課題研究報告書

意識に関する研究の調査
ー情報科学の視点からー

北陸先端科学技術大学院大学
情報科学研究科情報科学専攻

渡邊 大吾

2013 年 12 月

課題研究報告書

意識に関する研究の調査
ー情報科学の視点からー

指導教官　島津 明 教授

審査委員主査　島津 明 教授
審査委員　飯田 弘之 教授
審査委員　白井 清昭 准教授

北陸先端科学技術大学院大学
情報科学研究科情報科学専攻

0910951 渡邊 大吾

2013 年 11 月

概要

コンピュータ技術は、ヒトの知能を超えるところまで発達してきている。今のところ我々は、これらのマシンがヒトのように意識を持って考えたり振る舞ったりしているのではなく、コンピュータに埋め込まれたプログラムによって動作していることを知って使っている。一方で、ヒトの意識自体その複雑さから未だに解明されていない。しかし、ここ 10 年の間に多くの哲学者、心理学者および神経科学者たちは、コンピューターモデルを使用して、さらに意識に関する理論をテストし始め、ヒトの意識について取り組み始めている。Gamez によると、このような最近の傾向は、よりインテリジェントなマシン構築に結び付くかもしれないという憶測もあり、マシン意識（Machine consciousness）、人工意識（Artificial consciousness）や合成意識（synthetic consciousness）の研究としても知られるようになりさまざまな角度から研究がすすめられている。そして、Gamez の論文では、以下の４つの異なるクラスにマシン意識（MC：Machine Consciousness）の研究を区別している：（MC1）意識と関連づけられた外見的な振舞を備えたマシンの研究、（MC2）意識と関連付けられた認識的特徴を備えたマシンの研究、（MC3）ヒトの意識の根拠もしくは相関現象であると主張されるアーキテクチャを備えたマシン意識の研究、（MC4）現象的に意識的なマシン意識の研究。この分類は、ヒトの振る舞いの様相を模写するシステムで始まり（MC1）、実際の人工意識を作り出すことを試みるシステム（MC4）へと移行している。「マシン意識の学際的な性質は、哲学、心理学および神経科学からインスピレーションをとり、強い AI や一般的な人工知能の目的の多くを共有するので混乱の源となっている。これらのカテゴリ分けの適用はマシン意識と他のフィールドの関係を明確にする。」とこの論文では主張する。

また、コンピュータを使っているという前提であれば我々は、「マシンが意識を持っている」とは考えないが、この前提が隠されている場合、声や言葉の言い回し、振舞、外観などがヒトに近づけば近づく程、その区別が困難になってくる。チューリングは、1950 年の論文「計算するマシンと知性」の中で、２０００年までには 10^9 程の記憶容量を持ったデジタルコンピュータをうまくプログラムして、平均的な質問者が 5 分間やり取りしてもヒトもしくはマシンであるかを正しく判断できるのは、70%を超える見込みがない（つまり、マシンが 30%以上の割合で人をだませる）と予想した。実際５分間だまされ続けた人もおり、ELIZA やインターネットチャットボット MGONZ は人々を欺いて話し相手がプログラムかもしれないと気付かせなかった。ALICE というプログラムは２００１年のロブナー賞競技会（Loebner Prize competition）において審査委員の一人を欺いた[29]。Alan Turing 生誕１００年にあたる２０１２年の６月には、イギリスの Milton Keynes で開催された競技会で、チャットボット Eugene Goostman は、２９％欺いたという記録を残した。チューリングの予言した年までには間に合わなかったが、彼の予言は現実化しつつあり、ヒトと

区別がつかないマシンの登場も夢ではなくなってきた。このようにチューリングテストは、コンピュータのようなマシンがヒトに近づいてきている事を示すひとつの指標になるかもしれないが、マシンが本当に意識を持っているかどうかを判断することができない。

また、神経科学の分野においては、脳という物理的に単純な細胞の集合で思考、行為、自意識の実現が可能であることは、驚異的なこととされている。 Russell and Norving[29]によると、もし、それが間違いであるならば、これに代わる考えは、「意識は物理的な世界の範囲を超えたところで動作している」とする神秘主義しかない。このような課題に科学的に取り組んできたものとして、認知科学や脳科学等が挙げられるが解明には至ってない。チャーマーズは、この分野で解明された問題はイージープロブレムであって、解決の糸口も得られていないハードプロブレムの存在を主張している。

また、ヒトの意識について取り組み始めている研究が増えてきたのはここ最近のことで、それまでは“意識”の問題が正面から取り上げられることは少なく、その理由の一つに「意識の定義の困難さ」が挙げられている。もし、「意識のようなもの」をマシンに構築できれば計り知れない程の貢献が期待されるが、肝心の意識についての定義も研究者ごとに異なりはっきりしない。

本課題研究の位置付けとしては、（未だに解明されていない）ヒトの意識をマシンで実現しようとする研究者たち、のこれまでの試みを整理することにある。これは、どのようなものをマシンに組み込むことで意識のようなものが実現できるか、といったヒトに近づいたマシン構築の助けになる。意識に関する研究、特に「意識」について情報科学の視点から整理する。

第１章はじめに

１．１　背景及び目的

近年、コンピュータは自動車、家電、情報端末機器など我々の身近なものに入り込み、社会的にも必要不可欠な存在となってきた。コンピュータの能力は、ヒトの知能（Intelligence）に近づきそれを超えるレベルにまで達してきている。いくつかの例を示すと、1997 年にスーパーコンピュータ「ディープブルー」はチェス世界チャンピオンのカスパロフを２勝１敗３引き分けで破った。さらに、２０１３年 第２回電王戦においても、プロ棋士 １勝、コンピュータ３勝、１引き分けとコンピュータが勝利している。IBM の質問応答システムワトソン君においては、２０１１年２月 米国の人気クイズ番組に挑戦し、最高金額を獲得している。このように、ヒトの知能を超えるところまでコンピュータ技術は発達してきている。

今のところ我々は、これらのマシンがヒトのように意識を持って考えたり振る舞ったりしているのではなく、コンピュータに埋め込まれたプログラムによって動作していることを知って使っている。ヒトの意識自体その複雑さから未だに解明されていないが、ここ 10 年の間に多くの哲学者、心理学者および神経科学者たちは、コンピューターモデルを使用して、さらに意識に関する理論をテストし始め、ヒトの意識について取り組み始めている。Gamez[15]によると、このような最近の傾向は、よりインテリジェントなマシンの構築に結び付くかもしれないという憶測もあり、マシン意識（Machine consciousness）、人工意識（Artificial consciousness）や合成意識（synthetic consciousness）の研究としても知られるようになりさまざまな角度から研究がすすめられている。そして、Gamez[15]では、以下の４つの異なるエリアにマシン意識の研究を区別している：(MC1) 意識と関連づけられた外見的な振舞を備えたマシンの研究、(MC2) 意識と関連付けられた認識的特徴を備えたマシンの研究、(MC3) ヒトの意識の根拠もしくは相関現象であると主張されるアーキテクチャを備えたマシン意識の研究、(MC4) 現象的に意識的なマシン意識の研究。この分類は、ヒトの振る舞いの様相を模写するシステムで始まり (MC1)、実際の人工意識を作り出すことを試みるシステム (MC4) へと移行している。「マシン意識の学際的な性質は、哲学、心理学および神経科学からインスピレーションをとり、強い AI や一般的な人工知能の目的の多くを共有するので混乱の源となっている。これらのカテゴリ分けの適用はマシン意識と他のフィールドの関係を明確にする。」とこの論文では主張する。この分類については、４章で後述する。

また、コンピュータを使っているという前提であれば我々は前述のように、「マシンが意識を持っている」とは考えないが、この前提が隠されている場合、声や言葉の言い回し、

振舞、外観などがヒトに近づけば近づく程、その区別が困難になってくる。チューリングは、1950 年の論文「計算するマシンと知性」の中で、２０００年までには 10^9 程の記憶容量を持ったデジタルコンピュータをうまくプログラムして、平均的な質問者が 5 分間やり取りしてもヒトもしくはマシンであるかを正しく判断できるのは、70%を超える見込みがない（つまり、マシンが 30%以上の割合で人をだませる）と予想した。実際５分間だまされ続けた人もおり、ELIZA やインターネットチャットボット MGONZ は人々を欺いて話し相手がプログラムかもしれないと気付かせなかった。ALICE というプログラムは２００１年のロブナー賞競技会（Loebner Prize competition）において審査委員の一人を欺いた[29]。Alan Turing 生誕１００年にあたる２０１２年の６月には、イギリスの Milton Keynes で開催された競技会で、チャットボット Eugene Goostman は、２９％欺いたという記録を残した。チューリングの予言した年までには間に合わなかったが、彼の予言は現実化しつつあり、ヒトと区別がつかないマシンの登場も夢ではなくなってきた。

このようにチューリングテストは、コンピュータのようなマシンがヒトに近づいてきている事を示すひとつの指標になるかもしれないが、マシンが意識を持っているかどうかを判断することができない。その点については、Haikonen[21]も指摘している。さらに、Russell and Norving[29]では Turing[34]を引用して以下のように議論を進めている。

> *ある機械がチューリングテストに合格したとしてもそれは本当に考えているのではなく考えることをまねているにすぎない、と多くの哲学者が主張してきた。*
>
> *この反論もまた Turing によって予見されていた。彼は Geoffrey Jefferson 教授の演説（1949）を引用している：*
>
> *「たまたま記号が降ってくる事によってではなく、感覚された思考と感情に基づいて機械がソネットやコンチェルトを書くまでは、機械が脳に等しい　－　つまり、書くだけでなく書いたことを知っている－　ということに同意できない。」*
>
> *Turing はこれを、意識（Consciousness）に基づく議論　－機械が自分自身の心的状態と行為を自覚しなければならない－　と呼ぶ。*

ここで、Russell and Norving[29]は、意識は重要な主題であると述べ、さらに Jefferson と他の論者の注目点について以下のように述べている。

> *意識が重要な主題である一方で、Jefferson の論点は実は現象学（Phenomenology）、あるいは直接的経験－　機械は本当に感情を持たなければならない－　の研究と関係している。他の論者は志向性、　－つまり機械の信念や欲求、およびその他の概念作用が本当に、現実世界に存在する何かについてのものなのかどうかという問題－　に注目する。*

そして、Turing は以下のように反論すると Russell and Norving[29]は述べている。

> *この反論に対する Turing の返答は興味深い。彼は機械が実際に意識（あるいは現象学または志向性）を持てる理由を提出することもできただろう。その代わりに彼は、この問いは“機械は考えられるか”という問いと同様に定義不良だと主張する。それに、機械向けの基準を人間向けの基準より高くしようと言い張る必要もないだろう。何といっても、日常生活において我々は、他人の内部の心的状態についてのいかなる直接的証拠も持たないのだ。それでもなお Turing は“この点に関する議論を続けるのはやめて、皆が考えているという上品な慣習（polite convention）を持つのが普通だ”と言う。*

このような“上品な慣習”を持つ事がチューリングテストの前提となっている。このためチューリングテストでは本当に意識を持っていることを判断できないが、Haikonen[21]は意識のテスト（Test for Consciousness）として他のテスト方法等も含めて紹介している。（しかし、これらのテストの中にも本当に意識を持っていることを判断することはできない。）参考のため、本稿では意識のテスト方法についての記載のある Haikonen[21]の 9.1 章から 9.3 章についての内容を Appendix B に掲載した。

また、神経科学の分野においては、脳という物理的に単純な細胞の集合で思考、行為、自意識の実現が可能であることは、驚異的なこととされている。 Russell and Norving[29]によると、もし、それが間違いであるならば、これに代わる考えは、「意識は物理的な世界の範囲を超えたところで動作している」とする神秘主義しかない。このような課題に科学的に取り組んできたものとして、認知科学や脳科学等が挙げられるが解明には至ってない。チャーマーズは、この分野で解明された問題はイージープロブレムであって、解決の糸口も得られてないハードプロブレムの存在を主張している。この点については、3 章で取り上げる。また、参考のため脳科学の視点から、澤口の論文[47]「意識とは何か」についてまとめた内容を Appendix A に掲載する。

また、ここ 10 年の間にヒトの意識について取り組み始めている研究が増えてきていると前述したが、川口[46]によるとそれ以前は“意識”の問題が正面から取り上げられることは少なく、その理由の一つに「意識の定義の困難さ」を挙げている。さらに、川口[46]は「より包括的な定義を行うためには、例えば意識がわれわれの認知システムにおいてどのような機能を担っているかを明らかにしない限りむずかしい」と主張している。しかし、近年の脳科学の進歩は目覚ましいものがあり、その成果から脳内の情報処理に学んだ新しい信号処理アルゴリズムや機械学習手法の開発が進み、広く応用されるようになった。例えば、

脳波を使ってコンピュータを操る研究などがある。計測した脳波のデータをコンピュータに取り込んで ICA(Independent component analysis：独立成分分析）を駆使して解析し、カーソルを約 70%の確率で左右の望む方向へ動かすことに成功している[11]。さらには、脳の学習原理の一部について理論的な解明が進んだことを基礎として、感覚運動機能を持ち模倣・学習するロボットの設計、試作、応用が進んできた。Takeno, Inaba and Suzuki[31]によると、ヒトの意識のようなものをコンピュータやロボットなどに構築するという手段は、より一層ヒトに近づいたマシン構築を実現して社会にも貢献でき、こうした研究はヒトの脳のモデル構築にも寄与すると主張している。さらに、その知識は脳の謎を解く貴重な情報源となって新しい AI に基づくマシンの開発がはじまり、世界中のマシンがより高いレベルの発展を遂げるとその論文の中で期待している。さらに、以下のような例を挙げている。

> *例えば、従来の自動車が主として運転者の指示にただ従うように設計されているのに対して、新 AI を搭載した自動車はまるで専属の運転者がいるように機能する。その自動車は、運転者の指示に従いながらも絶えず安全を確保しながら走行できる。例え運転手が運転中に突然意識を失っても周囲の安全を確保しながら運転を続ける。そして、近くの安全なところまで走行し停止することができる。このとき、到着時間を気にしなければ目的地まで安全に走行することすら可能となる。*

このように意識のようなものをマシンに構築できれば計り知れない程の貢献が期待されるが、肝心の意識についての定義も研究者ごとに異なりはっきりしない。本課題研究の位置付けとしては、未だに解明されていないヒトの意識をマシンで実現しようとする研究者たちのこれまでの試みを整理することにある。これは、どのようなものをマシンに組み込むことで意識のようなものが実現できるか、といったヒトに近づいたマシン構築の助けになる。意識に関する研究、特に「意識」というものについて情報科学の視点から整理することを本課題研究では目的とする。

１．２　本稿の構成

本稿の構成は以下の通りである。２章では、辞書にみる意識の意味についてあらためて確認しその内容を整理する。３章では、意識問題の難しさについて整理する。４章では、意識研究の分類について Gamez[15]をとりあげ整理する。５章では、意識に関する理論について文献等をとりあげ整理する。6 章では、本稿のまとめと今後の課題について述べる。

第２章　辞書等にみる意識の意味

表２－１に、いくつかの辞書から抜き出して調べた意識の意味について掲載する。一般的な国語辞書では、大きく分けて「心の状態」、「気づき」、「ある物事に対してもっている見解」、仏教用語としての「自覚的意識」の４つに分かれる。

1	心の状態	あることをしているとき、またはある状態に置かれているとき、それに気づいている心の状態・働き。	学研国語大辞典 学習研究社
		目覚めているときの心の状態。	国語大辞典 言泉 小学館
2	気づき (awareness)	あるものごとや状態に気づくこと。はっきりとそれと知ること。自覚。	学研国語大辞典 学習研究社
		何事か気に留めること。	国語大辞典 言泉 小学館
3	ある物事に対してもっている見解	階級・社会などに対する意識。例:意識の低い…	学研国語大辞典 学習研究社
		ある物事に対してもっている見解、感情、思想など、社会的、歴史的な影響を受けて形成される心の内容。	国語大辞典 言泉 小学館
4	自覚的意識 [仏]	六識または八識の一つ。ものごとを見分け、考える心。	学研国語大辞典 学習研究社
		六識、八識の一。	国語大辞典 言泉 小学館
5	覚醒*2、アウェアネス、高次自己意識の 3 層構造[情]	基盤となる覚醒（生物的意識）、覚醒を基礎とするアウェアネス（知覚・運動的意識）、さらに高次自己意識（自己認識の意識）の3層を形成すると考えられる。	認知科学辞典 共立出版

表２－１　辞書にみる意識の意味

１から３までは、我々が日常的に使っているような一般的な意識の意味となっている。一方で、４のように仏教用語など専門的な分野で使われている場合もある。仏教用語として使われる場合、六識や八識等の中の一つという意味になる*１。また、認知科学・心理学、脳科学、情報科学では、意識の意味としては、覚醒*1、アウェアネス*2（知覚、運動的意

識、感知)、自意識（自己意識）の３つの意味で用いられるようである。

注：*1

ウィキペディア「唯識」[39]によると、このような考え方は唯識思想からきている。その説によると、各個人にとっての世界はその個人の表象（イメージ）に過ぎないと主張され、前五識（五感)、意識、末那識、阿頼耶識の八種の「識」を八識説（図２－１[39]）として仮定している。五識はさらに眼識（げんしき：視覚）耳識（じしき：聴覚)、鼻識（びしき：嗅覚)、舌識（ぜつしき：味覚)、身識（しんしき：触覚）に分けられる。これらは知覚に対応する。意識を通して我々は、これらの知覚を経験することができるので、後述する情報処理的にみたアウェアネスに対応していると考えられる。ウィキペディア[39]によると、意識は自覚的意識で「第六意識」と呼ぶことがある。

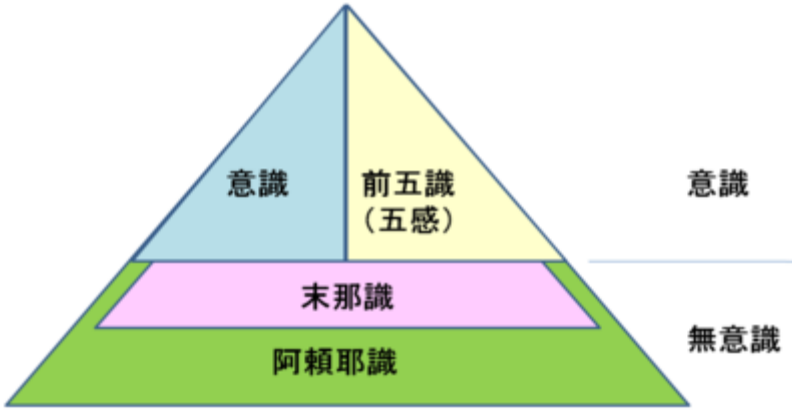


図２－１　八識説の概念図の一例[39]

注：*2

情報処理的にみたとき意識は基盤となる覚醒（生物的意識)、覚醒を基礎とするアウェアネス（知覚・運動的意識)、さらに高次自己意識（自己認識の意識）の３層を形成すると考えられる（図２－２)。[認知科学辞典 共立出版]

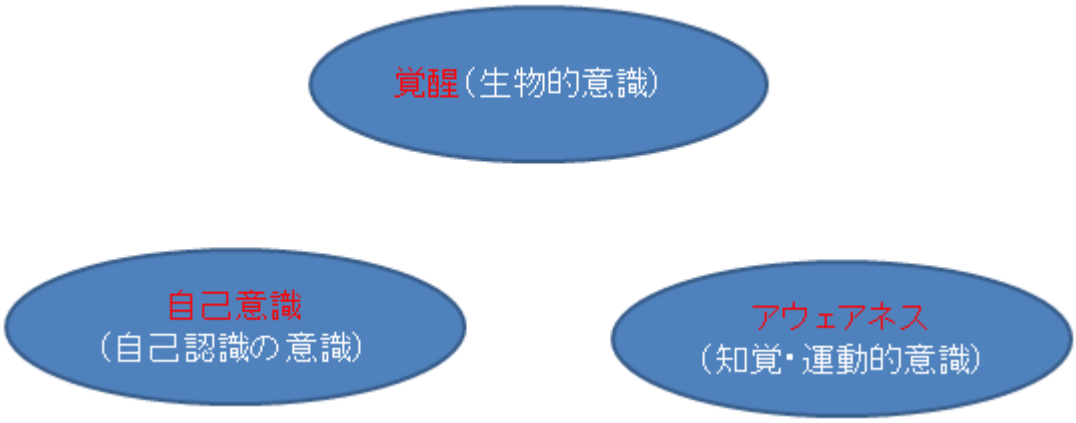


図２－２　意識の３層構造

実際、ヒトには仏教用語でいう自覚的意識、情報処理的にみた自己意識というものを持っており、各個人で主観的に感じることができる。

澤口[47]によると脳科学の分野では、意識を支えるベースとしての意識---覚醒---があるとしている。脳の基本原理は分業にあり、覚醒という脳の基本機能もその分業の一つと澤口により主張されている。それに関わる脳の重要部位は、青斑核（神経細胞集団）と言われ、大脳新皮質を含めた各部位へ神経繊維を延ばし、神経伝達物質ノルアドレナリンを各部位へ運び行動上意味のある情報を処理する状態になる 。青斑核はノルアドレナリン系で、注意に関係していると言われ、覚醒している状態で自身を意識しつつ、行動・思考・感情のコントロールなどをしていると考えられている。

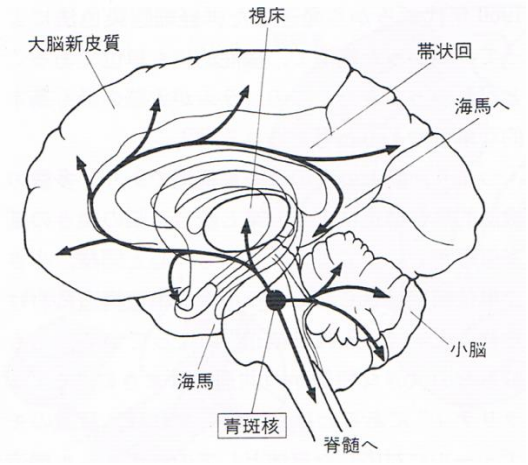


図２－３　青斑核とノルアドレナリン系[47]

意識の３つの意味については、認知心理学研究の川口[46]からも同様の見解が示されている。（川口は日本認知科学会（運営委員）を務めており、**認知科学辞典はこの学会による編集であるためと思われる。ただし、認知科学辞典「意識」は、苧阪直行担当。**）ここで、川口[46]p137 からの「意識」という言葉の意味を整理した内容を以下に掲載する。

覚醒	**目覚めているかどうかを指す**
アウェアネス（感知）	**外界の事象に対する“気づき”**
	内部の記憶情報に対する“気づき”
自意識	**情報のコントロール機能としての“意識”（意図的注意）**

表２－２　川口[46]による「意識」という言葉の意味

> *…ここで簡単に“意識”という言葉の意味を整理しておこう。その一つは覚醒である。“夜寝ているときには意識がなく、昼間おきているときには意識がある”という表現においては、“意識”は目覚めているかどうかを指している。第２に“意識”には“感知（awareness）”、つまり“気づく”という意味がある。例えば、隣の人の話し声を意識するとか、昔の出来事を思い出して意識している、といった場合に用いられる。つまり、外界の事象に対する“気づき” と内部の記憶情報に対する“気づき”の２種類がある。第３に“意識”は自意識の意味でも用いられる。これは、もっとも高次な意味の“意識”である。ある行動や思考を“気づく”かどうかは説明できても、なぜその行動を行おうとしたかという意図を説明する事は容易ではない。さらに情報のコントロール機能として“意識”をとりあげることもある。例えば、“意識的注意を向けて処理を行う”といった表現は特定の処理を意図的にコントロールしていることを意味している。*

また、ご参考のため下記に示す２つの英英辞典：

・The Oxford English Dictionary, second edition Clarendon press/Oxford.

・Webster's Third New International Dictionary, Merriam Webster.

で「consciousness」を調べた内容を表２－３に掲載する。前述した覚醒、アウェアネス、自意識については、網羅されていると考える。ここでは、英英辞典での内容と日本語訳を示し、キーワードを追記した。

The Oxford English Dictionary, second edition Clarendon press/Oxford.			キーワード
1	Joint or mutual konowledge.	結合・相互の知識。	結合・相互意識
2	Internal knowledge or conviction; knowledge as to which one has the testimony within oneself; esp. of one's own innocence, guilt, deficiencies,etc. one's own innocence, guilt, deficiencies,etc.	内部の知識あるいは確信;何に対して自分自身内の証言をしているかというような知識;特に、自分自身の潔白さ、罪、不足などの(知識)	内的知識・確信
3	The state or fact of being mentally conscious or aware of anything.	精神的に意識的であるもしくは何かに気が付いている状態もしくは事実。	気づき
4a	The state or faculty of being conscious, as a condition and concomitant of all thought, feeling, and volition;' the recognition by the thinking subject of its own acts or affections'(Hamilton).	ある条件として、そして、すべての考え、感覚および決断の相伴う物として、意識していることの状態もしくは機能;それ自身の行為あるいは好みの主題を思考することによる認識(ハミルトン)。	思考・感覚・決断を伴う意識の状態・機能、認識
4b	State of consciousness.	意識の状態。	意識の状態
5a	The totality of the impressions, thoughts, and feelings, which make up a person's conscious being.	人の意識の存在から成る、印象、考えおよび感情の全体性。	印象・考え・感情の全体性
5b	Limited by a qualifying epithet to a special field, as the moral or religious consciousness.	モラルや宗教的意識として、特別のフィールドに対して形容語句を与えることによって制限されること。	モラル・宗教的意識
5c	Attributed as a collective faculty to an aggregate of men, a people, etc., so far as they think or feel in common.	人が共通に考えるもしくは感じる限り、人々、民族などの集合体に集合的な機能として起因すると考えられるもの。	集合意識
6	The state of being conscious, regarded as the normal condition of healthy waking life.	健全な目が覚めている生命の通常の状態として見なされた、意識していることの状態。	覚醒

表２－３（１）　英英辞典での”consciouness”の意味：
The Oxford English Dictionary, second edition Clarendon press/Oxford.の場合

Webster's Third New International Dictionary, Merriam Webster.			キーワード
1a	Awareness or perception of an inward psychological or spiritual fact:intuitibly perceived knowledge of something in one's innerself.	心理学的もしくは精神的な内部への事実の意識あるいは知覚：内的自己の直観的に知覚された何かに関する知識。	知覚・知識
1b	Inward awareness of an external object, state or fact.	外部オブジェクト、状態あるいは事実に関する内的意識。	内的意識
1c	Concerned awareness: INTEREST, CONCERN- ofen used with an attributive noun.	関心を持つ意識:興味､関心-しばしば限定名詞と共に使用される。	関心・興味
2	The state or activity that is characterize by sensation, emotion, volition, or thought: mind in the broadest possible sense: sometihing in nature that is distinguished from the physical.	感覚、感情、決断あるいは考えによって特徴をなす状態もしくは活動：可能な限り広い感覚のマインド：物理的なものと識別される自然界におけるもの。	感覚・感情・決断などの状態・活動、マインド
3	The totality in psychology of sensations, perceptions, ideas, attitudes, and feelings of which an individual or group is aware at any given time or within a particular time span - compareSTREAM OF CONSCIOUSNESS.	個人またはグループが何時も、あるいは特別のタイムスパン内に知っている感覚、知覚、考え、姿勢および感情の心理的全体性。	集合意識
4	Waking life (as that to which one returns after sleep, trance, fever) wherein all one's mental powers have returned.	すべてのメンタルパワーが戻ったところの、(睡眠、トランス、興奮状態の後に戻る場合の)目が覚めている生命。	覚醒
5	The part of mental life or psychic content in psychoanalysis that is immediately available ot the ego.	自我に対して直ちに利用可能な精神分析でのメンタルな生命もしくは精神的な内容部分。	メンタルな生命・精神

表２－３（２）　英英辞典での”consciouness”の意味：

Webster's Third New International Dictionary, Merriam Webster.の場合

第３章　意識問題の難しさ

３．１　意識のハードプロブレム

意識は、脳の神経活動をベースとして我々に思考などをもたらしているが、その際に起こる脳内の神経活動（神経発火）のパターンについてヒト自身把握することができない。その代わり感覚や思考という形で、我々は知覚することができる。「神経活動から意識がどのようにしてもたらされるのか？」ということは、意識についての基本的な問題となっている。

ヒトの神経活動は自身で知覚できない代わりに環境や物体のクオリティとして直接発生し、周りの世界や体の感覚で知覚されたこのようなクオリティは、クオリア（Qualia）と呼ばれる内部の主観的（各個人のヒトの）経験から構成される（クオリアは、主観的意識体験とも呼ばれる）。物質としての脳から心的現象やクオリアがどのように生まれるかという問題は、1994 年にオーストラリアの哲学者デイヴィッド・チャーマーズによって、これからの科学・哲学の課題として提起された。この問題は、「意識のハードプロブレム（Hard problem of consciousness：意識の難しい問題）」と呼ばれている（図３－１）。

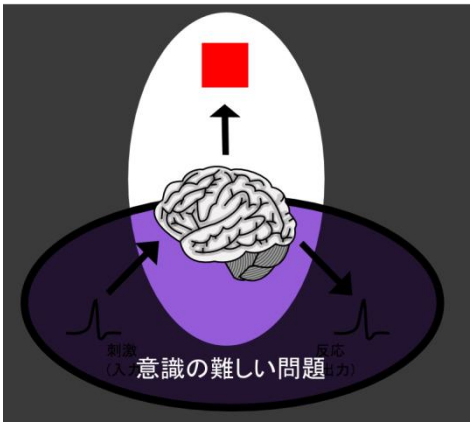


図３－１　意識のハードプロブレム[37]

意識という言葉は使う人によって全く違った意味を持たせることがあり、意識を扱う場合に混乱した状況を招いてしまう。このような状況においては、正しく議論もすることができない。チャーマーズは上述の「ハードプロブレム」という用語を考案し、意識の「イージープロブレム（やさしい問題）」（図３－２）と対比させた。「イージープロブレム」は、物理的に存在するヒトの脳がどのように情報を処理しているのか、という一連の問題のことをいう。20 世紀後半の神経科学の発展によって、意識に関する問題はもう残されていないと神経科学者や認知科学者たちから考えら

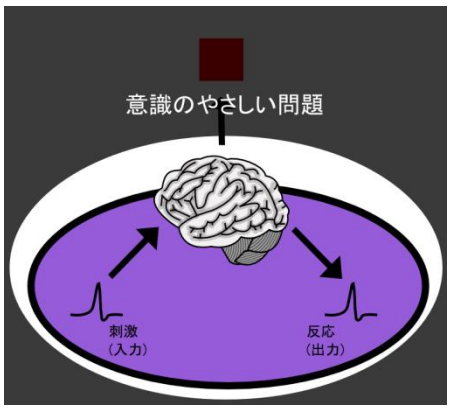


図３－２　意識のイージープロブレム[37]

れていた。しかしながら、彼らの扱っていた問題はイージープロブレムで、ハードプロブレムに関しては議論さえされてないとチャーマーズは主張した。

３．２　クオリア

クオリアは、知覚された色（Perceived color）、音色・音質（Timbre of sound）、食べ物の味、水の感触（wetness of water）、ティッシュの柔らかさ、氷の冷たさ、痛みや喜びの感覚など知覚表象（Percepts）の質（Quality）を表わす。クオリア自体そもそも一体何なのかということは、「意識のハードプロブレム」になっているが、ここでは具体的な例をいくつか挙げて説明する。

３．２．１　TV エンジニアリングでの色の例

TV スクリーン上には、赤、緑、青色の 3 原色からなる発光体から構成され、赤色と緑色の発光で TV スクリーン上では黄色に見える。つまり、ヒトは赤と緑で黄色の知覚（Percept）を得ることができる。そこには赤と緑の波長があるだけで、黄色の波長は存在しない。これは目の色検出方法によるもので、３つの視覚波長帯を検出するレセプタ（受容体）を利用している。この場合、黄色のクオリアは光学上の知覚システムで作られた事になる 。

３．２．２　蝸牛インプラント技術での例

この技術で用いられる蝸牛インプラントは、マイクロフォンを通して音を受信し、音のスペクトラムに応じた聴覚神経を刺激する。これにより音のクオリアの知覚（Perception）が実現し、聴覚神経においては人工的な刺激による聴覚クオリアが可能となる。

一般的にクオリアは実世界のプロパティ（Property）ではない 。しかし、色の場合のようにある単一の可視スペクトラムであるクオリアとして矛盾なく現れる 。つまり、実際のカラーサンプルのような視覚的なものだけでなく、赤、青、黄色のように書かれたラベルなども対応するクオリアのシンボルとなる。そのラベルは、その色を見たときの経験のクオリアを語る事になる。ただし、クオリアはシンボルではなく、脳の中では経験情報の直接的なもので、ヒトは自身のクオリアを知ることはできるが、他のヒトのものは知ることができない。このため「生物的な類似性」や「各個人で同じものに対する経験が異なる」ということは証明できていない。

明確に言えることは、ヒトのような現象的意識についての原理は存在するという

事である。そのことはクオリアベースの主観的内部経験（Qualia-based subjective inner experience）すなわち、神経もしくは電気的活動による内部状況（the internal appearances：意識）の存在要求となっている。いくつかの精巧な方法が発明されなければ、その意識の特徴は被験者自身でしか観測できない。このため現在では、意識の存在は間接的な方法で決定される。「意識に関するテスト」については、Haikonen [21]の 9.1 から 9.3 に記載があり、本稿ではその内容を Appendix B に掲載したので参照されたい。

第４章　マシン意識の研究

マシン意識の領域は、明確に定義されたものではなく、哲学、心理学および神経科学からインスピレーションをとり、強い AI および人工の一般的な知能の目的の多くを共有するので、マシン意識の学際的な性質は多くの場合混乱の源となっていると、Gamez[15]は指摘する。意識に関する科学的な研究はまだ前範例のような状態で、このエリアで調査されている研究の方向性について規定するには早すぎるという Metzinger[26]の意見もある。しかし、Gamez[15]は意識研究の分類を試みており、大きく分けて MC1 から MC4 の４つのクラスに分類し、マシン意識の研究がこのように分離されるならば、マシン意識と他分野との関係はよりクリアになると提案している。

この４つのクラスは、本稿の１章の１．１節で前述したように簡単に述べると以下のようになっている。

・MC1：意識と関連づけられた外見的な振舞を備えたマシンの研究
・MC2：意識と関連付けられた認知的特徴を備えたマシンの研究
・MC3：ヒトの意識の根拠もしくは相関現象であると主張されるアーキテクチャを備えたマシン意識の研究
・MC4：現象的に意識的なマシン意識の研究

以下の章で、それぞれの意識研究について説明する。

４．１　マシン意識のクラス MC1

MC1 は、「一般的な人工知能」に分類される。意識的なヒトの振舞をシステム上で模写することは、マシン意識の研究領域の一つとなっている。振る舞いを生成するには大きなルックアップ表か first-order ロジックを使用することができる。現象的な状態の私たちのただ一つのガイドがシステムの表面上の振る舞いであるので、現象的経験が MC1 マシンに起因すると考えなければならないと Harnad[17]は主張している。このポイントを支持して、Moor[28]は、私たちがそれらを理解するためにそのようなシステムに対してクオリアに帰着する必要があるだろうということを提言している。チューリングテストに取り組む人たちや人工汎用知能(Artificial General Intelligence：人間レベルの知能の実現を目指したもので、他の AI プロジェクトと区別するために AGI と呼ばれている)についての研究は、MC1 の一部であると考えることができると Gamez[15]では述べている。

４．２　マシン意識のクラス MC2

MC2 はマシン意識における主要テーマであり、単純なコンピュータ・プログラムからニューロンに基づいてシミュレートされたシステムまで、種々様々な方法で実行されている。MC2 について、Gamez[15]は以下のように説明している。

> *このエリアにしばしばカバーされる認知の特性は、イマジネーション、感情、グローバルな作業スペースアーキテクチャ(global workspace architecture)およびシステムのボディおよび環境の内部モデルを含む。いくつかのケースで認知状態のモデリングは、MC1 モデルのようなよりリアルな意識的な振舞を目的にしたり、MC3 モデルのように意識に関係したアーキテクチャを使ったりする。しかし、MC2 システムも MC1 または MC3 なしで作り出すことができる。例えば、表面上の振る舞いを持たない感情やイマジネーションのコンピューターモデルなどがこのような場合に当てはまる。また、MC2 と MC4 についても、例えばシミュレーションによる怖れと実際の恐れとは違うことから、これらの研究が関連する必要性はない。*

イマジネーション、感情、自己のように、意識と認知的な特徴との間に多くの関係が存在していると Gamez[15]は述べている。また、Metzinger [26]は、意識的な経験に 11 の制約を適用した。一方で、Aleksander[2]は、意識に最小に必要な 5 つの認知メカニズムを提案している。(５．１.「アレクサンダ公理」 参照)。意識の状態については、Husserl [20]、Heidegger[18]、Merleau-Ponty[27]などの現象学者が述べている。

４．３　マシン意識のクラス MC3

MC3 は「意識と相互に関連されるコンピュータの神経モデルの作成」に分類され、意識に関係する神経理論や認知理論をモデル化しテストする要求から発生する。この領域は、マシン意識の最も特徴的なエリアのうちの１つで、Gamez[15]は：

> *意識に関連したアーキテクチャをベースにしたシステムが意識もしくは意識的な振る舞いの認識の特性をつくりだすために使用される場合、MC3 の研究は MC2 と MC1 にオーバーラップする。さらに、もし意識に関連したアーキテクチャのインプリメンテーションが現象的な状態ができるであろうことが考えられれば、それは MC4 とオーバーラップするだろう。*

と述べ、各領域との重複がみられる。

Baars[8]のグローバル作業スペース（global workspace）、Crick[12]の神経同期（neural synchronization）、Tononi[32] [33]の高度な情報統合を備えたシステム（systems with high

information integration）など多くの研究者がこのエリアに取り組んでいる。しかしながら、マシン中の「意識的な」アーキテクチャのシミュレートは、実際にマシンが意識しているようになるのには十分ではないかもしれないと Gamez[15]は述べている。

４．４　マシン意識のクラス MC4

マシン意識への最初の 3 つのアプローチ（MC1 から MC3）は、現象的な状態に関する要求を伴わない意識にリンクしたものをモデル化しているのであまり論争の的になっておらず、マシン意識の 4 番目のエリア（MC4）は、現象的経験を持つ（意識研究での単なるツールではなく、それら自身意識している）マシンに関係があるので、より哲学的な問題になっていると Gamez[15]は述べている。MC1 から MC3 に関する研究とは無関係に現象的な状態を可能とする生物学のニューロンに基づいたシステムを作り出すことは可能かもしれず、サーモスタットにさえ単純な意識的な状態があるかもしれないことは Chalmers[9]によって主張されている。もし、彼の主張することが正しければ、マシンで実現されている現象的な状態の存在は、それが実行されているより高いレベルの機能の大部分が独立したものになると Gamez[15]は述べている。さらに Gamez[15]は、実際の意識を備えたシステムは現象的な状態を測定およびデバッグする方法なしでは開発することはできず、したがって、MC4 および合成現象学(synthetic phenomenology)の間に密接な関係があると主張している。

合成現象学(synthetic phenomenology)は、マシン意識の研究から現われた新しいエリアで、この用語は、現象的な状態の合成に当てはめるために使用した Jordan[22]によって最初につくられた。さらに、それは合成認識論(synthetic epistemology)と関係があり、Chrisley および Holland (1994, p. 1)[10]によって、天然、人工両方のエージェントがどのように世界を描写するかの説明で生じる哲学的な問題を明確にするための人工システムの創造および分析として定義される。Husserl [19]の現象のプロジェクトはヒトの意識の描写で、合成の現象のプロジェクト(the synthetic phenomenological project)はマシン意識の描写となり、合成現象学は MC4 に取り組んでいるほとんどの研究者に関係があると Gamez[15]は述べている。

４．５　マシン意識研究の分類整理

ここまでマシン意識の研究について、Gamez[15]の MC1-4 の各クラスに関して述べた。さらに、Gamez[15]は、マシン意識の研究について MC1-4 のクラスに分類していくつかの文献を紹介している。本論文では、その内容について表４－１にまとめて整理した。

マシン意識のモデル	文献	MC
Agent-based conscious architecture	Angel, L. (1989). How to build a conscious machine. Boulder, San Francisco & London: Westview Press.	2
Cog	Dennett, D. C. (1997). Consciousness in human and robot minds. In M. Ito, Y. Miyashita, & E. T. Rolls (Eds.), Cognition, computation and consciousness. Oxford: Oxford University Press.	4
	Brooks, R., Breazeal, C., Marjanovic, M., Scassellati, B., & Williamson, M. (1998). The Cog project: Building a humanoid robot. In C.Nehaniv (Ed.), Computation for metaphors, analogy and agents. Lecture notes in artificial intelligence (Vol. 1562). Berlin: Springer-Verlag.	2
Dehaene et al.'s neural simulations of the global workspace	Dehaene, S., Kerszberg, M., & Changeux, J. P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. Proceedings of the National Academy of Sciences of the United States of America, 95, 14529-14534.	2,3
	Dehaene, S., Sergent, C., & Changeux, J.-P. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. Proceedings of the National Academy of Sciences of the United States of America, 100, 8520-8525.	
	Dehaene, S., & Changeux, J. P. (2005). Ongoing spontaneous activity controls access to consciousness: A neuronal model for inattentional blindness. Public Library of Science Biology, 3(5), e141.	
Inner speech	Steels, L. (2001). Language games for autonomous robots. IEEE Intelligent Systems and Their Applications, 16(5), 16-22.	2
	Steels, L. (2003). Language re-entrance and the 'inner voice'. In O. Holland (Ed.), Machine Consciousness. Exeter: Imprint Academic.	

表４－１（１）　マシン意識研究の分類

マシン意識のモデル	文献	MC
Neural schemas（神経の図式）	McCauley, L. (2002). Neural schemas: A mechanism for autonomous action selection and dynamic motivation. In Proceedings of the third WSES neural networks and applications conference, Switzerland.	2,3
IDA(Intelligent Distribution Agent：Global workspace models)	Franklin, S. (2003). IDA: A conscious artifact. In O. Holland (Ed.), Machine consciousness. Exeter: Imprint Academic.	1,2,3,4
CyberChild	Cotterill, R. (2003). CyberChild: A simulation test-bed for consciousness studies. In O. Holland (Ed.), Machine consciousness. Exeter:Imprint Academic.	1,3,4
公理と神経情報表現モデル(Axioms and neural representation modelling)	Aleksander, I., & Dunmall, B. (2003). Axioms and tests for the presence of minimal consciousness in agents. In O. Holland (Ed.), Machine consciousness. Exeter: Imprint Academic.	2,4
	Aleksander, I. (2005). The world in my mind, my mind in the world: Key mechanisms of consciousness in people, animals and machines. Exeter: Imprint Academic.	3
Khepera models（ケペラロボット：汎用ロボット）	Holland, O., & Goodman, R. (2003). Robots with internal models. In O. Holland (Ed.), Machine consciousness. Exeter: Imprint Academic.	2,4
	Ziemke, T., Jirenhed, D. A., & Hesslow, G. (2005). Internal simulation of perception: A minimal neuro-robotic model. Neurocomputing, 68, 85-104.	2

表４－１　マシン意識研究の分類（２）

マシン意識のモデル	文献	MC
A cognitive approach to conscious machines	Haikonen, P. O. (2003). The cognitive approach to conscious machines. Exeter: Imprint Academic.	1,2,3,4
	Haikonen, P. O. (2006). Towards the times of miracles and wonder: A model for a conscious machine. In Proceedings of BICS 2006, Lesbos, Greece.	
Schema-based model of the conscious self	Samsonovich, A. V., & DeJong, K. A. (2005a). Designing a self-aware neuromorphic hybrid. In K. R. Thorisson, H. Vilhjalmsson, & S.Marsela, (Eds.), AAAI-05 workshop on modular construction of human-like intelligence, Pittsburg, PA, AAAI Technical Report WS-05-08 (pp. 71-78). Menlo Park, CA: AAAI Press.	1,2
	Samsonovich, A. V., & DeJong, K. A. (2005b). A general-purpose computational model of the conscious mind. In M. Lovett, C. Schunn, C. Lebiere, & P. Munro (Eds.), Proceedings of the sixth international conference on cognitive modeling ICCM-2004 (pp. 382-383). Mahwah, NJ: Lawrence Erlbaum Associates.	
Shanahan's brain-inspired global workspace models	Shanahan, M. P. (2006). A cognitive architecture that combines internal simulation with a global workspace. Consciousness and Cognition, 15, 433-449.	1,2,3
	Shanahan, M. P. (2008). A spiking neuron model of cortical broadcast and competition. Consciousness and Cognition, 17(1), 288-303, forthcoming.	2,3
CRONOS	CRONOS:Holland, O., Knight, R., & Newcombe, R. (2007). A robot-based approach to machine consciousness. In A. Chella & R. Manzotti (Eds.),Artificial consciousness. Exeter: Imprint Academic.	2,3,4

表４－１　マシン意識研究の分類（３）

４．６　マシン意識研究と他分野との関係

ここまで示したGamez[15]の４つの意識に関する研究のクラス分けのほかに、Searl[30]の強いAIと弱いAI、Franklin[13]の機能的意識と現象的意識、Chalmers[9]のイージープ

ロブレムとハードプロブレムなどがある。これらとの関係について Gamez[15]では以下のように述べている。

> *サールによれば、弱い AI はマインドが働くのと同じ方法で機能する（ヒトが解釈できる）シンボルを使ってマインドをモデリングするプロセスであるのに対し、強い AI は、私がマインドであるという感覚のマインドである何かをつくりだす試みである。これは、フランクリン(2003)によってつくられた現象的・機能的な意識との分類に似ており、さらにチャーマーズ(1996)によるイージープロブレムとハードプロブレムの問題の違いとも関係する。 … MC1 から MC3 がサールの感覚における弱い AI の例となり、このことは MC4 および強い AI の間の合理的に明瞭な写像を示唆する。*

この記述は、サール、フランクリンの主張の類似性、チャーマーズの主張の関係性についてである。ここに出てくるいくつかのキーワード及びここまでの章で説明してきたこととの関係も含めて、図４－１のようにまとめた。

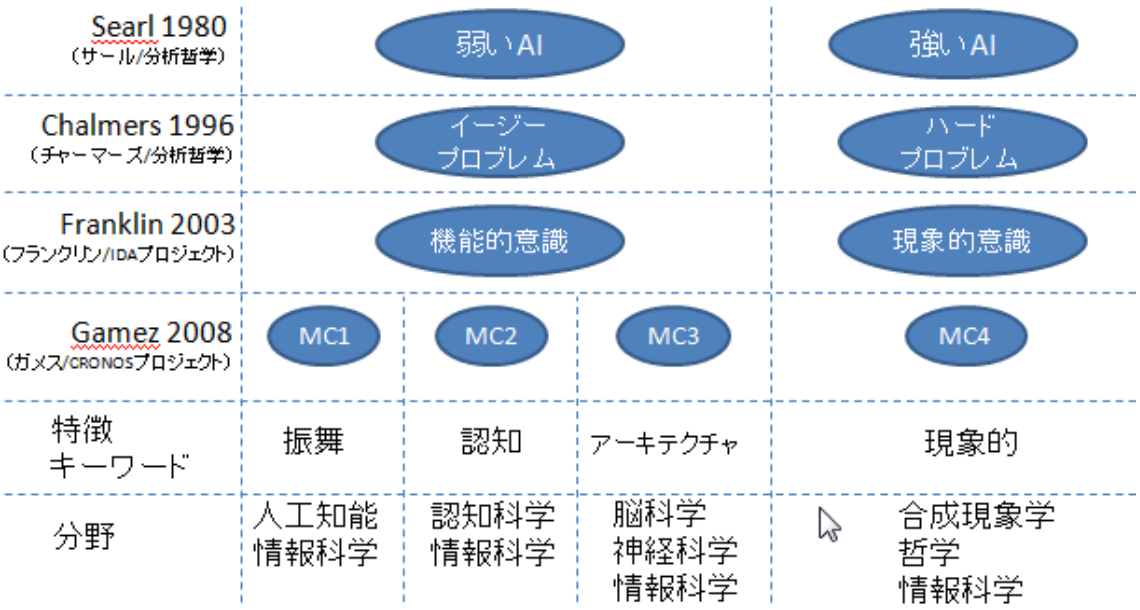


図４－１　マシン意識研究と他分野との関連

サールとチャーマーズは哲学者の立場から、IDA（Intelligent Distribution Agent）プロジェクトのフランクリンはマシン意識を構築するする立場からの主張となる。フランクリンは、Bernard Baars の GWT(Global Workspace Theory：1988 年、1997 年）に定義された意識の機能の一部を備えた場合に、機能的意識を持っていると定義した。彼の生み出した IDA（Intelligent Distributed Agent）は GWT のソフトウェアによる実装であり、

その定義により機能的意識を備えている。ガメスは、2008 に CRONOS プロジェクトのホランド(Holland)の指導でドクターを取得して彼自身もこのプロジェクトに参加している。表４－１に MC1-4 に関する文献を示したが、MC4 に関する文献については、可能性を示唆したものになっており、MC4 のマシンは未だに実現されていない。

第５章　意識に関する定義と理論

意識に関する定義について、アレクサンダ公理と ConsScale が Haikonen[21]によって紹介されている。この章では、アレクサンダ公理についてこの文献の中で紹介されている内容と、ConsScale については他のいくつかの文献も交え整理する。さらに、意識の理論について、関係する文献などを提唱者、その専門とする分野を示して整理する。

５．１　アレクサンダ公理（Aleksander’s Axioms）

Aleksander と Dunmall[1]は、意識的な生物もしくは非生物についての具体的な必須条件の最小限のセットを定義した、５つの公理を提案している。

１．描写（Depiction）

エージェントは、外部の世界とエージェント自身を描写する内的知覚状態を持たなければならない。

２．イマジネーション

エージェントは、公理１のスタイルで呼び出されたかもしくはイメージされたものを描写する内的状態を持たなければならない。

３．注意（Attention）

エージェントは、与えられた時間でこの世界もしくはエージェント自身がどのパートを内的に描写するかを選ばなければならない。

４．プランニング

エージェントは、将来のアクションをイメージすることができ、つくりだされた選択肢の中から選択できなければならない。

５．感情

エージェントはプランされたアクションを評価し、実行されるアクションの選択を決定する感情的状態を持たなければならない。

明らかに通常ヒトはこれらの要求を満足し、意識的でもある。最初の要求の描写について、内的な知覚描写がなければ認識はおこらないので最も重要にみえる。しかしながら、イマジネーション、注意、プランニング、感情などの制限された機能を伴うヒトは聡明ではないとはいえ、意識的なヒトとしてカウントされる。一方で、本当の意識を持ち合わせ

ていないが、認識マシンはこれらのキャパシティを持っているようにみえる。これは、内的描写の性質とスタイルに依存する。このようにアレキサンダー公理は、意識のための必須条件について役立つリストとして考えられることができる。しかしながら、これらの必須条件を満足してもシステムに意識があることを保証するものではない。

５．２　The ConsScale

Arrabales,Ledezma, and Sanchis [4][5][6][7]は、意識は非常に弱い微かな最小の意識からスタートして超意識（Super-conscious）の状態まで拡張する継続的な現象だと図５－１のように提案している。Arrabales らによると、機能的意識レベルの評価のために生物学的にインスパイアされたスケールを、インプリメンテーションされるマシン意識の指標のために作成した。このスケールは ConsScale と呼ばれ、認知スキルのために特別な依存関係に基づくものとなっている。マシンの意識レベルを計算するツールも http://www.consscale.com/で公開されている。この計算ツールでは、チェック項目があってそれをチェックしていくと表５－１の CQS（CQS:ConsScale Quantitative Score）が計算されマシンの意識レベル（CLS：ConsScaleLevel)が結果として割り当てられる。その結果からマシンがどの意識レベルまで達しているかを判断することができるようになっている。

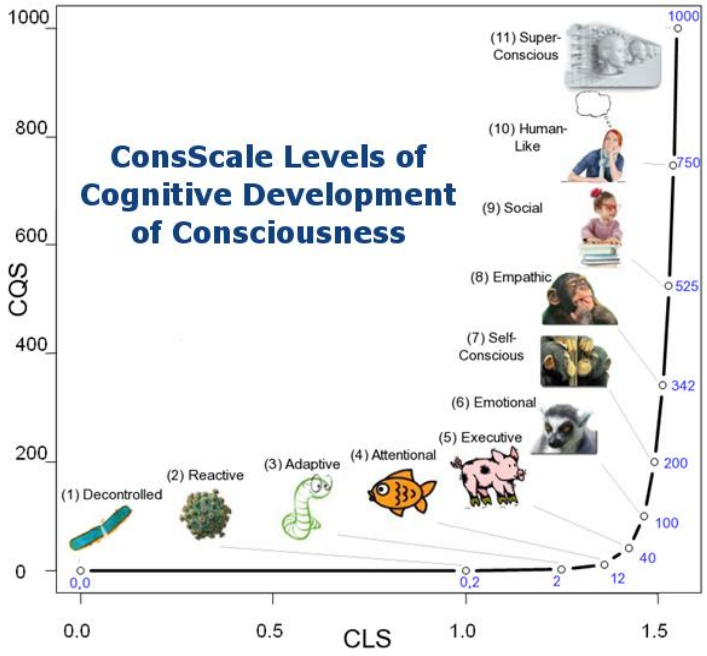


図５－１　人工的な意識レベルの割り当て[44]

Level	Description	CQS
1	*Decontrolled*	0.00
2	*Reactive*	0.18
3	*Adaptive*	2.22
4	*Attentional*	12.21
5	*Executive*	41.23
6	*Emotional*	101.08
7	*Self-Conscious*	200.03
8	*Empathic*	341.45
9	*Social*	524.54
10	*Human-Like*	745.74
11	*Super-Conscious*	1000.00

表５－１　CQS[44]

意識に関して計算機の観点からの研究		ConsScale level
複雑な認識アーキテクチャーの最近の例	Haikonen's cognitive architecture:P. O. A. Haikonen, Robot Brains. Circuits and Systems for ConsciousMachines. UK: John Wiley & Sons, 2007.	6
	LIDA:U. Ramamurthy, B. Baars, S. K. D 'Mello and S. Franklin, "LIDA: A working model of cognition," in Cognitive Modeling, 2006, pp. 244–249.	6
	CRONOS project:O. Holland. A strongly embodied approach to machine consciousness.Journal ofConsciousness Studies 14 pp. 97–110.2007.	4

表５－２　認知アーキテクチャを ConsScale でテストした結果

表５－２は、最近報告されている認知アーキテクチャについて、Arrabales らが公開した ConsScale Level 結果である。図５－２は、ELIZA と表５－２に掲載した CRONOS との比較である。

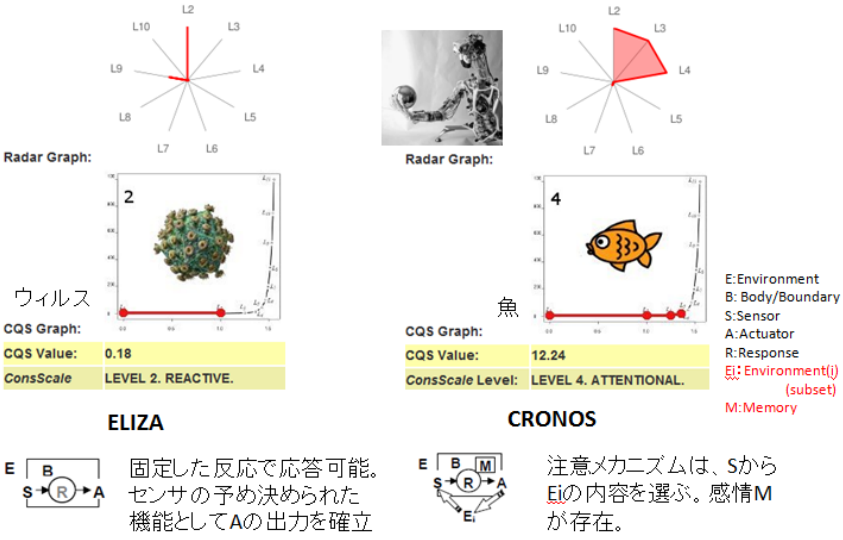


図５－２　ELIZA と CRONOS との比較（一部画像[44]使用）

ELIZA は固定した反応での応答が可能で、この場合刺激に対して反応できるウィルスレベル、一方、CRONOS の場合、部分集合としての環境 Ei の内容をセンス（S）して選ぶことができ生物学的には魚レベルの意識にまで達している。

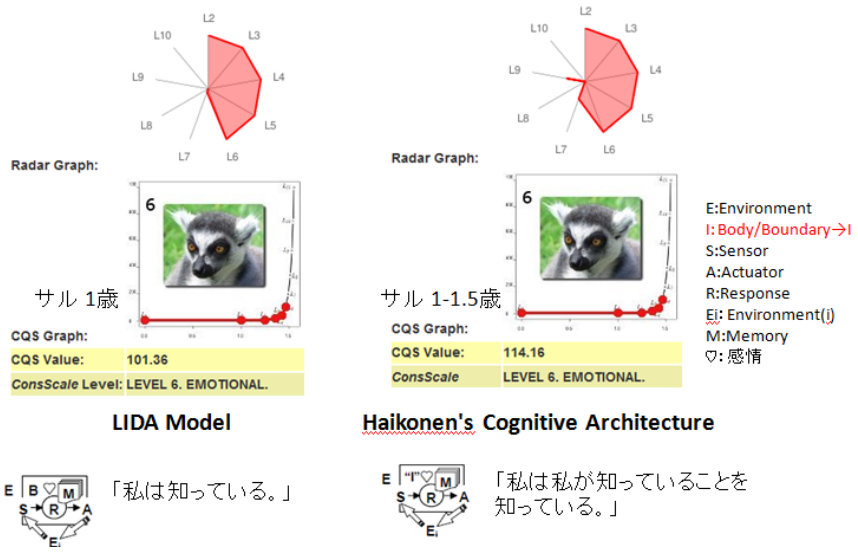
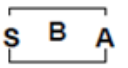
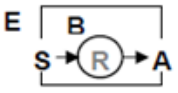
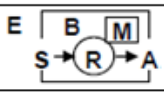
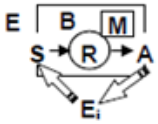
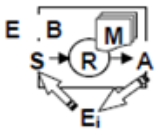


図５－３　LIDA と Haikonen の認知アーキテクチャとの比較（一部画像[44]使用）

図５－３は、前述の IDA の進化型である学習機能を搭載した LIDA（Learning Intelligent Distribution Agent）と、Haikonen の認知アーキテクチャとの比較を示したものである。LIDA と Haikonen の両方とも Level6 で同じであるが、認知レベルとしては、LIDA は「私

は知っている。」というレベルに対して、Haikonen のものは「私は私が知っていることを知っている。」という認知レベルを持ち LIDA よりも高度なレベルにまで達している。

上述のように例を交えて ConsScale について簡単に紹介したが、ConsScale は意識の機能的アスペクトに関係すると考えられ、表５－３のように－１から１１の１３段階に意識が定義される。レベルが上に行くほど、たくさんの能力が増え、リストされる項目も増えていく。これらをベースに ConsScale は、認知マシンの機能的意識の形式的評価用レーティングシステムを提供する。しかしながら、ConsScale はダイレクトな方法で意識の現象的発生をアドレスするものではない。(ヒトの中に)存在すると思われる意識の範囲は ConsScale によって評価できるが、ConsScale での人工エージェントの評価は、評価されるエージェントの認知能力の範囲についてだけとなる。ConsScale テストをパスしても人工エージェントが現象的に意識的だとは検証できない。

マシンの意識レベル		生物学的系統	アーキテクチャ	説明
-1	Disembodied	実態のない状態 (蛋白質の一部としてのアミノ酸)		エージェントの境界(B)は定義されてない。それは環境(E)と混同される場合がある。
0	Isolated	孤立した状態(染色体レベル)		ボディ(B)と環境(E)に明らかに違いがあるが、自立的プロセスは持たない。
1	Decontrolled	コントロールできない (バクテリア)		センサ(S)or/andアクチュエータ(A)を持つ。
2	Reactive	反応できる (ウィルスレベル)		固定した反応で応答可能。センサの予め決められた機能としてAの出力を確立
3	Adaptive	対応可能 (ミミズ:1か月)		アクションは、メモリおよびSによって獲得した現在の情報の両方の動的関数
4	Attention	注意 (魚:5か月)		注意メカニズムは、SからEiの内容を選ぶ。感情Mが存在。
5	Executive	実行的 (哺乳類:9ヶ月)		メモリの中で明示的に表わされる複合的なゴール、Iに対する処理ができる。

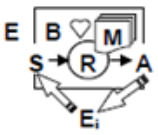
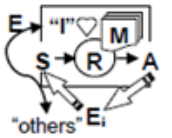
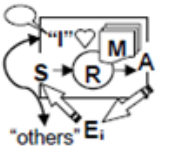
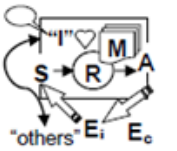
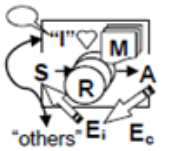
E:Environment　　R:Response

B: Body/Boundary　　Ei：Environment(i)　subset：部分集合

S:Sensor　　M:Memory

A:Actuator　　♡：感情

表５－３　ConsScale レベル（１）[4]

マシンの意識レベル		生物学的系統	アーキテクチャ	説明
6	Emotional	他の人の感情と繋がる (サル:1歳)		感情。ステージ1:「私は知っている。」MIに対する支援(扁桃体)
7	Self-Conscious	自己認識可能 (サル:1.5歳)		ステージ2:「私は私が知っていることを知っている。」MIに対する支援。
8	Empathic	他の人の感情や経験が理解可能 (チンパンジー:2歳)		ステージ3:「私はあなたが知っていることを知っている。」MIに対する支援。
9	Social	社会と繋がる (ヒト:4歳)		ステージ4:「私はあなたが私が知っていることを知っている。」MIに対する支援。
10	Human-Like	ほぼ人間と同じ (Adult:ヒト)		ヒトのような意識。環境Ecや文化に対応可能。
11	Super-Conscious	超意識 (n/a:n/a)		自身の中に意識のいくつかのストリームを持つ

E:Environment　　Ei：Environment(i)　部分集合

I: Body/Boundary→I　　M:Memory

S:Sensor　　♡：感情

A:Actuator　　Others：他者との関係

R:Response　　Ec：Environment(c)

表５－３　ConsScale レベル（２）[4]

５．３　意識に関する理論

意識に関する理論については、Andreae [3]でいくつか取り上げられている。表５－４は、それらの理論について、提唱者、分野、及び文献を追加し整理した内容である。Baars、Crick、Tononi は、「４．３　マシン意識のクラス MC3」で紹介したように主に MC3 の領域で研究をすすめている。特に、Baars の GWT（Global Workspace Theory）は、Franklin、Dehaene、

Shanahan に意識の理論として採用されている。なお、Haikonen Cognitive Architecture (HCA) については、本稿の著者が追加した。

意識に関する理論	提唱者、分野、及び文献
Dennett's Multiple Drafts theory*3	D. C. Dennett：哲学、認知科学
	D. C. Dennett, Consciousness Explained. London, U.K.: The Penguin,1992.
40 Hz Bindiheoryng t*4	F. Crick：分子生物学・生物物理学・神経科学
	F. Crick, The Astonishing Hypothesis. New York: Simon & Schuster,1994
the Global Workspace of Baars*5	B. J. Baars：神経生物学
	B. J. Baars, In the Theater of Consciousness. New York: Oxford Univ. Press, 1997.
the Sensorimotor Contingency theory of visual consciousness*6	O'Regan：実験心理学
	A. Noë：哲学
	J. K. O'Regan and A. Noë, "A sensorimotor account of vision and visual consciousness," Behav. Brain Sci., vol. 24, pp. 939–1031, 2001.
Axioms for Minimal Consciousness*7	I. Aleksander：電気・電子工学
	I. Aleksander and B. Dunmall, "Axioms and tests for the presence of minimal consciousness in agents," in Machine Consciousness, O. Holland, Ed. Exeter: Imprint Academic, 2003, pp. 7–18.

表５－４　意識に関する理論（１）

意識に関する理論	提唱者、分野、及び文献
Reentrant Loop theory*8	G. M. Edelman：1972 年ノーベル生理学賞・医学（抗体分子の構造の発見）
	G. M. Edelman, Wider Than the Sky: The Phenomenal Gift of Consciousness.New Haven, CT: Yale Univ. Press, 2004.
the higher order thought (HOT) theory*9	D. M. Rosenthal：哲学
	D. M. Rosenthal, Consciousness and Mind. New York: Oxford Univ.Press, 2005.
the Conscious Reasoning theory*10	P. Johnson-Laird：認知心理学
	P. Johnson-Laird, How We Reason. New York: Oxford Univ. Press,2006.
Information Integration theory*11	G. Tononi：神経科学、精神病医
	G. Tononi, “The information integration theory of consciousness,”in The Blackwell Companion to Consciousness, M. Velmans and S.Schneider, Eds. Oxford, U.K.: Blackwell, 2007, pp. 287-299.
Strange Loop theory*12	D. Hofstadter：認知科学
	D. Hofstadter, I Am a Strange Loop. New York: Basic Books, 2007.
Haikonen Cognitive Architecture（HCA）*13	Pentti O Haikonen：電子工学
	Pentti O. Haikonen Consciousness and robot sentience, World Scientific, 2012

表５－４　意識に関する理論（２）

注：*3

意識をつかさどる中央処理装置「カルテジアン劇場」(Cartesian Theater)の存在を否定し、それに代わるものとして意識の「多元的草稿理論」(Multiple Drafts Theory)モデルを提唱している。意識とは「カルテジアン劇場」のような中央処理装置をもたない、空間的・時間的に並列した複数のプロセスから織り出され構成されるものだとデネットは論じる（意識のパンデモニアム・モデル）。以上のようなプロセスを経て構成される意識を、デネットは「物語的重力の中心」(Center of

Narrative Gravity)と呼んでいる。デネットは、人間の思考プロセスはコンピュータ（ジョン・フォン・ノイマン・マシン）によってシミュレートすることが原理的に可能なものだと考える。したがって彼はチューリング・テストの意義を認めている。（ウィキペディア「ダニエル・デネット、多元的草稿モデル(Multiple Drafts Model)とカルテジアン劇場批判の項」[38]より）

注：*4

意識的なプロセスの鍵は、40 ヘルツ(35～75Hz)のまわりの頻度で皮質に生じる同期されたニューロンの発振にあるとする、という説である。（ウェブサイト「THE BRAIN FROM TOP TO BOTTOM」WHAT IS CONSCIOUSNESS? [45]より一部抜粋）

注：*5

GWT に関して考える最も簡単な方法は、「劇場隠喩」である。「意識の劇場」では、「選択的注意のスポットライト」は、ステージに明るいスポットを光らせる。意識の内容、登場したり掃けたりしてスピーチあるいは互いとの対話をする俳優たち、を示す。聴衆はライトアップされない --- 暗闇の中（例えば無意識で）それは演劇を観賞している。陰で、しかも暗闇の中で、舞台係、台本作者、シーンデザイナなどそういったものが存在する。それらは、明るいスポットの中の目に見える活動を形作るが、自らは姿を現さない。「誰か」が劇を観賞する暗黙の二元的な仮定に基づかず、そして、マインドの一つの場所に位置しないので(2005 年 Blackmore)、これはデカルトの劇場の概念とは異なると Baars は主張する。（ウィキペディア「Global Workspace Theory」の「The theater metaphor」の項[40]より）

注：*6

見ることが行動する手段であることを提案するもので、周囲を調査する特別の方法であると主張。外部の世界はそれ自身のものとして役に立ち、私たちが感覚運動の偶発性の準拠法と呼ぶものをマスターする場合、見た経験が生じる。このアプローチの利点は、それが視覚的な意識、および異なる感覚形式の知覚経験の質の差を説明する自然で理に適った方法を提供する。（Regan and A. Noë[35]の Abstract より一部抜粋）

注：*7

本稿「５．１節　アレクサンダ公理」を参照。

注：*8

Edelman はプライマリ(primary)意識と高階層(higher-order consciousness)意識とを区別する。プライマリ意識はあるシーンにメンタルに気づいていることに関係があり、彼は、(ウィリアム・ジェームズの「もっともらしいプレゼント"specious present"」を参照して)「記憶されたプレゼント"remembered present"」のように斬新な言い方している。これをレム睡眠に関連した意識の形式と比較することができるかもしれない。プライマリ意識は、(いつなのかは不明瞭なままだが)人類が行ったように存続の機会を増加させたので、進化しているのだろう。エーデルマンによれば、主要な意識を備えた動物はクオリアを経験するが、意識していることに気づいていない。エーデルマンは、その最も高度に発展した形式中の高階層意識が言語の獲得を要求すると考える。(THE NEWENGLAND JOURNAL OF MEDICINE「Wider Than the Sky: The Phenomenal Gift of Consciousness」Book review より一部抜粋)

注：*9

ローゼンタールは、意識に関する高階層思考(HOT : higher-order-thought)理論で最もよく知られている。彼は、メンタルな状態がその状態に気づいていない場合、意識していないと主張；したがって、メンタルな状態がそれ自身その状態であることに気づいている場合のみ、意識していることを主張している。([41]ウィキペディア「David M. Rosenthal (philosopher)、Higher-order thoughts の項」より一部抜粋)

注：*10

メンタルモデルは伝統的様式で、つまり、モデルの一部はそれぞれが示す各部分に相当するという説。([42]ウィキペディア「Mental models and reasoning の項」より一部抜粋)

注：*11

この論文は、意識は何か、また、それをどのように測定することができるかについての推測を示す。この理論によれば、意識は、情報を統合するシステムのキャパシティに相当する。このクレームは意識の 2 つの重要な現象の特性によって動機づけられる：1 つは分化－非常に多くの意識的な経験の有効性；もう 1 つは統合－個々のそのような経験の統一。この理論は、要素の複合体のΦ価値としてシステムに利用可能な意識の量を測定することができると述べる。Φは、要素の部分集合の情報の最も弱いリンクを横切って統合することができる、有効な情報量となる。1 つの、複雑さは、Φ>0 を備えた要素の部分集合であり、それはより高いΦの部分集合の一部ではない。この理論は、さらに複合体の(それはそれらの中の有効な情報の値によって指定される)要素の情報の関係によって意識の質が決定されると主張。最終的に、特別の意識的な経験はそれぞれ、複合体の要素中の情報の相互作用を仲介する変数に何時も価値によって指定される。(Tononi[32]の「Presentation of the hypothesis」より)

注：*12

専門的には混乱した階層意識(tangled hierarchy consciousness)と呼ばれる奇妙なループ(a strange loop)は、この階層システムを通って単に上方へあるいは下方へ移動することによって、スタートした場所に自身が戻ることを見つけた場合に生じる。奇妙なループは自己言及と逆説を含んでいるかもしれない。([43]ウィキペディア「Strange loop」より一部抜粋)

注：*13

Haikonen アーキテクチャは、グローバルワークスペース (Global Work space) や劇場ステージ（Theater stage）を伴わない並列分散型となる。執行部と注意部の機能は、分散型となる。これらのモジュールは、それぞれほぼ同様なもので知覚/応答フィードバックループ原理に基づく。各センサ形式は、膨大な並列信号をハンドルする自身知覚/応答フィードバックループを持つ。(Haikonen[21] 20.4.2 より一部抜粋)

第６章　おわりに

６．１　まとめ

本稿でも述べたとおり、意識という言葉は、使われる分野やその状況や使う人によって全く異なった意味を持つ。既に紹介したように、先人たちはそれを整理するためにさまざまな分類を試みた。サールの「強い AI と弱い AI」、チャーマーズの「ハードプロブレムとイージープロブレム」、フランクリンの「現象的意識と機能的意識」、そしてさらに細分化されたガメスの MC1 から MC4 等である。意識について哲学者の主張からだけではなく、実際にマシン意識を構築する研究者からの主張も説得力をもって世の中に出てくるようになってきている。これは、実際にものづくりをして実証している強みであろう。さらに、ConsScale は構築されたマシン意識の段階を定義したもので､その開発段階の指標が提示されている。マシン意識で主に使われている理論としては本稿でも紹介した、神経生物学の Baars が提唱した Global Workspace であり、ConsScale のレーティングにおいては、この理論を使った IDA の進化型である学習機能を搭載した LIDA が好成績を挙げている（IDA、LIDA とも本稿にて既に紹介）。さらにそれを上回るのが独自の理論を主張している、本稿でも紹介した Haikonen cognitive architecture である。彼の研究においては、MC1-4 の研究の複合型で MC4 を見据えた研究を行っている。マシンが 30%以上の割合で人をだませるというチューリングの予言が達成されつつある現在では、MC1-3 の成果をベースに MC4 マシンへの挑戦が行われている。

６．２　今後の課題

３章で取り上げたチャーマーズの指摘のように、意識に関する研究には「ハードプロブレム」と「イージープロブレム」の二種類に分けられる。後者の中にも未解決な問題はたくさんあるが、脳科学や神経科学などで研究されている脳内にある何らかの神経回路、シナプスの状態、化学物質の状態などが、結果として思考、記憶、判断といった心的機能を可能としているであろうことは基本的に疑いの余地はなく、こうした問題を科学的に研究するにあたって方法論的な困難はない。一方で「ハードプロブレム」に関するような主観的意識、クオリアと呼ばれる内的経験については、Appendix B で意識に関するテストをいくつか紹介したが、その（クオリアと呼ばれる内的経験）存在の有無を確実に確認する手段が存在しない。この点は、「意識」を科学的に解明しようとする際の核心になるものである。この点についてチャーマーズは以下のように述べている。「意識の科学の核心にあるのは、一人称の視点を理解することだ。科学の視点から世界を見るときには三人称の視点を使う。」三人称の世界から一人称の視点を理解することは可能なのであろうか？この点が、

意識研究を難しくしている点で、主観的意識にどう近づいて（どう定義し）、どのように意識を実現できるかが今後の課題となる。

謝辞

本課題研究を進めるにあたり、親身にご支援、ご指導を頂いた島津明教授に深くお礼申し上げます。中間審査において貴重なご意見を頂いた飯田弘之教授、白井清昭准教授に厚く感謝致します。また、JAIST の各関係スタッフの方々には、何かと大変お世話になりました。この場を借りて御礼申し上げます。

参考文献

[1] Aleksander, I. and Dunmall, B. Axioms and Tests for the Presence of Minimal Consciousness in Agents, in O. Holland(ed.), Machine Consciousness(Imprint Academic), pp.7-18　2003.

[2] Aleksander, I. (2005). The world in my mind, my mind in the world: Key mechanisms of consciousness in people, animals and machines.
Exeter: Imprint Academic.

[3] John H. Andreae: A Multiple Context Brain for Experiments With Robot Consciousnes, IEEE 2011.

[4] R. Arrabales, A. Ledezma and A. Sanchis, “Criteria for consciousness in artificial intelligent agents, “ in *ALAMAS&ALAg Workshop at AAMAS 2008, 2008, pp. 57-64*

[5] R. Arrabales, A. Ledezma and A. Sanchis, "ConsScale: A plausible test for machine consciouness? “ in Proceedings of the *Nokia Workshop on Machine Consciousness. 2008, pp. 49-57.*

[6] Arrabales, R. Ledezma, A. Sanchis: A. Establishing a roadmap and metrics for conscious machines development, Proc.8th IEEE int. Conf. on Cognitie Informatics(ICCI’09), 2009.

[7] Arrabales,Ledezma, and Sanchis. [02010] The Cognitive Development of Machine Consciousness Implementation, IJMC 2(2),213-225

[8] Baars, B. (1988). A cognitive theory of consciousness. Cambridge: Cambridge University Press.

[9]　Chalmers, D. (1996). The conscious mind. Oxford: Oxford University Press.

[10] Chrisley, R. J., & Holland, A. (1994). Connectionist synthetic epistemology:

Requirements for the development of objectivity. COGS CSRP, 353, 1-21.

[11] A.Cichocki. Riken news No.29 August 2005.
http://www.bsp.brain.riken.jp/~cia/aboutANDRZEJ/RikenNews-05.pdf#search='%E7%8B%AC%E7%AB%8B%E6%88%90%E5%88%86%E5%88%86%E6%9E%90+%E8%84%B3%E7%A7%91%E5%AD%A6'

[12] Crick, F. (1994). The astonishing hypothesis. London: Simon & Schuster.

[13] Franklin, S. (2003). IDA: A conscious artifact. In O. Holland (Ed.), Machine consciousness. Exeter: Imprint Academic.

[14] Gallup, G. Jr., 1970 Chimpanzees: Self-recognition, Science 176, 86-87

[15] David Gamez(2008). Progress in machine consciousness. Consciousness and Cognition 17 (2008) 887-910

[16] Harnad, S. 1992 The Turing Test Is Not a Trick: Turing Indistinguishability Is a Scientific Vriterion, SIGART Bulletin 3(4), 9-10.

[17] Harnad, S. (2003). Can a machine be conscious? How? In O. Holland (Ed.), Machine consciousness. Exeter: Imprint Academic.

[18] Heidegger, M. (1995) Being and time (J. Macquarrie & E. Robinson, Trans.). Oxford: Blackwell.

[19] Husserl, E. (1960) Cartesian meditations (D. Cairns, Trans.). The Hague: Nijhoff.

[20] Husserl, E. (1964). The phenomenology of internal time-consciousness. The Hague: Nijhoff.

[21] Pentti O. Haikonen Consciousness and robot sentience, World Scientific, 2012.

[22] Jordan, J. S. (1998). Synthetic phenomenology? Perhaps, but not via information processing. Talk given at the Max Planck Institute for Psychological Research, Munich, Germany.

[23] Koch, Ch. And Tononi G.(2011a) A test for consciousness, scientific American 304(6), 26-29.

[24] Koch, Ch. And Tononi G.(2011b) Testing for consciousness in machines, scientific American mind 22(4), 16-7.

[25] Lewis, M., Sullivan, M. W., Staner, C. and Weis, M.[1989] Self-development and self-conscious emotions, Child Development 60(1), 146-156.

[26] Metzinger, T. (2003). Being no one. Cambridge Massachusetts: The MIT Press.

[27] Merleau-Ponty, M. (1995) The visible and the invisible (A. Lingis, Trans.). In C. Lefort (Ed.), Evanston: Northwestern University Press.

[28] Moor, J. H. (1988). Testing robots for qualia. In H. R. Otto & J. A. Tuedio (Eds.), Perspectives on mind. Dordrecht/Boston/Lancaster/
Tokyo: D. Reidel Publishing Company.

[29] Stuart J. Russell, Peter Norving： エージェントアプローチ 人工知能 第 2 版 共立出版、2003. [Stuart J. Russell, Peter Norving: Artificial Intelligence A Modern Approach Pearson Education (US) 2nd International, 1998.]

[30] Searle, J. R. (1980). Minds, brains and programs. Behavioral and Brain Sciences, 3, 417–457.

[31] Takeno, J., Inaba K. and Suzuki T. 2005 "Experiments ad examination of mirror image cognition using a small robot", in Proc. 6th IEEE International Symposium on Computational Intelligence in Robotics and Automation(CIRA2005)、pp. 493-498.

[32] Tononi, G., 2004 An information Integration Theory of Consciousness, BMC Neuroscience 2004, 5:42.doi:10.1186/1471-2202-5-422

[33] Tononi, G. 2008 Consciousness as Integrated Information: Aprovisional Manifesto, Biological Bulletin 215(3),216-142

[34] Turing, A.M. 1950 Computing Machinery and Intelligence, Mind LIX no 2236 433-460.

[35] J. K. O'Regan and A. Noë, "A sensorimotor account of vision and visual consciousness," Behav. Brain Sci., vol. 24, pp. 939–1031, 2001.

[36] THE NEWENGLAND JOURNAL OF MEDICINE「Wider Than the Sky: The Phenomenal Gift of Consciousness」Book review :
http://www.nejm.org/doi/full/10.1056/NEJM200504213521626

[37] ウィキペディア「意識のハードプロブレム」:
http://ja.wikipedia.org/wiki/%E6%84%8F%E8%AD%98%E3%81%AE%E3%83%8F%E3%83%BC%E3%83%89%E3%83%BB%E3%83%97%E3%83%AD%E3%83%96%E3%83%AC%E3%83%A0

[38] ウィキペディア「ダニエル・デネット、多元的草稿モデル(Multiple Drafts Model)とカルテジアン劇場批判の項」:
http://ja.wikipedia.org/wiki/%E3%83%80%E3%83%8B%E3%82%A8%E3%83%AB%E3%83%BB%E3%83%87%E3%83%8D%E3%83%83%E3%83%88

[39] ウィキペディア「唯識」: http://ja.wikipedia.org/wiki/%E5%94%AF%E8%AD%98

[40] ウィキペディア「Global Workspace Theory」:
http://en.wikipedia.org/wiki/Global_Workspace_Theory

[41] ウィキペディア「David M. Rosenthal (philosopher)、Higher-order thoughts の項」:
http://en.wikipedia.org/wiki/David_M._Rosenthal_(philosopher)#Higher-order_thoughts

[42] ウィキペディア「Mental models and reasoning の項」
「http://en.wikipedia.org/wiki/Mental_model#Mental_models_and_reasoning」

[43] ウィキペディア「Strange loop」:
http://en.wikipedia.org/wiki/Strange_loop

[44] ウェブサイト「ConsScale」:

http://www.consscale.com/

[45] ウェブサイト「THE BRAIN FROM TOP TO BOTTOM」WHAT IS CONSCIOUSNESS? :
http://thebrain.mcgill.ca/flash/i/i_12/i_12_p/i_12_p_con/i_12_p_con.html#crickkoch

[46] 川口潤：意識と無意識のはざま：－脳から心へ－、136-144、岩波書店、1995.

[47] 澤口俊之：意識とは何か：－脳から心へ－、126-135、岩波書店、1995.

Ａｐｐｅｎｄｉｘ Ａ　「意識とは何か」

脳科学者の視点から、澤口の論文[47]「意識とは何か」pp127-132 の内容について、主要な部分をまとめた内容を以下に掲載する。

「現象学の祖　フッサールは“意識とは、なにものかについての意識である”といった。“なにものか”が無限である以上、無限の意識があることになる。しかし、その性質・特徴から、意識はいくつかのカテゴリに分けることができる。」澤口[47]は、このように前置きをしながら、Gardner が提唱した“多重知性論”での知性の分類を持ち出して認知科学の成果を参考にしつつ対象ごとに分ければ、意識はさしあたって言語的意識、空間的意識、論理数学的意識、音楽的意識、身体運動的意識、自己意識のように区分できるだろうとしている。認知科学の分野では、こうした要素的な単位をふつう“モジュール（module)”といい、そうした単位が多数集まっている性質のことを“モジュラリティ（modularity)”とよぶ。こうした考えは Gardner だけではなく、Ornstein : 多重マインド説、Fodor : 心の モジュラリティ説、Minsky : 心の社会説など同様な理論を展開している。この区分以外にも、社会的意識、時間的意識、絵画的意識など図 A－１のようにさまざまな意識が無限に考えられ、意識は要素的意識・心の入れ子構造で現代認知科学が明らかにしたこの点は脳と意識・心との関係を考察する上でも重要であると澤口[47]は述べている。

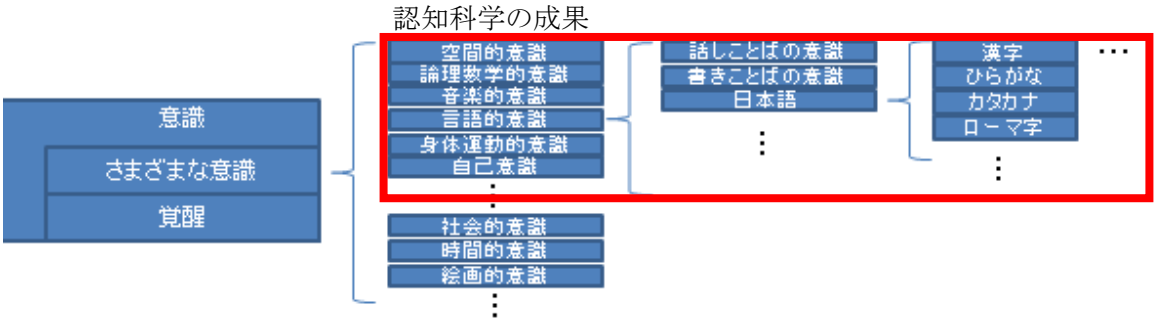
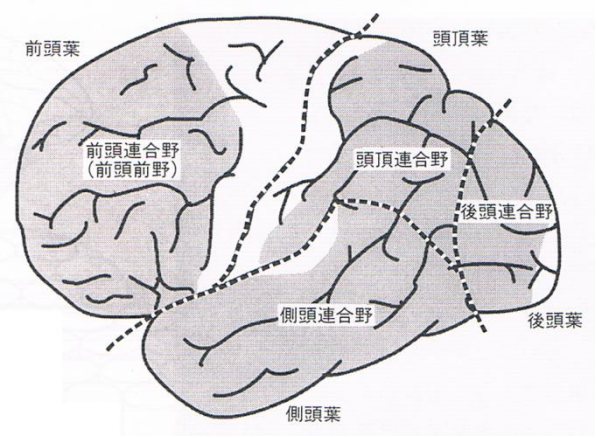


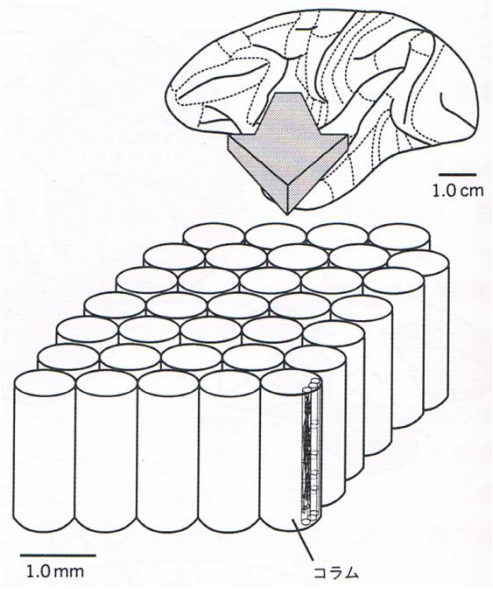
図 A－1　さまざまな意識のカテゴリ分け

大脳は前頭葉、側頭葉、頭頂葉の４つの脳葉から成り、それぞれの脳葉には図 A－２の灰色部に示す連合野（高次な働き）それ以外は第１次感覚・運動野　・連合野の各領野はコラムの集合体（図 A－３）から成る。澤口[47]によると意識とは、連合野コラム群の並列的・

逐次的活動であり、ある特定の意識には一連のコラム群がつくるフレームの活動が対応し、大脳新皮質の多重フレーム説で無限の意識の原理説明ができる（図Ａ－４）。



図Ａ－２　大脳の大まかな区分



図Ａ－３　大脳新皮質のコラム構造

実際、澤口はヒトの意識について下記のように述べている。

> *意識とは連合野コラム群の並列的・逐次的活動なのである。ある特定の意識にはある一連のコラム群がつくるフレームの活動が対応する。フッサールのいう意味での意識は無限にあるとしても、それはこの仮説と矛盾しない。コラムが並列的、かつ逐次的に活動する場合、その組み合わせは事実上無限に近いからだ（26文字のアルファベットを逐次的に並べただけで、膨大な単語ができ、無限の小説が書けることを想起してほしい）。しかも連合野コラムは学習や経験によって変化するはずだ。運動野や体性感覚野（どちらも）非連合野のモジュール構造が手足の運動野そこからの感覚情報の変化によって変わるとうデータはすでにかなりある。連合野はこうした非連合野よりも可塑性（かそせい：変化しやすさ）に富むという証拠は多い。連合野コラムの働きが学習や経験によって変化することはほぼ間違いない。こうした変化を含めた連合野コラムのダイナミックな活動こそが我々意識の実体なのである。*

澤口が脳科学から考察した上述の点は、現代認知科学が明らかにした「意識は要素的意識・心の入れ子構造」である点で一致しており、連合野コラムの変化やコラムの並列的・逐次的活動がヒトの持つ意識に無限性をつくりだしていることを説明している。

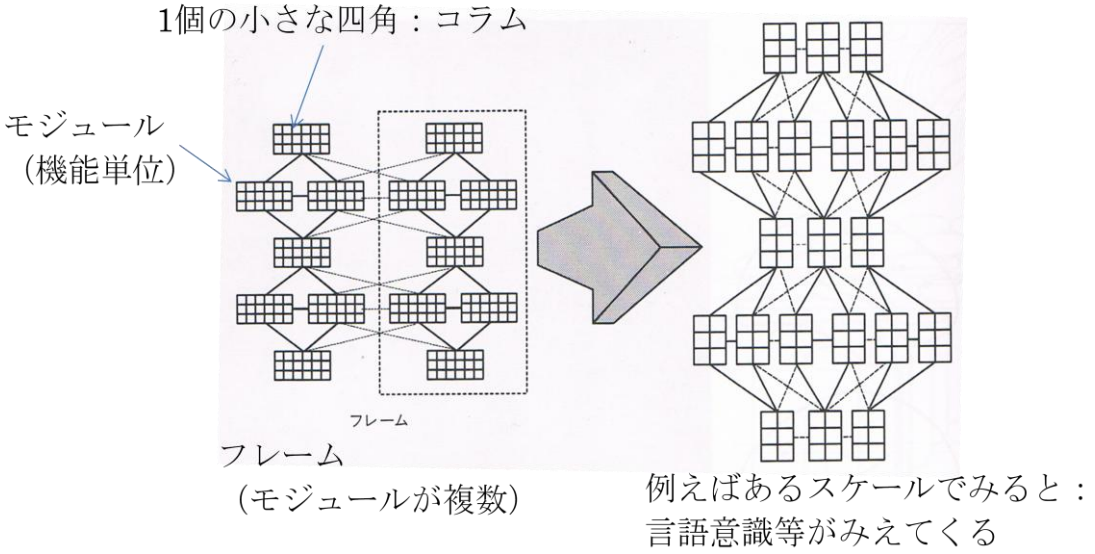


図 A－4　大脳新皮質の多重フレーム説

Ａｐｐｅｎｄｉｘ Ｂ 「意識のテスト方法」

意識のテスト方法について、Haikonen[21]の 9.1 章から 9.3 章について主要な部分をまとめた内容を以下に掲載する。

ヒトについては、意識があることが前提となっている。あきらかに眠っている場合や、麻酔やその他の医学的な条件によって意識を失っていたりした場合、以下のようなテストで意識の存在を確認する。

・被験者は機能的な知覚プロセスを持つか
・被験者は痛みを感じてないか
・被験者は見たり聴いたりできるか、刺激に反応するか
・口頭もしくは身体で何かをレポートできるか
・レポートは気の利いたものになっているか、単なる反射的なものになってないか
・被験者の状況で気づいている範囲は何か？
・被験者は自身の名前、どこにいるか、何曜日かを知っているか？

レスポンスがあれば通常これらのテストは機能し、被験者は多かれ少なかれ 意識を持っていると言える。失敗した場合、限られた状態で意識的かもしれない。例えば、眠っているヒトはこれらのテストに反応することができないかもしれないが、まさに経験している進行中の夢に気づいているかもしれない。あきらかに無意識のヒトはロックされた状態にあり、実際に経験しているけれども刺激に反応したり痛みをレポートしたりすることができない。

前述のようにヒトの場合、意識がある事が前提となっている。それに対し、マシンが意識を持っている事は前提となっていない。このため、ヒトに対して使われる意識テストをマシンに適用したとしても、テストに合格しなければマシンが意識的ではない事は明らかである。一方で、このテストに合格しても、意識が無くても機械的な手段でパスすることができるため、必ずしも意識的なマシンとは言えない。ヒトに対して行うようなダイレクトなテストの代わりに、認知ロボットでは意識の存在や範囲を決定する間接的な方法をとらなければならない。次の章でその方法について、いくつか紹介する。

Ｂ１　マシンに対する意識テスト

Ｂ１．１　チューリングテスト

チューリングテスト[34]は、修正を伴ってマシン意識のためのテストとして時に推奨される。50 年代アランチューリングは、「コンピュータは考えることができるか」という疑問に

ついて、「もし考えることができるのならどのようにして我々はそれを知ることができるか」ということを考えていた。その頃、コンピュータは、躊躇なく脳と結び付けられ、以前から、考えることができるヒトのマインドによって唯一実行されるメンタルアクションを実行する電子脳だった。例えば、条件付き IF-THEN プログラムコマンドは素朴にコンピュータの決定する能力のように見られる。チューリングは、もしテストする人と会話を交わして人として信じさせることができれば、コンピュータは「考える」と言われることを提案した。このテストには欠陥があることは明らかで、なぜなら信じることは真実であることと異なるからである。一方でこの種のだましは、実際の意識の要求つまりクオリアベースの主観的内部経験（Qualia-based subjective inner experience）について何も言っていない。

いわゆる Harnad のトータルチューリングテスト[16]は、オリジナルのチューリングテストの欠陥を直すものとして考えられている。このテストでは、経験的にテスト可能な認知チャレンジ全てについてヒトのように正確に振る舞うことを要求される。このテストの全体性は、人工物がヒトと同じように実際に考え、気づく存在であるということを証明すると考えられている。しかしながら、トータルチューリングテストは論理的に誤った考え（Logical fallacy）をベースにしている。認知的にヒトと等価な人工物はこのテストをパスするということは論理的に正しいが、ヒトと認知的に同じで意識を持たない人工物でもこのテストにパスすることは論理的に可能である。トータルチューリングテストは、テストされる人工物がどのグループに含まれるかを示すものではない。

チューリングテストとトータルチューリングテストは、クオリアベースの内部経験の存在を直接テストするものではない。

Ｂ１．２　Picture understanding test

意識情報は、統合され統一される(Information integration theory) [32][33]。この統一化は、脳の異なる部分からの多くの相互作用から上がってくる。この相互作用がなければ、深い眠りに入ったように意識は消える。意識的であるために被験者は、アクティブに結合された膨大な知識が要求される。これに基づき Koch と Tononi は、絵を理解することをトライすることによってマシンの意識をテストするのにこの条件は使えると提案した Koch and Tononi[23][24]。以下に、具体的に説明する。

まず、二枚の絵を用意する。1 つはコンピュータディスプレイとその前にキーボードが描かれている。もう 1 つは、同じコンピュータディスプレイとその前に花瓶が描かれている。Koch と Tononi によると、マシンビジョンを持った無意識なコンピュータは二つの絵の違いを理解しない。なぜならコンピュータは、絵の中のアイテムの関係について必要なバッ

クグランド情報を呼び起こさないため。このため、コンピュータは本当に何を見ているかを理解しない。それに対してヒトは、膨大なバックグランド情報と意味的にインテグレートする能力によってこの絵を見て直ちに普通でない（つまりコンピュータディスプレイと花瓶の組み合わせはおかしい）ことを理解する。

このテストをパスしたマシンはヒトのように絵を理解する能力と同じものを示すが、このテストではマシンが実際に意識を持っているかを証明することはできない。このテストは、コンピュータのビジョンアルゴリズムと、利用できるバックグランド情報とのインテグレーションに関する問題を扱うことになる。この例の問題については、人工的な神経分類（Artificial neural classification：コンピュータディスプレイとキーボードの膨大な絵を保存しておくこと）によって容易にパスできる。この中からコンピュータディスプレイと花瓶の組み合わせはこのクラスからは目立って例外となる。ここには、意識は含まれていない。このように、このテストはクオリアと内部経験の本質的な問題が無視されているので、意識とはあまり関係ない。

Ｂ１．３　The Cross-Examination Test

ロボットのクオリアベースの内部経験を見つけることができる方法は、ひとつだけ。ロボット自身に尋ねるか尋問する。この方法は、自然言語を使ったり理解したりできないロボットには使えない。これは、動物が何らかの内部経験を持つかどうかを見つけだそうとする場合の問題と同じである。

以下の条件でロボットはメンタルコンテンツを内省できて、それをレポートできると判断できる。つまり、意識を持っていると言える。

１．内的な心象とインナースピーチ（inner speach）を持っていることをレポートできる
２．このレポートが予めプログラムされたものでない
３．ロボット脳の内的アーキテクチャがこれらをサポートする

ある種のスピーチのような形でインナースピーチを持っていることがレポートできれば、ロボットは何らかの意識（internal appearances）やクオリアを持っていると決めるべき。

Cross-Examinaton Test は以下についてフォーカス：

１．マシンは何らかのメンタルコンテンツ（精神内容）のフロー（スムースな流れ）を持ってるか
２．マシンは、メンタルコンテンツ（Percepts、考え、Inner speech、その他）をそれ自身（もしくは他者）にレポートでき、同じオーナーシップ（所有者）を認識できるか。
３．マシンは、いくつかのクオリアを描写できるか

４．マシンは直近の過去を覚えてるか
５．現象的な気づきの”ハンマーテスト（hammer test）：マシンは痛みを感じるか。どのように感じるか。

Cross-examination test はチューリングテストの変形ではないかと考える人もいるかもしれない。チーリングテストでは、テストされる機械類の実際の実現したものは考えられてない。しかしながら、Cross-examination test は、潜在的に意識的なマシンの基本要件を満たすために知られるアーキテクチャのために重要である。

Ｂ２ 自己意識のテスト

Ｂ２．１　自己意識（Self-Consciousness）

自己意識は、ディスクリートな存在としての自身の存在を認識するための主観的能力である。自己意識の実体は、自身の存在に気づき、いくつかの種類のボディーイメージとメンタルな自己イメージを持つ。自己意識は「First person ownership」の以下のような：

・私は自身のボディが自分のものと気づいている
・私の考えは私のもの
・私の記憶は私のもの
・私の決定は私のもの
・私のスピーチは私のもの
・私の行動は私のもの

といった概念も含む。

ボディーイメージは、形、サイズとボディ機能、そしてボディと環境の違いに関係している。ヒトと動物では、自身と環境の境界は通常クリアになっている。ボディは被験者を含むが環境は含まない。ボディは通常脳に接続され、一方で環境は接続されていない。

神経的に接続されたさまざまなボディーセンサのネットワークは、体知覚システム（somatosensory system）と呼ばれる。このシステムは、身体部分の位置(proprioception：自己受容性感覚)、タッチ（taction：接触)、痛み（nociception：侵害受容）や温度をセンスする膨大なレセプターを伴って、いくつかの知覚の形式から構成されている。この体知覚システムは、脳が体のステータスを継続的にモニターすることを可能にする。このシステムによって運ばれた情報は、いくつかのシンプルなルールを導く：①もし痛むなら、それは自分のもの、自分で食べるものではない。②私が感じることができれば、それは私のもの。最初のルールはおかしく見えるが、先天的無感覚症（Congenital insensitivity to pain：CIPA）に見られるもので、おかしくはない。これらの患者は痛みを感じることができず、舌や唇を噛んでしまう。二つ目のルールは例えば、寝相の悪さによる手の感覚異常

（paresthesia）でみられる。この時、被験者は手を異物と認識するようなことがおこる（手が痺れて無感覚になる状況）。

メンタルセルフイメージは、被験者はそれ自身について持っているというアイデアに関連している。私は誰で、自分のことをどのように感じ、何がほしいか、どのようにして他の人をレスペクト（respect）するかなど。

このグループの（体知覚システムに対して物質としての自己に基づくボディーイメージやメンタルセルフイメージを含むという）アイデアは、被験者のセルフコンセプト（self-concept）と呼ぶ事ができる。自己意識の実体は、このアイデアのセルフコンセプトグループ内でたくさんの方法で自身について参照できる。

自己意識についての必要条件は以下にリストできる。意識についての通常の要求に加えて、自己意識を持つエージェントは下記をもつ：

① 体知覚システム（Somatosensory system）
② Body image
③ メンタルセルフイメージ
④ 記憶された個人的ヒストリ（Personal history）

ロボットの自己意識についてのテストは、内省（introspection）と回顧（retrospection）の機能を伴う体知覚システムとメモリシステムと等価な存在の立証からスタートするべきで、これはロボットの認知アーキテクチャの検査によって確認できる。もしこれらがロボットにインプリメント（implement）されていれば、ボディーイメージとメンタル自己イメージの存在と範囲は決められる。ときどき、認知システムのブロックダイアグラムは”body image”と”self-image”のようにラベルされたボックスを含む。これらのボックスはその表示通りにとられるべきではない。Body image とメンタル self-image の存在は、機能的な手段によってテストされるべき。例えば、ボディと自己に関係されるメンタルコンテンツを参照するロボットの能力をチェックするなど。

Ｂ２．２　ミラーテスト（The mirror Test）

いわゆるミラーテストは、赤ん坊や最近では動物などの自己意識（Self-consciousness）をテストするのに使われることもある[14]。テストの被験者（subject）は、ミラーの前につれてかれ被験者の行動が観察される。被験者がミラーの中の自身を認識すると、このテストにパスする。多くの動物や非常に幼い赤ん坊ではミラーの中の自身のイメージを他の誰かと認識して、ミラーの裏側を見ようとする。このようなケースでは明らかにこのテストをパスしない。しかしながら、被験者がミラーの中の自分を認識できているかを直接判

断することが難しいこともあり、このテストの変形版が使われることもある。ルージュテスト（rouge test）では本人に知らされずに赤いドットが被験者の額に付けられ、被験者の前にミラーが置かれる。ミラーイメージの赤いドットが自身の顔に描かれていると被験者自身が認識できれば、このテストはパスしたことになる。18 か月もしくはそれ以上のヒトの赤ん坊は、通常ミラーテストをパスする。これは、社会的に自己の気づき（self-aware）が起こる時期でもある[25]。チンパンジーや象、ハトもこのテストにパスする。厳密にはいくつかのケースで被験者はミラーイメージの自分を認識できることがミラーテストで示されるだけである。これは、被験者がいくつかの種類の自己概念（self-concept）を持っている場合でも失敗する認知タスクである。ミラーテストは、ロボットにも採用された（例：Tononi[32]）。

Ｂ２．３　ネームテスト（The Name Test）

ネームテストは、被験者が自己概念（self-concept）を持っていると仮定している事がベースになっている。このため、これはネームと関係される。小さい子供は自分の名前を早くから学び､自身のスピーチで自身を持ち出すときＩの代わりに名前を使う｡多くの動物は、与えられた名前を学ぶことができる。しかし、それは自身もしくはイベントと関連付けられているかはクリアではない。猫は名前を呼ばれると寄ってくるが、猫のマインドでは食べ物もしくは可愛がられるなど猫自身の名前でないものに実際は関連づけられる。しゃべる能力が無い動物は、スピーチで名前を使うことができず、どんな種類の関係がそこで起きているかは我々には分からない。このため、ネームテストをパスするのに、自己意識の信頼できる表示は必ずしも必要ではない。

聴覚の形式が少なくともしゃべり言葉を認識できるかについて、ネームテストはロボットにも用いられる。もし高度なロボットが自然言語の機能を持っているなら、ロボットが自身の名前を学習し、自身を参照するときに使われるかが観察できる。それから自身の概念についてロボットに尋ねたりすることもでき、このような方法でロボットの自己意識の問題は解決される。

Ｂ２．４　オーナーシップテスト（The Ownership Test）

自己意識（Self-consciousness）を持つ被験者（subject）は、ボディ、考え、メモリ、決定、スピーチ、行動や取得可能な持ち物のオーナーシップの概念を持つべきで、もし自然言語でコミュニケーションがとれないのなら、オーナーシップの概念の存在は間接的な方法で決めなければならない。被験者は、自身がある行動をおこしたことを認識するかを確認する。この決定は、必ずしも難しくないかもしれない。例えば、一匹で遊んでいて家の物を壊した犬のことを考えてみる。その犬は罪（guilt）と恥（shame）を示すようなサインを示すことがある。

Ｂ２．５　反対尋問テスト（The Cross-Examination Test）

このテストは、口語でコミュニケートできるロボットの自己意識の範囲を決定するのに使われることがある。例えば、以下のような問題が考えられる。

・環境とマシン自体の違いをつくりだすことができるか
・マシンは自身の存在に気づいているか。どのように？
・マシンは個人的なヒストリや予測される未来に対して今の状況を拘束するか。
・自己描写；ボディーイメージやメンタルセルフイメージ。どのようにマシンは自身を知覚するか。
・オーナーシップ；マシンは何かしらのオーナーシップを宣言できるか。