

⁴⁵ Ribeiro (2011) は、認識的アクラシアが成立する事例として、懐疑論を信奉する哲学者を挙げている。彼によると、懐疑論者であれば、たとえば、私には二つの手があるという命題に対する懐疑論的な議論が完全に説得的であると判断することによって、私には二つの手があるとも信じないし、そうでないとも信じないという判断保留の態度をとるべきだと判断する。しかし、そうであるにもかかわらず、懐疑論者は自分の手が二つあるかどうかということについて、判断を保留することなしに、二つの手があると信じているのである。これはまさに認識的アクラシアの事例であると Ribeiro は主張する。こうして彼は、P と信じているという態度と P でないと信じているという態度に加えて、どちらとも信じないという判断保留の態度を P に関する態度の選択肢として付け加えることによって、認識的アクラシアの議論に新たな局面を切り開いている。だが、判断保留状態の場合、P を支持する証拠とそれに反対する証拠が互いに拮抗して、P とも P でないと決定できない状態なので、どちらの証拠もその力が無化されていないのではないだろうか。もしそうだとすれば、判断を保留すべきだと思いながら、P だと信じたとしても、P の反対証拠が無化されずにその力を保っているのだから、心から P を信じていることにならないだろう。つまり、そのとき形成された信念は十全な信念ではなく、部分的な信念に留まるだろう。実際、夫が浮気をしているかどうかに関して判断を保留すべきだと思いながら、夫が浮気をしていないと信じている場合のような通常の非懐疑論的なケースであれば、部分的な信念に留まるだろう。しかし、これに対して懐疑論的なケースは別だと主張されるかもしれない。すなわち、懐疑論的なケースでは、懐疑論者が懐疑的な議論に基づいて判断を保留すべきだと思いながら、それでも自分には二つの手があると信じている場合、心からそう信じているのだと言われるかもしれない。懐疑論者は確かに懐疑論的な議論に十分説得力を感じているが、だからといって自分に二つの手があるという信念が部分的なものに留まることはないように思われる。だが、もしこれが部分的な信念ではなく、十全な信念であるとしたら、それはもともと懐疑論的な議論がいくら説得的に思っても、実は通常の信念形成には全く影響を及ぼさない特殊なものだからではないだろうか。つまり、単なる懐疑のための懐疑という特殊な議論だからではないだろうか。その特殊性をここで詳細に論じることはできないが、そうだとすれば、懐疑論的な議論に基づいて判断を保留すべきだと思っても、それは実際に判断を保留させるような働きをするわけではないのである。従って、二つの手があると心から信じているとしても、それは判断を保留すべきだという思いに逆らってそう信じているわけではない。その思いとは無関係にそうするのである。従って、このケースは認識的アクラシアには相応しないように思われる。

7.5 欲求の影響

これまでの考察によれば、証拠からすると P だと信ずべきだと判断すれば、P に反対する証拠の力が無化され、それゆえ P ではないと信じることはありえないということになりそうである。つまり、認識的アクラシアは不可能であるということだ。しかし、日常的な場面では、証拠だけではなく、欲求が信念の形成に影響を及ぼし、それによって認識的アクラシアが生じているかのように思われるケースがある (cf. Adler 2002, pp. 8-10)。この節では、欲求の影響によって認識的アクラシアが生じうるかどうかを検討していこう。

たとえば、恋人が浮気しているかどうかという場合、浮気をしていることを示す証拠 (彼の外泊が増えたことと最近の彼の様子がおかしいことなど) と、浮気をしていないことを示す証拠 (彼は私に対してやさしいなど) があつたとしよう。純粋に理論的な推論を行うなら、証拠を比較考量して、そこから彼が浮気をしていると信ずべきであると判断し、実際にそう信じるようになるはずである。しかし、浮気をしていると信ずべきであると思いつつ、浮気をしていないと信じてしまう。というのも、浮気を支持する証拠やそれに反対する証拠とは別に、自分だけを愛してほしいという欲求が、理論的推論過程やそのあとの実際の信念形成過程に影響を及ぼすからである⁴⁶。ここで、欲求がそれらに影響を及ぼす仕方には、欲求が理論的推論過程に働きかけるか、それとも実際の信念形成過程に働きかけるか、また意識的に働きかけるか、それとも無意識的に働きかけるかで、四通りの場合が想定できるだろう。

まず、自分だけを愛してほしいという欲求が理論的推論過程ではなく、実際の信念形成過程に働きかけ、かつ、無意識的に働きかける場合を考えてみよう。この場合、証拠を比較考量している理論的推論には欲求は影響を及ぼさないが実際に信念形成を行う段階になると、突如としてこの欲求が無意識的に働き出す。そして、理論的推論により、証拠からすれば彼が浮気をしていると信じるべきだと思いつつも、その欲求のゆえに、彼は浮気をしていないと信じてしまう。

このようなケースは、果たして認識的アクラシアだと言えるだろうか。Adler (2002, p.11) によると、これはそうではありえない。なぜなら、彼は浮気をしていないと信じるようにさせた欲求は証拠を比較考量している理論的推論の段階では考慮に入っていないからである。つまり、彼が浮気をしていると信じるべきだと判断する際にすでに考慮されていた理由によって、彼は浮気をしていないと信じるようになったわけではない。しかも、欲求が無意識的であるため、本人からすれば、どういうわけか突如、彼は浮気をしていないと信じてしまったのである。従って、ここではそう信じるのが本人の観

⁴⁶ ここで、理論的推論過程というのは信念を形成させる力とは別に理論的に証拠を考慮する過程であり、その結果はどんな信念を形成すべきかの判断であるが、それに対して、信念形成過程というのは信念を形成させる力が働いている過程であり、その結果としては何らかの信念が実際に形成される。

点から不合理だと理解されているわけではなく、それゆえ一人称的不合理性が成り立っていない。それゆえ、このケースは認識的アクラシアとはいえないのである。

次に、自分だけを愛してほしいという欲求が理論的推論過程に無意識的に働きかける場合を検討してみよう。理論的推論は、欲求が理由として働く実践的推論とは異なり、あくまでも証拠（つまり認識的理由（epistemic reason））のみに基づいて何を信じるのが最善かの判断を下す（Adler 2002, pp.2-5）。しかし、欲求は証拠の重みに影響を与えることによって、理論的推論に影響を与えうる。自分だけを愛してほしいという欲求は、たとえば彼が外泊したという証拠を軽視させ、彼が自分にやさしいという証拠を重視させるかもしれない。そうだとすれば、私は証拠からすると、彼が浮気をしていないと信じるべきだと判断することになる。そして実際、そう信じるだろう。このケースは認識的アクラシアだと言えるだろうか。明らかにそうではない。なぜなら、何を信じるべきかの判断と実際の信念が一致しているからである。

では、自分だけを愛してほしいという欲求が実際の信念形成過程に意識的に働きかける場合はどうであろうか。これは最初の事例と同様に、証拠を比較考量している理論的推論には欲求が影響を及ぼさないが、そのあとの信念形成において突如として欲求が影響を及ぼす。しかし先ほどの場合と違って、欲求は意識的に影響を及ぼす。そのため、彼は浮気をしていると信じるべきだと思いつつも、浮気していないと信じてしまうのは、自分だけを愛してほしいという欲求の影響であることに気付いている。しかし、そうだとすると、やはり最初の場合と同じく、その欲求は理論的推論において考慮されていたのではない。従って、この場合にも、一人称的不合理性が成立していない。それゆえ、認識的アクラシアとは言えないのである。

最後は、自分だけを愛してほしいという欲求が理論的推論過程に意識的に働きかける場合である。この場合は、その欲求が彼の外泊を軽視させ、自分へのやさしさを重視させることに私は気付いている。そのため、欲求によって証拠の重みが左右されてはならないと考えて、欲求の影響を排除できるかもしれない。そうだとすれば、結局、証拠からすると彼が浮気をしていると信じるべきだと判断することになる。ここで、自分だけを愛してほしいという欲求が実際の信念形成過程にも意識的に働きかけたとしたらどうであろうか。その場合は、彼は浮気をしていると信じるべきだと思いつつも、浮気をしていないと信じることになる。これは認識的アクラシアではないか。しかし、この場合でも、浮気をしていないと信じさせた欲求は理論的推論のなかで理由として考慮されたわけではなく、むしろ理由にならないとして、そこから排除されたものである。従って、理論的推論において考慮された理由に基づいて浮気をしていないと信じるようになったわけではない。つまり、ここでも一人称的不合理性は成り立っていないのである。それゆえ、このケースも認識的アクラシアとは言えないのである。

これまで見てきたように、欲求の影響によって認識的アクラシアは生じえない。たとえ一見するとそう思われるものであっても、実際に信念を形成させた要因である欲求が

理論的推論過程のなかで理由として考慮されているわけではないからである。だがもし、理論的推論過程における比較考量の段階で認識的理由だけではなく欲求も考慮にいれることが可能であるとしたらどうだろうか。次の節では信念の本性からして、そのようなことは不可能であり、だからこそ欲求の影響による認識的アクラシアは決して生じえないのだということを確認しよう。

7.6 信念の本性

純粋な理論的推論の場合、証拠（認識的理由）だけが信念形成の理由となると想定される（Adler 2002, pp. 2-5）。しかし、理論的推論において欲求も信念形成の理由となることはないのだろうか。この節では、欲求が信念形成の理由となることが信念の本性からしてありえないことを確かめ、それゆえ欲求の影響による認識的アクラシアが生じえないことを確認しておこう。

恋人が浮気をしていると信ずべき諸々の証拠や欲求と、浮気をしていないと信ずべき諸々の証拠や欲求を比較考量して、浮気をしていると信ずべきだと判断しながら、実際には浮気をしていないと信じることはありうるだろうか。恋人の外泊や行動のおかしさからすると、恋人は浮気をしていると信じるべきだが、恋人のやさしきや自分だけを愛してほしいという欲求からすると、恋人は浮気をしていないと信じるべきである。これらの理由を比較考量して、浮気をしていると信じるべき理由の方が重いと考えて、最終的に恋人は浮気をしていると信じるべきだと判断する。しかし、実際には、自分だけを愛してほしいという欲求に負けて、浮気をしていないと信じてしまう。このようなことはありうるだろうか。もしありうるとすれば、これは欲求の影響による認識的アクラシアといってよいのではなからうか。この場合、一人称的不合理性と動機上の理解可能性はともに成り立っている。

しかし、欲求が信念形成の一つの理由となって信念が形成されるということは、信念が通常的行為と同様に意志的に形成されるということである。だが、そもそも私たちは意志的に信念を形成することなどできるのだろうか。

たとえば、地球は四角形であると信じるならば殺さないが、そうでなければ殺す、と言われたとしよう。しかし、そう言われても、おそらく私たちは地球が四角形であると信じることはできないだろう（信じるふりはできても）。地球が丸いことの証拠と殺されたくないという欲求を比較考量すれば、地球は四角形だと信じるべきだと私たちは判断するかもしれない。しかし、それでも、私たちは地球が四角形だと信じることはできないだろう。それは私たちの意志が弱いからではない。信念はそもそも意志的なものではないからである。もちろん、前節で考察したように、信念が欲求の影響をうけることはある。殺されたくないという欲求によって、地球が四角いと信じてしまうことはありうる。しかし、これは殺されたくないという欲求を信念形成の理由として考慮することによって信念を形成したものではない。それは単に欲求が信念形成に因果的に影響を与

えたと過ぎない。それゆえ、それは信念を意志的に形成したものではない。このように、欲求の影響によって歪んだ信念が形成されることはあるが、欲求を理由として考慮することによって意志的に信念が形成されることはありえないと思われるのである。

Williams (1973, ch. 9) や Owens (2002, pp. 392-5) によると、信念は真であることのみを目的としており、純粹に認識的な観点から信念の内容が真であるかどうかだけを問題にしている⁴⁷。それゆえ、私たちは与えられた証拠に基づいて信じたり信じなかつたりできるだけで、証拠に反した信念や証拠に基づかない信念を意志的に形成することはできないのである。地球が四角であると信じなければ殺すと言われたとしても、地球が四角形であると信じることはできないし、仮にそうすることができたように見えたとしても、それはもはや信念とはいえない。なぜなら、信念は真であることの証拠に基づいて形成されるべきものであり、本質的にもっぱら真理をめざすものだからである。従って、欲求によって意志的に形成された信念は、仮にそのようなものがありえたとしても、実は本質的には信念ではありえないのである⁴⁸。

恋人が浮気をしていることに関係する証拠と自分だけを愛してほしいという欲求を比較考量して、浮気をしていると信じるべきだという判断をしたにも関わらず、浮気をしていないと信じてしまうということが起こったとすれば、実はそこで問題になっているのは信念ではなく、それとは別の心的状態（意志的に形成できる心の状態）である。そしてそこで生じたことは、認識的アクラシアではなく実践的アクラシアである。つまり、欲求が信念とは別の心の状態を形成する際に、理由として考慮されることはありうるとしても、欲求が信念形成において理由として考慮されることはありえないのである。それゆえ、欲求の影響による認識的アクラシアはそもそも原理的に不可能なのである。

7.7 信念形成における自己制御の喪失

これまで認識的アクラシアが可能かどうかを検討するために、認識的アクラシアの不可能性を唱える論者の主張と可能性を探る論者の主張を比較することにより、認識的アクラシアが不可能であることを示した。また、欲求が信念形成に影響する場合に認識的アクラシアが成立しうるかどうかということも検討したが、それは結局、信念の本質からして不可能であることを示した。以上より、認識的アクラシアは信念の本質からして不可能であると結論付けることができよう。

⁴⁷ Williams (1973) によると、信念には満たさなければならない五つの特徴があり、信念は真であることを目指すということと、信念は証拠に基礎づけられるということはその内の二つの特徴である。意のままに形成された信念が成立不可能であるのは、その二つの特徴を満たさないからである。また、Adler (2002, pp. 19-20) でもこの Williams の論点については言及されている。

⁴⁸ 本論文では、信念が意志的に形成されるものではないという観点から欲求を考慮に入れた新たなタイプの認識的アクラシアの可能性を退けているが、Adler (2002, pp. 10-2) も、人は認識的理由と、欲求のような認識的理由でないものを区別する能力があり、欲求は理論的推論の中では考慮され得ないと主張している。

だが、たとえ認知的アクラシアが不可能だとしても、信念形成における自己制御の喪失だといえるケースは存在する。それは、ひとつには、欲求や情動⁴⁹が意識的にせよ無意識的にせよ実際の信念形成過程に働きかけ、何を信じるべきかの判断と一致しないような信念が形成される場合である。たとえば、恋人が浮気をしていると信じるべきだと判断しても、自分だけを愛してほしいという欲求や恋人を失うのではないかという不安によって、浮気をしていないと信じてしまうような場合である。この場合、信念形成過程に影響を与えた欲求や情動は、証拠を比較考量している理論的推論の段階では考慮に入っておらず、その点で一人称的不合理性は成立していない。従って、このケースは認知的アクラシアではない。だが、それは自己制御の喪失のケースだとは言えよう。恋人が浮気をしていると信じるべきだと判断するなら、実際にそう信じるべきである。自分だけを愛してほしいという欲求や恋人を失うのではないかという不安の影響で、浮気をしていないと信じそうになってしまったら、そのような欲求や情動の影響を排して、浮気をしていると信じるようにすべきである。そうするのが合理的であり、そうしないのは不合理である。それゆえ、そうするのが自己制御の遂行であり、そうしないのは自己制御の喪失である。これを二重システム理論の枠組みで捉えたとすると、システム2がシステム1を制御し損ねているケースだということになる。システム2は浮気をしていないと信じさせようとするシステム1の影響を抑制しきれないのである。

また、理論的推論過程において欲求や情動が無意識的に影響を及ぼして、何を信じるべきかの判断を歪める場合も、信念形成における自己制御の喪失だとみなせる。この場合、欲求や情動が理論的推論過程において証拠の重みに影響を与えるものの、何を信じるべきかの判断に従って実際に信念が形成されるので、一見すると自己制御の喪失は生じていないように見える。だがこの場合、実は利用可能な証拠の比較考量が不適切な欲求や情動の影響によって正しく行われず、何を信じるべきかの判断が歪められたという点で自己制御の喪失が生じているのである。これを二重システム理論の枠組みで捉えたとすると、システム1の過活動によってシステム2がシステム1にとって都合のよい信念の正当化をさせられているのだと言えよう。理論的推論過程は欲求や情動に支配されて、欲求や情動にとって都合のよい信念を正当化するようにさせられているのであり、そのせいで何を信じるべきかの判断が捻じ曲げられているのである。そうだとすれば、その判断に従って実際に信念が形成されたとしても、それは理論的推論過程の自律的な働きによって利用可能な証拠を比較考量し、それに基づいてなされた判断とは言えない。従って、そのような判断に従って実際に信念が形成されても、それは信念形成における自己制御の喪失だと言わざるをえないのである。

⁴⁹ 認知的アクラシアを巡る従来の議論では、欲求の影響は考慮されるものの、情動の影響については明示的に論じられていなかったもので、前節までは欲求の影響だけを問題にしてきた。しかし、信念形成における自己制御の喪失に関しては、欲求と並んで情動の影響も重要なので、ここでは情動についても考察する。

このように信念形成においても、不適切な欲求や情動の影響によって、自己制御の喪失が起こりうる。このような喪失が起こらないようにするためには、システム2によって欲求や情動の影響を受けないようにすることが重要であるが、もっと重要なのは、そもそもそのような影響を受けないようにするために、不適切な欲求や情動を持たないようにすることである。逆に、適切な欲求や情動を持つことによって、合理的な信念形成が可能になることもある。つまり、適切な欲求や情動が信念形成における自己制御に貢献することがあるのである。

たとえば、次のような場合を考えてみよう。私は得られた諸々の証拠からすると A 氏が犯人だと信じるべきだと思うが、そう結論することには強い不安を感じる。そして、その不安のゆえに、A 氏は犯人ではないと信じるべきだと思うようになり、実際にそう信じるようになる。さらに、実際にこの信念は正しく、A 氏は犯人ではなかった。但しここで、A 氏が犯人だと結論付けることへの不安は、A 氏が犯人であってほしくないという欲求によるものではなく、得られた証拠には反映されていない重要な情報（たとえば、A 氏の微妙な態度）を捉えたものである。しかし、私はこの不安がどんな情報を捉えているかは分からず、ただ不安を感じるのである。それでも、その不安が何か重要な情報を捉えているだろうと思い、その不安のゆえに、A 氏は犯人ではないと信じるべきだと思うようになるのである。このケースでは、私の不安が、A 氏が犯人でないという信念を合理的に形成するのに貢献しているように思われる。このように情動が信念形成における自己制御にそれなりに寄与することもあると言えるのである。

私たちは時として、いったんやると決意したにもかかわらず、目の前の利益に惑わされて、一度決意したことを反故にする。たとえば、少し太ってしまったと感じるので、今週末は甘いものを控えようという決意を週初めにしたとする。週末までたつぷりと時間のある間は、その決意は揺るがない。しかし、いざ週末になって甘いものを目の前にするとどうしても食べたくなくなってしまう。そして、甘いものを控えようという決意を翻して食べてしまう。あるいは何とか決意は保っても、それでも食べてしまう。このような現象は「意志の弱さ」と呼ばれる。逆に、甘いものを控えるという決意を堅持し、しかもその決意通りに甘いものを控えることができれば、それは「意志の強さ」と呼ばれる。

だが、そもそも「意志」とは何だろうか。意志そのものが何であるのかを説明するために、従来、様々なモデルが提示されてきたが、ここでは、そのうちの二つのモデルを特に取り上げて検討したい。一つは、意志とは器官のようなものであると主張する「器官理論 (Organ Theory)」であり、もう一つは意志とは器官などではなく交渉状態であると主張する「異時点間交渉モデル (Intertemporal Bargaining Theory)」である。この二つのモデルを比較検討することにより、どちらがより意志のあり方を説明するモデルとして相応しいのかを明らかにし、意志とは何かを少しでも明確化したい。

意志の弱さ（つまり自己制御の喪失）と価値の時間割引については、すでに第六章で触れたが、ここではその点についてももう少し詳しく考察しておきたい。というのも、意志の異時点間交渉モデルは、価値の時間割引と深い関係にあり、それについての詳細な理解が必要とされるからである。本章では、まず、意志の弱さと価値の時間割引について詳しく考察し、それに続いて意志に関する器官理論と異時点間交渉モデルの比較検討を行う。

8.1 価値の双曲的時間割引と意志の弱さ

私たちは報酬が手に入る時点が今からどれくらい先なのかによって、その報酬の価値を割り引く傾向にある (Ainslie 2001, ch.3、邦訳、3章；佐伯 2011, 2章)。すなわち、人々は報酬が手に入る時点が今からずいぶん先ならその価値を低く見積もり、逆に今にも手に入りそうになると高く見積もるのだ。そして報酬に対する価値の割り引き方として大別して二つの仕方が想定できる。一つは、価値の双曲的時間割引である。この場合、たとえば報酬が手に入るのが今日から明日へと遅れるときには、価値を 10%割り引くが、同じ 1 日遅れるのでも 50 日後から 51 日後へと遅れるときには、1%しか割り引かない。つまり、報酬の割引率が次第に小さくなっていく。これをグラフに表すと双曲線になるので、双曲割引と言われる (図 8.1 を参照)。これに対して、もう一つは、価値

⁵⁰ この章は Nishitsutsumi (2011) を加筆修正したものである。

の指数的時間割引である。この場合、報酬が手に入る時点が現在から近かろうが遠かろうが割引率は常に一定であり、グラフに表すと指数曲線になる（図 8.2 を参照）。人々は実際のところ、どちらの仕方で割引を行うのだろうか。近年有力だとされているのは双曲的時間割引の方であり、それを支持する行動経済学者らによると、人々は報酬が手に入る時点が遅くなるのに反比例して、すなわち双曲的な時間割引に応じて、報酬の価値を割り引くのである（Laibson 1997; Harris and Laibson 2001）。

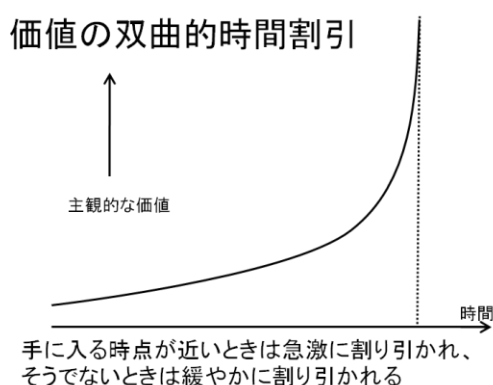


図 8.1

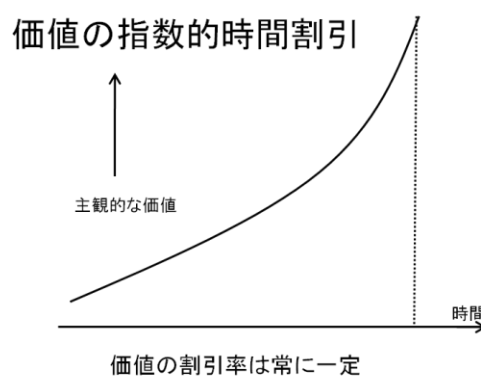


図 8.2

価値の双曲的時間割引と指数的時間割引の決定的な違いは、報酬に対する選好の逆転が生じるかどうかということである。時間的に先に（より早く）手に入る小さな報酬と、それより後で手に入るより大きな報酬がある場合、二つとも現在からかなり先の時点で手に入るなら、現時点では後で手に入る大きな報酬の価値の方が大きく見積もられることになる。従って、後の大きな報酬の方が選好される。だが、双曲的時間割引では、小さな報酬が手に入る直前になると、その報酬の価値が後の大きな報酬の価値を上回ることになるということが起こりうる。そうすると、先の小さな報酬の方が選好されることになる。これが「選好逆転」と呼ばれる現象である。

具体的に見ると、ケーキを食べることは、小さな報酬（つまり食べることの快樂）であるとはいえ、食べるとすぐ報酬が得られるのに対し、ケーキを食べないでダイエットすることは、大きな報酬（つまりダイエットの効果）であるとはいえ、食べずにダイエットしてもすぐ報酬が得られるわけではないので、ケーキを食べることは先に得られる小さな報酬であるのに対し、ダイエットは後で得られる大きな報酬である。この両者について、実際にケーキを食べることが現時点からするとかなり先である場合、ダイエットの方をケーキを食べることよりも大きな価値があると見積もるだろう。従って、ダイエットの方が選好される。だが、ケーキが食べられる時点が近づくにつれ、その価値はダイエットによって得られる報酬の価値よりも高く見積もられることになる。ケーキを目の前にしてしまうと、ダイエットよりもケーキを食べることの方が、価値は大きいと

感じてしまうのである。従って、ケーキを食べることの方が選好されることになる。こうして報酬に対する選好の逆転が生じる（図 8.3 を参照）。

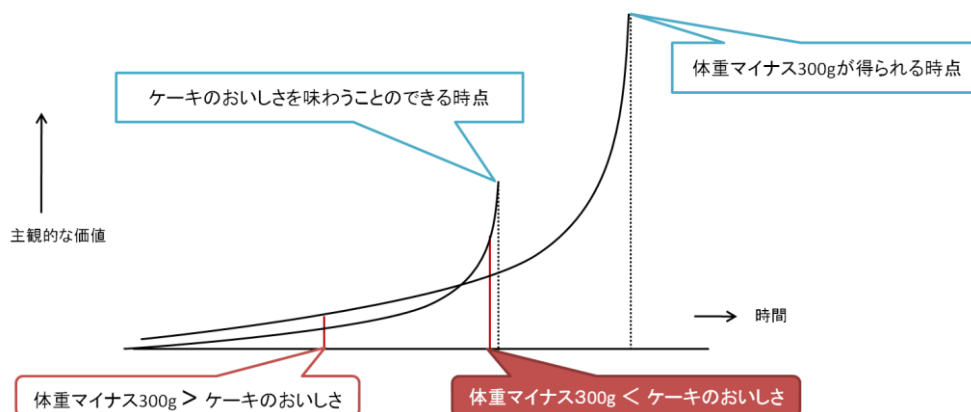


図 8.3

これに対して、このような選好逆転の現象は、指数的時間割引では決して生じない。なぜなら、指数的時間割引の場合、報酬が手に入る時点とは関係なく、価値の割引率が常に一定であるからだ。時間的に先に手に入る小さな報酬も、それより後で手に入る大きな報酬も、どちらも同じ率で価値が割り引かれるから、いったん前者の価値が後者の価値より低く見積もられれば、常にそう見積もられることになり、見積もりが逆になることは原理的に決して生じないのである（図 8.4 を参照）。従って、選好逆転は決して生じない。このように、双曲的時間割引と指数的時間割引は、報酬に対する選好逆転が生じるかどうかという点で決定的に異なる。

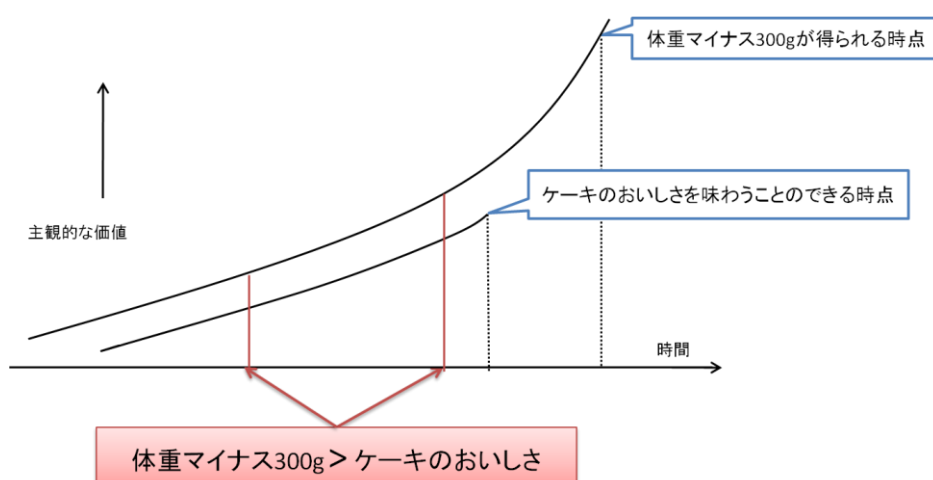


図 8.4

私たちはしばしば、実際に選好逆転を体験する。そのことからすると、私たちの実際の価値の時間割引をよりよく表しているのは、双曲的時間割引の方であると言える。実

際、双曲的時間割引の正しさは多くの実験によって確かめられている。双曲的時間割引は、選好逆転を起こしうるといふ点で不合理な価値評価であり、それに対して指數的時間割引はそのような選好逆転を決して起こさないのが合理的な価値評価である（報酬が得られるのが遅くなるほど、得られる確実さがそれだけ低くなるので、価値の時間割引自体は決して不合理ではない）。しかし、人々の実際の価値の時間割引によりよく当てはまるのは双曲的割引なのである。

但し、情動的評価と知的評価の区別から言えば、価値の双曲的時間割引を起こすのは情動的評価の方だと考えられる。ケーキを食べることとダイエットをすることの価値が逆転して選好逆転が生じるのは、両者を情動的に評価した場合である。知的評価では、そのような逆転は生じず、ケーキを食べられる時点が近づいてきても、相変わらずダイエットの方がケーキを食べることよりも高く評価される。ダイエットをすべきだとあくまでも判断しながら、それでもケーキを食べてしまうということが起こるのは、知的評価では価値の逆転が生じないのに、情動的評価ではその逆転が生じるからだと考えられる。情動的評価が価値の逆転を起こす双曲的割引を示すのに対し、知的評価はそのような逆転を起こさない指數的割引を示す（または、そもそも時間割引を示さない）と考えられる。つまり、知的評価は価値を割り引くにしても、常に一定の率で割り引くのである。

では、このような価値の双曲的時間割引はなぜ生じるのだろうか。Ersner-Hershfield et al. (2009) によると、私たちが将来得られる報酬の価値を不当に割り引いてしまうのは、将来の自分に対してうまくイメージが描けないことに起因する。つまり、将来の自分をまるで他人のように感じてしまい、そのため、たとえば将来の備えとして貯蓄をしなければならないにもかかわらず、それに失敗してしまうのである。Ersner-Hershfield らは fMRI を用いて、現在の自己と将来の自己をそれぞれ考慮している間、どの脳部位が活性化するかを計測した。従来の研究によると、自己を考慮する場合と他者を考慮する場合とでは、自己を考慮する場合の方が内側前頭前皮質（medial prefrontal cortex: MPFC）から吻側前帯状回皮質（rostral anterior cingulate cortex: rACC）におよぶ領域で高い活性化がみられていた。彼らはこの研究を踏まえた上で、現在の自己を考慮する場合と将来の自己を考慮する場合とでは、現在の自己を考慮する方が MPFC と rACC が高く活性化することを見出し、しかもこれらの活性化の違いには個人差があり、活性化の違いが大きい人ほど時間割引も大きいことを明らかにしたのである。つまり、将来の自己をまるで他人のように思う傾向が強いほど、時間割引も大きいというわけである。

ところで、私たちは報酬（プラスの価値）を得る場合だけではなく、損失（マイナスの価値）を被る場合もある。損失を被る場合、その価値は時間によってどのような割引かれ方をするのだろうか。これまでの研究成果によると、価値の時間割引率は一般にその価値がプラスかマイナスかで異なり、マイナスの価値の割引率の方が、プラスの価

値の割引率よりも小さいのである。たとえば、美味しいものを食べることの価値（プラス）は、食べるのが3日後だと20%割引かれるのに対し、掃除をすることの価値（マイナス）は、掃除をするのが3日後でも10%しか割引かれない、といったようになる。言い換えれば、将来の損失（苦）は、将来の報酬（快）ほど割引かれないのである。これは符号効果（sign effect）と呼ばれる（Ikeda et al. 2010）。しかし、このようにプラスの価値とマイナスの価値で割引率が違うにせよ、マイナスの価値もまた、通常、双曲割引を示すことには変わりがない。つまり、今の損失は急激に大きなマイナスに感じられ、将来の損失はそうではないのである⁵¹。

さて、価値の双曲的時間割引は、二重システム理論のもとでの自己制御とどのように関係するであろうか。二重システム理論のもとでは、自己制御が実質的に必要となるのは、システム2の知的評価とシステム1の情動的評価が食い違う場合である。このような食い違いが生じるのは、知的評価が価値の双曲割引という不合理な時間割引を示さないのに対し、情動的評価はそれを示すからだと考えられる。この点を具体例で見てみよう。

たとえば、ミカンよりマンゴーのほうが美味であるなら、マンゴーを食べることの報酬のほうがミカンを食べることの報酬より大きいであろう。しかし、いま、ミカンは2時間後に食べられるが、マンゴーは3時間後にしか食べられないとする。つまり、ミカンは先の小さな報酬であり、マンゴーは後の大きな報酬である。このとき、まずシステム2では、ミカンを食べることの価値とマンゴーを食べることの価値が知的に評価される。この知的評価は一般に物事の客観的な価値をほぼ正しく評価するので、従って、物事の価値を時間的に割引かないか、あるいは割引くとしても、それが手に入る確からしさの低下に応じて割引く（指数割引）だけである。つまり、双曲割引のように、今すぐ手に入る場合は過大に評価し、将来手に入る場合は過小に評価するということはない。それゆえ、ミカンもマンゴーも、今すぐ食べられる場合と、後で食べられる場合とでほとんど評価は変わらないか、指数的に割引かれる。このように、システム2の知的評価においては、評価の逆転が起こらない。今から2時間後に食べてよいミカンと、3時間後に食べてよいマンゴーを知的に評価したとき、3時間後のマンゴーの方が高く評価されるが、2時間経ってミカンは直ちに食べられるのに、マンゴーはまだ1時間後だということになっても、ミカンの知的評価とマンゴーの知的評価は2時間前と変わらないか、指数的に割引かれるから、依然としてミカンよりマンゴーの方が高く評価されることになる。

これに対してシステム1においては、物事の価値は情動的に評価される。情動的評価

⁵¹ この符号効果は池田らによると、痩せている傾向の人には見られるが、太っている傾向の人には見られないようである（Ikeda et al. 2010）。つまり、太っている人は、将来の損失を将来の報酬と同じくらい割引くか、もしくはそれどころか将来の損失の方を将来の報酬よりも大きく割引くのである。池田らは、肥満がこのような個人的な時間割引率の違いによって引き起こされると結論付けている。

(欲求ないし選好)では、一般に物事が今すぐ手に入る場合は過大に評価され、将来手に入る場合は過小に評価される。つまり、価値が双曲的に割り引かれる。情動的评价が双曲的割引を示すので、先の選好逆転の説明で見たように、今から2時間後のミカンと3時間後のマンゴーでは、3時間後のマンゴーのほうが情動的に高く評価されても、2時間経って今のミカンと1時間後のマンゴーということになると、今のミカンの方が情動的に高く評価されるようになるのである。

システム2での知的評価とシステム1での情動的评价が食い違うことがあるのは、情動的评价のほうが双曲的割引を示すことによる。システム2では、2時間経って、ミカンがすぐ食べられるようになっても、依然としてマンゴーの方が知的に高く評価されるが、システム1では、双曲的割引の為に、ミカンの方が情動的に高く評価される。こうしてシステム2は一時間後のマンゴーを選択するのに、システム1は今すぐのミカンを選択する。このため、システム2がシステム1を実質的に制御する必要が生じてくるのである。

このように、二重システム理論においてシステム2とシステム1の行為決定が食い違い、それゆえ実質的にシステム2による自己制御が必要となるような事態が生じるのは、システム2における知的評価が双曲的な時間割引を示さないのに対し、システム1における情動的评价は双曲的な時間割引を示すからだと考えられる。

さてこのように、システム2がシステム1を実質的に制御するような仕方で自己制御が必要となるときに、それがうまく行えると、意志が強いと言われる。逆に、それがうまく行えないと、意志が弱いと言われる。しかし、このように強いとか、弱いとか言われる意志とは、いったい何なのであろうか。次に、意志とは何かという問題を、意志についての二つの有力なモデルを比較検討しながら、考察していきたい。

8.2 意志のモデル

意志に関しては、これまで様々なモデルが提示されてきたが、その中でも代表的なのは以下の四つである (Ainslie 2001, p. 118、邦訳, p.178)。まず哲学者 G. Ryle や従来の効用理論家 G. Becker などによって提示された「虚無理論 (Null Theory)」である。この理論によると、実は意志などというものは存在せず、すべては効用によって決まる。つまり意志の理論でありながら、意志の存在を否定する理論である。第二に、認知心理学者 R. F. Baumeister や M. T. Gailliot などによって提案された「器官理論 (Organ Theory)」がある (この理論では、意志力の発揮に関わる脳の領域を一つの器官とみなしていると考えられる)。この理論によれば、意志は行為選択において情動や欲求とは異なる独自の力を及ぼす一つの器官である。第三の「確固たる選択理論 (Resolute Choice Theory)」は、哲学者 E. F. McClennen や M. E. Bratman によるものであり、意志とは一度決めたことについての再考を合理的に避ける能力であるとされる。最後に、第四の「パターン探求理論 (Pattern-Seeking Theory)」は行動心理学者 H. Rachlin に

よるものであり、意志とは内在的に動機付ける力をもつ規則的パターンを把握する能力であるとされる。これらの伝統的な意志のモデルは、意志が通常持つだろうとされる諸性質の説明を試みた理論である。

Ainslie はこのような伝統的な意志のモデルを意志の諸性質を正確に説明できないということから批判する (Ainslie 2001, pp. 120-125, 邦訳, pp. 182-88)。そして彼はこれらに代わる新たな意志のモデルとして「異時点間交渉モデル (Intertemporal Bargaining Theory)」を提案する。それによると、意志とは異時点間の交渉状態に他ならない。彼は伝統的な意志のモデルとこの異時点間交渉モデルを比較し、意志を説明するモデルとして彼のモデルの方がより尤もらしいと結論づける。なぜなら、異時点間交渉モデルが、意志に関する一切の諸性質を説明することができるからである。Ainslie は伝統的な意志のモデルの中で、とりわけ器官理論に対して批判的である。なぜなら器官理論は、意志の弱さが限定的ではないと主張するからである (Ainslie 2001, p. 120, 邦訳, pp. 82-3)。そこで、伝統的なモデルの中でも器官理論に焦点を絞り、この理論と Ainslie の異時点間交渉モデルの比較を試みることにしよう。そのために以下では、まず器官理論を検討し、それに続いて器官理論に対する Ainslie の批判を取り上げて考察してみよう。

8.2.1 意志の弱さの非限定性：器官理論

まず、意志の弱さの非限定性を唱える器官理論について考察してみよう。器官理論によると、意志は意志力を発揮する器官である。また意志は、忍耐力や決意の持続力のような認知資源を備え、それを使用することでまさに意志力を行使する。もし認知資源がある課題で消耗すると、それ以外の課題でも意志の弱さが見られるようになる。なぜならもう認知資源が残っていないからである。このことから伺えるように、器官理論によれば、意志の弱さは特定の領域に限定されないのである。

このような意志の弱さの非限定性は、自己制御に関する認知心理学研究において繰り返し実証されてきた。意志はまさに自己制御においてその力を行使する。自己制御に関する実験事例によると、自己制御の喪失は認知資源の消耗によって引き起こされる。もしある領域の課題で自己制御の力を行使して認知資源を使い果たすと、別の領域の課題で自己制御があまりできなくなる。つまり、ある領域で意志の弱さが生じると、別の領域でも意志の弱さが生じるということである。これが心理学における多くの研究において実証されてきたことである。たとえば Gailliot et al. (2007) によると、被験者は畜殺のビデオを二分間か、あるいはコメディのビデオを二分間、情動的な感情や反応を抑制しながら見るように求められる。そのあとで、被験者は図形トレース課題を遂行する。この課題は、与えられた図形を一筆書きでなぞるもので、被験者は解けたか、あるいはあきらめたかすれば、それを実験者に伝える。合計 5 つの図形を 20 分以内に解かなければならないが、一番目の図形は解決不可能な形となっている。実験の間、実験者は、ビ

ビデオ課題の最初と最後に被験者の血中グルコース値を測定した。それらのグルコース値の差でもって、実験者らは被験者らがどのくらい認知資源を第一課題のビデオ課題で消耗したのかを確認する。そして実験者らは、ビデオ課題によって認知資源を消耗した被験者らと、そうではない被験者らを比較検討した。その結果、第一課題で畜殺のビデオを見た被験者の方が、続く一筆書き課題において、第一課題でコメディのビデオを見た被験者に比べ、早くあきらめるということを見出したのである（この種の実験については、次章で「リソースモデル」を検討する際にもう少し詳しく見ていく）。

この実験が実証したように、自己制御をすると血中グルコース値が引き下げられる。この自己制御課題後の低グルコース値は、それに続く異なるタイプの自己制御課題において、パフォーマンスの低下を引き起こす。このことが示唆しているのは、自己制御に必要な認知資源は、ある一定量のグルコースに依存しているということである。グルコースは私たちにとっては限りあるエネルギー資源である。グルコースがほとんど消費されてしまったということは、認知資源がほとんど消費してしまったということである。それゆえ、もし多くのグルコースが一つの課題で消費されたら、続く別の課題で必要な自己制御がうまくできないことになる。このように自己制御の喪失は、一つの課題でのみ生じるというわけではない。それは別の種類の課題へと広がっていくのである。これは、意志の弱さは一つの領域にとどまる限定的なものではないということであり、他の領域へと広がっていくものであるということを示している。

8.2.2 Ainslie の器官理論批判

Ainslie は伝統的な意志のモデルの中で、とりわけ器官理論に対して批判的である。なぜなら器官理論は、意志の弱さが限定的ではないと主張するからである（Ainslie 2001, p. 120、邦訳, pp. 82-3）。異時点間交渉モデルによると、ある領域での意志の弱さが、別の領域の意志の弱さをもたらすとは限らない。ある領域での意志の弱さは、ドミノ倒しのように次々に他の領域に広まる場合もあるが、そうではなくその部分だけにとどまる場合もある。器官理論を除く一切のモデルは、意志の弱さが限定的かどうかについて、特に何の想定もしていない。しかし、8.2.1 で見たように、器官理論だけは、意志の弱さが非限定的であると想定している。器官理論によると、意志は単一の認知資源を持ち、従って、もしその認知資源がある領域において枯渇すると、その認知資源はどの領域でも枯渇することになると想定している。つまり、ある領域で意志の弱さが引き起こされると、他の領域でも意志の弱さがみられると言うのである。これに対して Ainslie は意志の弱さは限定的であると主張し、そのような事実を説明できるのは、異時点間交渉モデルのみであると主張するのである。

8.2.3 異時点間交渉モデルによる意志の弱さの限定性

Ainslie の異時点間交渉モデルはどのようにして意志の弱さの限定性を説明できるの

だろうか。この点を理解するために、まず、異時点間交渉モデルがどのようなものかを概観し、その上でそのモデルがどのように意志の弱さの限定性を説明できるかを見ていこう。

8.2.3.1 異時点間交渉モデルとは何か

私たちは日々の生活の中で、時間的に先に（より早く）手に入る小さな報酬とそれより後で手に入るより大きな報酬のどちらを選ぶかという状況にしばしば直面する。Ainslieによると、私たちの意志が機能するのはそのような状況においてである (Ainslie 2001, ch.5、邦訳、5章)。そのような状況は、一回きりの場合もあるし、同じような状況が何度も繰り返される場合もある。

一回きりの場合を考えてみよう。価値の双曲的時間割引によると、そのような場合には、私たちは先の小さな報酬が手に入る直前になると、それより後で手に入るより大きな報酬よりも先の小さな報酬の方が価値が大きいと感じるようになる。それゆえ先の小さな報酬の方を選んでしまうということになる。Ainslie は双曲的に割引かれるような価値の見積もり以外に選択を左右する力を認めないので、選択はそのような見積もりによって完全に決まると考えているのである。

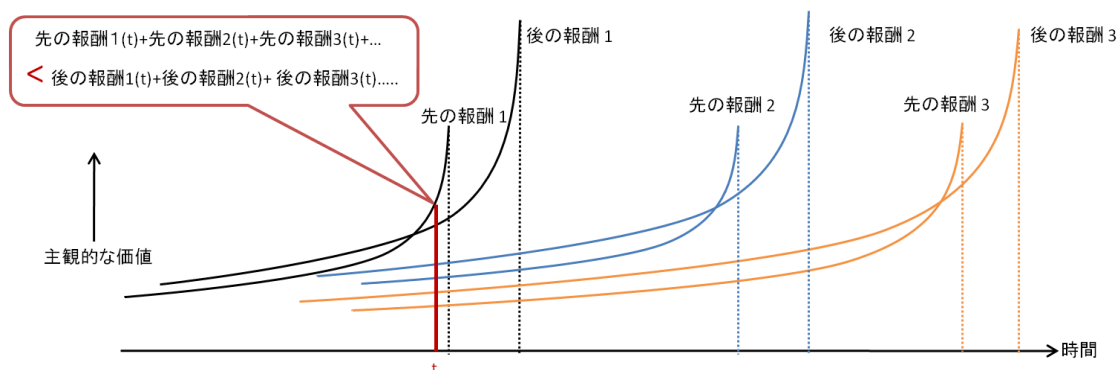


図 8.5

一方、そのような選択を行う状況が繰り返し生じる場合はどうだろうか。この場合、どの機会にも一貫して後で得られる報酬を選ぶことにすれば、それは一貫して先の小さな報酬を選ぶことよりも全体としての価値が大きく見積もられる。というのも、目の前の機会ではなく、後の機会については、後の大きな報酬が先の小さな報酬よりも大きく見積もられ、それらが足し合わされることで、目前での小さな報酬に対する過大な見積もりに打ち勝つことができるからである (図 8.5 を参照)。従って、後で手に入るより大きな報酬を確実に手に入れるためには、一貫して後で手に入る報酬を選ぶのだという個人的規則を立てて、それをしっかり守ればよい。但し、時々規則違反は、厳格に規則

に従うよりもしばしば利益が大きくなる。なぜなら、ある機会ではすぐに手に入る報酬を得ることができ、別の機会にはそれより後で手に入る報酬が得られるというように両方の報酬を得ることができるからである。だがこの場合、私たちは、規則を破ることについて、未来の自分と交渉することになるのである。未来のそれぞれの時点の自分は、まさに自分の選択の機会において規則を破りたがるが、そういうわけにはいかない。なぜなら、どの時点の自分も規則を破ると、結局、先に手に入る小さな利益しか得られないことになるからである。後の大きな報酬を手に入れるためには、現在と未来の自己の内、ごく一部の自己しか規則を破ることができない。そうだとすれば、どの時点の自己が規則破りをしていいのかについて、現在の自己と未来の自己の間で交渉状況が生じる。Ainslieによると、意志とはこのような交渉状況に他ならないのである。この交渉によって、多くの時点の自己が規則を守り、それを破る自己があっても、それはごく少数だということになれば、意志が強いということになるし、そうでなければ、意志が弱いということになる。彼は意志をこのような交渉状態として捉えて、自身の意志の理論を「異時点間交渉モデル」と呼んでいる。

以上の点を具体的な例に即して見ておこう。もし私たちが、先に手に入る小さな報酬とそれより後で手に入るより大きな報酬の間で一度きりの選択をするなら、先の小さな報酬が今にも手に入りそうな時点になると、先の報酬の方を高く評価するようになる。たとえば、甘いものを食べることは先に得られる小さな報酬であり、甘いものを控えてダイエットすることは後に得られる大きな報酬である。一度きりの選択では、甘いものがすぐ食べられる時点になると、私たちはダイエットよりも甘いものを食べる方を選びがちである。一方、そのような選択機会が繰り返し生じる場合はどうだろうか。その場合、意志は今後の同じような選択の機会を束ねて、以下のような二つの規則を立てるだろう。一方の規則は、一貫して先の小さな報酬を選ぶという規則であり、もう一方の規則は、一貫して後の大きな報酬を選ぶという規則である。先の事例でいうと、一方は一貫して甘いものを食べることを選ぶという規則であり、もう一方は一貫してダイエットすることを選ぶという規則である。この場合、今すぐ甘いものを食べるのとダイエットするのとを比べれば、甘いものを食べる方が価値が大きく見積もられても、今後も一貫して甘いものを食べるという規則に従うことと、今後も一貫してダイエットするという規則に従うこととを比べれば、後者の方がその時々々の価値の現時点の見積もりを足し合わせた総和が大きくなる。従って、一貫してダイエットするという規則が選択されることになる。

ところが Ainslie は、より大きな報酬を得るために規則を立ててそれを守ることを意志の働きの一つの側面として認めつつも、一貫して規則に従うよりも時々規則を破る方が結果的に価値の総和は大きくなると指摘する (Ainslie 2001, ch.6、邦訳、6章)。たとえば、一貫してダイエットするのではなく、概ねダイエットしつつ、たまに甘いものを食べるようにすれば、ダイエットによる報酬に加えて、甘いものによる報酬も得られる

ことになる。なぜなら、概ねダイエットをしさえすれば、一貫してダイエットするのと同じだけの報酬が得られるからである。もし私たちが価値の総和を最大にしたいなら、一部の例外を設け、それ以外は規則に従うというやり方が最も望ましい方法なのである。そうだとすると、私たちは規則を設けても、例外なくそれに従うのではなく、むしろ時々それを破りつつ規則に従うことを選択するようになる。

しかし、規則の中で例外を設けることは、報酬の価値の総和を最大化させるためには必要なことかもしれないが、それにより規則が容易に破綻する恐れがある。たとえば、ダイエット中で一切甘いものを食べないという規則に従っているとしよう。その規則の中で、疲れていたらケーキを食べてもよいという例外を設けたとする⁵²。そうすると、友人と一緒にケーキを食べていい、よく働いた後ならケーキを食べてもいい...といった具合に、ある例外が呼び水となって他の例外が容易に作られやすくなる。一つの例外から他の例外が容易に引き起こされるなら、おそらく私たちはいつまでたっても先に手に入る小さな報酬の方ばかり選び、それしか得ることができない。だとすれば、結局、後のより大きな報酬を得る最もよい方法は、規則の中にいかなる例外も認めないことではないだろうか。つまり、人々は厳格にあるいは強迫的に規則に従うべきなのである。ダイエットの事例でいえば、もし一貫してダイエットするという規則に従うと決めたなら、いかなる時も甘いものを食べてはいけないのである。だが、先に指摘したように、このように強迫的に規則に従うことは、必ずしも報酬の価値を最大化させることにはならないのである。なぜなら強迫的な規則遵守は、甘いものを食べるという快楽を決して得られないものにしてしまうからである。では、どうすればよいのだろうか。

Ainslieによると、私たちは「明確な境界線」によって区切られた例外を持つべきなのである。明確な境界線とは、例外と例外でないもの間をはっきりと区別することを可能にするようなものであり、例外が例外をよぶことがないようなものである (Ainslie 2001, pp. 94-100、邦訳, pp. 141-9)。たとえば、家族や親しい友人の誕生日ケーキなら食べてもいいという例外は、時々ならケーキを食べてもいいという例外よりも明確な規定である。この規定なら、ケーキを食べていい機会とそうではない機会をはっきり見分けることができる。このような明確な境界線を引けば、例外が例外をよぶことなく、従って、甘いケーキのような先に手に入る小さな報酬とダイエットのような後で得られるより大きな報酬の両方の報酬を得ることが可能になる。このように、明確な境界線によって区切られた例外を持つ規則は、最大の報酬をもたらすと考えられるのである⁵³。

⁵² このような例外を設けるということは、疲れていたらケーキを食べてもよいが、そうでない限り一切ケーキを食べないという新しい規則に変えたということであるともいえる。このように捉えるなら、本文でこれに続いて述べた、例外が例外をよぶという事態は、アドホックに新たな規則を作ってしまうという事態として捉えられることになる。

⁵³ どんな例外でも明確な境界線によって区切りさえすれば、報酬が最大になるというわけではない。重要なのは、最大の報酬をもたらすような例外を明確な境界線によって区切るということである。

異時点間交渉モデルによると、意志とは異なる時点の自己の間の交渉であり、その交渉のあり方によって意志が強いということになったり、あるいは意志が弱いということになったりする。意志が弱いというのは、規則を立てても、ある時点の自己が規則を破れば、他の時点の自己もみな規則を破ってしまい、規則の崩壊を招くような場合である。対照的に、最も意志が強いというのは、いくつかの時点の自己が規則を破りつつも、それ以外の時点の自己はすべてきちんと規則に従うような場合である。その次に意志が強いのは、どの時点の自己も必ず規則に従うような場合である⁵⁴。

8.2.3.2 意志の弱さの限定性

Ainslie によると、明確な境界線によって区別された例外領域は、「墮落領域 (lapse district)」と呼ばれる (Ainslie 2001, p. 149, 邦訳, p. 222)。墮落領域における意志の弱さは限定的であり他の領域に及ばないと言う。具体的な事例として、Ainslie はヴィクトリア朝の都市における悪徳地帯を挙げている。そこでは抑圧できない悪徳を容認していたが、その悪徳は決して他の領域へと波及することはなかった。悪徳領域における意志の弱さはその領域にとどまる限定的なものだったのである。

意志は墮落領域においては弱いものである。だが、ある意味では意志はそこで弱くないのだと言うこともできる。なぜなら、人々はそこで自分自身に規則を破らせているとも言えるからである。そもそも人々はそこで最初から規則に従う気はないのである。また、この領域では、規則を破ることが、意志が誘惑に屈するということを意味しない。なぜなら、墮落領域では、意志は誘惑に屈しないようにしようとするのではなく、むしろ積極的に誘惑に従おうとするからである。このような意味で、墮落領域では、意志は決して弱くないとも言えるのである。

8.2.4 異時点間交渉モデルと器官理論の調停

これまで見てきたように、意志の弱さに関する異時点間交渉モデルと器官理論の間には、意志の弱さが限定的かそうでないかで大きな食い違いがある。異時点間交渉モデルにおいては意志の弱さは限定的でありうるのに対し、器官理論においては非限定的である。しかしながら、これは真の不一致なのだろうか、それとも単に表面的なものなのだろうか。これら二つの理論は両立しえないのだろうか、従って、少なくともどちらか一方を否定すべきなのだろうかについて考察してみよう。

ここで注目すべきなのは、異時点間交渉モデルと器官理論が意志の弱さの異なる側面に焦点を当てているのではないかという可能性である。器官理論は比較的短い時間幅の中で引き起こされる意志の弱さを問題にしている。もし何らかの課題で認知資源の消耗

⁵⁴ 異時点間交渉モデルは、二重システム理論の枠組みでは、それぞれの機会に規則を破ろうとするのはシステム1であり、明確な境界線で区切られた例外付の規則を立てそれを遵守しようとするのがシステム2であると解釈できよう。

のゆえに意志の弱さが生じるなら、次に続く別の課題でも意志の弱さが生じる。ここで重要なことは、第一の課題のすぐ後、まだ認知資源が回復しない内に、二番目の課題が実行されるということである。これはほぼ同じ時点に実行されると言っても差し支えないだろう。器官理論における意志の弱さの非限定性が意味しているのは、ある課題における意志の弱さはほぼ同じ時点の別の課題へと広がるということである。

一方、異時点間交渉モデルが注目しているのは、繰り返し生じる異なる時点においての意志の弱さである。もし意志の弱さがある時点の選択機会において生じた場合、他の未来の時点の同じような選択機会においても引き起こされるかもしれないが、そうだとすると、それはいくつかの時点に限られており、すべての時点で意志の弱さが生じるとは限らない。異時点間交渉モデルが想定している意志の弱さの限定性は、このような異なる時点での選択に関わるものであり、器官理論が想定しているほぼ同じ時点での異なる課題ないし選択に関わるものではない。つまり、異時点間交渉モデルが問題にしている時間幅は器官モデルが問題にしているそれよりも相当長いものである。しかも、異時点間交渉モデルが想定している意志の弱さの限定性は器官モデルの観点からも理解可能である。すなわち、たとえ意志に関わる認知資源がある時点で消耗されても、それは次の時点では回復する。従って、ある時点での意志の弱さは必ずしも別の時点の意志の弱さを生まないのである。

器官理論がほぼ同じ時点での選択について意志の弱さを問題にしているのに対し、異時点間交渉モデルは繰り返し生じる異なる時点（器官理論が想定しているよりもはるかに時間の幅が大きいある一定の期間）での選択について意志の弱さを問題にしている。もしそうであるなら、器官理論と異時点間交渉モデルは、意志の弱さの異なる側面を問題にしていることになる。しかも、異時点間交渉モデルにおける意志の弱さの限定性は器官モデルの観点からも理解可能である。それゆえ、器官理論における意志の弱さの非限定性と、異時点間交渉モデルにおける意志の弱さの限定性は矛盾しないことになる。つまり、これら二つの理論は両立しうるのであり、どちらから一方を否定する必要はないのである。

そうだとすれば、結局、意志とは異時点間交渉モデルが言うように、異時点間の自己の交渉という側面と、器官理論が言うように、情動や欲求に対抗しうる独自の力を発揮する器官であるという側面を併せ持つということになる。このような二つの側面をいかに兼備しうるかという問題は、さらなる考察を要するが、ここでは異時点間交渉モデルの捉えた側面と器官モデルが捉えた側面は両立しうるということを指摘することにとりあえず満足することにしたいと思う。

第九章 意志力の枯渇⁵⁵

前章では、意志についての二つのモデルを考察したが、本章では、自己制御を行うために意志が発揮する力、すなわち意志力とは何であり、それがどのようにして枯渇して自己制御の喪失（意志の弱さ）に至るのかを考察しよう。従来、自己制御は意志の力によって為され、情動はむしろ自己制御を妨げる要因として捉えられてきたが、その描像が正しくないことはすでに論じてきた。情動は自己制御を支える要因でもありうるのである。しかし、ここでいったん情動を離れ、意志の力による自己制御のあり方を論じることにしよう。

ここで扱う自己制御は、システム1とシステム2の行為決定が一致する場合の自己制御（実質的な制御なしにおのずとなされる自己制御）ではなく、それらの決定が食い違い、それゆえ実質的な制御が必要な自己制御（誘惑との葛藤がある自己制御）である。自己制御の喪失（意志力の枯渇）については、それが「自己資源」の消耗によって起こるとする「リソースモデル」が有力である。すなわち、自己制御を可能にする認知的な資源である自己資源というものがあり、それが消耗すると、自己制御が失われるというわけである。このリソースモデルは前章で検討した意志の器官理論に立脚したモデルだと言える。すなわち、意志は自己資源を備えた一つの器官であり、その資源を用いて自己制御を行うが、そのため資源が消耗すると、自己制御ができなくなるというわけである。しかし、このリソースモデルに対しては、自己資源というような何か実体的な資源があるわけではなく、ある種の認知的な過程によって自己制御の喪失が起こるとする「プロセスモデル」が反論として唱えられている。本章では、これらのモデルを検討しながら、自己制御の喪失について考察していく。

本論文では、基本的に意志に関するリソースモデルを受け入れることにするが、それによれば、意志の力は限りある資源であることになる。そしてもしそれが正しいのであれば、ここに自己制御に対して情動が持つもう一つの役割が明らかになるだろう。すなわち、限りある資源としての意志力を節約し補うという役割である。常に意志力を発動して自己制御を行うことは、人間には不可能である。それゆえ、日常の多くの場面で自己制御はむしろ情動によってもたらされることになるのである。

では、リソースモデルとプロセスモデルの検討に入ろう。しかし、その前に、まず、自己制御およびその喪失に関与する脳部位が確かにあるのだということを明らかにしようとした Hare たちの興味深い実験があるので、それを参照しておこう。これはまさに意志力と呼ばれるものの神経基盤の一端をそれなりに明らかにした非常に重要な成果であり、意志力が自己制御とその喪失という現象を説明するための単なる便利な道具的存在ではなく、脳の神経活動に基盤をもつ実在的なものであることを示した点で参照しておくに値する。

⁵⁵ この章は西堤（2014）第四節に加筆修正を加えたものである。

9.1 自己制御（意志力の行使）の神経基盤

Hare らは、ダイエットを行っていると自ら宣言している人々37名を被験者とし、fMRIを用いて彼らが食品選択を行っている最中の脳の状態を調べることにより、自己制御を司る脳部位を見出そうとした (Hare et al. 2009)。Hare らは被験者に fMRI の中で以下の三つの試行を順に行ってもらった。まず、50個の食品についてそれぞれ「健康に良いと感じられるかどうか」を五段階で被験者に評価してもらった。「とても良い、良い、どちらでもない、悪い、とても悪い」の五段階である。被験者は、目の前のスクリーンに最大4秒間提示される一つの食品に対して、その画面が消えた直後0.5秒以内に五段階で評価を下すことを求められる。ある食品の評価が済んで、次の食品が提示される間に4秒から15秒間の固視点注視の時間（脳を平静な状態に戻すための時間）が設けられており、その後、次の食品に対しても同様の提示と評価が行われることになる。たとえば、スクリーンにリンゴが提示された直後、五段階評価を下し、固視点をしばらく見つめた後、チョコバーが提示され、それについてまた評価を下すといった具合である。被験者はこのようにして、50個の食品すべてに対して五段階の評価を行った。

次の試行では、同じ50個の食品について「美味しいと感じられるかどうか」を同様の五段階で評価することが求められた。この場合も先ほどと同じように、まず食品が提示され、次に評価がなされ、固視点注視の時間をはさんで次の食品が提示されるという手順が繰り返された。すると、それぞれの被験者について、健康と美味しさの両方の面で「どちらでもない」と評価される食品が出てくるが、このような食品のうちの一つをその被験者にとっての「準拠食品」とした。この準拠食品は、次の第三の試行で、ある食品を食べるかどうかを決定する際の比較基準となるものである。

第三試行では、被験者は、それぞれの食品を準拠食品と比較して、準拠食品よりもそれの方を選んで食べるかどうかについて、五段階（「絶対に選ぶ」「選ぶ」「どちらともいえない」「選ばない」「絶対に選ばない」）で判断を求められた。この試行もこれまでの試行と同様の提示と評価、固視点注視の時間が順に設けられた。被験者は絶対に選ぶとした食品の中から、ランダムに選ばれた一品を実験終了後に実際に食べなければならず、そのことをあらかじめ告げられていた。これは被験者の選択を本気の選択（つまり現実的な意思決定）にするための措置である。

第三試行での被験者の選択から、被験者は二つの群、すなわち主に健康に基づいて意思決定を行った被験者らの群（19名）と、主に美味しさに基づいて意思決定を行った被験者らの群（18名）に分かれた。すなわち、前者は、健康に良いけれども美味しくなくない食品を選び、また美味しいが健康に良くない食品を選ばない傾向を示した。これは自己制御が行われた結果だと考えられる。それに対して後者は、逆に、美味しいが健康に良くない食品を選び、健康に良いが美味しくなくない食品を選ばない傾向を示した。これは自己制御が行われなかった結果だと考えられる。それゆえ、前者は自制心のある群とされ、後者は自制心のない群とされた。

この実験の結果明らかになったのは、第一に、どちらの群においても、それぞれの食品についてそれをどれくらいの強さで選ぶのかの判断と、VMPFC (ventromedial prefrontal cortex : 前頭前野腹内側部) の活動強度が相関したことである。すなわち、強く選ぶ場合ほど、VMPFC は大きく活動したのである。また、第二の重要な結果として、自制心あり群の被験者が健康に基づいて食品を選んだとき、VMPFC の活動に加えて、DLPFC (dorsolateral prefrontal cortex : 前頭前野背外側部) も活動していることが確認された。

Hare らはこれらの結果から、自制心あり群の被験者が健康に基づいて食品を選ぶことができたのは、VMPFC の活動を DLPFC が調節していたからだと考えた。たとえば、健康に関して「とても悪い」で、美味しさに関して「とても美味しい」と評価された食品の場合、自制心なし群の被験者であれば、美味しさにより VMPFC が活動するだけで、その活動が健康を考慮した DLPFC の活動によって抑制されないためにその食品を選ぶであろうが、自制心あり群の被験者であれば、いくら美味しさによって VMPFC が活動したとしても DLPFC がそれを抑制するので、その食品を選ぶことはないのである。逆に、健康に関して「とても良い」で、美味しさに関して「とても美味しくない」と評価された食品の場合、自制心なし群では、DLPFC が VMPFC の活動を高めないために、その食品を選ばないが、自制心あり群では、DLPFC が VMPFC の活動を高めるために、その食品を選ぶのである。このように DLPFC による VMPFC の調節が自己制御にほかならないと解釈されるのである。

この実験は、私たちが実質的に自己制御を行っているとき、すなわちシステム2がそれと食い違う行為決定を行うシステム1を意志力によって制御しようとするとき、DLPFC が重要な役割を果たしていることを示している。確かに、この実験から意志力とは DLPFC の活動に他ならないと結論付けることは到底できないだろうし、何らかの確固たる結論を導くにはもっと多くの神経科学的な探究が必要であろうが、少なくとも意志力の何らかの神経基盤が確かにありそうだということは言えるだろう。

9.2 リソースモデル

意志力については、心理学においても以前から様々な探究がなされてきた。それによって明らかになった一つの重要な事実、意志力を行使してある誘惑に耐えたとき、そのすぐ後で別の誘惑にさらされると、その誘惑に負けやすくなるということである。このような事実から、意志力の行使に必要なある限られた認知資源があり、その資源が消耗することによって自己制御の喪失が起こるという考えが提唱された。この考えは「リソースモデル (Resource Model)」と呼ばれ、それによると、意志力の行使に必要な資源は、どんな意志力の行使にも共通に用いられる資源であり、誘惑に抗して行われるいかなる自己制御にもこの同じ資源が用いられる。この共通の資源は「自己資源 (self-resource)」と呼ばれ、それが何らかの原因によって消耗すると、自己制御の喪

失が起こる。この自己資源を消耗した状態は、「自己消耗 (ego-depletion)」と呼ばれる (Baumeister et al. 1998; Inzlicht and Schmeichel 2012)。

自己制御とその喪失に関するこのリソースモデルを支持する主要な証拠は、前章でも簡単に触れたように、時間的に連続して行われる二つの課題を用いた実験によって生み出されてきた。この実験パラダイムでは、被験者はまず自己制御を要するような課題を最初に行い、それに続いてその課題とは関連がないがやはり自己制御を要するような第二の課題を行う。そして第二の課題において自己制御の低下が起こるかどうかを見るのである。この実験パラダイムを用いて実験を行い、自己制御がある限られた共通の資源に基づくことを最初に主張したのは、Baumeister たちである。彼らの実験を概観しておこう。

Baumeister らは、心理学を専攻する学生を被験者として実験を行った (Baumeister et al. 1998)。被験者はまず、第一の課題において、チョコクッキーを目の前にしてそれを食べたいという誘惑に晒されながら、生のラディッシュを食べ続けることを求められ、続く第二の課題では、実は解決不能のパズルをできるだけ早く解くように求められた。ただし、比較対象として、一部の被験者は第一の課題において、チョコクッキーを我慢して生のラディッシュを食べる必要はなく、ただチョコクッキーを食べただけ食べれば良かった。この実験の結果、第一の課題で、チョコクッキーの誘惑に晒されながら生のラディッシュを食べるように強いられた被験者らは、そうでなかった被験者らに比べ、第二の課題の解答不能のパズルを早く投げ出してしまうことが見出された。

この実験の結果は、Baumeister らにより、次のようにリソースモデルを支持するものとして解釈された。チョコクッキーの誘惑に晒されながらラディッシュを食べなければならなかった被験者らは、意志力の行使によって限られた自己資源を消耗したため、第二課題においてその資源があまり残っておらず、そのためパズル解きを早々に断念してしまった。一方、ただチョコクッキーを食べればよいだけの被験者らは誘惑に耐える必要がないので、自己資源を消耗させず、それゆえ第二の課題において、十分にある自己資源を用いて粘り強くパズル解きを続けることができた。つまり、第一の課題で自己資源の消耗が生じると、第二の課題で自己制御の喪失が起こり、そうでなければ、第二課題で十分な自己制御ができるのである。

二重システム理論からは、このリソースモデルはどのように位置づけられるだろうか。まず、自己資源はシステム2の資源であると考えられよう。システム2とシステム1の行為決定が食い違うとき、システム2はシステム1を制御して自身の決定した行為を実行しようとするが、その際システム2はそれの資源である自己資源を用いてシステム1を制御すると考えられる。そのようなシステム2の自己資源が消耗すると、システム2はシステム1を制御できなくなり、システム1へと主導権の移行が生じる。こうして、システム1が決定した行為がなされることになる。これが自己制御の喪失に他ならない。第二課題でパズル解きが早々に断念されたのは、第一課題でシステム2の自己資源が消

耗したために、第二課題でパズル解きを行おうとするシステム2がそれよりもっと楽なことをしようとするシステム1を制御できなくなったことによると考えられる。

さて、意志力の行使に必要な自己資源とは、具体的にはどのようなものなのであろうか。前章でも触れたように、自己資源を脳内で利用されるグルコースに求める考えがある。GailliotとBaumeisterたちによると、私たちの脳において利用可能なグルコースが減少すると、自己資源の消耗が起こる (Gailliot et al. 2007)。たとえば、糖尿病や低血糖症の人、あるいはダイエット中の人、利用可能なグルコースが減少しており、誘惑に負けやすい傾向がある。このような事実から、グルコースこそが自己資源だという考えが生まれてくる。しかし、このような考えは恐らく単純すぎるだろう。自己資源はグルコースだけではなく、もっと様々なものからなる複合的なものであろう。とは言え、グルコースが自己資源の重要な要素であることは恐らく間違いあるまい。

意志力の行使を自己資源の活用とみるリソースモデルにおいて一つ特徴的なことは、システム2の意志力をシステム1の欲求ないし情動とは別種の力だとみなしている点である。システム1は欲求ないし情動を動機付けの力として、それが決定した行為を遂行しようとするが、システム2はシステム1と食い違う行為決定をしたとき、意志力によってシステム2の動機付けの力と対抗しようとする。この意志力はシステム2の自己資源に支えられたものであって、欲求ないし情動によるものではない。このように、リソースモデルは、欲求ないし情動による動機付けの力には還元できないような意志の力を想定する点に一つ注目すべき特徴がある。

リソースモデルはBaumeisterらの実験結果をよく説明できる点でなかなか魅力的であるが、すでにいくつかの問題点が指摘されている。これらはいずれも、自己資源というような何か実体的なもの（蓄えられたエネルギーのようなもの）があるというリソースモデルの根幹となる考えに疑問を投げかけるような問題である。まず、Muravenらの実験によると、第二課題において自己制御を行うように何か動機を与えると、第二課題の成績が良くなる (Muraven and Slessareva 2003)。たとえば、パズルが解ければ、大きな報酬を与えることにすると、パズルを解いている時間が長くなる。これは、自己制御の喪失が自己資源の消耗ではなく、自己制御への動機の欠如によるのではないかという疑問を提起する。つまり、大きな報酬が与えられれば、自己制御への動機が高まって、パズルを解いている時間が長くなるが、それが与えられないと、動機が低いままなので、時間が短くなるというわけである。

また、Jobらの実験は、意志力（自己制御を行使する能力）が限られた資源であると思っている人ほど、第二課題での成績が悪いことを明らかにした (Job et al. 2010)。すなわち、Jobらは被験者に六つの項目、たとえば、「難しい精神的な作業をした後、あなたのエネルギーは消耗し、それを取り戻すために休まなければならない。」とか「精神的な活力は衰えない。難しい精神的な作業をした後でさえ、さらにもっとやり続けることができる。」といった項目を提示して、それぞれどの程度そう思うかを六段階で評

価値してもらい、この評価に基づいて、被験者がそれぞれ、意志力がどの程度限られた資源であると思っているのかを決定した。これに続いて、彼らは被験者に意志力の消耗を調べるための二つの課題を行わせた。すなわち、まず意志力を要するような第一課題を行わせ、さらにそれとは別種のやはり意志力を要する第二の課題を行わせた。この第二の課題の成績が、意志力を限られた資源だと思っている人ほど悪かったのである。

これらを通して得られた実験結果は、意志力が実際に限られた資源によるのではないということ、従ってそのような限られた資源が実際にあるわけではないということを示しているように見える。なぜなら、意志力が実際に限られた資源であるとする、その資源がどれだけ残っているかで第二課題の成績がほぼ決まるはずで、限られた資源だと思うかどうかで第二課題の成績がそれほど大きく左右されることはないはずだからである。

さらに、Tice たちの実験では、第一課題を行った後、被験者に面白いビデオを見せるなどしてポジティブな気分させると、第二課題での成績が改善されることが明らかにされた (Tice et al. 2007)。この結果は、自己制御が気分によって左右されることを示しており、自己資源という実体的なものがあるという考えと折り合いが悪いように見える。なぜなら、そのようなものがあるとする、面白いビデオを見たりすることが直ちに自己資源を回復させるということになるが、なぜ直ちに回復させることができるのか疑問だからである。

リソースモデルには、このようにいくつかの重大な問題点があるように思われる。そこで、リソースモデルに対抗して登場したのが「プロセスモデル (Process Model)」である。次節ではプロセスモデルを見ていこう。

9.3 プロセスモデル

プロセスモデルは、Inzlicht たちによって主張されている。それによると、自己制御の喪失とは、自己資源というその本性が謎のままであるような実体的なものが無くなったということではなく、動機や注意がシフトするといったある一定の過程が生じたということである (Inzlicht and Schmeichel 2012)。たとえば、ある課題で自己制御を行うと、次の第二課題で自己制御を行うことへの動機が減り、衝動的な欲求を満たすことへの動機が増える。つまり、自己制御から欲求充足へ動機のシフトが起こる。また、それと並行して、第二課題において、自己制御の必要性に対する注意が低下し、衝動的な報酬に対する注意が増加する。こうして注意もシフトする。このように、動機や注意がシフトすることにより、自己制御の喪失が起こるのである。

Inzlicht らは、このプロセスモデルを支持するような実験結果をいくつか引用している。たとえば、Muraven の実験である (Muraven et al. 2002)。この実験によると、ある課題で自己制御を行うと、次の課題において、喫煙やアルコールの摂取といった衝動的な行動を行うようになる。これは、最初の課題で自己制御を行うことによって、第

二課題において動機が自己制御から衝動的な行動へシフトしたためと理解できる。

また、Schmeichel らによれば、自己制御を行うと、衝動的な報酬刺激に対する注意が高まる。彼らの実験では、被験者は第一のライティング課題で、自由に何でも書いていい人たちと、用いてもいい単語を制限された人たちに分けられた。そして次の第二課題では、どの被験者も、衝動的な報酬と関連した視覚的な記号 (\$) か、あるいは衝動的な報酬と関連しない視覚的な記号 (%) を見せられ、どちらの記号かを瞬時に判断するように求められた。その結果、第一課題で単語を制限された被験者群 (自己制御を行った群) は、そうでない群よりも、衝動的な報酬と関連した記号をより正確に検出した (Schmeichel et al. 2010)。これは、自己制御を行ったために、衝動的な報酬と関連した知覚が敏感になったためと考えられる。この事例は、Inzlicht らによれば、自己制御を行うことにより、注意が自己制御の必要性から衝動的な報酬の方へシフトしたものとして解釈できる。

また、Inzlicht たちは、従来、リソースモデルを支持すると解釈されてきた実験結果も、プロセスモデルによって十分理解可能だと言う。たとえば、Inzlicht 自身らによるかつての実験 (Inzlicht and Kang 2010) では、女性は男性よりも数学ができないという固定観念が自己制御に及ぼす影響が調べられた。被験者はすべて女性で二つのグループに分けられた。一方のグループはこの固定観念を見直してそうではないという新たな認識のもとで数学の問題を解いた。それに対してもう一方のグループは、この固定観念を持ったまま、自分の数学の成績がそのような固定観念のもとで見られるかもしれないという脅威を感じながら、そうならないようにしようと頑張って数学の問題を解いた。その後でどちらのグループもアイスクリームの試食を行った。その結果、後者のグループの方が、前者のグループよりもたくさんアイスクリームを食べる傾向が見られた。Inzlicht らはこの実験を行った時点では、この結果をリソースモデルに基づいて理解している。すなわち、固定観念を見直したグループはその効果で数学の問題を解く際に自己制御が少なく済み、それゆえ自己資源の消耗が少なく、その結果アイスクリームを食べる量が少なかったのだというわけである。

だが、Inzlicht たちはその後、この結果をプロセスモデルによって解釈することも可能だと考えを改める (Inzlicht et al. 2012)。すなわち、第一課題でより多くの自己制御を行うと、第二課題においてアイスクリームという美味しい食品を食べないようにしようという動機 (自己制御への動機) から、それを食べようという動機へと動機がシフトしたからだと解釈できるというわけである。

こうして、Inzlicht らは、Baumeister らが考案した実験パラダイムによる諸々の実験結果、すなわち自己制御の遂行による自己制御の喪失を示す諸結果は、自己資源という得体の知れないものを措定するリソースモデルによらなくても、動機や注意のシフトという明瞭な過程のみに基づくプロセスモデルによって十分説明できると主張するのである。

9.4 どちらのモデルが正しいか

これまで見てきたように、プロセスモデルは自己制御の喪失という現象を、リソースモデルが想定しているような自己資源の消耗とみなすのではなく、私たちの動機や注意のシフトのゆえとみなし、自己資源と呼ばれるものを用いずに自己制御の喪失の説明を試みている。プロセスモデルとリソースモデルの最も異なる点はここにある。そして、プロセスモデルはリソースモデルのように、自己資源というその本性が明らかでないままのものを用いないという点では確かに優れていると考えられる。だが、果たしてプロセスモデルはリソースモデルよりも本当に優位なモデルと言えるのだろうか。

プロセスモデルを二重システム理論の枠組みで捉えると、動機や注意のシフトにより自己制御の喪失が起こるとするのは、動機や注意がシステム2に有利なあり方からシステム1に有利なあり方へと移行し、それゆえシステム1主導の行為が行われるようになるということとして捉えられる。たとえば、チョコクッキーを我慢してラディッシュを食べた人が、次のパズル解きの課題で成績が悪かったのは、システム2が行おうとする粘り強いパズル解きに有利に働くような動機や注意のあり方が、システム1が行おうとするもっと楽なことに有利に働くような動機や注意のあり方へと移行したからである。もし動機や注意がそのように移行することがなければ、システム2によって意志力を発揮し、もっと長い時間パズル解きに取り組むことができたであろう。

他方、リソースモデルの方は、システム2が意志力によってシステム1を制御するために必要となる自己資源というものの存在を認め、その自己資源が消耗することにより、自己制御の喪失が起こるという見方として捉えられる。このモデルでは、意志力は自己資源という独自の資源を用いるものであり、欲求ないし情動には還元できない独自の力であるとされる。それに対して、プロセスモデルでは、自己資源のような独自の資源を認めない。従って、それゆえ恐らく欲求ないし情動に還元されないような独自の意志力も認めないであろう。そうだとすれば、プロセスモデルでは、どのような行為が起こるか、特にシステム2とシステム1の行為決定が異なるときどちらが決定した行為が生じるかは、もっぱら欲求や情動のあり方によって決まることになる。動機や注意のシフトはこの欲求ないし情動のあり方を変えるに過ぎない(何に注意を向けるかによって欲求ないし情動は変わる)。動機や注意のシフトにより、システム2が決定した行為を動機付ける欲求や情動が最も強力であるようなあり方からそうでないあり方に変化すると、自己制御の喪失が起こるのである。

しかし、そうだとすると、システム2とシステム1の行為決定が食い違うとき、必ず自己制御の喪失が起こることになるだろう。なぜなら、システム1が決定する行為は最も強力な欲求や情動に支えられた行為であり、それゆえシステム2が決定するそれとは別の行為はそのような最も強力な欲求や情動に支えられないからである。プロセスモデルにおいては、自己制御が成立し、システム2が決定した行為が実行されるのは、システム2の知的な評価に一致するよう情動的评价がシステム1においてなされている場

合にかぎる。これは意志力という独自の力を認めないことから必然的に帰結することである。だが、私たちの実行する行為は、常に欲求や情動の大きさによって決まるのだろうか。たとえ弱い欲求や情動にしか支えられていない行為でも、まさに意志の力によってその行為がなされるという可能性もあるのではないだろうか。そうだとすれば、自己資源という正体不明の資源を持ち出すとはいえ、リソースモデルにもそれなりの魅力があることになる。

また、リソースモデルへの反論として出された実験結果も、リソースモデルの側からそれなりの解釈を与えることも可能である。たとえば、Job らの実験では、意志力が限られていないと信じる人は、そうでない人と比べて、第二課題での成績が良かった。このように自己制御が意志力についての信念に左右されることは、確かに意志力を裏打ちする自己資源のようなものがあるという考えに疑問を投げかける。しかし、自己資源が実際にどれくらいあるかということと、その資源をどれくらい利用できるかということとを分けて考えれば、この実験結果をリソースモデルの観点から理解することも可能である。すなわち、第一の課題で自己資源を消耗して自己資源がほとんど残っていても、自己資源が限られていないと思う人は、まだ自己資源はあると思って、そのわずかに残った資源を利用しようとするし、自己資源が限られていると思う人は、もう自己資源がないと思って、資源を利用しようとしな。そのため第二課題での成績に差が出るのである。同様に Tice たちの実験結果、すなわちポジティブな気分が自己制御を回復させるという事実も、ポジティブな気分が自己資源の利用を促進するとみなせば、リソースモデルの立場から十分理解可能である。このように、リソースモデルにとって不利であるとされた実験結果も、リソースモデルによってそれなりに解釈することが可能であり、リソースモデルにとって決定的な困難とはならないのである。

逆に、リソースモデルの側からプロセスモデルに疑問を呈することも可能である。プロセスモデルでは、自己制御の喪失は動機や注意のシフトによって起こるとされるが、このシフトはなぜ生じるのだろうか。それは自己資源が消耗したからではないだろうか。動機や注意のシフトがなぜ起こるのかを説明するために、自己資源の想定が必要であるように思われる。プロセスモデルでは、第一課題でチョコクッキーを我慢してラディッシュを食べると、第二課題のパズル解きにおいて、粘り強いパズル解きに有利な動機や注意のあり方から、もっと楽なことに有利な動機や注意のあり方へのシフトが生じるとされる。しかし、なぜこのようなシフトが起こるのだろうか。また、第一課題でチョコクッキーを好きなだけ食べればよい場合は、第二課題において、そのようなシフトが起きないとされるが、なぜそうなのだろうか。それは、自己資源が消耗して少なくなると、そのようなシフトが起こり、そうでなければ起こらないということではないだろうか。動機や注意がどのようなあり方になるかは、自己資源の有無によって左右されるように思われる。粘り強いパズル解きに必要ない動機や注意のあり方を維持しようとするれば、自己資源が必要であり、自己資源が消耗して少なくなると、もっと楽なことに有利な動機

や注意のあり方になるのではないだろうか。そうだとすれば、動機や注意のシフトがなぜ生じるのかということは、リソースモデルでは自己資源の消耗によって説明できるが、プロセスモデルでは全く説明されないままということになる。

以上より、リソースモデルもまだ十分に生き残る余地があると考えられる。いや、それどころか、むしろリソースモデルのほうがプロセスモデルより有力ではないだろうか。リソースモデルはプロセスモデルの言う動機や注意のシフトが自己資源の消耗によって起こるとすることで、プロセスモデルにその基礎を与えることができる。確かに、リソースモデルには、自己資源の正体が明らかではないという弱みが残るが、いずれその正体が明らかにされる可能性もある。リソースモデルとプロセスモデルのいずれが正しいかは、最終的には自己制御についての科学的な探求がさらに発展するのを待つ必要があるが、少なくとも現時点では、リソースモデルのほうが有力だと思われる。そうだとすれば、意志や意志力というのは単なる説明の道具ではなく、それなりの実体のあるものだということになる。

第十章 責任の二つの面

ここまでは行為主体が意思決定をしたり意志の力を発揮する場面を考察してきたが、最後に行為を評価する場面、すなわち行為者にその行為の責任を帰したり、その行為のゆえに行為者を非難したりする場面を取り上げよう。責任には大きく分けて二つの面がある。そしてこの二つの面の区別はシステム1とシステム2の区別と密接に関わっている。本章ではこのことを論じていこう。

通常、他人の物を盗んだり人を殺したりといった悪いことをすると、それについて責任を問われる。だが、事情によっては行為者の弁解が成立し、その行為を引き起こしたことについては責任があるとされるにもかかわらず、非難には値しないとみなされる場合がありうる。たとえば、大きな池で母親がボートに乗っているときに、そのボートから離れたところの一方で自分の子供が溺れかけていて、他方で5人の他人の子供が溺れかけているとしよう。その母親はどちらか一方にしか行くことができず、そのため、溺れかけている自分の1人の子供を助けるか、溺れかけている他人の5人の子供を助けるかのどちらかしかできない状況に立たされていた。彼女は自分の子供を救うことを選び、その結果他人の5人の子供は溺れ死んでしまった。彼女は1人の子供を救うために5人の子供を死なせたがゆえに、やはり悪いことをしたと言われるだろう⁵⁶。それでも、その1人の子供が自分の子供であるがゆえに、致し方ないこととして非難はされないに違いない。つまり、助けた子が自分の子であったということが弁解として成り立つのだ。だが、悪いことをしたなら、どのような事情があれば、それ相応の非難を受けるべきではないだろうか。そうであるにもかかわらず、なぜ弁解によって行為者が非難されずに済むということが生じるのだろうか。

このような弁解による責任免除の問題に答えるために、まず責任に関する一つの重要な区別である「帰属責任 (attributability)」と「非難責任 (accountability)⁵⁷」という責任の二つの側面とその関係を明らかにする。次に、その区別を踏まえて、弁解が責任実践の中でどのような位置を占めるかについて、正当化と弁解の違いを浮き彫りにし、最後に、弁解がなぜ成立可能であるのかについて、情動の観点から探究を試みる。

⁵⁶ 彼女は救った子供の母親であるがゆえに、彼女の行為は決して悪いとは言えないという考えもあるだろうが、ここでは彼女は悪いことをしたとして話を進めることにする。

⁵⁷ 「accountability」はふつう日本語では「説明責任」と訳され、自分の行った行為について責任があり、かつ、その理由を尋ねられればその説明を行わなければならないというような意味で用いられる。しかしここでは、Fischer and Tognazzini (2011) において彼らがこの語を用いた意味を汲んで「非難責任」と訳すことにした。Fischerらがこの語を非難責任の意味で用いたのは、Watson (2004) に倣ってであり、Watsonは責任の二つの面、すなわち、その行為を誰が行ったのかという面 (帰属の面) とその行為がどのような非難に値するかという面 (非難の面) を区別するために、accountabilityを通常とは異なる特別な意味で用いている。

10.1 責任の二つの側面とその関係

まず、責任の二つの側面とその関係について見ていこう。いかにして弁解が可能かという問題に取り組むのに、Fischer and Tognazzini (2011) の責任論の枠組みが非常に有益である。彼らは、Watson (2004) が区別した道徳的責任の二つの面を踏まえて、道徳的責任を帰属責任 (attributability) と非難責任 (accountability) に区別している。帰属責任というのは、道徳的に悪であるような行為が、ある人に帰属させられるという意味での責任である。つまり、その行為は確かにその人が行ったことだとされるということである。この帰属責任が成立するためには以下の三つの条件が必要となる。第一に、行為者が行った行為が意図的であること、第二に、行為者が目標を立てそれを確実に行う賢明さ (prudence) の能力を有すること、第三に、行為者が反応的態度 (reactive attitude) ⁵⁸ (非難や称賛) の対象となりうるような道徳的能力 (善悪を弁別し、善をなし、悪をなさないようにする能力) を持ち合わせていることである⁵⁹。このような三つの条件がすべて成立しているときにのみ、行為者に対して、帰属責任が成立しているといえる。つまり、帰属責任があるということは、道徳的能力のある行為者が、その能力の行使の結果として、ある道徳的行為を賢明な仕方で行ったということなのである。責任の帰属面が成立するためには、行為者は単にある行為を意図的にを行い、且つ、目的を立てそれを確実に行う賢明さの能力を有するだけでなく、そもそも非難の対象になりうるような道徳的能力を持ち合わせていなければならない。このよう道徳的能力のある行為者が、その能力の行使の結果として、ある善または悪の行為を意図的に行ったというのが帰属責任である。

一方、道徳的責任のもう一つの側面である非難責任とは何であろうか。これは道徳的に悪い行為に対する行為者の帰属責任が確定された後で、その帰属責任のゆえに共同体から行為者が非難されるということである。共同体から行為者がどれくらいの非難に値するのかが評価され、それに応じた非難がなされるが、これが非難責任である。行為者に向けられる非難は、単に一人で憤慨している程度の軽いものから、非常に厳しい制裁を加えるといった程度まで、様々である⁶⁰。

⁵⁸ この用語は Strawson (1962) の論文において用いられていたものである。

⁵⁹ Fischer and Tognazzini (2011) は第二の条件と第三の条件を明確に区別するが、Watson (2004) は第二の条件には第三の条件が含まれると考える。たとえば、サイコパスが免責されるとすれば、それは Fisher によると、彼らは目的を立て達成する能力はあるが、非難に値するような道徳的能力を持ち合わせていないとみなされるからであるが、Watson によると、彼らはそもそも善悪にかなった目的を立てることができないからということになる。ここでは、賢明さの能力と道徳的能力を区別できるかどうかという問題を論じる余裕はないが、サイコパスのような事例からすると、この二つの能力を分けることは可能だと思われるので、本論文ではそれらは区別できるとして話を進めることにする。

⁶⁰ Fischer and Tognazzini (2011) は行為者に向けられる非難を軽いものから順に五つの段階に区別している。第一段階は、行為者に対して憤りなどの反応的態度を抱くことである。たとえば、友人が嘘をついたことを知って、その友人に対して憤りを感じるような場

さらに、非難責任の場合、そのあり方には現実的な側面と原理的な側面の二つがあると Fischer らは言う。現実的な側面とは、道徳的に悪いことをした人に対して誰かある現実の人が実際に非難をするのが正当である場合、非難責任が成立するということである。一方、原理的な側面とは、そのような人が現にいなくても、非難する資格のある人を想定することが可能である場合、非難責任が成立するということである。ここで重要なことは、現実的な非難責任が成立すれば、原理的な非難責任も成立するが、原理的な非難責任が成立するからといって、必ずしも現実的な非難責任が成立するとは限らないということである。例えば、悪口を言われた人に対して憤りの反応的態度を抱くことが現実的に正当であるのは（つまり現実的な非難責任が成立するのは）、悪口を言った人と同じ状況におかれても悪口を言ったことがないような人が現実にいる場合である。そのような場合には、当然、そのような人を想定することができるから、憤りの反応的態度を抱くことは原理的にも正当である（つまり原理的な非難責任も成立する）。しかし、憤りの反応的態度を抱くことが原理的に正当であるのは、悪口を言った人と同じ状況におかれても決して悪口を言わない人が単に想定できる場合であり、このような場合には、現実にもそのような人がいるとは限らない。従って、現実的な非難責任は必ずしも成立しないのである。

Fischer and Tognazzini (2011) は、帰属責任が確定すると、通常それに応じた非難責任が課されるとしつつも、そうでないこともありうるとしている。というのも、ときにその行為に対して弁解が可能だからである。弁解は、帰属責任と非難責任の間の必然的なつながりにいわば「ギャップ」を設けるのである。ただし、ここで「弁解」と「正当化」は区別されねばならない。次ではその違いを明らかにしながら、なぜ弁解が成立しうるのかを考察していこう。

10.2 正当化と弁解

もし行為が行為者に帰属させられない場合、その行為者はいかなる非難も免れうるが、行為が行為者に帰属させられる場合、ふつうその行為者には何らかの非難が及びせられる。しかし、たとえ帰属責任が成立したとしても、非難責任には値しないとされる場合がある。なぜなら、そこには正当化や弁解の余地があるからである。このような正当化や弁解は非難責任を免除するためにどのような働きをするのだろうか。そしてさらに正

合がそうである。第二段階は、行為者に対して表立って反応的態度を示すことである。友人が嘘をついたことを知って、その友人に対して憤りを感じるだけでなく、文句を言う、つまり表立った仕方で憤りを表現する場合がそうである。第三段階は、行為者に対して、単に表立った反応的態度を示すというのではなく、何らかの制裁を加えることである。友人の嘘が原因で 10 万円を失った場合、それに対して文句を言うだけでなく弁償させるような場合がこれにあたる。第四段階は、行為者に対して身体的苦痛を与えることである。ある人が強盗を行ったなら、その人を刑務所に入れるというような場合がそうである。最後の第五段階では、行為者に対して永遠の苦しみ（たとえば、地獄における苦しみのような終わりのない苦しみ）を与えることとなる (cf. Strawson 1994)。

当化と弁解はどのような点で異なるのだろうか。まず正当化についての考察から始めよう。

行為者が自分の行為の正当化に成功すると、行為者の帰属責任が無化され、それゆえ一切の非難責任が無化される。正当防衛はその典型的な事例である。たとえば、Aが人を殺したが、その人を殺さなければ自分が殺される状況だったことを明らかにして、その状況での殺人は実は正当だったと主張し、それが受け入れられたとしよう。このような場合、結局、最初科せられた帰属責任は無化され、それゆえそれに応じた非難責任も無化されることになる。なぜなら、最初は殺人という悪い行為が帰属させられていたのだが、正当化により、その行為は殺人という行為ではあるものの、正当防衛であるために悪い行為でなかったことが明らかとなったので、悪い行為がその行為者に帰属させられなくなるからである。

一方、弁解の場合はどうだろうか。行為の弁解が成立すると、帰属責任は無化されないが、非難責任は無化される。たとえば、一人の子供を助けて五人の子供を助けなかった母親が、助けたのは自分の子だという弁解を行い、その弁解が受け入れられた場合、帰属責任は無化されないが、非難責任は無化される。つまり、母親が悪いことをしたという点は変わらないが、それでもその行為ゆえに非難されることはない。彼女の弁解が非難責任を無化するのである。行為の正当化は帰属責任を無化することによって非難責任を無化するが、弁解は帰属責任を無化することなく非難責任を無化する。だが、如何にして弁解は非難責任を無化するのだろうか。帰属責任があるなら、やはり非難責任が必ずあるのではないだろうか。一体如何にして帰属責任があるのに、非難責任がないということが可能なのだろうか。次の節で、弁解が如何にして非難責任を無化するのかについて考察してみよう。

10.3 なぜ弁解は可能か

これまで見てきたように、弁解は帰属責任が成り立つにもかかわらず、非難責任を成り立たなくさせる。なぜこのようなことが可能なのだろうか。この問題を考察するには、まず P. F. Strawson の責任論を参照しておくことが有益である。

Strawson (1962) の責任論によると、責任とは人々が基本的な善意を要求し合うことによって結びついた共同体の中で生じるものであり、その基本的な善意の要求が満たされるかどうかに応じて共同体の成員から行為者にある反応的態度が向けられる。このことが私たちの責任実践の中核をなすのである。たとえば、ある人が他人の財布を盗めば、その被害者や共同体の成員はその行為者に対して憤りを示す。だが、いったんその行為者にそもそも道徳的な責任を果たす能力（基本的な善意の要求を満たす能力）がないことが見出されたり強制的にやらされたりしたことが判明すれば、行為者に非難の反応的態度が向けられることはなくなるだろう。つまり、その行為者は非難に値しないとみなされ完全に免責されるのである。なぜこのようなことが起こるのかというと、

Strawson の責任論によれば、そもそも行為者に対して非難の心情が湧かなくなるからである。非難の心情はそれ自体が反応的態度の一つであるが、それはその心情の表立っての表現や制裁などの他の反応的態度を支える基盤ともなる。非難の心情が湧かなければ、如何なる非難の反応的態度も向けられなくなり、それゆえ非難責任が問われることもなくなる⁶¹。

Strawson の責任論に従うと、Fischer らのいう非難責任とは、非難の反応的態度を行為者に向けるということである。この反応的態度の基盤となるのは心情（つまり情動）であり、それゆえ非難責任があるかどうかは、行為者に対して情動が生じるかどうかの問題として捉えなおすことが可能である。ここで、情動の観点から弁解の問題を考察するには、Watson の議論を参照するのがおおいに役に立つ。Watson は弁解の問題を考察するために、非常に印象的なロバート・ハリスの事例を持ち出している。ハリスは幼少期に度重なる虐待を両親から受け続けるという劣悪な環境下で育った。彼の家族の証言によると、もともとはとても心優しい少年だったが、その性格は次第に変化し、少年時代には少年院に度々収監されるような不良少年となった。ハリスは正常な知能を持ち、道徳的な善悪も理解していたが、自ら進んで悪いことをするようになった。そして大人になった彼は、残虐な仕方二人の少年を殺すという大きな罪を犯してしまった。このハリスの犯した罪に弁解の余地はあるだろうか。

Watson によると、私たちはハリスの養育環境を知ることによって、彼に対して同情の念を抱くようになる。だが、それは一方で憤り（という反応的態度）を抱きつつ、もう一方で（憤りを緩和するような）同情（という反応的態度）を抱くというアンビバレントな状態に陥るということである。私たちがアンビバレントな状態に陥るのは、ハリスの不幸な養育環境を知ることが私たちに同情を抱かせはするが、その同情によって私たちの憤りが完全に鎮静化されたりはしないからである。しかし、そうだとすると、ハリスへの同情が彼への憤りを緩和することによって、非難責任をある程度小さくするであろう。それゆえ、ハリスの養育環境はそれなりに弁解の役割を果たすといえるだろう。

では、先の母親のケースはどうだろうか。私たちは彼女が母親であることを知ることによって、彼女に対して同情を抱くようになるが、この場合はハリスの場合と違って、私たちは彼女に対する憤りと同情のアンビバレントな状態になるのではなく、もっぱら同情を抱くだけであろう。そうだとすれば、彼女に科される非難責任は完全になくなるだろう。こうして母親であるということは弁解の役割を完全に果たすことになる。

これまで見てきた二つの事例からすると、弁解は同情によって可能になると考えられる。では一体、弁解において、何がこのような同情を私たちに抱かせるのだろうか。弁解を聞いたとき、私たちはどのような考慮によって同情を抱くようになるのだろうか。

まず、第一に考えられるのは、反実仮想である。「自分もハリスと同じ環境で育った

⁶¹ Strawson (1962) は帰属責任と非難責任を区別していない。ここでは Strawson の議論のうち、非難が情動に深く関わっているという指摘のみに注目している。

ら、同じことをするだろう」という思いが同情を抱かせる。これは、自分もハリスと本質的に同じ人間だという思いがあるからである。もう一つは、この反実仮想と密接に関連するが、道徳的運である。自分とハリスは確かに異なるが、その違いは、本人の努力ではどうしようもない養育環境の違いだけである。つまり、私は良い環境で育てられて運が良かったが、ハリスは劣悪な環境で育てられて運が悪かったのである。このような道徳的運に関する考慮が彼への同情を抱かせる。同情を抱かせるのは反実仮想と道徳的運という互いに密接に関連する二つの考慮であるように思われる。そして、この同情によって、非難の反動的態度が抑制され、非難責任が無化される。こうして弁解が可能となるのである。

しかしながら、まだ残されている問題がある。弁解が成立する場合、帰属責任は成立するが、非難責任は成立しない。しかし、Fischer によると、非難責任には現実的な非難責任と原理的な非難責任の区別がある。弁解が成立する場合、現実的な非難責任だけではなく、原理的な非難責任も成立しないのだろうか。現実には誰も非難を行えないとしても、原理的には誰か（たとえば聖人）が非難を行いうることは想定可能ではないだろうか。

これまでの検討によると、弁解が可能な場合、現実の誰もが本質的に同じであり、違いはただ道徳的な運によるものに過ぎないから、現実の誰もが行為者に同情せざるをえず、それゆえ行為者を非難することができないということであった。だが、同じ状況下でも同じ悪いことをしないような人を想定することは可能である。たとえば、聖人のような人を想定することができる。そうだとすれば、原理的な非難責任はなくなってしまうように思われる。つまり、非難責任を科す資格のある人が現実には存在しないとしても、原理的には存在しうるとすれば、現実的な非難責任が無化されたとしても、原理的な非難責任は無化されないように思われるのである。同情によって弁解が成立する場合、現実的な非難責任は無くなるが、原理的な非難責任はなお残ると考えるべきであろう。

10.4 まとめ

これまで見てきたように、道徳的責任は、道徳的に悪であるような行為がある人に帰属させられるという意味での「帰属責任」と、行為者の帰属責任が確定された後でその帰属責任のゆえに共同体から行為者に非難が向けられるという意味での「非難責任」という二つの面から成る。また、非難責任には、現実的なそれと原理的なその二つがある。通常は帰属責任が確定するとそれに応じた非難責任が科されるが、行為の正当化によって行為者に科せられた帰属責任が無化されると、非難責任も無化される。また、弁解が成立すると、帰属責任は無化されないものの、非難責任は無化される。そしてこのような弁解がなぜ可能かと言えば、それは行為者に関する反実仮想と道徳的運の考慮によって行為者への同情が生じ、非難の情動が抑制されるからである。弁解はこのようにして帰属責任を無化しないものの、非難責任を無化する。但し、非難責任を無化すると

いっても、それは現実的な非難責任であり、原理的な非難責任は無化されずに残る。弁解は同情により非難責任を無化するが、それゆえに現実的な非難責任を無化するだけで、原理的な非難責任まで無化するわけではないのである。

以上の話が二重システム理論の枠組みでどのように捉えられるのか見ておこう。まず現実的な非難責任はシステム1の働きによるものだと言えよう。現実的な非難責任とは、非難の反応的態度を行為者に向けることであり、この反応的態度の基盤となるのは情動である。それゆえ現実的な非難責任があるかどうかは、行為者に対して情動が生じるかどうかの問題として捉えられる。また、弁解によって現実的な非難責任が無化されるのは、同情によってそのような情動が生じなくなるからである。このように、現実的な非難責任は、情動の働きによって成立しているとみなすことができる。そしてシステム1は情動的なシステムである。それゆえ、現実的な非難責任はシステム1の働きによるものだと考えられよう。

これに対して、原理的な非難責任はシステム2の働きによるものだと言えよう。原理的な非難責任は、憤りのような情動を実際に持つことによって成立するのではなく、そのような情動を抱くことが正当であるような人を想定することができさえすれば成立する。このような想定は知的な働きによるものだと言えよう。また、原理的な非難責任は、弁解によって行為者への同情が生じる場合でも無化されない。この点でもそれは情動ではなく、知的な働きに関係すると言えよう。そしてシステム2は、知的なシステムである。それゆえ、原理的な非難責任はシステム2の働きによるものだと考えられよう。

また、帰属責任もシステム2の働きによるものだと言えよう。帰属責任は、道徳的に悪であるような行為がある人に帰属させられるという意味での責任である。この悪い行為の帰属は、情動的な非難に先立って為されるものである。このような帰属は知的な判断によってなされる。従って、帰属責任はシステム2の働きによるものだと言えよう。また、このことは、帰属責任が原理的な非難責任と不可分な関係にあるということからも伺える。帰属責任が成立するなら、必ず原理的な非難責任は成立する。帰属責任が成立しているときに、弁解によって現実的な非難責任が無化されても、原理的な非難責任は決して無化されない。そしてもちろん、原理的な非難責任が成立するなら、必ずその前提として帰属責任が成立している。このように、帰属責任は原理的な非難責任と不可分な関係にある。そして、原理的な非難責任は、システム2の働きによる。このことから、帰属責任もシステム2の働きによることが伺える。

以上のように、責任には帰属責任と非難責任の二つの面があり、また非難責任には現実的な非難責任と原理的な非難責任がある。そして帰属責任と原理的な非難責任は、システム2の働きによるものであり、現実的な非難責任はシステム1の働きによるものだと捉えることができる。

むすび

本論文は、自己制御と情動を主題とし、自己制御において情動が果たす役割と意義を明確にすることを中心としつつ、さらに信念形成における自己制御の問題、および意志による自己制御のあり方を論じ、また、私たちの現実の実践における情動の役割と意義を責任という場面に即して論じた。

特に本論文で明らかにしたことは、自己制御において適切な情動（合理的な情動）がいかに関与しているかということである。情動は確かに不適切な（合理的でない）場合もあるが、適切な場合もある。情動が不適切な場合、システム1とシステム2の行為決定が食い違い、システム2がシステム1を実質的に制御することによって、自己制御が為される。従って、システム2がシステム1をうまく制御できなければ、自己制御の喪失が起こる。

だが、不適切な情動のゆえに自己制御の喪失がもたらされるだけでなく、情動の欠如のゆえにそれがもたらされることもある。情動は自己制御において積極的な役割を果たしており、行為を動機付ける力を持つ。それゆえ、情動が欠如してシステム2しか働かない状態になると、システム2による知的な評価を行為に結び付けることができなくなり、むしろ自己制御も意思決定も行えなくなるのである。

また、自己制御はシステム1とシステム2が食い違う場面でなされるだけではない。情動が適切な場合は、情動が合理的なので、情動によるシステム1の行為決定はシステム2の行為決定と一致する。この場合は、自己制御がおのずとなされることになる。すなわち、意志の力によって情動を制御する必要がなく、情動のままに（システム1の行為決定どおりに）行為すれば、システム2の決定どおりの行為が為される。このように自己制御には、情動が不適切な場合に実質的になされる自己制御と、情動が適切な場合におのずとなされる自己制御の二つの種類がある。そして、おのずとなされる自己制御の場合は、むしろ情動に積極的な役割が認められることになる。意志の力は消耗しやすいため、不適切な情動を制御する必要のある自己制御は失敗に終わりやすい。従って、概ね自己制御ができているというようになるためには、概ね適切な情動を抱くことができ、それゆえたいいていの場合に、意志の力を行使せずにおのずと自己制御ができているということが必要となる。そしてそのような場面では、情動が意思決定や行為を導く役割を担っているのである。

本論文は、自己制御における情動の重要性を明らかにするために、まず第一章で、自己制御を考察する際の枠組みとして本論文で用いた二重システム理論の考察を行った。そして前半の第二章から第五章で、適切な情動によっておのずと自己制御が為される場合の自己制御について考察し、さらに後半の第六章から第九章で、適切な情動がないために意志の力によって自己制御が為される場合の自己制御について論じるとともに、信念形成における自己制御の問題についても論じた。最後に第十章では、私たちの現実の

実践において情動が果たす役割と意義を責任という場面に即して論じた。これらを簡単に振り返っておこう。

適切な情動によっておのずと自己制御が為される場合の自己制御については、まず、第二章で、情動がどのような役割を果たすのかに先立ち、そもそも情動とは何かについて、その本性を明らかにする試みを歴史的な背景を踏まえつつ行った。その結果、情動の本質とは身体性であるとする見方が最も優れていると結論付けた。情動の本質は外界の対象によって引き起こされる身体状態にあり、この身体状態は外界の対象の価値を反映する状態である。情動に関するこのような見方を代表するのが **Damasio** のソマティック・マーカー仮説である。この仮説によると、外界の価値を反映する身体状態が、意思決定場面において各選択肢の有利・不利を予測するための因子として働き、常に選択に対してバイアスを掛けるのである。

第三章では、このようなソマティック・マーカー仮説がどの程度経験的に裏付けられているかを、その仮説を支持する証拠とされるアイオワ・ギャンブル課題の実験結果の検討を通じて考察した。その結果、次のような点をさらに考慮してソマティック・マーカー仮説を部分的に修正する必要があることを明らかにした。すなわち、①ネガティブな情動とポジティブな情動の区別、②目先の報酬への非情動的な傾向性、③以前の行動の結果として生じた事後的な情動の残存、④意思を実現する執行プロセスに対して情動が影響を及ぼす可能性、の四点である。しかし、外界の対象の価値を反映する身体状態が意思決定においてある一定の影響を及ぼしていることは間違いないと考えられるので、ソマティック・マーカー仮説の基本的な考えは正しいと言えよう。

次に第四章において、自己制御における情動の役割について、その最も本質的な要素と考えられる身体状態が、実際にどのような役割を果たすのかについてさらに具体的に考察を行った。ここでは三つの価値評価、すなわち、身体的情動、あたかも身体的情動、知的価値評価が、自己制御が必要な場面でそれぞれどのような役割を果たすのかについて比較検討した。その中で、身体状態を含まないあたかも身体的情動が、それを含む身体的情動なしには成立しないという点で、やはり身体性が情動にとって本質的であることを確認した。また外界の対象についての価値評価についても、身体的情動は外界の対象に応じて柔軟な価値評価を行うのに対し、あたかも身体的情動は過去の価値評価と同じような紋切型の価値評価しか行うことができず、やはり価値評価においても身体性が重要な役割を果たしていることを示した。さらに、実際に行為を引き起こす動機付けの力については、身体的情動が外界の対象の価値評価に見合うだけの十分な動機をもたらす一方、あたかも身体的情動はそれに見合うほどの動機付けの力をもたらさず、また、知的価値評価についてはそれ自体としてはそもそも動機付けの力を持たないことを示した。このように、身体状態を含む情動が、自己制御が必要な場面において、外界のあり方に応じた柔軟な価値評価とその評価に見合う動機付けの力をもたらし、それゆえ外界のあり方に応じて適切な意思決定を可能にするという点で、重要かつ独自の役割を果

たすことを明らかにした。

第五章では、情動が自己制御に貢献しうることをさらに明らかにする為に、Helmの情動論を援用して、情動の合理性について論じた。ここでは情動が三つの対象、すなわち標的、形式的対象、焦点を持つこと、そして情動が共通の焦点に基づいて合理的なパターンを形成しており、さらにそのような情動が欲求や知的判断とも合理的なパターンを形成していることを示した。このように、情動の合理性を示すことにより、情動が自己制御においてそれを促進する役割を果たしうることを明らかにした。

第六章から第九章では、自己制御の喪失（意志の弱さ）に焦点を当てて論じた。まず第六章で、自己制御の喪失のあり方を二重システム理論の枠組みの中で考察した。自己制御の喪失は、最善の判断に合致しない意思を形成してしまう意思決定過程で起こるものと、意思に合致しない行為をしてしまう執行過程で起こるものとの二つのタイプがある。前者はシステム2が最善の判断に従って意思決定を行おうとする段階で生じる自己制御の喪失であり、後者はシステム2が最善の判断に従って意思決定をし、その意思通りの行為を行おうとする段階で生じる自己制御の喪失である。前者はシステム1における情動の異常ゆえに生じた知的評価と動機付けの力の乖離によって引き起こされ、後者は、システム1の過活動やシステム2による事後的正当化、自己資源の消耗、価値の時間割引によって引き起こされることを提示した。

第七章では信念形成における自己制御の喪失について論じた。ここではまず、認知的アクラシアが可能かどうかを詳しく検討し、十全な信念の形成においては反対の証拠の力が無化されるということから、それが不可能であると結論付けた。また、欲求を加えた新たな認知的アクラシアについては、信念の本性上、意志的に信念を形成することはできないということから、やはりそれが不可能だとの結論を下した。そして認知的アクラシアが不可能であるとしても、信念形成における自己制御の喪失だといえるようなケースは存在することを指摘した。そのケースとは、欲求や情動が意識的にせよ無意識的にせよ実際の信念形成過程に働きかけ、何を信じるべきかの判断と一致しないような信念を形成する場合と、欲求が理論的推論過程に無意識的に働きかけ、何を信じるべきかの判断を歪める場合である。

第八章では、意志の弱さの非限定性を唱える器官理論と意志の弱さの限定性を唱える異時点間交渉モデルを比較検討した。その結果、器官理論は比較的短い時間幅で引き起こされる意志の弱さを問題にし、異時点間交渉モデルは繰り返し生じる異なる時点における意志の弱さを問題にしていることを指摘し、これら二つの理論が矛盾しないことを明らかにした。それにより、意志が、異時点間の自己の交渉という側面と、情動や欲求に対抗しうる独自の力を発揮する器官であるという側面を併せ持つことを指摘した。

第九章では、意志の力とは何かを明らかにするために、意志力が枯渇して自己制御の喪失に陥るといった事態がどのようにして生じるのかについて、自己資源という実体的なものを想定してその消耗によって生じるとするリソースモデルと、ある種の認知過程

(動機や注意のシフト)によって生じるとするプロセスモデルを比較検討した。その結果、リソースモデルが意志力そのものだと主張する自己資源はその正体が明らかではなく、その点でリソースモデルはプロセスモデルよりも不利であるが、それでもリソースモデルはプロセスモデルの言う動機や注意のシフトがなぜ起こるのかを自己資源の消耗によって説明しうる点で優れており、むしろ現段階ではリソースモデルの方が有力そうだと言えることを示した。

最後の第十章では、私たちの現実の実践において情動が果たす役割と意義を責任という場面に即して考察した。その結果、悪い行為であるにもかかわらず、弁解により責任が免除されることがありうるということについて、弁解の場合、正当化と違って、悪い行為をなしたことの帰属責任を免れないが、それでも同情という情動の働きにより、非難責任は免れうることを論じた。但し、非難責任を免れうるといっても、それは現実的な非難責任であり、原理的な非難責任は免れえないことを指摘した。そして、二重システム理論の枠組みからは、帰属責任と原理的な非難責任はシステム2の働きによるものであり、現実的な非難責任はシステム1の働きによるものであることを示した。

以上のように、本論文では自己制御における適切な情動の重要性を主題的に論じてきた。ここで取り上げた問題の中には今後のさらなる探究が必要なものが数多く含まれる。たとえば、本論文中でも指摘したように、意志が異時点間交渉モデルの捉えた側面と器官モデルが捉えた側面を併せ持つとしても、実際どのようにそれらを兼備しうるのかということや、リソースモデルが主張する自己資源の正体は何かなどである。加えて、本論文ではほとんど触れられなかった大きな問題もたくさん存在する。たとえば、自己制御の能力を高めるためにはどうすればよいか、その際、情動をどう育てていけばよいかということや、そもそも私たちの自己制御の能力は発達段階の中でどのように獲得されていくのか、また、自己制御が要請される実践的場面において具体的に自己制御の能力をどう発揮すればよいか、などである。このような問題については今後の検討課題としたい。本論文では、自己制御が概ねできているようになるためには、情動が極めて重要であるということを示すという課題に取り組んできた。それにより、自己制御とは何かという問題についても、一つの重要な探求の方向性を示すことができたのではないと思われる。

参考文献

- Adler, J. E. (2002). Akritic Belief. *Philosophical Studies* 110: 1-27.
- Ainslie, G. (2001). *Breakdown of Will*. Cambridge: Cambridge University Press.
(G・エイズブリー『誘惑される意志』山形浩生訳、NTT 出版、2006 年)
- Arnold, M. B. (1960). *Emotion and Personality* Vol. 1, *Psychological Aspects*. New York: Columbia University Press.
- Bard, P. (1928). A Diencephalic Mechanism for the Expression of Rage with Special Reference to the Sympathetic Nervous System. *American Journal of Psychology* 84: 490-513.
- Baumeister, R. E., Bratslavsky, E., Muraven, M., and Tice, D. M. (1998). Ego Depletion: Is the Active Self a Limited Resource? *Journal of Personality and Social Psychology* 74: 1252-1265.
- Baumeister, R. E. and Vohs, K. D. (2007). Self-Regulation, Ego Depletion, and Motivation. *Social and Personality Psychology Compass* 1: 115-28.
- Bechara, A. (2003). Risky Business: Emotion, Decision-Making, and Addiction. *Journal of Gambling Studies* 19(1): 23-51.
- Bechara, A., and Damasio, A. R. (2005). The Somatic Marker Hypothesis: A Neural Theory of Economic Decision. *Games and Economic Behavior* 52: 336-372.
- Bechara, A., Damasio, H., Tranel, D., and Damasio, A. R. (1997). Deciding Advantageously before Knowing the Advantageous Strategy. *Science* 275: 1293-1295.
- Bechara, A., Dolan, S., and Hinds, A. (2002). Decision-making and Addiction (part II): Myopia for the Future or Hypersensitivity to Reward? *Neuropsychologia* 40(10): 1690-1705.
- Burns, K., and Bechara, A. (2007). Decision Making and Free Will: A Neuroscience

Perspective. *Behavioral Sciences and Law* 25: 263-280.

Cannon, W. B. (1927). The James-Lange Theory of Emotion: A Critical Examination and an Alternative Theory. *American Journal of Psychology* 39: 106-124.

Colombetti, G. (2008). The Somatic Marker Hypotheses, and What the Iowa Gambling Task Does and Does not Show. *British Journal for the Philosophy of Science* 59: 51-71.

Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Putnam Publishing. (A・R・ダマシオ『生存する脳—心と脳と身体の神秘』田中三彦訳、講談社、2000年)

Damasio, A. R. (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Florida: Harcourt Publishing Company. (A・R・ダマシオ『無意識の脳 自己意識の脳』田中三彦訳、講談社、2003年)

Damasio, A. R. (2003). *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain*. London: Vintage. (A・R・ダマシオ『感じる脳—情動と感情の脳科学 よみがえるスピノザ』田中三彦訳、ダイヤモンド社、2005年)

Damasio, H., Bechara, A., and Damasio, A. R. (2002). Reply to 'Do Somatic Markers Mediate Decisions on the Gambling Task?' *Nature Neuroscience* 5 (11): 1104.

Davidson, D. (1980). How is Weakness of the Will Possible? In D. Davidson, *Essays on Actions and Events*. Oxford: Oxford University Press, 1980. (D・デイヴィッドソン「意志の弱さはいかにして可能か」『行為と出来事』服部裕幸・柴田正良訳、勁草書房、1990年、所収)

De Neys, W. (2006). Dual Processing in Reasoning: Two Systems but One Reasoner. *Psychological Science* 17: 428-433.

De Sousa, R. (2011). *Emotional Truth*. Oxford: Oxford University Press.

遠藤利彦 (2013) 『「情の理」論 情動の合理性をめぐる心理学的考究』東京大学出版会

Ersner-Harshfield, H., Wimmer, G. E., and Knutson, B. (2009). Saving for the Future Self: Neural Measures of Future Self-Continuity Predict Temporal Discounting. *Social Cognitive and Affective Neuroscience* 4: 85-92.

Evans, J. St. B. T., Barston, J. L., and Pollard, P. (1983). On the Conflict between Logic and Belief in Syllogistic Reasoning. *Memory and Cognition* 11(3): 295-306.

Evans, J. St. B. T., and Curtis-Holmes, J. (2005). Rapid Responding Increases Belief Bias: Evidence for the Dual-Process Theory of Reasoning. *Thinking and Reasoning* 11: 382-389.

Evans, J. St. B. T., and Frankish, K. eds. (2009). *In Two Minds: Dual Processes and Beyond*. Oxford: Oxford University Press.

Evans, J. St. B. T and Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science* 8(3): 223-241.

Fischer, J. M., and Tognazzini, N. A. (2011). The Physiognomy of Responsibility. *Philosophy and Phenomenological Research* 82(2): 381-417.

Gailliot, M. T., Baumeister, R. F, DeWall, C. N., Maner, J. K., Plant, E. A, Tice, D. M., Brewer, L. E., and Schmeichel, B. J. (2007). Self-Control Relies on Glucose as a Limited Energy Source: Willpower is more than a Metaphor. *Journal of Personality and Social Psychology* 92 (2): 325-336.

Goel , V., and Dolan, R. J. (2003). Reciprocal Neural Response within Lateral and Ventral Medial Prefrontal Cortex during Hot and Cold Reasoning. *Neuroimage* 20: 2314-2321.

Goldie, P. (2000). *The Emotions: A Philosophical Exploration*. Oxford: Oxford University Press.

Hare, T. A., Camerer, C. F., and Rangel, A. (2009). Self-Control in Decision-Making Involves Modulation of the VMPFC Valuation System. *Science* 324: 646-648.

Harris, C., and Laibson, D. (2001). Dynamic Choices of Hyperbolic Consumers. *Econometrica* 69 (4):935-957.

Helm, B. W. (2010). Emotions and Motivation: Reconsidering Neo-Jamesian Accounts. In P. Goldie (ed.), *The Oxford Handbook of Philosophy of Emotion*. Oxford: Oxford University Press.

ヒューム (2010) 『人性論』 土岐邦夫・小西嘉四郎訳、中央公論新社

Ikeda, S., Kang, M., and Ohtake, F. (2010). Hyperbolic Discounting, the Sign Effect, and the Body Mass Index. *Journal of Health Economics* 29: 268-284.

Inzlicht, M., and Schmeichel, B. J. (2012). What Is Ego Depletion? Toward a Mechanistic Revision of the Resource Model of Self-Control. *Perspective on Psychology Science* 7(5): 450-463.

Inzlicht, M., and Kang, S. K. (2010). Stereotype Threat Spillover: How Coping With Threats to Social Identity Affects Aggression, Eating, Decision Making, and Attention. *Journal of Personality and Social Psychology* 99(3):467-481.

James, W. (1884). What Is an Emotion? *Mind* 9: 188-205.

Job, V., Dweck, C. S., and Walton, G. M. (2010). Ego Depletion — Is It All in Your Head? Implicit Theories About Willpower Affect Self-Regulation. *Psychological Science* 21(11): 1686-1693.

Kahneman, D. (2011). *Thinking, Fast and Slow*. London: Penguin. (D・カーネマン 『ファースト&スロー 上・下』 村井章子訳、早川書房、2012年)

Kahneman, D., and Frederick, S. (2002). Representativeness Revisited: Attribute Substitution in Intuitive Judgement. In T. Gilovich, D. Griffin, & D. Kahneman (eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge: Cambridge University Press.

Kahneman, D., and Tversky, A. (1984). Choices, Values, and Frames. *American*

Psychologist 39(4): 341-350.

加藤忠史 (2013) 「記憶と可塑性」理化学研究所脳科学総合研究センター編『脳科学の教科書 神経編』岩波書店

Kruglanski, A. W., and Gigerenzer, G. (2011). Intuitive and Deliberative Judgements are Based on Common Principles. *Psychological Review* 118: 97-109.

Laibson, D. (1997). Golden Eggs and Hyperbolic Discounting. *Quarterly Journal of Economics* 62(May): 443-477.

Lange, C. G. (1885). (trans. by Krella and I. A. Haupt) The Emotions: A Psychophysiological Study. Reprinted in K. Dunlap (ed.) *Psychology Classics*. Baltimore: Williams and Wilkins, 1922.

Lazarus, R. S. (1984). On the Primacy of Cognition. *American Psychologist* 39(2): 124-129.

LeDoux, J. E. (1996). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York: Brockam, Inc. (J・ルドゥー『エモーショナル・ブレイン 情動の脳科学』松本元他訳、東京大学出版会、2003年)

Levy, N. (2004). Epistemic Akrasia and the Subsumption of Evidence A Reconsideration. *Croatian Journal of Philosophy* 4:149-156.

Levy, N. (2007). *Neuroethics: Challenges for the 21st Century*. Cambridge: Cambridge University Press.

Levy, N. (2011). Resisting Weakness of the Will. *Philosophy and Phenomenological Research* 82 : 134-155.

McClure, S. M., Laibson, D. I., Loewenstein, G., and Cohen, J. D. (2004). Separate Neural Systems Value Immediate and Delayed Monetary Rewards. *Science* 306: 503-507.

Mischel, W., Shoda, Y., and Rodriguez, M. L. (1989). Delay of Gratification in

Children. *Science* 244: 933-937.

Muraven, M., Collins, R. L., and Neinhuis, K. (2002). Self-Control and Alcohol Restraint: An Initial Application of the Self-Control Strength Model. *Psychology of Addictive Behaviors* 16(2): 113-120.

Muraven, M., and Slessareva, E. (2003). Mechanisms of Self-Control Failure: Motivation and Limited Resources. *Personality and Social Psychology Bulletin* 29: 894-906.

村上郁也編 (2010) 『イラストレクチャー認知神経科学—心理学と脳科学が解くこころの仕組み—』 オーム社

Newstead, S. E. (2000). Are There Two Different Kinds of Thinking? *Behavioral and Brain Sciences* 23: 690-691.

西堤優 (2010a) 「ソマティック・マーカー仮説について—アイオワ・ギャンブル課題の解釈をめぐる問題—」『科学哲学』、43 巻 1 号、pp. 31-44

西堤優 (2010b) 「二つの行為形成システムにおける自己制御の喪失」『脳科学時代の倫理と社会』、pp. 63-85

西堤優 (2011) 「身体状態と意思決定」『哲学・科学史論叢』、第十三号、pp. 77-101

Nishitsutsumi, Y. (2011). Hyperbolic Time Discounting and the Will. In *The Proceedings of the Fifth Annual Philosophical Meeting for Young Scholars of BESETO, Peking University*. 251-263.

西堤優 (2014) 「自己制御と誘惑」信原幸弘・太田絃史編『シリーズ新・心の哲学Ⅲ 情動篇』勁草書房

Nussbaum, M. C. (2001). *Upheavals of Thought: The Intelligence of Emotions*. Cambridge: Cambridge University Press.

Osman, M. (2004). An Evaluation of Dual-Process Theories of Reasoning. *Psychonomic Bulletin & Review* 11: 988-1010.

Owens, D. (2002). Epistemic Akrasia. *The Monist* 85: 381-397.

J・ピネル (2005) 『バイオサイコロジー 脳—心と行動の神経科学』佐藤敬他訳、西村書店

プラトン (2014) 『パイドロス』藤沢令夫訳、岩波書店

Prinz, J. (2004). *Gut Reactions: A Perceptual Theory of Emotion*. Oxford: Oxford University Press.

Reber, A. S. (1967). Implicit Learning of Artificial Grammars. *Journal of Verbal Learning and Verbal Behavior* 6: 855-863.

Riberio, B. (2011). Epistemic Akrasia. *International Journal for the Study of Skepticism* 1: 18-25.

Roskies, A. (2005). A Case Study of Neuroethics: the Nature of Moral Judgment. In J. Illes (ed.), *Neuroethics*. Oxford: Oxford University Press. (A・ロスキース「脳神経倫理学のケーススタディー：道徳判断の本性」田口周平訳、『脳神経倫理学』高橋隆雄・糸和彦監訳、篠原出版新社、2008年、所収)

佐伯大輔 (2011) 『価値割引の心理学—動物行動から経済現象まで』昭和堂

Schachter, S., and Singer, J. E. (1962). Cognitive, Social, and Physiological Determinants of Emotional State. *Psychological Review* 69: 379-399.

Schmeichel, B. J., Harmon-Jones, C., and Harmon-Jones, E. (2010). Exercising Self-Control Increases Approach Motivation. *Journal of Personality and Social Psychology* 99(1): 162-173.

Sloman, S. A. (1996). The Empirical Case for Two Systems of Reasoning. *Psychological Bulletin* 119: 3-22.

Smith, E. R., and DeCoster, J. (2000). Dual-Process Models in Social and Cognitive Psychology: Conceptual Integration and Links to Underlying Memory Systems. *Personality and Social Psychology Review* 4: 108-131.

Stanovich, K. E. (2009a). Distinguishing the Reflective, Algorithmic and Autonomous Minds: Is It Time for a Tri-Process Theory? In Evans and Frankish 2009.

Stanovich, K. E. (2009b). *What Intelligence Tests Miss: The Psychology of Rational Thought*. New Haven: Yale University Press.

Stanovich, K. E. (2011). *Rationality and the Reflective Mind*. Oxford: Oxford University Press.

Strack, F., and Deutsch, R. (2004). Reflective and Impulsive Determinants of Social Behavior. *Personality and Social Psychology Review* 8(3): 220-247.

Strawson, G. (1994) . The Impossibility of Moral Responsibility. *Philosophical Studies* 75: 5-24.

Strawson, P. F. (1962). Freedom and Resentment. *Proceedings of the British Academy* 48: 1-25. (P・F・ストローソン「自由と怒り」法野谷俊哉訳『自由と行為の哲学』門脇俊介・野矢茂樹編、春秋社、2010年、所収)

Tice, D. M., Baumeister, R. F., Shmueli, D., and Muraven, M. (2007). Restoring the Self: Positive Affect Helps Improve Self-Regulation Following Ego Depletion. *Journal of Experimental Social Psychology* 43: 379-384.

Tomb, I., Hauser, M., Deldin, P., and Caramazza, A. (2002). Do Somatic Markers Mediate Decisions on the Gambling Task? *Nature Neuroscience* 5 (11): 1103-1104.

von Neumann, J., and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.

渡邊正孝 (2008) 「行動の認知科学」 田中啓治編『認識と行動の脳科学 シリーズ脳科学 2』 東京大学出版会

Watson, G. (2004). Two Faces of Responsibility. In G. Watson, *Agency and Answerability*. Oxford: Oxford University Press.

Williams, B. (1973). Deciding to Believe. In B. Williams, *Problems of the Self*. Cambridge: Cambridge University Press.

Zajonc, R. B. (1984). On the Primacy of Affect. *American Psychologist* 39(2): 117-123.