

監視制御系におけるオペレータの認知・推論・

決定の支援に関する研究

1999年3月

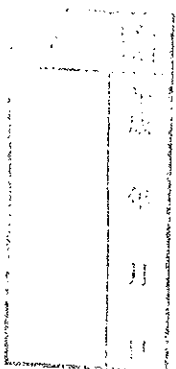
伊藤 誠

1510  
1210  
1310

監視制御系におけるオペレータの認知・推論・  
決定の支援に関する研究

1999年3月

伊藤 誠



99012464

# 目次

## 第1章 緒論

1.1 研究の背景 . . . . .	1
1.2 研究目的 . . . . .	4
1.3 研究の位置づけ . . . . .	6
1.4 本論文の構成 . . . . .	11

## 第2章 情報の解釈と判断の証拠理論的支援

2.1 はじめに . . . . .	13
2.2 不確実環境下での情報の獲得 . . . . .	15
2.2.1 確信度の表現方法 . . . . .	15
2.2.2 情報統合 . . . . .	17
2.2.3 情報更新 . . . . .	20
2.2.4 Dempster の条件付けの一般化 . . . . .	25
2.2.5 情報更新の拡張とその利点 . . . . .	30
2.3 プラント安全制御における合理的状況判断および意思決定 . . . . .	35
2.3.1 問題設定 . . . . .	35
2.3.2 決定者の行動に伴う損失 . . . . .	36
2.3.3 安全制御方策とその下での期待損失 . . . . .	37
2.3.4 情報統合の最適化 . . . . .	38
2.3.5 情報更新の最適化 . . . . .	42

2.4	混合戦略による安全制御での情報の統合と更新 . . . . .	47
2.4.1	問題設定 . . . . .	47
2.4.2	安全制御方策に対する混合戦略 . . . . .	47
2.4.3	混合戦略のもとでの期待損失 . . . . .	48
2.4.4	情報統合の最適化 . . . . .	49
2.4.5	純粋戦略に対する最適規則との同一性 . . . . .	51
2.4.6	情報更新の最適化 . . . . .	54
2.5	情報統合と情報更新との使い分け . . . . .	57
2.5.1	純粋戦略のもとでの情報統合と情報更新 . . . . .	58
2.5.2	混合戦略のもとでの情報統合と情報更新 . . . . .	59
2.6	むすび . . . . .	60
<b>第3章</b>	<b>意思決定と実行の状況適応自動化：必要性の実験的検討</b>	
3.1	はじめに . . . . .	62
3.2	マイクロワールド1 . . . . .	65
3.2.1	プラントの記述 . . . . .	65
3.2.2	ヒューマンインタフェース . . . . .	66
3.2.3	オペレータのタスク . . . . .	66
3.2.4	異常事象と事故 . . . . .	68
3.2.5	自動化システム . . . . .	70
3.3	実験の方法 . . . . .	71
3.3.1	計画と手順 . . . . .	71
3.3.2	評価指標 . . . . .	73
3.4	高レベルの自動化の必要性 . . . . .	74
3.4.1	人の判断の遅れ . . . . .	74
3.4.2	意思決定に利用できる時間余裕の不足 . . . . .	77

3.5	人と自動化システムの不整合	78
3.5.1	モード認識の混乱	78
3.5.2	自動化システムへの不信	79
3.5.3	自動化システムの不適切な使用	80
3.6	むすび	81

#### 第4章 安全確保への状況適応自動化：有効性と実現可能性

4.1	はじめに	82
4.2	マイクロワールド2	83
4.2.1	プラントの記述	83
4.2.2	ヒューマンインタフェース	84
4.2.3	オペレータのタスク	86
4.2.4	異常事象と事故	87
4.2.5	安全制御システム	88
4.3	実験の方法	90
4.3.1	実験計画	90
4.3.2	手順	91
4.3.3	評価指標	92
4.4	タスク遂行に関する被験者の特性	92
4.4.1	プラント運転成績の学習効果	93
4.4.2	異常診断に要する時間と自動化の介入時刻	94
4.5	状況適応自動化による安全確保	97
4.5.1	事故防止への自動化の貢献	97
4.5.2	ヒューマンオペレータの存在意義	98
4.6	trust, self-confidence のモデル	102
4.6.1	分散分析に基づく静的モデル	102

4.6.2	時間遅れを反映した動的モデル . . . . .	105
4.7	意思決定の自動化への許容 . . . . .	111
4.7.1	trust の動的変化 . . . . .	112
4.7.2	高レベルの自動化への trust . . . . .	114
4.7.3	モード認識を阻害しない自動化レベル . . . . .	116
4.8	むすび . . . . .	116
<b>第 5 章 結論</b>		
5.1	本論文の結果と意義 . . . . .	118
5.2	今後の研究課題 . . . . .	121
<b>謝辞</b>		123
<b>参考文献</b>		125

# 第1章 緒論

## 1.1 研究の背景

### 大規模システムの安全性

化学プラント，航空機，あるいは航空交通管制システムといった，産業や交通に関するシステムは，科学技術の進歩に伴って大型化・複雑化が進められてきた．システムの巨大化は，生産にかかるコストの削減や，大型化への純粋な欲求，場合によっては国家の威信などの理由によって必然的に進められたものである [127](pp. 16-20)．

こうした大規模複雑システムは，多くの恩恵を周辺地域や社会に対して与えた反面，事故が発生した場合には甚大な損失をもたらす可能性がある．たとえば，1984年にインドのボパールで発生した，ユニオン・カーバイド社の殺虫剤プラントのガス漏れ事故では，少なくとも2500人が死亡したといわれ，同社が支払いを命ぜられた損害賠償額は4億7千万ドルにのぼる [136]．また，1986年に発生したチェルノブイリ原子力発電所の事故では，直接の損害額だけで29.6～51.3億ドルになるとの試算もある [106]．このようにシステムが大規模であればあるほど，安全の確保は益々重要になる．

機器の品質や信頼性技術の発展に伴い，システムの安全性は格段に向上した．例えば，旅客機の100万飛行回数あたりの事故件数は，1960年には40件以上あったものが，'70年代以降には5件程度にまで減少している [7]．

しかし，事故が完全になくなったわけではない．旅客機事故の場合，1970年代以降，100万飛行回数あたりの事故件数はほとんど減少していない [7]．大規模システムにおける事故の主たる原因は，人間による操作や判断の誤りである．「誤るは人の常」といわれるよう

に、人間はしばしば操作や判断を間違えがちであり、機器の場合と比べ人間の信頼性は劇的には向上しない。実際、事故事例を見ても、システムを運転・管理する人間（オペレータ）が事故の直接の原因に関わった例は多い。旅客機では、過去に発生したすべての事故のうち、60－70%はフライトクルーによるエラーが主たる原因となっている [7]。離陸滑走時の事故に限定すると、約80%の事故は、パイロットが適切な操作や判断、準備をしていれば避けられたはずであるとされている [22]。自動車や航空の分野より比較的安全性の高い鉄道の分野でも、信楽高原鉄道事故 [103] のように、判断の誤りが事故の直接的な原因になりうる。

## 人間による監視制御

ところで、人間が犯す誤り、すなわちヒューマンエラーによる事故を防ぐ手段として直ちに思いつくのは、人間がなすべき操作をなくすることであろう。現実のシステムでも、人間を制御のループから排除し、運転・操作を自動化することが進められてきた。

自動化が進んだ結果、人間が直接何かを操作することは少なくなり、代わりにオペレータの立場は監督者・管理者となった。ここでのオペレータのタスクは、自動化システムやプロセスの状態を監視し、状況に応じて適宜運転計画を立案・変更したり、コンピュータにその計画を教示することである。もちろん、手動操作も時と場合に応じて行うこともある。このような形態によるシステムの制御は「監視制御」 [83] と呼ばれている。監視制御では、オペレータは HIC (Human Interactive Computer) から意思決定に必要な情報を入手し、これに基づいて判断を下し、自然言語に近い形で HIC に命令を与える。HIC は、オペレータから与えられた命令を実現すべく、個々の TIC (Task Interactive Computer) 群に命令を送る。各々の TIC は、HIC から与えられたタスクを遂行する。このように、監視制御系ではオペレータを頂点とする階層的な指揮系統が仮定される。

現在、自動化システムは高度な知能や自律性を有することができるようになっている。それでも、制御の形態は監視制御である。人間は制御系から完全には排除されない。その



理由の一つは、あらゆる事態に対処でき、つねに正しく稼動する完璧な機械やコンピュータが作れるとは考えられないことにある。さらに、つぎのような指摘がある。「安全性を高めようとするシステム全体がより複雑になり、そのことが事故につながりかねない」[105](p. 68),「ある事故から得られた教訓により施した防護策が、別の事故のきっかけとなりうる」[68] (P. 52), またソフトウェアを改善する場合に新たな問題が発生しやすい[46](p. 11)。これらの指摘が意味するのは、完全な安全性を目指して対策を施すほどシステムが複雑化し、信頼性を高く保つことが難しくなるということである。したがって、高度に知能化した自動化システムですら、最終的には人間の柔軟な対応能力に頼らざるを得ない(自動化の皮肉 [1])。

このように、機械やコンピュータの複雑化に伴い、システム安全に対する人間の重要性は相対的に大きくなった。Billings[89]が指摘するように、「人間は、人によってはシステムの救世主ともなり、アキレス腱ともなり、またその両方にもなる」ことから、オペレータの行動がシステムに及ぼす影響を反映させて、システム全体の信頼性や安全性を確保する方法を確立しなければならない。

## 人間中心の自動化

自動化の問題点を考える。これまでは、可能であるから自動化する、自動化できるところはすべてそうする、などの理由だけで自動化が進められていたきらいがある(技術中心の自動化[95])。このような自動化により、さまざまな問題点を生じさせたことが知られている[116]。

たとえば、作業負荷が低いところでより低く、高いところでより高くなったという問題がある。オペレータは自動化システムが正しく稼動しているか監視する責務を負うが、定常的な状態では自動化システムが十分うまく機能し、オペレータの介入を必要とするのはまれである。注目すべきことがほとんど発生しない、退屈な時間が長い間継続すると、やる気や士気が低下しやすく、もし異常事態が突如発生したとしても迅速に対応できると

は限らなくなる。逆に、システム状態が過渡的な場合こそ、人間が支援を必要とする。しかし、オペレータは制御に介入しつつプロセスを監視したり、自動化の設定を変更するなどせねばならない。この場合手動制御よりもかえって高い作業負荷がかかることもある。

また、モード認識の混乱・欠如 [76] やオートメーションサプライズ [74, 75] といったことも自動化が進んだ結果見られるようになった。オートメーション・サプライズとは、コンピュータが何をしているのか、なぜそのモードになってしまったかをオペレータが理解できなくなることである。コンピュータがあたかも人間の意図とは独立に、自らの意思に沿って行動しているかのようにオペレータには見える。

こうした、人間とコンピュータとが適切に協調できない問題は、新しいタイプの事故やインシデントとして顕在化するようになった [79, 121]。航空の分野においては、飛行するにはまったく支障のない機体が墜落する事故、いわゆる CFIT (Controlled Flight Into Terrain) [79] の問題などとして知られている。

今日の大規模システムの安全性を確保するには、できることは全て自動化に任せるという、技術中心の設計では不適切である。システムの救世主たる人間を中心に考え、人間がいかなる助けを必要としているかを反映させて自動化する必要があるとの認識が高まっている [4, 5, 6, 95]。すなわち、「人間中心の自動化」の確立が求められている。

## 1.2 研究目的

「人間中心の自動化」が具体的に指す内容や意味は、研究者により解釈がさまざまである [85] (Table 1.1)。具体的にどうすることが人間中心なのかは必ずしも自明でない。

ただし、人間中心の自動化において一つの原則としてほぼ認められているのは、「人間が制御の主体」 [95] であり、「人間が最終決定権を持つべき」である [4, 95] とする原則である。この原則に基づく場合、コンピュータがなすべきことは、機器の操作や運転に関わる意思決定を補助することである。

本論文では、人間が最終決定権を持つことを基本的には認めつつ、システムの目的の一

Table 1.1: Meanings of Human-Centered Automation[85]

- 
- (1) 人間に合う仕事を人間に，コンピュータに合う仕事をコンピュータにさせる
  - (2) 意思決定と制御のループから人間が逸脱しないようにする
  - (3) 人間に最終的な権限を持たせる
  - (4) 人間がする仕事を，簡単で，楽しむことができ，満足できるものにする
  - (5) 人間の能力を最大限に引き出す
  - (6) 人間が自動化システムを適切に信頼できるようにする
  - (7) 人間が必要としている情報を人間に与える
  - (8) ヒューマンエラーを軽減し，事象に与える人間の反応の多様性を最小限にする
  - (9) 人間を，部下たる自動化システムの監督者とする
  - (10) 人間と自動化システムとの組み合わせを，システムの目的に関して最適にする
- 

つである安全確保のために人間とコンピュータとの組合せを最適にすることも視野に入れ，次の二つの観点から人間の意思決定を助けるための理論の構築を目指す。

#### 1. システムの大規模複雑性に伴う状況の不確実性

大規模システムにおいては，センサの物理的制約や情報収集に利用できる時間の不足などから，得られる情報が一般には不確実性・不完全性をともなう。異常時にはとくに意思決定が困難になりやすいにもかかわらず，適切で素早い対処をしないと多大な損害を生じさせることにもなりかねない。以上のことから，動的で不確実な環境において，何かが起こっていることやその原因の把握および近い将来における状況の予測といった，いわゆる状況認識 [20, 21] を確保させる方法を確立する必要がある。ここでは，情報の不完全さを表現するのに適している証拠理論を採用し，適切な意思決定を下すための状況把握に必要な，情報を随時獲得して信念を更新する機構を整備する。

## 2. システムの動特性がもたらす意思決定の時間制約

どのような場合・状況においても「人間が最終決定権を持つべき」との主張を字義どおり安直に受け入れてよいとは限らないとの指摘がある [30]。壊滅的な事態にいたるまでの時間が人間の情報処理能力などに比べ極端に短い場合（たとえば、航空機の離陸時におけるエンジン故障の発生など）では、人間が意思決定に利用できる時間の余裕がほとんどなく、判断自体をコンピュータが代行する方が結果的にはよいとも考えられる。ここでは、意思決定を自動化することの有効性を確かめ、その実現に向けてコンピュータが持つべき性質を明らかにする。

## 1.3 研究の位置づけ

認知科学 [113, 120] は、人間の認知過程をコンピュータの情報処理過程になぞらえてモデル化する。Sheridan[85] は、自動化を認知過程における情報の入力側 (afferent side) と出力側 (efferent side) とに分けて考えることを提案している。afferent side の自動化とは、情報の収集や蓄積・解釈および選択肢の列挙など、状況を把握する部分の自動化のことであり、意思決定の「支援」に当たるものと解釈できる。efferent side の自動化は、行動の選択自体やその実行を自動化するものである。なお、本論文の以下では、特に断らない限り自動化とは efferent side のものを指すこととする。

本論文も Sheridan の考えと同様に、情報の入力側の支援と出力側の支援とに分けて考える。ここでは、前者を不確実性下の意思決定の研究としてとらえ、後者の研究を人間とコンピュータとの役割分担のそれと考え、各分野におけるこれまでの研究を俯瞰することにより本研究の位置づけを明確にする。

## 不確実性の下での意思決定

人間中心の自動化の実現を目指すにあたり、様々な対策が挙げられている [134]。意思決定の支援はその一つであり、不確実環境での情報の表現や認識の更新およびそれらに基づく意思決定の研究は、システム工学 [125] や人工知能 [73] の分野を中心に進められてきている。人工知能の領域における知識・信念の取扱に強い影響をもたらしたのは、Hintikka [102] である。Hintikka は、知識 (knowledge) と信念 (belief) を区別している。すなわち、ある個人  $a$  が正しいと思っていた事柄  $p$  が、新たに情報を獲得することによって誤りであると判明しうる場合、 $a$  は  $p$  が正しいと知っている (know) のではなく、正しいと信じている (believe) にすぎず、この信じている状態が信念 (belief) とよばれる。大規模システムの状態に対する人間の認識は、新たに情報を獲得することによって修正されうることから、信念とみなすことができる。

信念の表現や更新を理論的に扱う試みは、さまざまな視点から行われている。たとえば、古典論理に基づく方法 [52] や、様相論理、とくに認識論理などに基づく方法 [131] などの形式的アプローチがある。

これに対し、確率論や可能性理論、あるいは証拠理論 (Dempster-Shafer 理論) などに基づく定量的アプローチがある [17]。なお、形式的なアプローチと定量的なアプローチとは、互いに無関係ではない。例えば、Murai et al. [61, 62] は、種々のファジィ測度のクラスに関して健全かつ完全な論理体系を示している。

定量的アプローチにおいては、確率論に基づく方法がもっとも基本的である。確率論をベースとした意思決定支援システムとしては、政治軍事的な意思決定のための確率情報システム [19] などがある。確率論的な方法は、人間が合理的に判断すると仮定する場合に正当化される。人の信念が確率測度で表現されるべきであるとする理由の一つは、信念の表現形態が確率測度の公理を満たさない場合、必ず損をする賭を設定することができる (Dutch Book [3, 42, 66]) ことである。また、確率論は豊富な評価・解析手法を提供することから、確率論に基づく意思決定論が整備されており [119, 137]、統計的決定論のプラン

ト安全監視システムへの応用も数多くなされている (たとえば, [34, 112, 115, 117, 118]).

確率論で信念を記述する場合, 信念の更新は普通 Bayes の定理や, 得られる情報が曖昧さを伴う場合には Jeffrey の規則 [42] が用いられる. しかし, 人間の確率評価は Bayes の定理が示すほどには修正されないという保守性 (conservatism) [19] があるなど, 確率測度は人間の信念を記述するには不都合がある [8]. Bayes の定理以外にも,  $A$  が正しいとの情報を入手したときに,  $A$  と矛盾する要素  $w$  の確信度を,  $A$  に関して  $w$  から最も「近い」要素  $w_A$  へ分配する, imaging[53] という方法がある [17]. ただし, imaging では, 獲得情報が不確実な形態を取る場合に一般化されておらず, あまり用いられていない.

確率測度に基づかない不確実性の表現方法を利用した意思決定支援システムの構築もさまざまに行われており, 医療用のエキスパートシステム MYCIN[8] が構築されるなどしている. MYCIN では, 不確実性を表すものとして, CF(Certainty Factor) が用いられていた. CF 以外の方法に, 確率測度が要請する加法性を緩和したファジィ測度 [128] に基づくアプローチもある. これに関し, ファジィ測度の一つである, 可能性測度で表現された信念に対する更新のメカニズムは Dubois と Prade[17, 18] により構築されている.

証拠理論 [81] は, その測度が可能性理論におけるそれを特殊なケースとして含み [128], 情報の欠落を明示的に表現可能であるという特徴を持つ. これにより, 無知による命題の不支持と証拠に基づく命題の否定とを区別できたり, 命題への確信度を区間によって表現できる. また, 複数の情報を統合するという独特の手法を持ち, 証拠を積み重ねて推論することを容易にモデル化できる理論である.

証拠理論を利用して, 多くの研究がなされている [24, 97]. たとえば, 画像認識 [138, 139], エキスパートシステム [98, 99], 故障診断 [28, 32, 129], 意思決定問題 [27, 33, 97, 133, 135], 警報システムの設計 [88, 90, 91, 92, 118], などの研究がある. これらの研究の多くは, 複数の情報を並列的に結合する情報統合 [13, 28, 81, 96] を適用している.

本論文で想定する人間がおかれている環境は, 時間の経過とともに徐々に状況が変化する動的な環境であり, 意思決定に必要な情報収集に利用できる時間余裕は必ずしも十分で

はない。ここでは、部分的にでも得られた情報に基づいて逐次信念を更新する必要がある、これには情報統合が適しているかは自明ではない。情報統合は、複数の情報を同等に扱うものであり、「一方の情報を他方の情報で更新する」のではないからである。しかし、証拠理論では、信念を動的に更新することの研究は十分にはなされていない。Bayesの定理やJeffreyの規則の証拠理論への拡張がいくつか試みられている(たとえば, [26, 87, 130])ものの、そこで提案されている規則は、本論文2.2節において示すように適用可能性が厳しいことなどにより応用しにくいという問題がある。本論文では、証拠理論における情報更新に注目し、これを整理・拡張し、情報統合と比較することによりその性質を解明する。

## 人間とコンピュータの役割分担

システムを安全かつ効率的に運用するために状況認識の確保とならび重要なのが、人間とコンピュータの役割分担の適正化である。ここでは、タスクの一部を人間が遂行し、残りをコンピュータに任せる形態 (sharing control) とコンピュータが人間のバックアップとして機能したり、人間に置き換わってタスクを遂行する形態 (trading control) とがある [83](p. 65)。

人間とコンピュータとの間での役割分担の問題は、1951年にいわゆる“Fitts lists”が提案されたことをきっかけとして注目されるようになった [43, 82]。Fitts listsは、MABA-MABA (Men Are Better At and Machines Are Better At) lists とよばれることもあるように、人間とコンピュータがそれぞれいかなるタスクに長じているかを列挙したものである。

当初、複数あるタスクそれぞれの遂行に関して人間とコンピュータのいずれが適しているかを MABA-MABA lists に基づいて吟味し、タスクを配分することが考えられていた。しかし、技術レベルの変化によりリストの内容が変わりうる等の問題がある。たとえば、今日、パターン認識のタスクは人間の方が優れていると単純には言い切れなくなっている。各タスクが人間とコンピュータのいずれに適しているか断定することは困難であり

[84], 役割を固定するのは適切でないといわれている [72].

Fitts lists に基づくアプローチへの批判から, 人間とコンピュータとが共に各タスクを遂行でき, その役割を動的に配分することが提案されるようになった [72]. たとえば, Rouse[70] は, 適応支援 (Adaptive Aiding) という概念を提案している. Rouse[71] は, 適応支援を「人間とコンピュータの資源を最適利用することによって, 人間のニーズと能力に合うよう支援を変更すること」と説明している (p. 61). また, 近年様々な分野において, 適応支援の一形態として適応自動化 (Adaptive Automation: AA) が提案されており, AA を支持する研究が数多く報告されている [57, 64, 65, 78]. Scerbo[78] は AA を, 「自動化のモードやレベル, 自動化の数を動的に変更する. これにより, 人間とコンピュータとが制御を共有 (share control) する」ものと説明している. Fitts lists によるタスク配分のみならず, 適応支援や適応自動化という場合も, 一般に, 複数タスクの存在が仮定されている. すなわち, そのタスクを人間とコンピュータにいかに関与させるかという sharing control が主に問題となる [78].

trading control に注目する考え方もある. 人間中心の自動化は, 最終決定権を人間に持たせるとすることによって, 権限の trading に制約を加えていると考えることができる. これに対し, Inagaki[27, 29, 30] は, コンピュータが自主的に権限を持つべきである状況が存在することを指摘し, 緊急度に応じて動的に役割を交代 (trade) する状況適応自動化 (Situation-Adaptive Automation: SAA) を提案している. さらに, Inagaki[30] は, システム安全確保のためには SAA が必要であることを数学モデルに基づいて示している.

しかし, 数学モデルは現実のあらゆる要素を反映できるものではなく, 実際に SAA が必要となりうるか, あるいは実現可能であるか自明ではない. 逆に, 事故事例, たとえば福岡空港の離陸失敗事故 [122] を引用して SAA の必要性を述べようとしても, 個々の事故事例が各々特殊な事情を含むことから, 一般性のある主張をするのは難しい. そこで本論文では, システム安全確保に対する SAA の必要性や有効性を, 認知工学実験を通じて明らかにする.



ところで、人間と自動化システムとがいかに役割を分担できるかは、その両者の間の信頼関係に依存すると一般に考えられる。

Muir[58, 59] は、人間が自動化システムに抱く信頼感 (trust) を、社会学的概念である人と人との間の trust[2, 69] をもとに定性的にモデル化した。Muir と Moray[60] は、自動化システムへの依存性を trust で説明できることを実験により示している。また、Lee と Moray[49, 50] は、trust が時間的性質を持つこと [69] から、時間遅れを反映させた trust のモデルを定量的に構築する方法を提案している。さらに、Lee と Moray[51] は、オペレータ自身のタスク遂行能力への自信 (self-confidence) を導入し、trust と self-confidence との差が正ならば自動化システムが利用され、負ならば自動化システムが使われなくなること明らかにしている。

Parasuraman の研究グループ [55, 63] は、自動化システムへの過度の信頼 (overtrust) があると、複数タスクなど負荷の高い環境において人間の監視能力が低下することを実験により明らかにしている。また、Parasuraman[65] は、自動化システムへの依存性について、使用 (use)、誤用 (misuse)、不使用 (disuse)、濫用 (abuse) の 4 つにわけて詳細に論じ、自動化システムの使用 (use) が trust や作業負荷などの複雑な相互作用に依存すること、自動化への過依存 (over-reliance) が自動化システムの誤用 (misuse) につながることを、警報システムの誤報が自動化システムの不使用 (disuse) をもたらすこと、設計者による自動化の濫用 (abuse) が誤用や不使用をもたらすことなどを示している。

本論文では、人間が SAA を許容するかという観点から SAA を実現するための方策を提案する。そのために、trust や self-confidence と自動化への依存性との関係に着目し、SAA に対する trust と self-confidence およびそれらの動的変動が持つ性質を明らかにする。

## 1.4 本論文の構成

本論文の構成を示す。

第 2 章では、不確実性のもとでの信念の更新と意思決定を支援するための、証拠理論的

手法を整備する。まず、証拠理論における情報処理の方法に情報統合と情報更新の2つがあることを示す。つぎに、従来提案されている情報更新の問題点を指摘し、これらの問題を解消するよう、情報更新を拡張することを提案する。また、証拠理論的な信念の更新方法をプラント安全制御の問題に応用し、規則の最適性の観点から情報統合と情報更新との違いを明らかにする。さらに、いかに情報統合と情報更新とを使い分ければよいかを示す。なお、2.2節は文献[107]、2.3節は文献[108]、2.4節は文献[109]、2.5節は文献[40]に、それぞれ基づいている。

第3章では、状況適応的な意思決定の自動化の必要性を検証する。時間的制約のもとでは意思決定を支援するだけでは必ずしも十分でないことは数学的に示されているが、状況に応じてコンピュータが権限を持つことが実際に必要とされるか認知実験により明らかにする。なお、本章は、文献[35, 39, 110]の成果に基づいている。

第4章では、システム安全確保に対する状況適応自動化の有効性を、認知実験により厳密に検証する。また、trustやself-confidenceと自動化への依存性の関係に着目して、trustやself-confidenceのモデルを構築し、そこから得られる知見に基づいてSAAが人間に許容されるために持つべき性質を明らかにする。なお、本章は、文献[37, 38, 56, 111]の成果に基づいている。

# 第2章 情報の解釈と判断の証拠理論的 支援

## 2.1 はじめに

人間の意思決定に対するコンピュータの支援としては、判断に必要な情報を収集し、これを適切な形態で人間に提供することが考えられる。この支援を実現するためには、人間のもつ知識や、観測・計測によって得られる情報を数値化し、システム状態に関する認識を新たなデータで修正・更新する機構を数学的に構築する必要がある。ここで、システムの構造に関する人間のメンタルモデルやシステム状態の指標となる観測情報は、一部が欠落していたり、場合によっては誤りを含むことがあり、人間のシステム状態の理解には不確かさが伴う。

情報の不確か性を扱う理論体系としては、確率論とこれに基づく意思決定論 [119, 137] が代表的である。確率論では、一般に、信念の更新には Bayes の定理や、得られる情報が不確実な場合には Jeffrey の規則 [42] が用いられる。ここで Jeffrey の規則とは、標本空間  $\Omega$  の分割  $\{B_1, \dots, B_n\}$  に対して、 $B_i$  の確率  $P_{pos}(B_1), P_{pos}(B_2), \dots, P_{pos}(B_n)$  が観測によって得られたとき、これらによって事前確率  $P_{pri}$  を更新するものである。任意の事象  $A$  の事後確率  $P_{pos}(A)$  は次式で与えられる。

$$P_{pos}(A) = \sum_{i=1}^n P_{pos}(B_i) P_{pri}(A|B_i) \quad (2.1)$$

ただし、 $P_{pri}(A|B_i)$  は、 $A$  の  $B_j$  による条件付き確率を表す。

ところで、確率測度が要請する加法性は、人間の主観を表現するには制約が厳しすぎ、ときに不自然な状況を作り出す。たとえば、命題  $A$  が成立することを支持する証拠  $e$  を入

手したものとし、 $e$ によって $A$ は確信度0.6で成り立つものとする。このとき、この証拠 $e$ を確率測度で表現しようとするならば、 $e$ 自身は命題 $A$ の不成立については何ら言及しないにも関わらず、確率測度の加法性により $e$ は同時に命題 $A$ を0.4の確信度で否定することにもなる。

これに対し、証拠理論では、命題を否定することと命題を支持する証拠が不十分であることを区別できる。すなわち、命題の真偽に関わる不特定性 (nonspecificity) を明示的に扱える [47]。そこで、ここでは人間の信念を表現する枠組みとして証拠理論 [81] を採用する。

現実には、とくに動的なシステムでは、どの行動を選択するかだけが問題ではなく、いつ意思決定するかも問題となる。確率論に基づく場合、動的に情報を獲得しつつ意思決定することは逐次的意思決定の問題として定式化できる [73] (17章)。これに対し、証拠理論では逐次的に意思決定するモデルは十分に確立されていない。たとえば、Jaffray [41] は証拠理論を上界・下界確率としてとらえ、動的な意思決定のモデルの構築を試みているが、そこで信念更新に用いる獲得情報には不確実なものを扱えない。証拠理論では、情報処理の方法として、独立に得られた複数の情報を同時に結合する情報統合 [13, 28, 81, 96] 以外にも、獲得した情報により逐次信念を更新する情報更新 [13, 17, 26] がある。しかし、従来の情報更新は、その適用可能性に対する制約が大きいなどの問題がある [107]。このように、信念を更新する方法の構築が不十分である。

本章では、証拠理論における信念の更新方法を整備する。2.2節では、証拠理論における情報処理の手法として、「情報統合」、「情報更新」の2種類があることを示し、情報更新の新たな手法を提案する。これにより、情報統合のみならず、情報更新も非可算無限個の規則を持ちうることになる。2.3節では、大規模複雑プラントの安全制御問題を取り上げ、ある安全制御方策のもとで期待損失を最小化する意味で最適な情報統合、情報更新の規則を明らかにする。2.4節では、安全制御方策の選び方を確率的とした場合、すなわち混合戦略の下での情報統合、情報更新規則の最適化を行う。2.5節では、情報統合と情報更新との相違点を明らかにし、これらの手法を使い分ける指針を示す。

## 2.2 不確実環境下での情報の獲得

有限個の命題の集合を  $\Omega = \{x_1, \dots, x_n\}$  とし, 識別空間 (frame of discernment: fod) と呼ぶ. ここで,  $x_i \in \Omega$  を基本命題,  $A \subset \Omega$  を命題と呼び, 空集合  $\phi$  は恒偽命題とする.  $x_i$  は互いに排他的である (mutually exclusive) とし, いずれか一つが必ず真である (totally exhaustive) ものとする. また,  $A \subset \Omega$  の補集合を  $\bar{A}$  と書き,  $\Omega$  のべき集合を  $2^\Omega$  で表す.

証拠理論においては, 各命題の真偽に関する確信の度合, すなわち信念は, ある条件を満たす実数値集合関数によって表わされ, いくつかの表現形態がある [81]. ここでは, それらのうち代表的な, 信念割当関数とベリーフ関数を取りあげる.

### 2.2.1 確信度の表現方法

公理 1 を満たす実数値集合関数  $m : 2^\Omega \rightarrow [0, 1]$  を, 信念割当関数 (basic belief mass function: bbm)[114] という.

#### 公理 1

$$(i) \quad m(A) \geq 0, \quad A \subset \Omega$$

$$(ii) \quad m(\phi) = 0$$

$$(iii) \quad \sum_{A \subset \Omega} m(A) = 1$$

$m(A)$  は,  $A$  に属する要素のうち, いずれが真であるか特定はできないが, 真である要素が  $A$  に含まれる度合を意味する.

公理 1 を満たす  $m$  を用いて, 命題の確信の度合, 可能性の度合としての実数値集合関数  $\text{Bel}, \text{Pl} : 2^\Omega \rightarrow [0, 1]$  をつぎのように定義する.

#### 定義 1

$$\text{Bel}(A) = \sum_{B \subset A} m(B) \quad (2.2)$$

$$\text{Pl}(A) = \sum_{B \cap A \neq \emptyset} m(B) \quad (2.3)$$

ここで, Bel, Pl は,

$$\text{Bel}(A) = 1 - \text{Pl}(\bar{A}), \quad A \subset \Omega \quad (2.4)$$

を満たす [81].

$m(A) > 0$  を満たす  $A \subset \Omega$  を焦点要素 (focal element) という. ある bbm  $m$  と  $\mathcal{W} = \{A : m(A) > 0\}$  の対  $(m, \mathcal{W})$  を信念構造 (belief structure) という. とくに,  $m(\Omega) = 1$  であるとき, vacuous な信念構造という.

信念は Bel または Pl から導入することも可能である [81]. ただし, Bel と Pl は (2.4) 式を満たすため, Bel と Pl のうち一方に注目すればよい. 本論文では Bel に着目する.

公理 2 を満たす実数値集合関数  $\text{Bel}: 2^\Omega \rightarrow [0, 1]$  をビリーフ関数 (belief function) という.

## 公理 2

$$(i) \quad \text{Bel}(\emptyset) = 0$$

$$(ii) \quad \text{Bel}(\Omega) = 1$$

$$(iii) \quad A_1, \dots, A_n \subset \Omega \text{ に対して,}$$

$$\text{Bel}\left(\bigcup_{i=1}^n A_i\right) \geq \sum_{I \subset \{1, 2, \dots, n\}} (-1)^{|I|+1} \text{Bel}\left(\bigcap_{j \in I} A_j\right)$$

公理 2 で定めた Bel によって, 関数  $m$  を次式で定義する.

## 定義 2

$$m(A) = \sum_{B \subset A} (-1)^{|A-B|} \text{Bel}(B), \quad A \subset \Omega \quad (2.5)$$

定義 2 で与えられる  $m$  は bbm である [81].

なお,  $\text{Bel}(\Omega) = 1$  を満たし  $\Omega$  の真部分集合  $A$  に対して  $\text{Bel}(A) = 0$  となる Bel はビリーフ関数であり, この Bel に対応する信念構造は vacuous である [81] ことから, 本論文では

このような Bel を vacuous であるとよぶ。また、vacuous な信念構造もしくは Bel により規定される信念を vacuous であるとよぶことにする。

なお、公理 2 でピリーフ関数を定義する場合、フローザピリティ関数  $Pl: 2^\Omega \rightarrow [0, 1]$  は (2.4) 式、すなわち、

$$Pl(A) = 1 - Bel(\bar{A}), \forall A \subset \Omega \quad (2.6)$$

で与えられる。

## 2.2.2 情報統合

複数の信念構造を一つにまとめる機構が情報統合 [13, 28, 81, 96] であり、二つの信念構造  $(m', W')$ ,  $(m_I, W_I)$  を統合して  $(m, W)$  を得ることは、形式的に次式で表現される。

$$m = m' \oplus m_I \quad (2.7)$$

情報統合は、二つのステップからなる。

### 1. 予備割当関数の導出

予備割当関数 (ground mass assignment: gma)  $q$  を

$$q(E) = \sum_{B \cap C = E} m'(B)m_I(C) \quad (2.8)$$

と定義する。なお、 $q$  を、(2.8) 式で定義しない方法もある [133]。すなわち、 $q$  を (2.8) 式の代わりに次式で定義するものである。

$$q(E) = \sum_{\psi(B,C)=E} m'(B)m_I(C) \quad (2.9)$$

$\psi$  は、二つの情報源が各々  $B, C$  を支持する場合に、 $(B, C)$  をシステム状態としてどのように解釈するかを与える写像である。 $\psi(B, C) = B \cap C$  ならば、(2.9) 式は (2.8) 式に一致する。(2.9) 式を用いると、命題の組  $(B, C)$  をより柔軟に解釈でき、よりよい統合規則が得られることもある [133]。しかし、ある解釈  $\psi$  を与えるためには、情

報源間の関係に関する知識を必要とするが、そのような知識がつねに合理的に得られるとは限らない。そこで、本論文では、これ以降(2.8)式を  $q$  の定義として用いる。

## 2. $q$ から $m$ への変換

$q(\phi)$  は一般には 0 ではない。つまり、 $q$  は bbm ではないので、 $q(\phi)$  の値を何らかの方法で処理し、bbm に直す必要がある。ただし、開世界仮説 (open-world assumption) をとることによって  $m(\phi) > 0$  を許し、正規化をしないアプローチもある [88]。

gma  $q$  を bbm  $m$  に変換するための方法としてよく知られているのは、次式の Dempster の規則 [81] である。

$$m(E) = \frac{q(E)}{1 - q(\phi)}, \forall E \subset \Omega \quad (2.10)$$

Dempster の統合規則は、二つの情報の間にある矛盾を、あたかも矛盾がないかのように処理するものであり、このことが数値的不安定性などの問題点のもととなっている [12, 96, 100]。Yager[96] は、そうした問題点を解消すべく、次式の情報統合の規則を提案した。

$$\begin{aligned} m(E) &= q(E), A \neq \phi, \Omega \\ m(\Omega) &= q(\Omega) + q(\phi) \end{aligned} \quad (2.11)$$

これらに対し、Inagaki[27, 28] は Dempster の規則や Yager の規則を特殊な場合として含む情報統合のクラスが存在することを示した。Inagaki[27] によると、 $q(\phi)$  の値の一部を  $\phi$  あるいは  $\Omega$  以外の焦点要素に対し、

$$\frac{m(E)}{m(F)} = \frac{q(E)}{q(F)}, E, F \neq \Omega, \phi \quad (2.12)$$

を満たすように分配し、残りを  $\Omega$  に割り当てるクラスの情報統合規則は、次式で与えられる。



$$\begin{cases} m(E) = \{1 + kq(\phi)\}q(E), & E \neq \Omega, \phi \\ m(\Omega) = \{1 + kq(\phi)\}q(\Omega) + \{1 + kq(\phi) - k\}q(\phi) \\ 0 \leq k \leq \frac{1}{1 - q(\phi) - q(\Omega)} \end{cases} \quad (2.13)$$

(2.13)式では、 $q(\phi)$ のうち $\Omega$ に割り当てられる分はパラメータ $k$ の値に依存する。たとえば、 $k$ の値を小さくすると、 $m(\Omega)$ の値は大きくなる。すなわち、 $k$ の値を定めることによって、二つの情報間の不整合の度合である $q(\phi)$ をいかに扱うかを設定できる。

(2.13)式において $k$ の値を一つ定めることにより、一つの統合規則が生成される。たとえば、 $k = 1/(1 - q(\phi))$ とすると Dempster の統合規則 (2.10) 式が得られ、 $k = 0$ とすると Yager の統合規則 (2.11) 式を得る。

なお、(2.13)式をさらに拡張した規則が、Tanaka[91, 92]により提案されている。これは、 $q(\Omega)$ をも $q(E)$ の比に応じて $\Omega$ の真部分集合へ分配するものであり、 $k$ の最大値が、 $k = \frac{1}{1 - q(\phi) - q(\Omega)}$ よりも大きく、 $k = \frac{q(\phi) + q(\Omega)}{q(\phi)} \cdot \frac{1}{1 - q(\phi) - q(\Omega)}$ となる。ただし、この拡張は、fodの大きさが2のとき以外には明確な意義が与えられていないことから、本論文では(2.13)式を用いることにする。

ところで、(2.13)式が定義されるのは、

$$q(\phi) + q(\Omega) < 1 \quad (2.14)$$

が成立するときに限られるが、本論文では、(2.14)式の成立を要請せず、より緩い制約のもとで情報統合を行えるように拡張する。すなわち、

$$q(\phi) + q(\Omega) = 1 \text{ のとき } m(\Omega) = 1 \quad (2.15)$$

と定義する。(2.15)式において、 $q(\phi) + q(\Omega) = 1$ ならば、 $E \neq \phi$ あるいは $E \neq \Omega$ なる $E$ については $q(E) = 0$ となり、(2.13)式より $m(E) = 0$ を得る。また、 $q(\phi) + q(\Omega) = 1$ の

とき、 $q(\phi) = 1 - q(\Omega)$  を (2.13) 式に形式的に代入すれば、 $m(\Omega) = 1$  を得る。したがって、(2.15) 式は (2.13) 式の自然な拡張となっている。

## 情報の並列性

(2.13) 式の情報統合は可換である。すなわち、任意の bbm  $m_1, m_2$  に対して

$$m_1 \oplus m_2 = m_2 \oplus m_1 \quad (2.16)$$

しかし、一般には結合的でない (Dempster の規則のときのみ結合的)[27]。つまり、

$$m_1 \oplus (m_2 \oplus m_3) \neq (m_1 \oplus m_2) \oplus m_3 \quad (2.17)$$

ただし、(2.8) 式を次式で置き換えることによって、 $n$  個の ( $n \geq 3$ ) 情報を同時に統合することが可能である。

$$q(E) = \sum_{B_1 \cap B_2 \cap \dots \cap B_n} m_1(B_1)m_2(B_2)\cdots m_n(B_n) \quad (2.18)$$

したがって、情報統合では、統合される情報は本質的に並列的かつ対称的に扱われる。

なお、対称性を意図的に崩した情報統合も提案されている [13]。すなわち、二つの情報  $m'$ 、 $m_I$  の相対的な信頼度を  $\lambda$ 、 $1 - \lambda$  とするとき、この二つの情報を次式で定義する。

$$m(A) = \lambda m'(A) + (1 - \lambda)m_I(A) \quad (2.19)$$

統合する情報が  $n$  個 ( $n \geq 3$ ) の場合でも、各情報源  $i$  の相対的な信頼度を  $\lambda_i$  とし ( $\sum_{i=1}^n \lambda_i = 1$ )、統合を次式で定義すれば、この統合規則も本質的には情報を並列的に扱うことになる。

$$m(A) = \sum_{i=1}^n \lambda_i m_i(A) \quad (2.20)$$

### 2.2.3 情報更新

Bel によって信念が表現されているとする。新しい情報が入手できた際に、その情報に基づいて信念をいかに更新すべきか考察する。

### (i) 条件付け

まず、ある単一の命題が真であるという情報を得たときの信念の更新を考える。

「命題  $B \subset \Omega$  が真である」という条件のもとで命題  $A$  の真偽を論ずるとき、これを命題  $B$  による条件付き命題とよび、 $A|B$  で表す。また、このような状況のもとでの信念  $\text{Bel}$  の更新を、命題  $B$  による条件付け (conditioning) という [81]。

条件付けの方法は、これまでにさまざまな立場から考察され、いくつかの方法が提案されている。

**geometric な条件付け** 確率空間  $(X, \mathcal{F}, P)$  が与えられたとき、条件付き確率  $P(A|B)$  は、

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad A, B \in \mathcal{F} \quad (2.21)$$

で与えられるが、これと同じ形式で  $\text{Bel}$  の条件付けを行うもの、すなわち

$$\text{Bel}(A|B) = \frac{\text{Bel}(A \cap B)}{\text{Bel}(B)} \quad (2.22)$$

を geometric な条件付けという [15]。また、これに対応して、命題  $B$  による  $\text{Pl}$  の条件付けは次式で与えられる [15]。

$$\text{Pl}(A|B) = \frac{\text{Pl}(A \cup \bar{B}) - \text{Pl}(\bar{B})}{1 - \text{Pl}(\bar{B})} \quad (2.23)$$

$\text{Bel}$  の geometric な条件付け (2.22) 式に対応する  $\text{bbm}$  の更新メカニズムは Dubois と Prade によって示されている [13]。  $\text{Bel}(A|B)$  に対応する  $\text{bbm}$  を  $\tilde{m}$  とすると、

$$\tilde{m}(A) = \begin{cases} \frac{m(A)}{\sum_{C \subset B} m(C)} & A \subset B \\ 0 & \text{otherwise} \end{cases} \quad (2.24)$$

geometric な条件付けが定義できるためには  $\text{Bel}(B) > 0$  の成立が要求される。しかし、常に  $\text{Bel}(B) > 0$  が成立するとは限らない。たとえば、もとの信念  $\text{Bel}(\cdot)$  が vacuous であるとき、 $B \neq \Omega$  ならば必ず  $\text{Bel}(B) = 0$  となる。つまり、vacuous な  $\text{Bel}$  に対しては、 $B$

をどのように選んでも条件付けできない。このことから、本論文では、geometricな条件付けは利用しないことにする。

**Dempster の条件付け** 次式による Bel の条件付けを Dempster の条件付け [81] という。

$$\text{Bel}(A|B) = \frac{\text{Bel}(A \cup \bar{B}) - \text{Bel}(\bar{B})}{1 - \text{Bel}(\bar{B})} \quad (2.25)$$

これは、対応する  $\text{Pl}(A|B)$  を

$$\text{Pl}(A|B) = \frac{\text{Pl}(A \cap B)}{\text{Pl}(B)} \quad (2.26)$$

によって定義したことを意味する。

(2.25) 式は、Dempster の情報統合規則において、統合すべき二つの信念構造のうちの一つを  $B$  の重みが 1 であるものとした場合として解釈できる [81]。その意味において、(2.25) 式に対応する bbm の更新メカニズムは明らかになっている。

なお、(2.25) 式が定義できるための制約は、 $\text{Bel}(\bar{B}) < 1$  が成立することである。すなわち、(2.25) 式は  $\text{Bel}(B) = 0$  の場合でも定義できる点で geometric な条件付けより一般的な状況において適用可能である。

**下界確率に基づく条件付け** 前田と市橋 [130]、Halpern と Fagin [25]、Campos ら [9]、Dempster [10] は、確率測度の下界としての Bel の観点から、 $\text{Bel}(A|B)$  を次式で提案している。

$$\text{Bel}(A|B) = \frac{\text{Bel}(A \cap B)}{\text{Bel}(A \cap B) + \text{Pl}(\bar{A} \cap B)} \quad (2.27)$$

ここで、Halpern と Fagin、Campos らは、 $\text{Bel}(B) > 0$  の場合にのみ (2.27) 式を定義するとしている。これは、Bel を確率の下界値とするためである [25]。さらに、 $\text{Bel}(\cdot|B)$  が公理 2 を満たすようにするために、(2.27) 式右辺の分母が 0 になるとき、左辺を 0 と定義している。

一方、前田と市橋[130]は、(2.27)式が定義できるための制約を  $\text{Bel}(\bar{B}) < 1$  に緩めている。さらに、 $\text{Bel}(\cdot|B)$  が公理2を満たすようにするために、(2.27)式の右辺の分母が0になるとき、左辺を1と定義している。前田と市橋による条件付けの制約の緩和によって、Dempsterの条件付けと同じ制約のもとで条件付けを行うことができる。

(2.27)式の条件付けは、bbmでは次のように説明される[9]。条件付き bbm  $m$  を次式で定義する。

**定義 3**

$$m(C|B) = \frac{\sum_{D \cap B \subset C} \alpha_D m(D)}{\sum_{D \subset \Omega} \alpha_D m(D)}, \forall C \subset \Omega \quad (2.28)$$

$\alpha_D \in [0, 1]$  は、fod を  $B$  に限定した場合の  $m(D)$  の重みを表す。この  $m(C|B)$  を用いて、Bel の条件付けを次式で定義する。

$$\text{Bel}(A|B) = \sum_{C \subset A} m(C|B), \forall A \subset \Omega \quad (2.29)$$

ここで、 $D \cap B = \phi$  もしくは  $D \cap \bar{B} = \phi$  ならば、

$$\alpha_D = \begin{cases} 1 & D \subset B \\ 0 & D \cap B = \phi \end{cases} \quad (2.30)$$

$D \cap B \neq \phi$  かつ  $D \cap \bar{B} \neq \phi$  ならば、

$$\alpha_D = \begin{cases} 0 & D \cap B \subset A \\ 1 & \text{otherwise} \end{cases} \quad (2.31)$$

とすると、(2.27)式を得る。

なお、 $\text{Bel}(A|B)$  を (2.27)式で定義する場合、 $\text{Pl}(A|B)$  は次式で与えられる。

$$\text{Pl}(A|B) = \frac{\text{Pl}(A \cap B)}{\text{Pl}(A \cap B) + \text{Bel}(\bar{A} \cap B)} \quad (2.32)$$

しかし、(2.27)式による条件付けでは、獲得される情報が不確実になった場合に、更新結果が下界確率にならなくなることが知られており[25]、信念の更新を不確実な情報に基づ

くケースに拡張できない。また、Dubois と Prade は、(2.27) 式の条件付けを focusing とし  
てとらえている [16, 17]。すなわち、 $A|B$  を、 $B$  という世界に限定した (focus) 場合に、現  
在の信念  $\text{Bel}$  により  $A$  をどの程度支持できるかを求める手続きとして、(2.27) 式を与えて  
いる。この考え方は、「情報を獲得して、信念そのものを書き換える」方式の revision [16, 17]  
と明確に区別される。本研究では、信念の「更新」に興味があることから、(2.27) 式の条  
件付けを採用しないことにする。

## (ii) 不確実な情報に基づく信念の更新

情報が不確実性を含む場合、すなわち情報が信念構造として得られる場合の信念  $\text{Bel}$  の  
更新方法を考える。

**Dubois と Prade の更新手法** あらかじめ  $\text{Bel}'$  をもっているとき、新たに情報  $(m_I, W_I)$   
を得たとする。このとき、 $(m_I, W_I)$  によっていかに  $\text{Bel}'$  を更新するかが問題となる。

更新後の信念を  $\text{Bel}$  と書くとき、 $(m_I, W_I)$  による  $\text{Bel}'$  の更新を、Dubois と Prade は次  
式で定義している [15]。

### 定義 4

$$\text{Bel}(A) = \sum_{B \in W_I} m_I(B) \text{Bel}'(A|B) \quad (2.33)$$

これは、いわゆる Jeffrey の規則 [42](2.1) 式を証拠理論の枠組みに拡張したものに相当す  
る。なお、 $\text{Bel}'(\cdot|B)$  は、Dempster の条件付け規則 (2.25) 式を用いて計算される。

ところで、(2.33) 式の情報更新は、少なくとも次の 2 つの問題点を持つ。

1. Dempster の条件付け規則の性質から、(2.33) 式を用いた  $\text{Bel}'$  の更新においては、次  
式の成立が前提とされなければならない。

$$\text{Bel}'(\bar{B}) < 1, \forall B \in W_I \quad (2.34)$$

しかし、 $\text{Bel}'$ に対応する bbm  $m'$  と  $m_I$  を (2.15) 式で統合する場合には、 $B \in \mathcal{W}_I$  について  $\text{Bel}'(\overline{B}) = 1$  の場合でも統合が定義できることがある。たとえば、 $\Omega = \{a, b, c\}$  とし、 $m'(\{a\}) = 0.7$ 、 $m'(\{c\}) = 0.3$ 、 $m_I(\{b\}) = 0.2$ 、 $m_I(\Omega) = 0.8$  とするとき、 $\text{Bel}'(\overline{\{b\}}) = \text{Bel}'(\{a, c\}) = 1$  であるが、 $q(\phi) = 0.2$  であるので通常どおり情報統合可能である。したがって、(2.34) 式は情報統合に比べ厳しい制約であるといえる。

2. 焦点要素の集合  $\mathcal{W}'$  と  $\mathcal{W}_I$  が共に等しく、 $\Omega$  の分割  $\{B_1, \dots, B_n\}$  であるとする。このとき、(2.33) 式を用いた  $\text{Bel}'$  の更新結果  $\text{Bel}$  は、 $\text{Bel}'$  の値によらず、

$$\text{Bel}(\cdot) = \text{Bel}_I(\cdot) \quad (2.35)$$

となる [15]。つまり、更新結果は、事後に得られた証拠そのものによって決まり、事前の証拠は全く反映されなくなるという問題が生じる。なお、 $\text{Bel}_I$  は  $m_I$  に対応するビリーフ関数を表す。

これらのことから、オペレータの信念の推移をモデル化するには、(2.33) 式のままでは不適切であるといえ、こうした問題点を回避できる情報更新の規則が必要となる。

#### 2.2.4 Dempster の条件付けの一般化

命題  $B$  による条件付けは、「命題  $B$  が真である」として行われるのが普通である [14, 23]。このことは、更新の際の  $\text{fod}$  を  $B$  に制限することに対応する。しかし、「命題  $B$  が真である」ことは、ある情報源が主張しているにすぎない。ある情報源の主張をどの程度受け入れるかは、情報の更新を行う主体の判断に委ねられてもよいとする考え方もある [27]。そこで、本論文では、 $\text{fod}$  を  $B$  に制限する通常の場合を含む新しい条件付けを提案する。なお、「命題  $B$  が真であることをある情報源が主張しているときの、命題  $A$  の真偽を論ずる」際に、 $\text{fod}$  が  $B$  に制限されない場合も含みうることから、本論文では、このような条件付けを  $A||B$  と表す。

定義 5

(a)  $\text{Bel}(\overline{B}) < 1$  のとき,

$$\begin{cases} \text{Bel}(A||B) = \{1 + k_B \text{Bel}(\overline{B})\} \{ \text{Bel}(A \cup \overline{B}) - \text{Bel}(\overline{B}) \}, A \subset \Omega \\ \text{Bel}(\Omega||B) = 1 \\ 0 \leq k_B \leq \frac{1}{1 - \text{Bel}(\overline{B})} \end{cases} \quad (2.36)$$

(b)  $\text{Bel}(\overline{B}) = 1$  のとき,

$$\text{Bel}(A||B) = \begin{cases} 1 & A = \Omega \\ 0 & \text{otherwise} \end{cases} \quad (2.37)$$

定理 1 (2.36), (2.37) 式で定義される  $\text{Bel}(\cdot||B)$  はビリーフ関数である.

(証明) (2.37) 式による  $\text{Bel}(\cdot||B)$  はビリーフ関数であることは明らか. 以下に, (2.36) 式で与える  $\text{Bel}(\cdot||B)$  がビリーフ関数であることを示す.

$\text{Bel}(\cdot||B) \in [0, 1]$  であることは, つぎのように示される.

$A = \Omega$  のとき (2.36) 式により,

$$\text{Bel}(\Omega||B) = 1 \quad (2.38)$$

$A \subset \Omega$  のとき,

$$\text{Bel}(A \cup \overline{B}) \geq \text{Bel}(\overline{B}) \geq 0 \quad (2.39)$$

ゆえ,  $0 \leq k_B \leq 1/(1 - \text{Bel}(\overline{B}))$  に対して,

$$\begin{aligned} \text{Bel}(A||B) &= \{1 + k_B \text{Bel}(\overline{B})\} \{ \text{Bel}(A \cup \overline{B}) - \text{Bel}(\overline{B}) \} \\ &\in [0, 1] \end{aligned} \quad (2.40)$$



公理 2(i) の成立はつぎのように確かめられる.

$$\begin{aligned}
\text{Bel}(\phi||B) &= \{1 + k_B \text{Bel}(\bar{B})\} \{ \text{Bel}(\phi \cup \bar{B}) - \text{Bel}(\bar{B}) \} \\
&= \{1 + k_B \text{Bel}(\bar{B})\} \{ \text{Bel}(\bar{B}) - \text{Bel}(\bar{B}) \} \\
&= 0
\end{aligned} \tag{2.41}$$

また, (2.38) 式は公理 2(ii) を満たすことを示している.

つぎに, 公理 2(iii) の成立を示す. いま,  $N = \{1, 2, \dots, n\}$  とし,  $A_i \subset \Omega$ ,  $i \in N$  とする.

( $\alpha$ )  $A_1 \cup \dots \cup A_n \neq \Omega$  のとき,

$$\text{Bel}\left(\bigcup_{i=1}^n A_i || B\right) = \{1 + k_B \text{Bel}(\bar{B})\} \left\{ \text{Bel}\left(\bigcup_{i=1}^n A_i \cup \bar{B}\right) - \text{Bel}(\bar{B}) \right\} \tag{2.42}$$

ここで,

$$\text{Bel}\left(\bigcup_{i=1}^n A_i \cup \bar{B}\right) - \text{Bel}(\bar{B}) \geq \sum_{I \subset N} (-1)^{|I|+1} \text{Bel}\left(\left(\bigcap_{i \in I} A_i\right) \cup \bar{B}\right) - \text{Bel}(\bar{B}) \tag{2.43}$$

であり, また, 2項定理より,

$$\begin{aligned}
&\sum_{I \subset N} (-1)^{|I|+1} \text{Bel}\left(\left(\bigcap_{i \in I} A_i\right) \cup \bar{B}\right) - \text{Bel}(\bar{B}) \\
&= \sum_{I \subset N} (-1)^{|I|+1} \left\{ \text{Bel}\left(\left(\bigcap_{i \in I} A_i\right) \cup \bar{B}\right) - \text{Bel}(\bar{B}) \right\}
\end{aligned}$$

よって,

$$\begin{aligned}
\text{Bel}\left(\bigcup_{i=1}^n A_i || B\right) &\geq \sum_{I \subset N} (-1)^{|I|+1} \{1 + k_B \text{Bel}(\bar{B})\} \left\{ \text{Bel}\left(\left(\bigcap_{i \in I} A_i\right) \cup \bar{B}\right) - \text{Bel}(\bar{B}) \right\} \\
&= \sum_{I \subset N} (-1)^{|I|+1} \text{Bel}\left(\bigcap_{i \in I} A_i || B\right)
\end{aligned} \tag{2.44}$$

( $\beta$ )  $A_1 \cup \dots \cup A_n = \Omega$  のとき,

( $\beta$ -1) すべての  $i \in N$  について  $A_i \subset \Omega$  のとき,

$$\sum_{I \subset N} (-1)^{|I|+1} \text{Bel}\left(\bigcap_{i \in I} A_i || B\right) = \sum_{I \subset N} (-1)^{|I|+1} \{1 + k_B \text{Bel}(\bar{B})\}$$

$$\begin{aligned}
& \cdot \left\{ \text{Bel}\left(\bigcap_{i \in I} A_i \cup \overline{B}\right) - \text{Bel}(\overline{B}) \right\} \\
& \leq \{1 + k_B \text{Bel}(\overline{B})\} \{1 - \text{Bel}(\overline{B})\} \\
& \leq 1
\end{aligned} \tag{2.45}$$

(β-2) ある  $j \in N$  について  $A_j = \Omega$  であるとき,

$J = \{j | A_j = \Omega\}$ ,  $I - J = \{l_1, l_2, \dots, l_{n'-1}\}$ ,  $I' = (I - J) \cup \{l_{n'}\} = \{l_1, \dots, l_{n'-1}, l_{n'}\}$  とする. ただし,  $A_{l_{n'}} = \Omega$ . このとき, 任意の  $X \subset \Omega$  に対し,  $X \cap A_{l_{n'}} = X$  であることに注意すると,

$$\begin{aligned}
\sum_{I \subset N} (-1)^{|I|+1} \text{Bel}\left(\bigcap_{i \in I} A_i | B\right) &= \sum_{I' \subset \{l_1, l_2, \dots, l_{n'}\}} (-1)^{|I'|+1} \text{Bel}\left(\bigcap_{i \in I'} A_i | B\right) \\
&= \text{Bel}(A_{l_{n'}} | B) \\
&\leq \text{Bel}\left(\bigcup_{i=1}^n A_i | B\right)
\end{aligned} \tag{2.46}$$

(証明終わり)

$k_B = 1/(1 - \text{Bel}(\overline{B}))$  のとき, (2.36) 式の  $\text{Bel}(A | B)$  は Dempster の条件付け規則 (2.25) 式の  $\text{Bel}(A | B)$  に一致する.

### 新たな条件付け規則の性質

(2.25) 式の  $\text{Bel}$  は, Dempster の情報統合規則に基づいた条件付けであり,  $\text{fod}$  を  $B$  に限定するものである. すでに述べたように, Dempster の情報統合規則は, 数値的不安定性などの問題点が指摘されており [12, 96, 100], Dempster の条件付け (2.25) 式も同様の問題が発生しうる.

**例 1**  $\Omega = \{a, b, c, d\}$  とし,  $m(\{a, b\}) = 0.01$ ,  $m(\{a, d\}) = 0.99$  とする.  $B = \{b, c\}$  によって条件付けたピリーフ関数は,  $\text{Bel}(\{b\}) = 1$  となる. これに対し,  $m(\{a, b\}) = 0.01$ ,  $m(\{a, d\}) = 0.98$ ,  $m(\{c, d\}) = 0.01$  とすると,  $\text{Bel}(\{b\}) = 0.5$ ,  $\text{Bel}(\{c\}) = 0.5$  となる.

本論文で提案する  $\text{Bel}(A||B)$  は  $\text{fod}$  を  $B$  に限定するものではないため、一般には、

$$\text{Bel}(B||B) \leq 1 \quad (2.47)$$

となる。ここで、等号が成立するのは、 $k = 1/(1 - \text{Bel}(\overline{B}))$  とするとき、すなわち Dempster の条件付けを行う場合に限られる。

条件付けのパラメータ  $k_B$  は、与えられた情報  $B$  をどの程度受け入れるかを表すものとして見ることができる。 $\text{Bel}(A||B)$  は、Dempster の条件付けと同様に、事前に持つ信念  $\text{Bel}$  と、 $B$  の重みを 1 とする情報との統合として考えられる。 $k_B = k_{B_{max}}$  とすることは Dempster の規則に相当し、 $B$  と矛盾する部分を無視することを意味する。すなわち、“ $B$  is true” を完全に受け入れることとなる。逆に、 $k_B = 0$  とすることは、事前信念において  $B$  と矛盾する部分の重み  $\text{Bel}(\overline{B})$  を全て  $\Omega$  に割り当てることにあたる。

さらに、パラメータ  $k_B$  と  $\text{Bel}(A||B)$  との関係を見る。 $\text{Bel}(\overline{B}) < 1$  とするとき、 $\text{Bel}(A||B)$  は、パラメータ  $k_B$  の一次式である。パラメータ  $k_B$  をある特定の値に固定したときの  $\text{Bel}$  の条件付けを  $\text{Bel}_{k_B}(A||B)$  で表すものとする、 $k_B$  の係数  $\text{Bel}(\overline{B})\{\text{Bel}(A \cup \overline{B}) - \text{Bel}(\overline{B})\}$  は正であるから次式が成立する。

$$\text{Bel}_{k_B=0}(A||B) \leq \text{Bel}_{k_B=k_{B_{max}}}(A||B) \quad (2.48)$$

$\text{Bel}(A \cup \overline{B}) = \text{Bel}_{k_B=0}(A||B) + \text{Bel}(\overline{B})$  と公理 2(iii) より、

$$\text{Bel}(A \cap B) \leq \text{Bel}_{k_B=0}(A||B) \quad (2.49)$$

となる。すなわち、(1)  $B$  で条件付けることによって  $B$  と排他的でない命題  $A$  の確信度は  $\text{Bel}(A \cap B)$  以上になる、また (2)  $k_B$  の値を増やすほど  $\text{Bel}_{k_B}(A||B)$  は大きくなり、“ $B$  is true” をより受け入れることとなる、ことがわかる。

## 2.2.5 情報更新の拡張とその利点

### 条件付けの適用可能性

条件付け規則 (2.36),(2.37) 式を Dubois と Prade の更新手法 (2.33) 式に適用することによって,  $Bel'$  を  $(m_I, \mathcal{W}_I)$  で更新し,  $Bel$  を得る状況を考える. すなわち,

$$Bel(A) = \sum_{B \in \mathcal{W}_I} m_I(B) Bel'(A||B) \quad (2.50)$$

このとき, (2.37) 式から, ある  $B \in \mathcal{W}_I$  に対して  $Bel'(\overline{B}) = 1$  である場合でも  $Bel'$  の更新が可能である.

さらに,  $Bel'$  に対応する bbm が構成する信念構造を  $(m', \mathcal{W}')$  とするとき, 任意の  $A \in \mathcal{W}', B \in \mathcal{W}_I$  について  $A \cap B = \emptyset$  となる場合でも  $Bel'$  を  $m_I$  で更新することができ, このとき更新結果は vacuous な信念になる. つまり, 事前の信念と獲得した情報が全く矛盾する場合にも, 信念を獲得情報によって更新することが可能である. したがって, Dubois と Prade の手法に基づいて信念を更新する場合には, ビリーフ関数と信念構造に課せられる制約はなく, あらゆる場合について更新が可能である.

### 事前信念の影響

2.2.3 節において, Dubois と Prade の更新手法 (2.33) 式に Dempster の条件付け (2.25) 式を適用する場合には, つぎの問題が発生し得ることを述べた.

「 $\mathcal{W}' = \mathcal{W}_I = \{B_1, \dots, B_n\}$  が fod の分割である場合, 更新結果  $Bel$  は  $Bel_I$  に一致し,  $Bel'$  の値の取り方に依存しない」

これに対し, 本論文で提案している規則 (2.36) 式を用いる更新では,  $k \neq 1/(1 - Bel(\overline{B}))$  の場合には更新された信念に事前の信念が反映されることをつぎの定理で示す.

**定理 2**  $Bel'$  を  $(m_I, \mathcal{W}_I)$  によって (2.50) 式で更新する.

$Bel'$  に対応する信念構造を  $(m', \mathcal{W}')$  とし,  $\mathcal{W}' = \mathcal{W}_I = \{B_1, \dots, B_n\}$  が  $\Omega$  の分割であるとき, 更新結果  $Bel$  に対応する信念構造  $(m, \mathcal{W})$  の焦点要素の集合  $\mathcal{W}$  は, 次で与えら

れる.

$$\mathcal{W} = \{B_1, \dots, B_n, \Omega\} \quad (2.51)$$

(証明)  $\text{Bel}'(B_i) = m'(B_i)$  に注意して,

$$\begin{aligned} \text{Bel}(B_i) &= \sum_{B_j} m_I(B_j) \{1 + k_{B_j} \text{Bel}'(\overline{B_j})\} \{\text{Bel}'(B_i \cup \overline{B_j}) - \text{Bel}'(\overline{B_j})\} \\ &= m_I(B_i) \{1 + k_{B_i} \text{Bel}'(\overline{B_i})\} \{1 - \text{Bel}'(\overline{B_i})\} \\ &= m'(B_i) m_I(B_i) \{1 + k_{B_i} - k_{B_i} m'(B_i)\} \end{aligned} \quad (2.52)$$

ここで,

$$0 \leq k_{B_j} \leq \frac{1}{1 - \text{Bel}(\overline{B_j})} \left( = \frac{1}{m'(B_j)} \right)$$

$C \subset B_i$  のとき  $\text{Bel}(C \cup \overline{B_i}) = \text{Bel}(\overline{B_i})$ , また  $j \neq i$  に対して  $C \cup \overline{B_j} = \overline{B_j}$  であることに注意すると,

$$\begin{aligned} \text{Bel}(C) &= \sum_{B_j} m_I(B_j) \{1 + k_{B_j} \text{Bel}'(\overline{B_j})\} \{\text{Bel}(C \cup \overline{B_j}) - \text{Bel}(\overline{B_j})\} \\ &= 0 \end{aligned} \quad (2.53)$$

以上から, 更新後の信念においても focal element であるのは  $B_i$  自身のみであり,  $B_i$  の真部分集合は focal element となり得ないことがわかる.

つぎに,

$$C \supset B_{C_1}, \dots, B_{C_l}, l < n, \forall B_{C_j} \in \{B_1, \dots, B_n\}$$

を満たす任意の  $\Omega$  の部分集合  $C$  をとると,  $B_j, j = 1, \dots, n$  は, (i)  $B_j \subset C$ , (ii)  $B_j \subset \overline{C}$ , (iii)  $B_j \cap C \neq \phi$  かつ  $B_j \cap \overline{C} \neq \phi$ , のいずれかの条件を満たす. ここで条件 (i) が成立している場合,

$$\begin{aligned} \text{Bel}(B_j) &= m_I(B_j) \{1 + k_{B_j} \text{Bel}'(\overline{B_j})\} \{\text{Bel}'(C \cup \overline{B_j}) - \text{Bel}'(\overline{B_j})\} \\ &= m_I(B_j) \{1 + k_{B_j} \text{Bel}'(\overline{B_j})\} \{1 - \text{Bel}'(\overline{B_j})\} \\ &= m(B_j) \end{aligned}$$

(ii) が成り立つ場合には,  $C \cup \overline{B_j} = \overline{B_j}$  であることから,

$$\begin{aligned} \text{Bel}(B_j) &= m'(B_j)\{1 + k_{B_j}\text{Bel}'(\overline{B_j})\}\{\text{Bel}'(C \cup \overline{B_j}) - \text{Bel}'(\overline{B_j})\} \\ &= 0 \end{aligned}$$

(iii) が成立する場合には,  $\text{Bel}'(C \cup \overline{B_j}) = \text{Bel}'(\overline{B_j})$  であるから,

$$\begin{aligned} \text{Bel}(B_j) &= m'(B_j)\{1 + k_{B_j}\text{Bel}'(\overline{B_j})\}\{\text{Bel}'(C \cup \overline{B_j}) - \text{Bel}'(\overline{B_j})\} \\ &= 0 \end{aligned}$$

したがって,

$$\begin{aligned} \text{Bel}(C) &= \sum_{B_j} m_I(B_j) \{1 + k_{B_j}\text{Bel}'(B_j)\} \{\text{Bel}'(C \cup \overline{B_j}) - \text{Bel}'(B_j)\} \\ &= \sum_{B_j} m(B_j) \end{aligned} \quad (2.54)$$

このことは, (i), (ii), (iii) いずれの場合においても  $C$  が焦点要素ではないことを示している.

さらに,  $\Omega$  自身が焦点要素としてなりうることは,

$$\sum_{B_j} m(B_j) \leq 1$$

であることからわかる.

(証明終わり)

## 数値例

Dubois と Prade の更新 (2.33) 式に Dempster の条件付け規則 (2.25) 式を適用した場合に発生しうる問題点を示したあと, 本論文で提案した条件付け規則 (2.36),(2.37) 式により, その問題点が改善されることを示す.

fod を  $\Omega = \{R, G, Y\}$  とする. 事前の信念を  $\text{Bel}', \text{Bel}'$  に対応する bbm を  $m'$  とし, 獲得情報を  $(m_I, W_I)$  で表す. 以下では, 信念の構造をわかりやすくするため, 信念を bbm で表す. 更新結果を  $m$  とする.

**例 2** 事前の信念構造が

$$m'(\{R, G\}) = 0.2, m'(\{G\}) = 0.8$$

と与えられているとき, これを獲得情報

$$m_I(\{R, G\}) = 0.2, m_I(\{Y\}) = 0.8$$

によって更新したいものとする. このとき, (2.33) 式に Dempster の条件付け (2.25) 式を適用すると, 次式を得る.

$$\text{Bel}(\{G\}) = m_I(\{R, G\})\text{Bel}'(\{G\}|\{R, G\}) + m_I(\{Y\})\text{Bel}'(\{G\}|\{Y\}) \quad (2.55)$$

$\text{Bel}'(\{Y\}) = \text{Bel}'(\{R, G\}) = 1$  により, 上式のうち  $\text{Bel}'(\{G\}|\{Y\})$  は定義されない. したがって, (2.55) 式は計算できない.

一方, (2.33) 式に本論文で提案した条件付け (2.36) 式を適用すると,  $\text{Bel}(\{G\}) = 0$ ,  $\text{Bel}(\{R\}) = 0$ ,  $\text{Bel}(\{R, G\}) = 0.9$  となることから,

$$m(\{R, G\}) = 0.2, m(\Omega) = 0.8$$

を得る.

このように, (2.36) 式を更新に用いることによって, 獲得情報の焦点要素としてあらゆる命題を用いることが可能である.

**例 3** 事前の信念構造

$$m'(\{R, G\}) = 0.2, m'(\{Y\}) = 0.8$$

を獲得情報

$$m_I(\{R, G\}) = 0.9, m_I(\{Y\}) = 0.1$$

によって更新する場合を考える. このとき, (2.33) 式に Dempster の条件付け (2.25) 式を適用すると,

$$m(\{R, G\}) = 0.9, m(\{Y\}) = 0.1$$

となる。つまり、更新結果  $m$  は、 $m'$  にはよらず、獲得情報  $m_I$  のみによって定まることになる。

このことは次の問題を示している。事前の認識  $m'$  が、これまでに蓄積した正しい情報を表現するものであるとしよう。また、獲得した情報  $m_I$  が全くの誤りであり、 $m'$  の所有者は  $m_I$  が誤りであることを知りえないとする。このとき  $m'$  を  $m_I$  で更新するならば、これまでの正しい情報が一度の更新によって全て失われることになる。

これに対し、(2.33) 式に (2.36) 式を適用すると、

$$m(\{R,G\}) = 0.18 + 0.144k_{R,G}$$

$$m(\{Y\}) = 0.08 + 0.016k_Y$$

$$m(\Omega) = 0.74 - 0.144k_{R,G} - 0.016k_Y$$

ここで、

$$0 \leq k_{R,G} \leq 5, 0 \leq k_Y \leq \frac{5}{4}$$

つまり、従来 Dempster の条件付けを用いていた際には情報  $(m_I, W_I)$  の解釈が一義的であったのに対し、本論文で提案した条件付け (2.36),(2.37) 式を用いる場合にはパラメータの値を選択することによって、柔軟に情報を解釈することが可能となる。実際の問題においては、更新後の信念をもとに意思決定を行う。意思決定に際し、選択肢、決定基準、決定に伴うリスクを考慮して、リスクを最小化する決定を下す必要がある。与えられた情報から、リスク最小の決定をするために、決定基準に応じて適切に更新のパラメータを設定しなければならない。



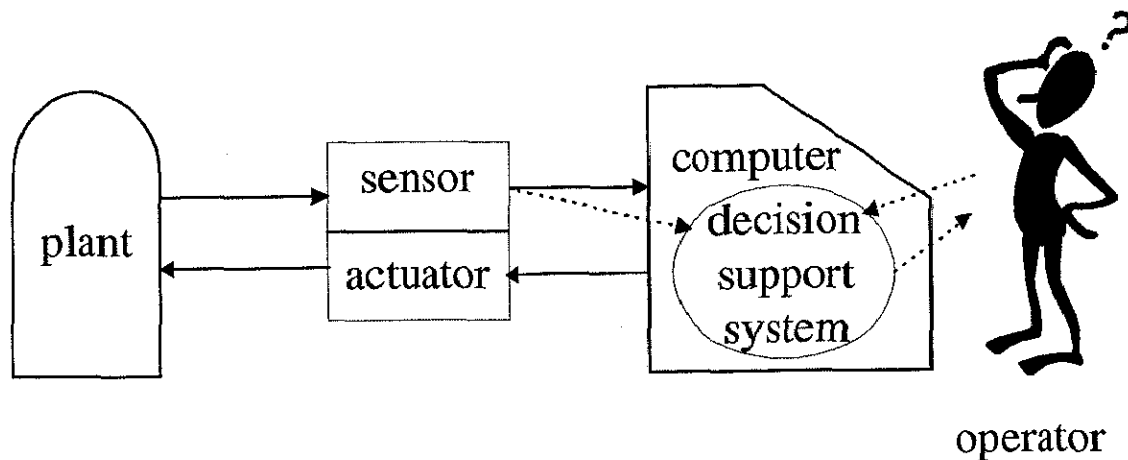


Figure 2.1: Supervisory Control System

## 2.3 プラント安全制御における合理的状況判断

### および意思決定

#### 2.3.1 問題設定

Fig. 2.1に示すような、プラント（制御対象）、オペレータ（意思決定者）、情報の収集やその提示などを行うセンサ系（情報源、意思決定支援のコンピュータなども含む）からなるシステムを考える。オペレータは、センサ系が示す情報に基づいてプラント状態を推定し、決定を下し、制御を行う。ここで、プラントは一般に大規模複雑であるため、センサ系から得られる情報は、一部が欠落したり、矛盾を含むことがある。このような状況のもとでプラントの安全制御を行うためには、情報の不完全さや不確かさをいかに意思決定に反映させるかが問題となる。

プラントの取りうる状態の集合を、 $\Omega = \{S, U\}$  とする。  $S$  はプラントが安全であること、 $U$  はプラントが危険であることを表す。プラント状態が  $S$  ならばプラントの稼働を継続してよいが、もしプラントの状態が  $U$  ならば、プラントは停止されるべきものとする。

オペレータが選択可能な行動の集合を  $A = \{GO, SD\}$  とする。  $GO$  はプラントの稼働

を継続すること、 $SD$ はプラントを停止することを表す。

なお、プラントの稼働を継続すべきか、あるいは停止すべきかは、プラント状態の観測情報に基づいて判断されるべきものとする。

### 2.3.2 決定者の行動に伴う損失

プラントの稼働に関するオペレータの決定を、プラントが被る損失の期待値によって評価する。

オペレータの行動が与えるプラントの損失を確率変数  $Z$  で表す。ここで、 $Z$  の実現値  $z$  は  $\Omega \times A$  上の非負実数値関数として定義し、安全なプラントを稼働しても損失は生じないことと、危険なプラントを適切に停止すれば事故による損害を免れること、すなわち  $z(S, GO) = 0$ ,  $z(U, SD) = 0$  を仮定する。決定の選択肢 ( $GO$  または  $SD$ ) への支持の度を確率測度  $P$  で表すと、損失の期待値  $E[Z]$  は次式で与えられる。

$$E[Z] = P(GO)\beta_{GO} + P(SD)\beta_{SD} \quad (2.56)$$

ただし、

$$\beta_{GO} = P(U|GO)z(U, GO) \quad (2.57)$$

$$\beta_{SD} = P(S|SD)z(S, SD) \quad (2.58)$$

(2.56)式において、 $P(GO)$ ,  $P(SD)$  はオペレータが持つプラント状態に関する認識に基づいて与えられる。プラントの構造が大規模複雑である場合には、プラントが安全であるか、危険であるかの判断が困難なことがある。このような状況は、証拠理論の測度を用いると表現しやすい。

### 2.3.3 安全制御方策とその下での期待損失

オペレータは、プラントに関する自身の知識および観測情報から  $\Omega$  上の bbm  $m$  を構成するものとする。また、オペレータは、プラントが安全とも危険とも判断できない場合でも、GO あるいは SD いずれかの決定を下さなければならないものとする。プラント状態をオペレータが断定できないときの結論の与え方には、Fig. 2.2 に示す二つのタイプ、すなわち安全確認型 (safety-preservation: SP) と危険検出型 (fault-warning: FW) の安全制御方策がある [27].

Safety-Control Policy	Belief		
	$S$	$\Omega$	$U$
Safety-Preservation Type	GO	SD	
Fault-Warning Type	GO		SD

Figure 2.2: Safety-Control Policies

SP 方策は、プラントの安全が確認できるときのみプラントの稼働を許す方策であり、リスク回避的な性質を持つ [27]. SP 方策のもとでは、意思決定の選択肢 (GO または SD) に対するオペレータの支持の度合は、 $A$  上の bbm  $a$  として次式で与えられる [27].

$$\begin{aligned}
 a(\{GO\}) &= m(\{S\}) \\
 a(\{SD\}) &= m(\{U\}) + m(\Omega)
 \end{aligned}
 \tag{2.59}$$

ここで bbm  $a$  の焦点要素  $\{GO\}$ ,  $\{SD\}$  はいずれも singleton であることから、 $a$  は  $A$  上の確率密度関数とみなすことができる [24]. 実際、 $P(GO) = a(\{GO\})$  を満たす確率測

度  $P$  の存在が証明できる [128].

$A$  上の確率測度  $P$  が  $P(GO) = a(\{GO\})$  を満たすとき,  $P(SD) = a(\{SD\})$  である. (2.59) 式より,  $a(\{GO\}) = \text{Bel}(\{S\})$ ,  $a(\{SD\}) = \text{Pl}(\{U\})$  であり, SP 方策のもとでプラント損失の期待値 (2.56) 式は次式で与えられる.

$$E[Z|SP] = \text{Bel}(\{S\})\beta_{GO} + \text{Pl}(\{U\})\beta_{SD} \quad (2.60)$$

一方, FW 方策は, プラントの危険が確信されないかぎりプラントを稼働させてよいとする方策であり, リスク愛好的な性質を持つ [27]. この方策のもとで  $A$  上の bbm  $a$  は次のようになる.

$$\begin{aligned} a(\{GO\}) &= m(\{S\}) + m(\Omega) \\ a(\{SD\}) &= m(\{U\}) \end{aligned} \quad (2.61)$$

(2.60) 式と同様に, FW 方策のもとでプラントが被る損失の期待値は次式で与えられる.

$$E[Z|FW] = \text{Pl}(\{S\})\beta_{GO} + \text{Bel}(\{U\})\beta_{SD} \quad (2.62)$$

### 2.3.4 情報統合の最適化

より良い意思決定を行うためには, すでに持っている情報に加えて, 新たに情報を獲得することも必要である.

2.2.2 節で示した情報統合を利用して, プラント状態に関してオペレータが持つ信念  $m'$  とセンサ系からの情報  $m_I$  を統合して  $m$  を得る場合を考える.

Inagaki の一般化統合規則 [27] を改めて示す.

$$\begin{cases} m(E) = \{1 + kq(\phi)\}q(E), E \neq \Omega, \phi \\ m(\Omega) = \{1 + kq(\phi)\}q(\Omega) + \{1 + kq(\phi) - k\}q(\phi) \\ 0 \leq k \leq \frac{1}{1 - q(\phi) - q(\Omega)} \equiv k_{max} \end{cases} \quad (2.63)$$

ただし,

$$q(E) = \sum_{B \cap C = E} m'(B)m_I(C), \forall E \subset \Omega \quad (2.64)$$

ここでは, (2.63) 式のパラメータ  $k$  の値をいかに決定すべきかを考える必要がある. 以下に示すように,  $k$  のとるべき値は, 安全制御方策のタイプ (SP あるいは FW) に応じて選択されなければならない.

### (1) SP 方策

(2.59), (2.63) 式から次式を得る.

$$E[Z|SP] = \beta_{SD} + (\beta_{GO} - \beta_{SD})q(\{S\}) + (\beta_{GO} - \beta_{SD})q(\{S\})q(\phi)k \quad (2.65)$$

$$0 \leq k \leq k_{max} \quad (2.66)$$

$E[Z|SP]$  を最小化する統合規則, すなわち SP 方策のもとでの最適統合規則を求めたい. パラメータ  $k$  の値を一つ定めるとそれに応じて統合規則が一つ定まる. そこで, (2.65) 式が  $k$  に関する一次式であることと,  $q(\phi) \geq 0$ ,  $q(\{S\}) \geq 0$  であることに注意すれば, パラメータ  $k$  の最適値は,  $\beta \equiv \beta_{GO} - \beta_{SD}$  の符号に依存して, つぎのようになる.

$$\begin{cases} \beta \geq 0 \text{ のとき } k = 0 \\ \beta < 0 \text{ のとき } k = k_{max} \end{cases} \quad (2.67)$$

ここで,

$$\beta \geq 0 \Leftrightarrow \frac{P(U|GO)}{P(S|SD)} \geq \frac{z(S, SD)}{z(U, GO)} \quad (2.68)$$

### (2) FW 方策

FW 方策のもとでの  $Z$  の期待値  $E[Z|FW]$  は, 次式で与えられる.

$$E[Z|FW] = \beta_{GO} - (\beta_{GO} - \beta_{SD})q(\{U\}) - (\beta_{GO} - \beta_{SD})q(\{U\})q(\phi)k \quad (2.69)$$

$$0 \leq k \leq k_{max} \quad (2.70)$$

したがって、 $E[Z|FW]$ を最小化する統合規則，すなわちFW方策のもとでの最適統合規則は、(2.63)式における $k$ を以下のように定めることによって得られる。

$$\begin{cases} \beta \geq 0 \text{ のとき } k = k_{max} \\ \beta < 0 \text{ のとき } k = 0 \end{cases} \quad (2.71)$$

ところでInagaki[27]によると、プラントの安全性を最大化するパラメータ $k$ は、SP方策のもとでは $k = 0$ 、FW方策のもとでは $k = k_{max}$ である。しかし、「プラントの安全性を最大限確保する」ことは必ずしも「期待損失を最小化する」ことを意味しない。実際、このことは(2.67)、(2.71)式の示すところでもある。

また、ある固定された $\beta$ に対してSP(FW)方策のもとで期待損失を最小化する $k$ の値は、FW(SP)方策のもとでは期待損失を最大化する最悪の選択になる。すなわち、 $k$ の値と安全制御方策とを協調させて選択する必要がある。

**例 4** 事故，誤停止による損失がつきのように与えられるとする。

$$z(S, SD) = 6000, z(U, GO) = 10000 \quad (2.72)$$

事前にオペレータが持っている情報が

$$m'(\{S\}) = 0.5, m'(\{U\}) = 0.1, m'(\Omega) = 0.4$$

なるbbmで表現できるとする。また、センサ系から得られた情報が

$$m_I(\{S\}) = 0.1, m_I(\{U\}) = 0.7, m_I(\Omega) = 0.2$$

で表現されるとする。また、 $P(S|SD) = 0.5$ 、 $P(U|GO) = 0.1$ とする。このとき、(2.63)式におけるパラメータ $k$ は次式を満たす。

$$0 \leq k \leq \frac{25}{14}$$

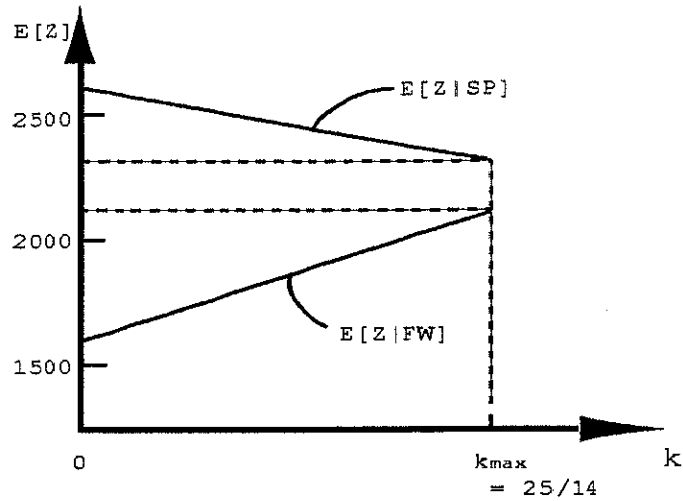


Figure 2.3: Optimal Value of  $k$  under  $z(S, SD) = 6000$  and  $z(U, GO) = 10000$

各決定方策のもとでの期待損失  $E[Z|SP]$ ,  $E[Z|FW]$  が  $k$  に依存する様子を Fig. 2.3 に示す. Fig. 2.3 において, SP 方策の下では  $k = 0$  とするとプラントの安全性をもっとも高めるものの, 期待損失も最大になる. FW 方策の下では,  $k = k_{max}$  とすると, プラントの安全性が最大限確保されるが期待損失を最大化する最悪の規則となる.

**例 5** 例 4 における (2.72) 式のうち,  $z(S, SD)$  のみを変更した次の状況を考えてみる.

$$z(S, SD) = 1000, z(U, GO) = 10000 \quad (2.73)$$

このとき,  $E[Z|SP]$ ,  $E[Z|FW]$  とパラメータ  $k$  との関係を表すと, Fig. 2.4 に示すようになる. この場合には, プラントの安全性を最も高く保つべく選択した  $k$  の値が, 期待損失も最小化するものとなっている. また, Fig. 2.3 ならびに Fig. 2.4 から, ある方策, たとえば SP(FW) 方策のもとで期待損失を最小化する  $k$  の値は, FW(SP) 方策のもとでは期待損失を最大化することもわかる. したがって, 期待損失を最小にする  $k$  を選択するには, オペレータの安全制御方策が SP, FW のいずれであるかが明確になっている必要がある.

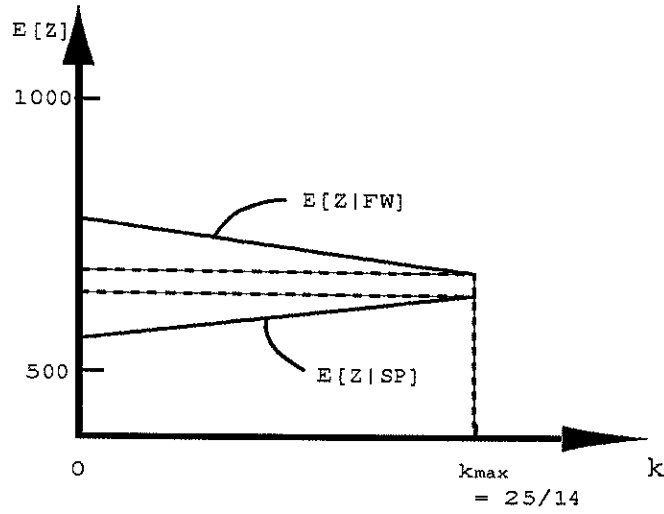


Figure 2.4: Optimal Value of  $k$  under  $z(S, SD) = 1000$  and  $z(U, GO) = 10000$

### 2.3.5 情報更新の最適化

プラントの状態に関する信念は、ビリーフ関数としても表される。2.2.4 節に示したように、オペレータがあらかじめプラント状態に関して抱いている信念  $Bel'$  は、センサ系から与えられる情報  $m_I$  によって、次式で与えられる  $Bel$  のように更新される。

$$Bel(E) = \sum_{B \subset \Omega} m_I(B) Bel'(E|B), \forall E \subset \Omega \quad (2.74)$$

なお、

(i)  $Bel'(\bar{B}) < 1$  のとき、

$$\begin{cases} Bel'(A|B) = \{1 + k_B Bel'(\bar{B})\} \{Bel'(A \cup \bar{B}) - Bel'(\bar{B})\}, A \neq \Omega \\ Bel'(\Omega|B) = 1 \\ 0 \leq k_B \leq \frac{1}{1 - Bel'(\bar{B})} \end{cases} \quad (2.75)$$

(ii)  $Bel'(\bar{B}) = 1$  のとき、

$$\begin{cases} Bel'(A|B) = 0, A \subset \Omega \\ Bel'(\Omega|B) = 1 \end{cases} \quad (2.76)$$



(2.75) 式に示されるように、上式の  $\text{Bel}'(E|B)$  は連続値パラメータ  $k_B$  を持つが、 $k_B$  は  $m_I$  の焦点要素  $B$  ごとに決められるものである。したがって、各焦点要素  $B$  に対応するパラメータ  $k_B$  の値の与え方が重要となる。ここでは、2.3.4 節と同様に、 $m'$  と  $m_I$  を用いて、オペレータの認識に基づく行動がプラントに与える損失の期待値を (2.60), (2.62) 式から計算する。

### (1) SP 方策

(2.74), (2.75), (2.76) 式より、

$$\text{Bel}(\{S\}) = \xi_S + \gamma_S k_S \quad (2.77)$$

$$0 \leq k_S \leq \frac{1}{1 - \text{Bel}'(\{U\})} \equiv k_{Smax}$$

ただし、

$$\xi_S = m_I(\{S\})(1 - m'(\{U\})) + m_I(\Omega)m'(\{S\}) \quad (2.78)$$

$$\gamma_S = m_I(\{S\})(1 - m'(\{U\}))m'(\{U\}) \quad (2.79)$$

このことと (2.60) 式から、次式を得る。

$$E[Z|SP] = \beta_{SD} + \beta\xi_S + \beta\gamma_S k_S \quad (2.80)$$

したがって、期待損失  $E[Z|SP]$  を最小化する更新規則、すなわち SP 方策のもとでの最適更新規則は  $\beta (= \beta_{GO} - \beta_{SD})$  の符号に依存し、(2.74) 式における  $k_S$  の値を次のように定めることによって得ることができる。

$$\begin{cases} \beta \geq 0 \text{ のとき } & k_S = 0 \\ \beta < 0 \text{ のとき } & k_S = k_{Smax} \end{cases} \quad (2.81)$$

## (2) FW 方策

同様にして, (2.74), (2.62), (2.75), (2.76) 式から  $E[Z|FW]$  は次式で与えられる.

$$E[Z|FW] = \beta_{GO} - \beta\xi_U - \beta\gamma_U k_U \quad (2.82)$$

$$0 \leq k_U \leq \frac{1}{1 - \text{Bel}'(\{S\})} \equiv k_{Umax}$$

ここに,

$$\xi_U = m_I(\{U\})(1 - m'(\{S\})) + m_I(\Omega)m'(\{U\}) \quad (2.83)$$

$$\gamma_U = m_I(\{U\})(1 - m'(\{S\}))m'(\{S\}) \quad (2.84)$$

$k_U$  をつぎのように与えると, FW 方策のもとでの最適更新規則を与えることができる.

$$\begin{cases} \beta \geq 0 \text{ のとき } k_U = k_{Umax} \\ \beta < 0 \text{ のとき } k_U = 0 \end{cases} \quad (2.85)$$

以上のことから, SP 方策, FW 方策のいずれの場合においても, 損失の期待値を最小化するパラメータの値は,  $\beta (= \beta_{GO} - \beta_{SD})$  の符号に依存することが明らかになった.

ビリーフ関数の更新規則 (2.74), (2.75), (2.76) 式を用いる際の特徴として, SP 方策のもとでは, 期待損失は  $k_S$  のみに依存し,  $k_U$  には無関係であることがあげられる. 逆に, FW 方策のもとでの期待損失は  $k_U$  のみに依存し,  $k_S$  には無関係である.

このことはビリーフ関数の条件付けを用いる「情報更新」と, bbm の統合規則に基づく「情報統合」との相違を明らかにするものである. 一般に, bbm の統合規則 (2.63) 式を用いて情報更新を行う場合には, Table 2.1 に示すように,

1.  $\beta$  の符号
2. 安全制御方策

の両者に応じて期待損失を最小化するパラメータ  $k$  を選択する必要がある. これに対し, ビリーフ関数の条件付けを利用し, (2.74), (2.75), (2.76) 式によって情報更新を行う場

Table 2.1: Optimal Value of  $k$ 

	SP Policy	FW Policy
$\beta \geq 0$	$k = 0$	$k = k_{max}$
$\beta < 0$	$k = k_{max}$	$k = 0$

合には、いずれの安全制御方策に対しても期待損失を最小化するように  $(k_S, k_U)$  の組を選択できる (Table 2.2) . したがって、オペレータの安全制御方策が SP, FW のいずれであるかわからない場合でも、更新パラメータ  $(k_S, k_U)$  を、期待損失最小化の意味で最適な値に設定することができる. さらに、状況によってオペレータが安全制御方策を随時変更することがあっても、パラメータの値をその都度変更する必要はない.

Table 2.2: Optimal Value of  $(k_S, k_U)$ 

	SP Policy	FW Policy
$\beta \geq 0$	$k_S = 0$	$k_U = k_{Umax}$
$\beta < 0$	$k_S = k_{Smax}$	$k_U = 0$

**例 6** 例 5 と同じ状況において、(2.74), (2.75), (2.76) 式を用いて  $m'$  を  $m_I$  によって更新する場合、 $k_S, k_U$  はそれぞれ次式を満たす.

$$0 \leq k_S \leq \frac{10}{9}, 0 \leq k_U \leq 2$$

このとき、 $k_S, k_U$  に対する  $E[Z|SP], E[Z|FW]$  の様子を Fig. 2.5 および Fig. 2.6 に示す. これらの図から明らかなように、 $E[Z|SP]$  は  $k_S$  に依存し、 $E[Z|FW]$  は  $k_U$  に依存する. そこで、 $E[Z|SP], E[Z|FW]$  をそれぞれ最小化するように  $k_S, k_U$  の組を選んでおけば、オペレータの安全制御方策が SP, FW のいずれであっても、期待損失を最小にすることが可能である.

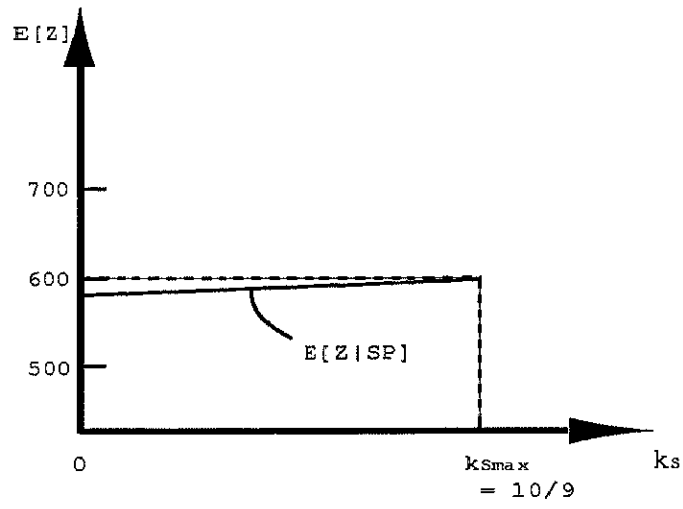


Figure 2.5: Optimal Value of  $k_S$

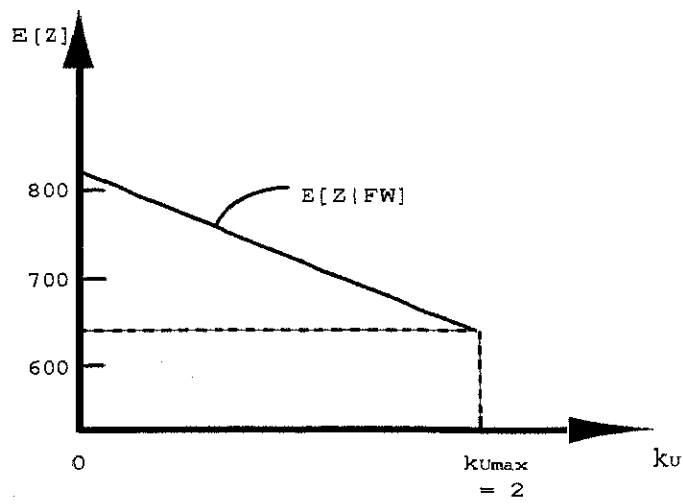


Figure 2.6: Optimal Value of  $k_U$

## 2.4 混合戦略による安全制御での情報の統合と更新

### 2.4.1 問題設定

2.3節の問題設定で示したのと同様に、大規模複雑プラントの安全制御の問題を考える。プラントの取りうる状態を $\Omega = \{S, U\}$ 、オペレータの行動の選択肢として $A = \{GO, SD\}$ を考え、期待損失を最小化する情報統合と情報更新の規則を明らかにする。

ただし、本節では、オペレータは安全制御方策を確率的に選択するものとする。

### 2.4.2 安全制御方策に対する混合戦略

2.3.3節で述べたように、プラントが安全とも危険とも判断できない場合の安全制御方策としてSP型とFW型があり、つぎの関係が成立する。

$$\begin{cases} P(GO|SP) = \text{Bel}(\{S\}) \\ P(SD|SP) = \text{Pl}(\{U\}) \end{cases} \quad (2.86)$$

$$\begin{cases} P(GO|FW) = \text{Pl}(\{S\}) \\ P(SD|FW) = \text{Bel}(\{U\}) \end{cases} \quad (2.87)$$

プラント状態を特定できない場合に、GO、SDいずれの決定を下すかをあらかじめ決めておくことは、オペレータにとって一般には困難であると考えられる。そこで、ここではオペレータが安全制御方策を確率的に与えるものと仮定する。

オペレータがSP方策を選択する確率を $\alpha (0 \leq \alpha \leq 1)$ とすると「確率 $\alpha$ でSP方策を選択する混合戦略」を考えることができる。この戦略を「混合戦略 $\alpha$ 」と呼ぶことにする。 $\alpha = 0$ とした純粋戦略がいわゆるFW方策、 $\alpha = 1$ がSP方策に対応する。

2.3節では、オペレータが選択する安全制御方策が固定されていると仮定していた。すなわち、混合戦略の特別な場合としての純粋戦略により安全制御方策を選択するという状況を考察していたことになる。

### 2.4.3 混合戦略のもとでの期待損失

プラントの稼働・停止に関するオペレータの決定の適否を，プラントが被る期待損失によって評価するものとしよう。

プラントの損失  $Z$  (確率変数) の実現値  $z$  を  $\Omega \times A$  上の非負実数値関数とする。なお， $z(S, GO) = 0$ ， $z(U, SD) = 0$  を仮定する。

混合戦略  $\alpha$  のもとで，プラント損失の期待値  $E[Z]$  は次式で与えられる。

$$E[Z] = \alpha E[Z|SP] + (1 - \alpha) E[Z|FW] \quad (2.88)$$

ここで，(2.86) 式に注意すると，

$$\begin{aligned} E[Z|SP] &= P(GO|SP) \sum_{X \in \Omega} P(X|GO, SP) z(X, GO) \\ &\quad + P(SD|SP) \sum_{X \in \Omega} P(X|SD, SP) z(X, SD) \\ &= \text{Bel}(\{S\}) \beta_{GO}^{SP} + \text{Pl}(\{U\}) \beta_{SD}^{SP} \end{aligned} \quad (2.89)$$

ただし，

$$\beta_{GO}^{SP} = P(U|GO, SP) z(U, GO) \quad (2.90)$$

$$\beta_{SD}^{SP} = P(S|SD, SP) z(S, SD) \quad (2.91)$$

同様に (2.87) 式から，

$$E[Z|FW] = \text{Pl}(\{S\}) \beta_{GO}^{FW} + \text{Bel}(\{U\}) \beta_{SD}^{FW} \quad (2.92)$$

$$\beta_{GO}^{FW} = P(U|GO, FW) z(U, GO) \quad (2.93)$$

$$\beta_{SD}^{FW} = P(S|SD, FW) z(S, SD) \quad (2.94)$$

このとき，改めて

$$\beta^{SP} = \beta_{GO}^{SP} - \beta_{SD}^{SP} \quad (2.95)$$

$$\beta^{FW} = \beta_{GO}^{FW} - \beta_{SD}^{FW} \quad (2.96)$$

とし, (2.4) 式に注意すれば,  $E[Z]$  は次式で表される.

$$E[Z] = \alpha [\beta_{SD}^{SP} + \beta^{SP} \text{Bel}(\{S\})] + (1 - \alpha) [\beta_{GO}^{FW} - \beta^{FW} \text{Bel}(\{U\})] \quad (2.97)$$

## 2.4.4 情報統合の最適化

### 情報統合規則

プラント状態に関してオペレータが信念  $m'$  を持っているとき, センサ系から情報  $m_I$  が得られたとする. このとき, プラント状態に関する新たな bbm  $m$  を,  $m'$  と  $m_I$  の情報統合によって得る場合を考える. ただし, 2.2.2 節で示したように, 情報統合は次式で行われるものとする.

$$\begin{cases} m(E) = \{1 + kq(\phi)\}q(E), E \neq \Omega, \phi \\ m(\Omega) = \{1 + kq(\phi)\}q(\Omega) + \{1 + kq(\phi) - k\}q(\phi) \\ 0 \leq k \leq \frac{1}{1 - q(\phi) - q(\Omega)} \equiv k_{max} \end{cases} \quad (2.98)$$

ここで,

$$q(E) = \sum_{B \cap C = E} m'(B)m_I(C), \forall E \subset \Omega \quad (2.99)$$

### 最適な情報統合規則

2.3 節の解析結果により, オペレータの安全制御方策が純粋戦略であるとき, 期待損失を最小化する意味で最適な情報統合規則は, 安全制御方策のタイプ (SP 型, FW 型) に依存することが明らかとなっている. 本節では混合戦略  $\alpha$  のもとで最適な情報統合規則を導出する.

(2.98) 式で与えられる  $m$  が定める Bel を用いるとき, (2.97) 式はつぎのように表すことができる.

$$E[Z] = c + \beta^* k \quad (2.100)$$

$$0 \leq k \leq k_{max} \quad (2.101)$$

ここで、定数  $c$ ,  $\beta^*$  は以下のように定義される.

$$c = \alpha [\beta_{SD}^{SP} + \beta^{SP} q(\{S\})] + (1 - \alpha) [\beta_{GO}^{FW} - \beta^{FW} q(\{U\})] \quad (2.102)$$

$$\beta^* = \{ \alpha [\beta^{SP} q(\{S\}) + \beta^{FW} q(\{U\})] - \beta^{FW} q(\{U\}) \} q(\phi) \quad (2.103)$$

$c$  はベースとなる期待損失の値,  $\beta^*$  は情報統合のパラメータ  $k$  に対する期待損失の感度を表す. すなわち,  $\beta^*$  の絶対値が大きければ, 統合パラメータ  $k$  を変化させることによる期待損失の変動も大きい.

(2.100) 式の期待損失  $E[Z]$  は, 統合規則を定めるパラメータ  $k$  に関して線形であることから,  $E[Z]$  を最小にする情報統合規則は  $\beta^*$  の符号, すなわち

$$\text{sign}(\beta^*) = \begin{cases} 1, & \beta^* \geq 0 \\ -1, & \beta^* < 0 \end{cases} \quad (2.104)$$

に依存し, 次式のように定まる.

$$k = \begin{cases} 0, & \text{sign}(\beta^*) = 1 \text{ のとき} \\ k_{max}, & \text{sign}(\beta^*) = -1 \text{ のとき} \end{cases} \quad (2.105)$$

なお, ここで与えた結果は 2.3 節の結果の一般化となっている. すなわち, 通常の SP 方策 ( $\alpha = 1$ ) のときは  $\text{sign}(\beta^*) = \text{sign}(\beta^{SP})$ , FW 方策 ( $\alpha = 0$ ) のもとでは  $\text{sign}(\beta^*) = -\text{sign}(\beta^{FW})$  であることから,  $k$  の最適値はそれぞれ (2.106), (2.107) 式のように定まるが, この結果は, 2.3 節が与える結果に一致している.

$$k = \begin{cases} 0, & \text{sign}(\beta^{SP}) = 1 \text{ のとき} \\ k_{max}, & \text{sign}(\beta^{SP}) = -1 \text{ のとき} \end{cases} \quad (2.106)$$

$$k = \begin{cases} k_{max}, & \text{sign}(\beta^{FW}) = 1 \text{ のとき} \\ 0, & \text{sign}(\beta^{FW}) = -1 \text{ のとき} \end{cases} \quad (2.107)$$



## 2.4.5 純粋戦略に対する最適規則との同一性

ところで,

$$\alpha^* = \frac{\beta^{FW} q(\{U\})}{\beta^{SP} q(\{S\}) + \beta^{FW} q(\{U\})} \quad (2.108)$$

とする。  $\alpha^*$  は、混合戦略  $\alpha$  が SP, FW いずれの純粋戦略に近い性質を持つかを判定するための指標である。  $\alpha \geq \alpha^*$  を満たす混合戦略  $\alpha$  のもとでは  $sign(\beta^*) = sign(\beta^{SP})$  が成立する。このことは、  $\alpha \geq \alpha^*$  なる混合戦略  $\alpha$  に対する最適な統合規則と、純粋な SP 方策に対する最適な統合規則が一致することを表す。

同様に、  $\alpha < \alpha^*$  のとき  $sign(\beta^*) = -sign(\beta^{FW})$  となることから、  $\alpha \leq \alpha^*$  なる混合戦略  $\alpha$  のもとでの最適情報統合規則は、純粋な FW 方策の場合に一致する。

これらのことは、対応する「情報統合規則の最適性」の観点から、混合戦略  $\alpha$  は純粋戦略と同一視できることを示している。すなわち  $\alpha \geq \alpha^*$  なる混合戦略は純粋な SP 方策とみなすことができ、  $\alpha < \alpha^*$  なる混合戦略は純粋な FW 方策として扱うことができる。

ただし、SP 方策を選ぶ確率  $\alpha$  が大きいからといって、その混合戦略  $\alpha$  を純粋な SP 戦略とみなしてよいとは限らない。FW 方策についても同様である。例えば、Fig. 2.7において、  $\alpha = 0.8$  の場合を考えてみる。これは、確率 0.8 で SP 方策をとり、確率 0.2 で FW 方策をとる混合戦略であるため、表面的には SP 方策的色彩が強いように見える。しかし、仮に  $\alpha^* = 0.85$  であれば、この混合戦略  $\alpha$  は FW 型の純粋戦略と同じ最適情報統合規則を持つ。すなわち、混合戦略  $\alpha$  は本質的には FW 方策であるといえる。このことは、  $\alpha^*$  が  $q$  に依存すること、すなわち  $m'$  と  $m_I$  の統合のために  $gma$  を計算する時点ではじめて  $\alpha^*$  の値が定まることに起因する。

まとめれば、最適情報統合規則は、  $\beta^{SP}$ ,  $\beta^{FW}$ ,  $\alpha$ ,  $bbm$   $m'$ ,  $m_I$  に依存して定まることになる。また、  $m'$ ,  $m_I$  は時刻により異なりうることから、情報統合の最適規則は時刻に依存するということができる。

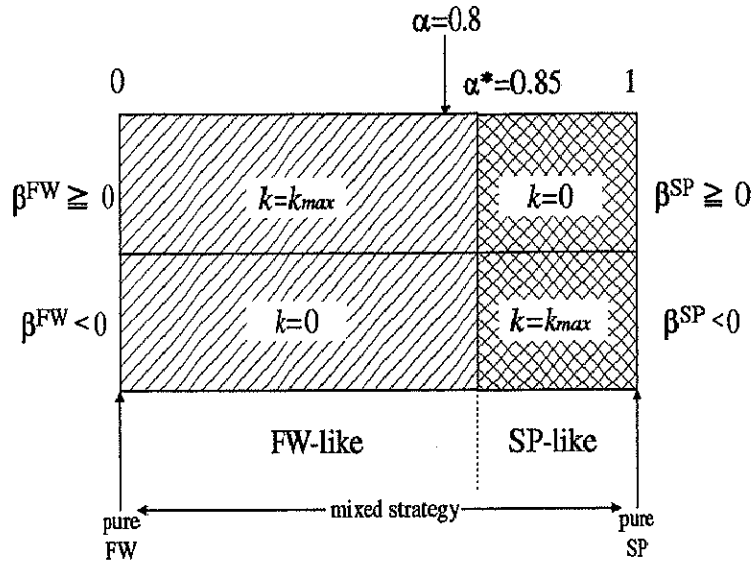


Figure 2.7: Relationship between  $\alpha$  and Optimal Value of  $k$

例 7 オペレータが,

$$m'(\{S\}) = 0.3, m'(\{U\}) = 0.6, m'(\Omega) = 0.1$$

なる bbm で表わされる信念を持っているとき, センサ系から

$$m_I(\{S\}) = 0.1, m_I(\{U\}) = 0.8, m_I(\Omega) = 0.1$$

なる情報が得られたとする. 仮に,

$$\alpha = 0.95$$

$$z(S, SD) = 1500, z(U, GO) = 10000$$

$$P(S|SD, SP) = 0.4, P(U|GO, SP) = 0.2$$

$$P(S|SD, FW) = 0.2, P(U|GO, FW) = 0.3$$

とすると,

$$\beta_{GO}^{SP} = 400, \beta_{SD}^{SP} = 2600, \beta^{SP} = -2200$$

$$\beta_{GO}^{FW} = 200, \beta_{SD}^{FW} = 3900, \beta^{FW} = -3700$$

を得る.  $m'$  と  $m_I$  を (2.98) 式の規則に基づいて統合すると,

$$\begin{aligned} \alpha^* &= \frac{\beta^{FW} q(\{U\})}{\beta^{SP} q(\{S\}) + \beta^{FW} q(\{U\})} \\ &= \frac{-3700 \times 0.62}{-2200 \times 0.07 - 3700 \times 0.62} \\ &\simeq 0.937 \end{aligned}$$

であるので,  $\alpha > \alpha^*$  である. したがって, 混合戦略 0.95 に対する最適情報統合規則は純粋な SP 方策に対するものと一致する.

実際, 得られた bbm によって期待損失を評価すると次式を得る.

$$E[Z] = 2448.4 - 9.48k \quad (2.109)$$

$$0 \leq k \leq \frac{100}{69} \quad (2.110)$$

この場合の最適統合規則は  $k = 100/69$  によって与えられる Extra 型 [133] である. この統合規則は,  $\beta^{SP} < 0$  のもとでの SP 方策に対する最適統合規則と同一である ((2.106) 式参照).

**例 8** 例 7 において,  $\alpha = 0.7$  とする. このときの混合戦略は, SP 型の安全制御方策をとる確率が 0.7 であることから, SP 的色彩の強いものといえる. しかし,  $\alpha < \alpha^*$  であるため, 混合戦略 0.7 に対する最適情報統合規則は FW 方策に対する最適規則に等しくなる.

実際, 期待損失

$$E[Z] = 2460.4 + 174.12k \quad (2.111)$$

の  $k$  の係数が正であるので, 最適な統合規則は  $k = 0$  によって与えられる Yager の規則になる. この最適規則は  $\beta^{FW} < 0$  のときの FW 方策における最適統合規則に一致する ((2.107) 式参照).

例 9 例 8 において,  $m'$  と  $m_I$  がそれぞれ

$$m'(\{S\}) = 0.5, m'(\{U\}) = 0.3, m'(\Omega) = 0.2$$

$$m_I(\{S\}) = 0.4, m_I(\{U\}) = 0.5, m_I(\Omega) = 0.1$$

である場合を考える. このとき,

$$\alpha^* \simeq 0.553$$

となる.  $\alpha = 0.7 > \alpha^*$  ゆえ, 最適情報統合規則は SP 方策の場合に一致する. 実際, 期待損失  $E[Z]$  は

$$E[Z] = 1605.6 + 101.528k \quad (2.112)$$

となり, 最適情報統合規則は  $k = 0$  が定める Yager の規則になる.

例 7, 8 は, 戦略  $\alpha$  に依存して最適情報統合規則が変わりうることを示している. 例 8, 9 は, 統合する情報  $m'$ ,  $m_I$  が変化すると最適情報統合規則も変化しうることを示している.

## 2.4.6 情報更新の最適化

### 情報更新規則

2.2.4 節で示したように, プラント状態に関するオペレータの認識  $\text{Bel}'$  は, センサ系から与えられる情報  $m_I$  によって,

$$\text{Bel}(E) = \sum_{B \subset \Omega} m_I(B) \text{Bel}'(E|B), \forall E \subset \Omega \quad (2.113)$$

で与えられる  $\text{Bel}$  として更新される. ただし,

(i)  $\text{Bel}'(\bar{B}) < 1$  のとき,

$$\left\{ \begin{array}{l} \text{Bel}'(A|B) = \{1 + k_B \text{Bel}'(\bar{B})\} \{ \text{Bel}'(A \cup \bar{B}) - \text{Bel}'(\bar{B}) \}, A \subset \Omega \\ \text{Bel}'(\Omega|B) = 1 \\ 0 \leq k_B \leq \frac{1}{1 - \text{Bel}'(\bar{B})} \equiv k_{Bmax} \end{array} \right. \quad (2.114)$$

(ii)  $\text{Bel}'(\bar{B}) = 1$  のとき,

$$\begin{cases} \text{Bel}'(A|B) = 0, A \subset \Omega \\ \text{Bel}'(\Omega|B) = 1 \end{cases} \quad (2.115)$$

### 最適な情報更新規則

混合戦略  $\alpha$  のもとで期待損失を最小化する最適更新規則を求める。

(2.113) 式が与える  $\text{Bel}(\{S\})$ ,  $\text{Bel}(\{U\})$  を (2.97) 式に代入すると次式を得る。

$$\begin{aligned} E[Z] &= \alpha [\beta_{SD}^{SP} + \beta^{SP} \text{Bel}(\{S\})] + (1 - \alpha) [\beta_{GO}^{FW} - \beta^{FW} \text{Bel}(\{U\})] \\ &= d + \alpha \beta^{SP} \gamma_S k_S - (1 - \alpha) \beta^{FW} \gamma_U k_U \end{aligned} \quad (2.116)$$

$$0 \leq k_S \leq \frac{1}{1 - \text{Bel}'(\{U\})} \equiv k_{Smax} \quad (2.117)$$

$$0 \leq k_U \leq \frac{1}{1 - \text{Bel}'(\{S\})} \equiv k_{Umax} \quad (2.118)$$

ただし,

$$\gamma_S \equiv m_I(\{S\})(1 - m'(\{U\}))m'(\{U\}) \quad (2.119)$$

$$\gamma_U \equiv m_I(\{U\})(1 - m'(\{S\}))m'(\{S\}) \quad (2.120)$$

$$d \equiv \alpha (\beta_{SD}^{SP} + \beta^{SP} \xi_S) + (1 - \alpha) (\beta_{GO}^{FW} - \beta^{FW} \xi_U) \quad (2.121)$$

$$\xi_S \equiv m_I(\{S\})(1 - m'(\{U\})) + m_I(\Omega)m'(\{S\}) \quad (2.122)$$

$$\xi_U \equiv m_I(\{U\})(1 - m'(\{S\})) + m_I(\Omega)m'(\{U\}) \quad (2.123)$$

$\gamma_S, \gamma_U$  は, それぞれ情報更新のパラメータ  $k_S, k_U$  に対する期待損失の感度を与える。また,  $d$  は情報更新規則を用いた場合でのベースとなる期待損失の値であり, (2.102) 式で定めた, 情報統合規則を用いた場合の期待損失のベース  $c$  と等しくなる。すなわち,  $d = c$  が成立する。

(2.116) 式は  $k_S, k_U$  について線形であり,  $k_S$  と  $k_U$  とは互いに独立に選択できることから,  $E[Z]$  を最小にする  $k_S, k_U$  の値は, 各々の係数の符号に依存して決まる。すなわち,

$0 \leq \alpha \leq 1$ ,  $\gamma_S \geq 0$ ,  $\gamma_U \geq 0$ に注意すれば,  $E[Z]$ を最小にする  $k_S$ ,  $k_U$ の値の組は Table 2.3のように定まる. この  $k_S$ ,  $k_U$ により定められる最適情報更新規則は,  $\beta^{SP}$ ,  $\beta^{FW}$ の符号のみに依存し,  $\alpha$ や  $m'$ ,  $m_I$ には依存しない.

Table 2.3: Optimal Value of  $(k_S, k_U)$  for a Randomized Safety-Control Policy

	$sign(\beta^{FW}) = 1$	$sign(\beta^{FW}) = -1$
$sign(\beta^{SP}) = 1$	$(k_S = 0, k_U = k_{Umax})$	$(k_S = 0, k_U = 0)$
$sign(\beta^{SP}) = -1$	$(k_S = k_{Smax}, k_U = k_{Umax})$	$(k_S = k_{Smax}, k_U = 0)$

最適な情報更新規則が安全制御方策の与え方 (戦略  $\alpha$ ) とは独立に定まるという結論は, 安全制御方策として純粋戦略 (SP型かFW型) のみを考察の対象とした場合の結果, すなわち 2.3節の結果と同じである.

**例 10** 例 7において, (2.113)式を用いた情報更新を利用して, 期待損失を評価すると

$$\xi_S = 0.07, \xi_U = 0.62, \gamma_S = 0.024, \gamma_U = 0.168$$

となることから

$$E[Z] = 2448.4 - 50.16k_S + 31.08k_U$$

$$0 \leq k_S \leq 2.5, 0 \leq k_U \leq \frac{10}{7}$$

になる. 最適な情報更新規則は,  $k_S = 2.5$ ,  $k_U = 0$ によって与えられる.

**例 11** 例 8において,  $\alpha = 0.7$ としてみると, 次式を得る.

$$E[Z] = 2460.4 - 36.96k_S + 186.48k_U$$

最適な情報更新規則は, 例 3と同じく  $k_S = 2.5$ ,  $k_U = 0$ によって与えられる.

例 12 例 9 において, (2.113) 式に基づいて情報更新を行うと,

$$\xi_S = 0.33, \xi_U = 0.28, \gamma_S = 0.084, \gamma_U = 0.125$$

となることから

$$E[Z] = 1682.6 - 129.36k_S + 138.75k_U$$

$$0 \leq k_S \leq \frac{10}{7}, 0 \leq k_U \leq 2$$

となる. このときの最適情報更新規則は,  $k_S = 10/7$ ,  $k_U = 0$  によって与えられる. この最適更新規則は,  $k_S$  については最大値 ( $k_{Smax}$ ),  $k_U$  については最小値 (0) によって与えられるという意味で例 11 と同じである.

例 7, 8 においては, 最適な情報統合規則を与える  $k$  の値が  $\alpha$  の値に依存して異なりうるといふ結果を得た. しかし, 例 10, 11 からは,  $\alpha$  の値が変わっても, 最適な情報更新規則は変化しないことがわかる.

また, 意思決定に際し必要な情報を得るために情報統合を用いる場合には, 例 8, 9 に示されるように  $m'$ ,  $m_I$  のとる値によって最適統合規則の構造が変わりうる. これに対し, 情報更新を採用すると, 例 11, 12 が示すように  $m'$ ,  $m_I$  のとる値が変更した場合でも, 最適更新規則の構造は変化しない.

このように, 情報統合規則の最適性は, 意思決定者の現時点での戦略や信念, システムに関する情報に依存するという動的な性質を持つのに対し, 情報更新規則の最適性はむしろ静的な性質を持つ. この点に, 情報統合規則と情報更新規則との間の, 本質的な相違点が見れている.

## 2.5 情報統合と情報更新との使い分け

観測や計測の結果得られる複数の情報に基づいて推論を行う際, 情報統合あるいは情報更新いずれの機構を用いるべきであろうか.

前節までに見たように、プラント安全制御問題においては、情報統合の最適規則は安全制御方策や現時刻の意思決定者の認識に依存するため、どの規則を採用すればよいかは情報統合を行う際にその都度定める必要がある。これに対し、情報更新の場合には、最適規則は安全制御方策と意思決定者の現在の信念にはよらないため、採用する規則をあらかじめ固定しておくことができる。

ここでは、情報統合・情報更新のパラメータの、期待損失への影響度を評価する。

### 2.5.1 純粋戦略のもとでの情報統合と情報更新

SP 方策のもとで、情報統合を用いる場合を考えよう。(2.65) 式において、 $\beta \geq 0$  ならば、 $k = 0$  が SP 方策の下で最適な統合規則を与えるが、 $k = k_{max}$  の場合には期待損失がもっとも大きい。逆に、 $\beta < 0$  ならば、 $k = 0$  が期待損失を最大にし、 $k = k_{max}$  が最適規則を与える。SP 方策の下では、パラメータ  $k$  の値の取り方によって、期待損失は次式で与えられる分だけ変化しうる。

$$|E[Z|SP]_{k=0} - E[Z|SP]_{k=k_{max}}| = |\beta| q(\{S\})q(\phi)k_{max} \quad (2.124)$$

FW 方策の場合には、情報統合のパラメータ  $k$  は、次式のように期待損失へ影響をもたらす。

$$|E[Z|FW]_{k=0} - E[Z|FW]_{k=k_{max}}| = |\beta| q(\{U\})q(\phi)k_{max} \quad (2.125)$$

情報更新の場合、意思決定方策が SP 型ならば、期待損失はパラメータ  $k_S$  の値に応じて変化し、次式で与えられる幅を持ちうる。

$$|E[Z|SP]_{k_S=0} - E[Z|SP]_{k_S=k_{Smax}}| = |\beta| m'(\{U\})m_I(\{S\}) \quad (2.126)$$

なお、FW 方策ならば、

$$|E[Z|FW]_{k_U=0} - E[Z|FW]_{k_U=k_{Umax}}| = |\beta| m'(\{S\})m_I(\{U\}) \quad (2.127)$$



(2.124), (2.125) 式は  $q(\cdot)$  で表現されているが, (2.126) 式と (2.127) 式は  $m'(\cdot)$  および  $m_I(\cdot)$  で記述されている. このことから, (2.124) 式も (2.125) 式も, 意思決定者の事前知識と観測情報が入れ替わっても計算結果に違いは現れないが, (2.126), (2.127) 式は,  $m'(\cdot)$  と  $m_I(\cdot)$  とが置き換わると式の値が異なる.

## 2.5.2 混合戦略のもとでの情報統合と情報更新

混合戦略  $\alpha$  のもとで, 情報統合, 情報更新のパラメータが期待損失に与える影響を評価する.

情報統合の場合, 期待損失は次式で与えられる範囲の値を取りうる.

$$\begin{aligned}
 |E[Z]_{k=0} - E[Z]_{k=k_{max}}| &= |\beta^*| k_{max} \\
 &= |\alpha\beta^{SP}q(\{S\}) - (1-\alpha)\beta^{FW}q(\{U\})| q(\phi) k_{max} \\
 &= |\alpha\beta^{SP}\kappa - (1-\alpha)\beta^{FW}(1-\kappa)| q(\phi) \quad (2.128)
 \end{aligned}$$

ただし,  $\kappa = q(\{S\})/(q(\{S\}) + q(\{U\})) \in [0, 1]$  である.

情報更新の場合, 期待損失のとりうる範囲は, つぎのようになる.

1.  $\beta^{SP}\beta^{FW} \geq 0$  の場合

$$\begin{aligned}
 |E[Z]_{k_S=0, k_U=k_{Umax}} - E[Z]_{k_S=k_{Smax}, k_U=0}| &= |\alpha\beta^{SP}\gamma_S k_{Smax} + (1-\alpha)\beta^{FW}\gamma_U k_{Umax}| \\
 &= |\alpha\beta^{SP}\tau + (1-\alpha)\beta^{FW}(1-\tau)| q(\phi) \quad (2.129)
 \end{aligned}$$

2.  $\beta^{SP}\beta^{FW} < 0$  の場合

$$\begin{aligned}
 |E[Z]_{k_S=k_{Smax}, k_U=k_{Umax}} - E[Z]_{k_S=0, k_U=0}| &= |\alpha\beta^{SP}\gamma_S k_{Smax} - (1-\alpha)\beta^{FW}\gamma_U k_{Umax}| \\
 &= |\alpha\beta^{SP}\tau - (1-\alpha)\beta^{FW}(1-\tau)| q(\phi) \quad (2.130)
 \end{aligned}$$

ここで、 $\tau = \gamma s k_{smax} / q(\phi)$  である。

(2.128)式からわかるように、情報統合の場合、 $\beta^{SP}$  と  $\beta^{FW}$  とが等しい符号を持つ場合には、この両者が期待損失の最大値と最小値との差を打ち消しあうように作用する。すなわち、(2.128)式の値は  $\min\{|\alpha\beta^{SP}q(\phi)|, |(1-\alpha)\beta^{FW}q(\phi)|\}$  よりも小さくなりうるのに対し、情報更新の場合は期待損失の最大値と最小値との差はつねに  $\min\{|\alpha\beta^{SP}q(\phi)|, |(1-\alpha)\beta^{FW}q(\phi)|\}$  以上である。したがって、パラメータの選択は、情報統合よりも情報更新の場合に、より慎重になる必要がある。

以上から、情報統合と情報更新との使い分けの方針をつぎのように与えることができる。

断片的に次々と得られた情報に基づいて随時判断する必要がある場合や、意思決定者の安全制御方策が時刻によって変化しうる場合などには、採用する規則を固定しておきたければ、情報更新規則を用いればよい。情報更新を用いるべき具体的なケースとしては、次の場合が考えられる。すなわち、システム安全制御の問題において、プロセスに異常が生じると、これが次第に拡大しうる状況を想定する場合である。このような状況では、入手された情報が、いつ得られたものであるかを把握しておく必要があるからである。

これに対し、情報獲得の時間間隔に比べプロセスの時定数が十分に大きいなど、入手した情報の時間的な順序を区別する必要がない場合や、システムを停止させた状態で保全作業や故障診断などを行う場合には、情報統合のメカニズムが有効であると考えられる。これは、情報統合を用いる場合、統合する情報を相互に置き換えても期待損失は影響されないのに対し、情報更新を用いる場合、これら二つの情報を入れ換えることにより期待損失の値が変化しうることからいえる。

## 2.6 むすび

本章では、不確実情報に基づく状況認識や意思決定に関して、証拠理論における情報処理の方法として情報統合と情報更新とがあることを示した。また、Dubois と Prade によ

る情報更新について問題点を挙げ、従来の情報更新では、オペレータが信念を更新する過程をモデル化するには不適切であることを指摘した。つぎに、従来の情報更新を拡張する規則を新たに提案した。本論文で提案した条件付けにおけるパラメータによって、獲得情報をどの程度受け入れるかを調整できるようになった。新しい情報更新規則によれば、あらゆる場合に情報更新の適用が可能であるなど、従来の情報更新規則が持つ問題点を解消できることも明らかにした。このことから、状況認識や意思決定の支援に、証拠理論の情報更新を実際に利用可能になったといえる。

さらに、プラント安全制御へ証拠理論的な情報処理機構を応用し、情報統合、情報更新のそれぞれについて期待損失を最小化する意味で最適な規則を与えた。この解析により、情報統合の最適規則がオペレータの安全制御方策の選び方に依存して定まるのに対し、情報更新の最適規則はどの方策が選ばれているかには依存せずに定まることを明らかにした。さらに、情報統合の最適規則は時刻に依存して異なりうるのに対し、情報更新の最適規則は時刻に依存しないことを示した。

これらの結果をうけ、情報統合と情報更新とを使いわけの指針を示した。すなわち、情報の入手順序を区別する必要がない場合には情報統合を適用する方がよいのに対し、時間の進行に伴い状況が徐々に変化する場合には情報更新を採用する方がよい。

本章で示した情報処理機構は、損失の期待値を最小化する「合理的」な推論を与えることができ、人間の意思決定を支援するツールの構築に役立つものと期待できる。

本論文では、情報統合と情報更新の比較を、fodの大きさが2である最もシンプルな問題設定で行った。ただし、情報統合の最適規則の動的な性質、情報更新の最適規則の静的な性質は一般的にいえることであると考えられる。情報統合は一つのパラメータにより定まるのに対し、情報更新は複数のパラメータで特定されることから、情報更新ではより柔軟に情報の解釈ができるからである。

## 第3章 意思決定と実行の状況適応

### 自動化：必要性の実験的検討

#### 3.1 はじめに

現実存在する大規模複雑システムでは、システムに異常が発生してから深刻な事故に至るまでに要する時間が短いことがあり、その場合オペレータが適切に判断し行動することは難しい。たとえば、航空機の離陸時におけるエンジン故障に伴う離陸中断 (Rejected TakeOff: RTO) もしくは継続の判断を考える。パイロットは、推力が失われたか (エンジンが停止したか)、機体の現在の速度がいくらかといったことを総合して、離陸を継続するか中断するか決断する。ただし、意思決定に許される時間余裕は数秒もなく、航空機の離陸時における判断の誤りに起因する事故は後を絶たない (たとえば、福岡空港での事故 [122])。このように、緊急性がきわめて高い状況においてシステムの安全性を確保するためには、状況の認識と判断を支援するだけでは必ずしも十分ではない。

安全管理の責任を負うのは基本的にオペレータであり、コンピュータや自動化システムは、オペレータの活動を支援する立場にある。しかし、その支援のあり方は、事態の緊急性に依存して変わりうるものであると考えられる。

Inagaki[30] は、緊急度の高い状況では自動化システムが意思決定の権限を持つ方がリスクが小さいこと、さらに、緊急度の低い状況ではヒューマンオペレータが意思決定の権限を持つ方がリスクが小さいことを、数理モデルの解析を通じて示している。このことから、Inagaki[30] は、大規模複雑システムにおいて、ヒューマンオペレータと自動化システムとの間で、状況に応じて動的に意思決定の権限を配分する形態を提案している。その

ような権限の配分に基づく自動化は、状況適応自動化 (Situation-Adaptive Automation: SAA) と呼ばれる。SAA の特徴は、Sheridan の10段階の自動化レベル [83](Table 3.1) の考え方をを用いて説明するとわかりやすい。すなわち、(1) 自動化レベルが状況に応じて動的に変化する、(2) レベル6以上の自動化をも許す場合がある、ことである。

Table 3.1: Scale of Degrees of Automation[83]

---

(1)	コンピュータの支援なしにすべてを人間が行う
(2)	コンピュータはすべての選択肢を提示する。 人間はそのうちの一つを選択して実行する
(3)	コンピュータは可能な選択肢をすべて人間に提示する
(4)	コンピュータは可能な選択肢の中から一つを選び、それを人間に提案する。 それを実行するか否かは人間が決定する
(5)	コンピュータは一つの場合を人間に提示する。人間が了承すれば、 コンピュータはそれを実行する
(6)	コンピュータは一つの場合を人間に提示する。人間が一定時間内に 実行中止を指令しない限り、コンピュータはその案を実行する
(7)	コンピュータがすべてを行い、何を実行したか人間に報告する
(8)	コンピュータがすべてを行う。人間に問われれば、 何を実行したか人間に報告する
(9)	コンピュータがすべてを行う。何を実行したか人間に報告するのは、 報告すべきであるとコンピュータが判断したときに限られる
(10)	コンピュータがすべてを行う

---

ところで、いわゆる人間中心の自動化 [4, 5, 95] は、「人間が制御の主体である」ことを前提とし、「人間が最終的な意思決定の権限を持つ」べきであるとの原則を持つ。これは、人間中心の自動化では6以上のレベルの自動化が許されないことを意味している。SAA

ではレベル6以上の自動化をも許しうることから、SAAの考え方は、人間中心の自動化の原則と部分的にはあるものの抵触する。ただし、このことはSAAが人間中心の自動化の立場からは否定されるということの意味しているのではなく、「人間が最終決定権を持つ」との原則を無条件に受け入れるべきではないことを主張するものである [30]。

ところで、数理モデルとそれの解析に基づくアプローチでは、モデル化の段階で、いくつかの仮定を設けたり、単純化するのが普通である。現実のシステムにおいて存在する、安全性に関わる要因のすべてが数理モデルに反映されているわけではない。したがって、SAAが実際に有効に機能するかは数理モデルに基づく解析だけでは必ずしも十分ではない。逆に、福岡空港での離陸失敗事故 [122] ではコンピュータが判断して自動的に離陸を継続すれば事故を防げたと考えられるからといって、事故事例のみから普遍性のある主張をすることは難しい。事故が発生したケースでは、そこに発生した事象はそもそも特殊なのであり、特定の事故から得られる教訓がどの程度一般性をもつかは自明ではないからである。

そこで、第3, 4章では、状況適応自動化の必要性や有効性を認知実験によって検証する。本章では、まず、状況適応自動化の現実的な環境における必要性を明らかにする。

3.2節では、本章での実験に利用した仮想的なプラントのシミュレータ（マイクロワールド）の概要を説明する。3.3節では、実験の計画とその方法を述べる。3.4節では、意思決定とその実行をコンピュータが代行する高レベルの自動化の必要性を、実験で観察された事例や得られたデータに基づいて示す。3.5節では、人と自動化システムとの間に発生しうるモード認識の混乱などの問題を明らかにする。

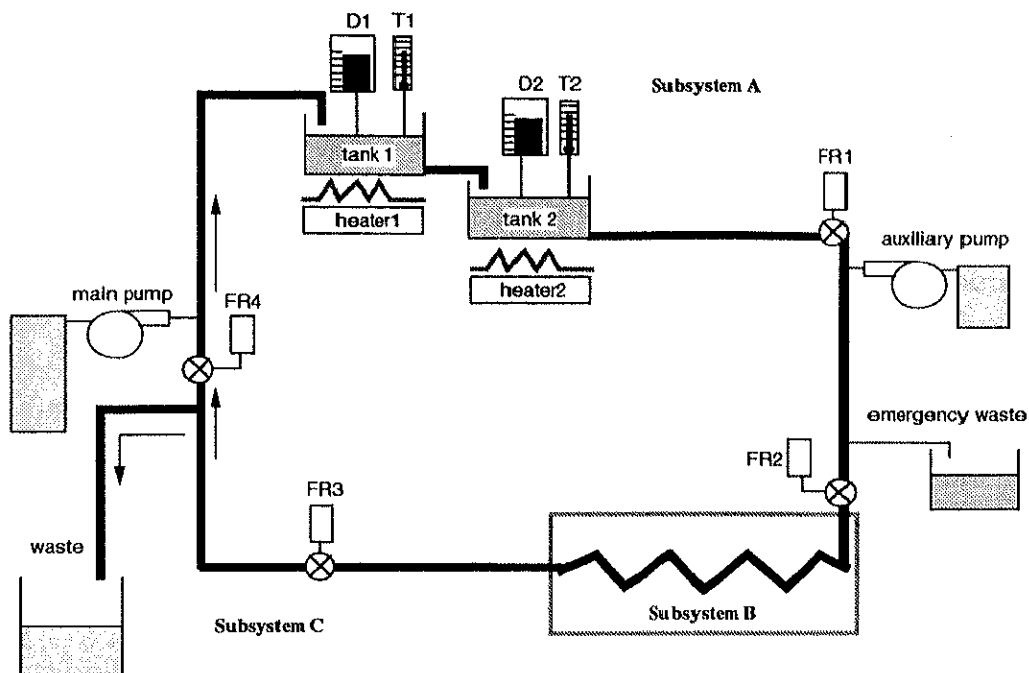


Figure 3.1: Microworld 1

## 3.2 マイクロワールド1

### 3.2.1 プラントの記述

本章で扱うプラントは、A, B, C 3つのサブシステムからなる (Fig. 3.1)。本プラントでは、ある流体が各サブシステム間を通り、循環している。サブシステムBは、その流体を利用して活動する部分である。サブシステムAは、Bに供給する流体の質を調整する系であり、サブシステムCでは、使用済み流体をAに循環させる。

サブシステムAには2槽のタンク (tank 1, tank 2) と、それぞれのタンクに一定の熱量を供給するヒーター (heater1, heater2) がある。各タンクには、水位計 (D1, D2) および温度計 (T1, T2) が装備されている。また、サブシステムAには、2つの流量計 (FR1, FR2) があり、パイプラインを通過する流体の量を知ることができる。

パイプライン中の流体の流量・温度は、通常、主ポンプ (main pump) の吐出の ON/OFF 制御により調整される。サブシステムBの入り口付近において流量が極端に多く/少なく

なった場合に備えて、補助廃棄タンク (emergency waste tank) および補助ポンプ (auxiliary pump) が用意されている。

サブシステム B はブラックボックスである。使用済みの流体は、サブシステム B から排出されるときにはその温度が室温まで下がっている。

サブシステム C には、再使用に耐えられない流体を棄てるための廃棄タンク (waste tank) がある。また、流量計が 2 つ (FR3, FR4) 備えられており、これら 2 つの流量計の示す値の差から廃棄された流体の量を知ることができる。

### 3.2.2 ヒューマンインタフェース

オペレータは、パイプライン中の流体の量や温度、タンクの水位などの値を、ヒューマンインタフェースの計器から読み取る。Fig. 3.2 は、その画面である。インタフェースのデザインは、予備実験を繰り返し、計器の形状や配置などを検討・改良した結果に基づき決定した。画面中にはパイプラインの絵が示されてはいないが、表示の意味は容易に理解できるようになっている。

### 3.2.3 オペレータのタスク

オペレータには、メインとサブの 2 つのタスクが課される。

**メインタスク** メインタスクは、サブシステム B への流体の流入量および温度が要求条件を満たすように、流体の品質を調整することである。要求条件は、つぎの 2 つである。(1) 流量が 14 から 22 (ある適当な単位で計る) におさまること、(2) 温度が 50 °C から 70 °C におさまることである。

**サブタスク** サブタスクとして、2 種類のタスクが用意されている。すなわち、(1) 英単語の書き写し (スキルベース [67] のタスク)、および (2) Fig. 3.3 に示すような問題解決



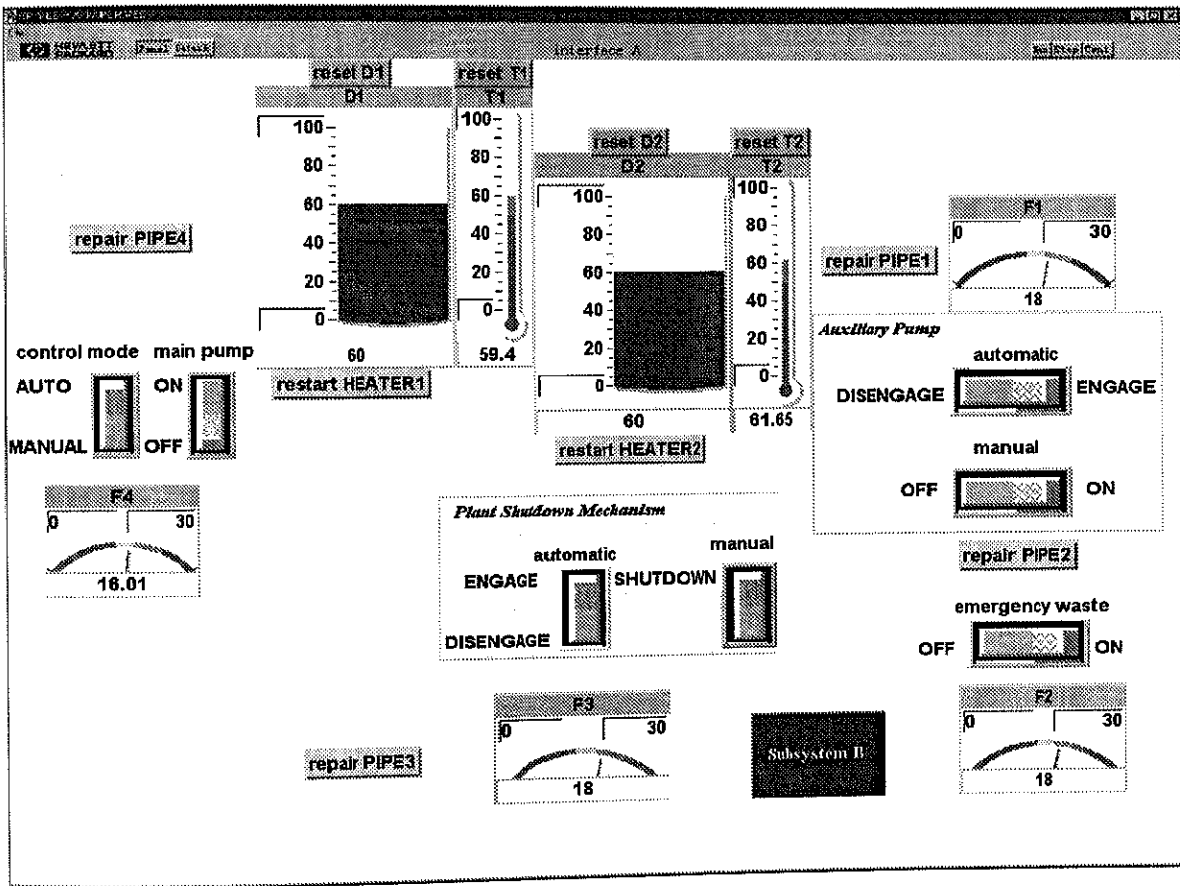


Figure 3.2: Human-Interface for Microworld 1

のタスク（ルールベース [67] もしくはナレッジベース [67] のタスク）である。オペレータは、メインタスクといずれか一方のサブタスクを同時に遂行しなければならない。

A,B,C,D,E,F,G の 7 つの命題がある。次は甲、乙、丙の 3 人が、これらの命題について研究した結果である。

1. 甲は命題 D が成立すると仮定して、これから C を証明し、さらにこれから G を証明した。
2. 乙は命題 F を仮定して、これから C を証明し、さらにこれから B を証明した。
3. 丙は命題 A を仮定して、これから G を証明し、さらにそれから D を証明し、次に F を証明した。

以上 3 人の証明はいずれも正しいものとして、たとえば D を仮定した場合、これからいくつかの命題を証明したのち G ができたとし、逆に G を仮定した場合、これからいくつかの命題を証明したのち D が証明できる場合、命題 D と命題 G とは同等であるという。甲、乙、丙の 3 人の結果を合わせて考えた範囲で、C と同等な命題の数は、C を含めずに全部でいくつあるか。

Figure 3.3: Example of a Problem Solving Task[124]

メインタスクに加えサブタスクを課すことは、次の理由による。すなわち、ここでのサブタスクは、オペレータのタスク遂行環境をより現実的なものとするためのもの、すなわち、主として遂行すべきタスクへの注意を阻害させるものと位置づける。現実のシステムでは、雑務などを含め様々な外乱によって、オペレータのワークロードや心理的な負担が増加し、本来成すべき仕事に対して集中できなくなることがあるからである。

### 3.2.4 異常事象と事故

ここでは、3 種類の異常事象が発生しうるものとする。

## 1. パイプ異常

パイプに関する異常として、流体がパイプラインから漏出するものを考える。パイプ異常は、つぎの3段階で進展する。(1) 第1段階では、パイプライン中のある箇所にひび割れが生じ、その裂け目を通過する流量の10%が失われる。(2) 第2段階では、漏出量が毎秒9%ずつ増加する。(3) 第3段階ではパイプが完全に破断し、破断箇所を通過する流体はすべて外部に漏出する。

第1段階から第2段階へは約80秒で遷移し、第2段階から第3段階へは約30秒で遷移する。また、パイプが完全に破断するなどして、サブシステムBへの流入量が0のまま一定時間がたつと、事故が発生したとみなされる。

第2段階までは、プラントの稼働中に修復が可能であるとする。コンピュータは、修復命令を受け取ったのち、約7秒で回復作業を完了する。この間、サブシステムBへの流入量が著しく減少することがあるので、必要に応じて補助ポンプを使用しなければならない。第3段階では、プラント稼働中にパイプを修復することはできず、事故の発生を防ぐためにはプラントを直ちに停止しなければならない。

## 2. 水位計の誤表示

水位計(D1, D2)の表示には誤差が含まれることがある。その誤差は正・負いずれの符号もとることがあり、誤差の絶対値は時間の進展に伴って一定の割合(w)で増加する。

オペレータが水位計の誤表示に気づいた場合、「リセット」ボタンを押せば、直ちに水位計は正常な状態に復帰する。

## 3. ヒーター故障

ヒーター(heater1, heater2)は故障により、熱をタンク中の流体に伝えられなくなることがある。タンクは断熱されていないので、ヒーターが作動しなくなった場合には、流体の温度は緩やかに室温へと降下する。

ヒーターは、「再起動」の命令を与えることによって、直ちに故障状態からもとの状

態へ復帰する。

### 3.2.5 自動化システム

このプラントでは、3つの自動化システム、(1)主ポンプ制御系、(2)補助ポンプ制御系、(3)プラント停止系、が用意されている。主ポンプ制御系はFR1が17～19を示す間一定量の水を供給し、補助ポンプ制御系はFR1が10～18.5の間一定量を供給し、プラント停止系は事故発生を防ぐためにプラント全体を停止する役割を持つ (Fig. 3.4).

オペレータは、主ポンプ等各操作要素の制御モードを選択する権限を持つ。オペレータがスキルや選好に応じて制御モードを自由に選択できるようにするために、各自動化システムの制御能力を人間のそれとほぼ同等なものにした。ここでは、自動化システムが制御に利用できる情報をFR1に限定することにより、平均的もしくはそれ以上の制御スキルを持つオペレータが自動化システムよりもよい制御成績をとり得るようにしている。

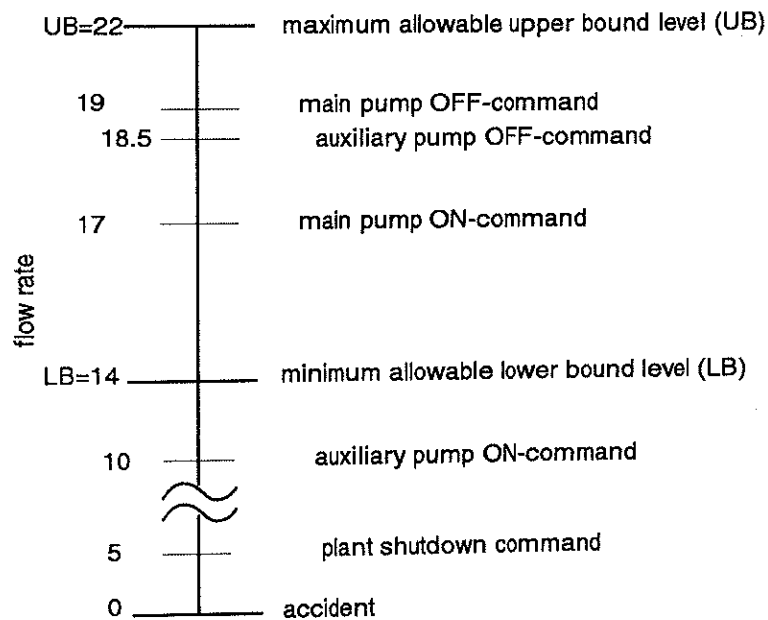


Figure 3.4: Activation/Deactivation Thresholds

## 3.3 実験の方法

### 3.3.1 計画と手順

#### 目的

複雑なタスク環境において、人がいかなる支援を必要としているか、とくに、意思決定とそれに伴う行動の遂行を自動化することが必要となるか明らかにする。また、人と自動化システムとの間に生じうる問題を示す。

#### 計画

サブタスク（英単語の書き写し、および問題解決）、異常事象の種類と数を様々な組み合わせ、多様な状況を用意し、それぞれの状況におけるオペレータの振る舞いを観察する。本実験は、探索型実験 [113] (pp. 22-29) である。

各被験者には実験参加への謝金の他、被験者のモチベーションを喚起するために、各試行において成績優秀者上位2名にボーナスを支払うこととした。このことは実験に先立ち、あらかじめ被験者に伝えておいた。

#### 手順

実験には、7名の被験者（男子大学生・大学院生）が参加した。各被験者につき、実験は5日間連続で行い、1日につき、1回300時刻（約6分半）の試行を4～6回行った。Table 3.2 は本実験のスケジュールを表している。

試行は、

1. トレーニング
2. データ収集

Table 3.2: Experimental Schedule

	#1	#2	#3	#4	#5	#6
Day 1	n	p(60)	h1(90)	p(60)	p(60)	-
Day 2	p(150)	n	h1(100)	n	p(40)	-
Day 3	h1(250)	p(130)	n	n	h2(120)	l2(100) w=+0.2
Day 4	h1(120),	h2(120),	l2(70),w2=-0.2	n	-	-
Day 4	l2(200) w=+0.2,	p(180),	p(100)			
Day 5	p stage 2(80)	h1(110)	l2(70) w=-0.1	n	l2(130)	
	l2(150) w=+0.2	p(180)	p(110)		p(60)	-
					h2(180)	

$p(x)$ : Pipe failure occurs at time point  $x$

$h_i(x)$ : Heater  $i$  fails at time point  $x$

$l_i(x) w=j$ : Level meter  $i$  begins to show erroneous value at time point  $x$ . The error of reading will increase  $j$  per a time point.

n: No any malfunction occurs in the trial.

に分かれる。

初日の全試行および2日目以降の最初の試行では、トレーニングとして、手動プロセス制御を行わせる。トレーニングの目的は、プラント制御に必要な技術を身につけ、向上させることにある。ただし、初日の5回目の試行では、各被験者に自動化システムによる制御の様子を観察させ、このシステムの能力を評価する機会を与えた。

データ収集においては、サブタスクとして、2、4日目には英単語書き写しを、3、5日目には問題解決のタスクを課す。なお、データ収集の試行では、単語書き写しタスクでは100語、問題解決のタスクでは2題をそれぞれ課すこととした。この負荷は、6分半以内に終わらせる程度の量として予備実験の結果に基づいて決定した。

メインタスクの遂行にあたっては、被験者は必要に応じて3つの自動化システムの制御モード（自動もしくは手動）を随時変更できるものとした。

### 3.3.2 評価指標

本実験では、つぎの点を評価する。

1. メインタスクの成績
2. サブタスクの成績
3. 異常事象発生から対処行動を開始するまでの時間 (Time To Detect fault: TTD)

ただし、本実験では被験者がいかなる振舞をするか観察することに主眼をおいており、メインタスク・サブタスクの成績は本章における解析では重視しない。タスクの成績は、被験者の士気を高めるためのものに過ぎない。

なお、メインタスクは、Fig. 3.5の基準によって評価する。各試行ごとに、流量、温度それぞれにつき2000点満点、合計4000点満点で評価する。事故が発生した場合には、その試行のメインタスクの評価を0点とする。

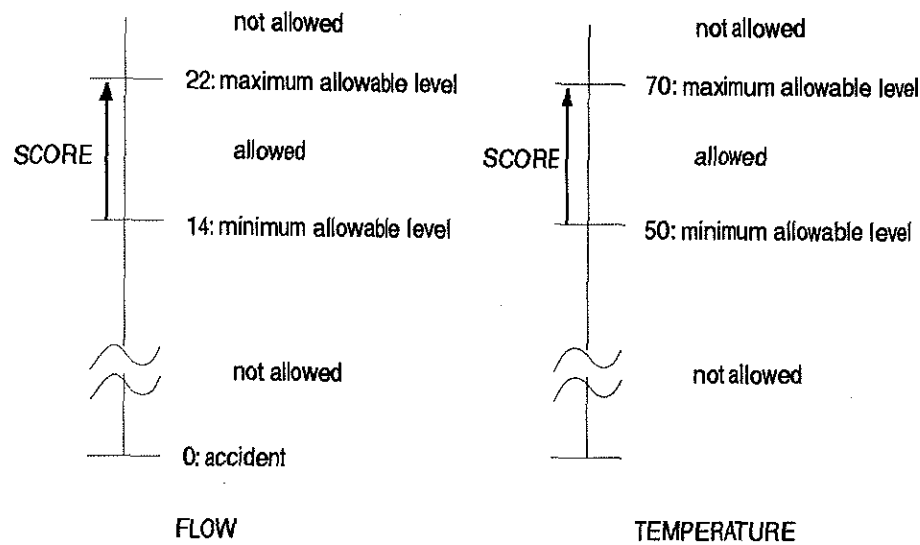


Figure 3.5: Performance Criteria

各試行では、いずれか一方のサブタスクが課せられる。サブタスクは、2000点満点で評価する。

### 3.4 高レベルの自動化の必要性

人間が状況判断のために情報収集しているうちに事故にいたってしまう場合、コンピュータが権限を持って事態に対処できるならば人間が権限を固持すべきではないと考えられる。本節では、いかに人の判断の遅れが生じるか、人間の判断の遅れがいかに事故に結びつくかを事例を通じて示す。

#### 3.4.1 人の判断の遅れ

注意が一点に集中してしまい（認知的視野狭窄：cognitive lockup, もしくは cognitive tunnel vision[80]) 状況を正しく把握するのが遅れたり失敗すると、適切な判断を下すの



が遅くなる。たとえば、つぎの事例が得られた。

#### 事例 1 被験者 TS, 2 日目第 2 試行 (異常事象なし)

この被験者は、主ポンプを 87 時刻もの間 ON にし続けた。通常は、これほど長い間ポンプを ON し続けることはない。その後主ポンプをオフにしたのは、サブシステム B への流入量が、要求条件の上限 (UB) を超過した後であった。被験者の様子を観察したところによると、被験者 TS は、注意を FR 2 に集中させすぎてしまい、主ポンプの状態を確認していなかったようである。

認知的視野狭窄は、複合事象など、複雑な状況でおこりやすい。

#### 事例 2 被験者 IN, 4 日目第 3 試行 (時刻 70 に水位計 2 誤表示 ( $w=+0.2$ ), 時刻 100 にパイプ異常)

タンク 2 の水位計誤表示が、試行開始 70 時刻 (約 90 秒) 後に発生した。表示された水位 ( $tl_2$ ) は、実際の水位 ( $tl_2$ ) と比べ低くなり、その誤差 ( $l$ ) が次第に大きくなっていった。水位計誤表示が発生した時点から 30 時刻後にパイプ異常が発生した (Fig. 3.6)。パイプ異常が発生していなければ、誤表示発生から 30 時刻たったころには、誤表示の発生を認知できるようになりはじめる (Table 3.3)。この被験者のケースでは、タンク 2 の水位計と FR 1 の指す値 ( $tl_2$  および  $fl_2 - l$ ) は、誤表示やパイプ異常により、本来の値 ( $tl_2$  および  $fl_2$ ) よりも小さくなっていった。しかし、その値の組合わせは、タンク 2 の水位と FR 1 での流量との関係を正しく示している、すなわち、「水位が下がっているから流量も下がっている」かのように見えた。被験者 IN は、水位計 1 と 2 の値が、通常見られる関係とは異なることに気づいてはいたものの、水位計 2 と FR 1 との関係が正常であるかのように見えたことから、異常事象は発生していないものとの判断を下した。結果として、パイプ異常が進行し、自動停止系が作動してプラントを停止するに至った。

このように、状況が複雑になるなどして、与えられた情報の持つ意味を的確に把握できないと、異常事象の認知とそれに対する判断に遅れが生じる。

Table 3.3: Time Elapsed before Taking Countermeasure against Level Meter Failure

name	Day 3, #6	Day 4, #3	Day 5, #3	Day 5, #4
MA	42	18	30	24
SH	172	181	226	80
KA	137	168	220	98
IN	108	75*	226	81
TO	49	48	88	43
TS	135	105*	192	56
MI	53	173	129 †	98

\* Accident was prevented by the automation.

† Accident occurred.

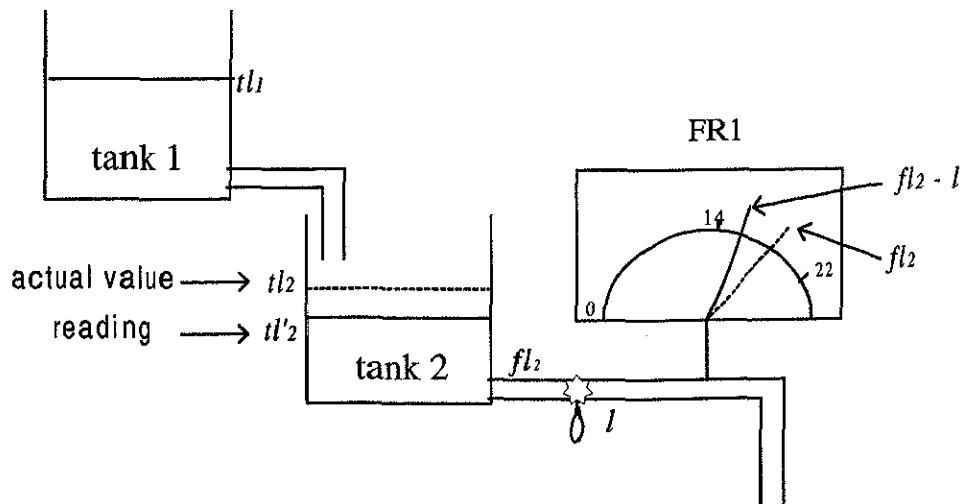


Figure 3.6: Hidden Malfunctions

### 3.4.2 意思決定に利用できる時間余裕の不足

認知と判断の遅れがシステム安全に与える影響をみる。

主ポンプの操作に関して、自動モードでは人間の意識がサブタスクに集中し、計器の数値を確認するのが少なくなる傾向があるのに対し、手動モードでは、サブタスクをこなしながらも頻繁に計器の読みを確認する傾向が見られた。

Table 3.4 は、主ポンプ制御系の制御モードと、パイプ異常に関する処置遂行の早さとの関係を表している。自動モードでは認知・判断が手動モードの場合と比べ遅くなる傾向がある。このことは、自動モードでパイプ異常を第一段階までに対処できたのが6件だけであることからわかる。

また、事故は自動モードでのみ発生しているが、これは、自動モードにおける認知・判断が遅れることによるものと考えられる。

なお、自動モードにおいては、22件中6件自動によりプラントが停止されている。もし、自動停止系が作動しなければ、この6件は事故に至ったものと考えられる。

以上をまとめると、次の3点がいえる。

1. パイプ異常は、これを修復できる時間には制約があり、異常の認識・判断の遅れが直接事故につながることもある。
2. 主ポンプの制御を自動化システムに任せることにより、主ポンプを手動制御する場合と比べてパイプ異常を認識するのが遅くなる。このことは、「自動化システムが異常事象を隠蔽しうる」[77]ことを表している。
3. プラント停止の自動化が、安全確保に貢献しうる。なお、自動停止系がプラントを停止する際には、オペレータが何の命令も与えていない点に注意する必要がある。この場合の自動化レベルは明らかに6以上である。すなわち、時間余裕が十分にはない場合、意思決定・判断自体を自動化することが、システム安全の確保には必要であると考えられる。

Table 3.4: Awareness of Pipe Rupture and Control Mode of Main Pump

	automatic mode	manual mode
repaired in the 1st stage	6	14
repaired in the 2nd stage	22(6*)	4
accident occurred	3(3†)	0

\* Automatic shutdown mechanism was activated for avoiding accident.

† Accident occurred while automatic shutdown mechanism was disengaged.

### 3.5 人と自動化システムの不整合

自動化によりもたらされるのは、必ずしも利点ばかりではない。中華航空 A300-600R 機墜落事故 [121] のように、人と自動化システムとの不整合が深刻な事態をもたらす。本節では、人と自動化システムとの間に生じうる、モード認識の混乱などの問題点を事例により示す。

#### 3.5.1 モード認識の混乱

メンタルワークロードが高い状況でエラーを犯したにも関わらず、エラーになかなか気づかないことが見られた。たとえば、つぎの事例のようである。

### 事例 3 被験者 T0, 2 日目第 2 試行 (異常事象無し)

被験者 T0 は, Day 2, # 2 において, 「プラント停止系」の自動化を選択するつもりで, 実際には「主ポンプ制御系」の自動化選択ボタンを押していた. 彼は手動による主ポンプの制御を好んでいたのである. この被験者は, 主ポンプが自動制御されていることに気付かないまま約 2 分半の間, 主ポンプの手動制御を試みていた. 実験後のインタビューにおいて, その被験者は, 「問題解決型のサブタスクの初めての試行であり, この問題を解くことにプレッシャーを感じていた」と語っている.

このプレッシャーは, 問題解決型のサブタスクの場合, タスク完了に要する時間を見積もるのが難しいことに起因していると考えられる.

試行開始直後などに, 補助ポンプを自動化モードに設定することがある. ある被験者の場合, 試行が始まりしばらくした後, パイプの異常が発生し, 第 2 段階にいたったケースがあった. この被験者は, 補助ポンプを起動させる必要があると考え, 補助ポンプの手動スイッチを ON にセットした. この時点では, 補助ポンプの制御モードが「自動」に設定されており, 自動化システムのロジックではまだ補助ポンプを起動するための条件が満たされていなかった. したがって, 補助ポンプは, 手動スイッチが ON を指しているにもかかわらず給水していなかった. 補助ポンプから流体が供給されないのを見て, その被験者は補助ポンプ設定スイッチを何度も ON に押した. すなわち, 補助ポンプの制御モードが自動モードになっていることに気づいていなかった. 同様の混乱は, 主ポンプの制御に関しても見られた.

### 3.5.2 自動化システムへの不信

複数の被験者が, 自動化システムの制御ロジックと, 被験者自身の制御方策との間に違和感を感じたと述べている. これらの被験者は, 「主ポンプや補助ポンプの自動化システムが, 流量の許容範囲の境界近くまで達してはじめてポンプを起動する」ことに不快な印象を抱いている.

人間は、自動化システムを信頼しなくなる (distrust) とそれを利用しなくなる (disuse) こともある。自動化システムを利用しない場合、メインタスクとサブタスクをすべて自分で遂行しなければならなくなり、作業負荷が高くなる。それでもあえて自動化を使用しない被験者が見られた (たとえば、被験者 TO)。

自動化は本来人間の作業負荷を高すぎなくするはずのものである。自動化システムは、人間から適切に信頼を受けなければならない。

### 3.5.3 自動化システムの不適切な使用

生物には、ものをどのように使うことができるかということ、すなわちアフォーダンス [104, 123] を発見する能力がある。人間も同様であり、当初のシステム設計の意図とは無関係にスイッチなどの使い方を「発見」する。本実験では、つぎの事例が見られた。

**事例 4** 被験者 MI は、主ポンプの制御に関し、設計者が意図しなかった方法で自動化を利用した。彼は、手動スイッチを ON に設定し、「制御モード切り替えスイッチで主ポンプの ON/OFF を制御した」。制御モードを「手動」にすれば、それは主ポンプを ON にすることを意味する。また、制御モードを「自動」にすることは、FR1 の値が 19 よりも大きければ (すなわち、自動化システムのロジックでは主ポンプを稼働させない状況ならば)、ポンプを OFF にすることに相当する。自動モードにおいて、FR1 の値が 17 よりも小さければ、その値が 19 を超えるまでポンプが ON になる。

アポロ 13 号の事故 [54] のように、人間のこのような柔軟性が、システムを壊滅的な事態から回避させるのに役立つことがある。したがって、設計時の想定とは異なる使い方をすること自体を責めるべきかどうかは自明ではない。

しかし、人間が自らシステムの安全性を奪うことがある。事故が発生した 3 つのケースでは、被験者はあらかじめ「プラント停止系」を自動モードにセットしていたものの、パイプ異常の第 3 段階に至る直前にプラント停止系を手動モードに切り替えた。その直

後パイプ破断に至り事故が発生している。実験後のインタビューによると、これらの被験者は、みな「プラントを停止するとメインタスクの成績が低くなるので、プラントをとめたくなかった」旨のコメントを残している。このことは、チェルノブイリ原子力発電所の事故においてみられた、タービンの惰力による発電の実験を遂行するために意図的に ECCS(Emergency Core Cooling System) などの安全系を作動しないようにした [127] ことと類似している。

以上から、緊急時においては人間に最後まで権限を与えておくことが得策であるとはいえず、むしろコンピュータ側に権限を渡す状況適応的な自動化が必要であるといえる。

### 3.6 むすび

本章では、意思決定に利用できる時間余裕が十分でない状況において、システム安全を確保するためには、判断自体を自動化することによる人間への支援が必要であることを明らかにした。

この結果は、Inagaki[30]が数理モデルの解析を通じて示した結果を実験的に裏付けるものである。判断自体の自動化は、Sheridan の 10 段階の自動化レベルにおいて、6 以上のレベルに相当する。人間中心の自動化が基本的には 5 以上のレベルを許さないものであることを考えると、「人間が最終決定権を持つべきである」とする人間中心の自動化の原則は検討し直す必要があるといえる。

# 第4章 安全確保への状況適応自動化： 有効性と実現可能性

## 4.1 はじめに

第3章では、システム安全を確保するために状況適応自動化 (SAA) が必要であることを実験的に確認した。しかし、SAA が本当に有効に機能するか厳密な検証が必要である。本章では、要因実験によって SAA の有効性を確かめる。

また、人間の SAA に対する許容性を検討し、SAA を実現するためにいかなる工夫が必要となるかを明らかにしなければならない。なぜなら、SAA は緊急時にはオペレータから権限を奪うことがあり、プロセスの監視者としての人間が SAA を受け入れようとするか定かではないからである。「人間中心の設計」[71] は、人間に受け入れられるシステムの構築を一つの目的としており、この目的を達成することは、監視制御系のシステム設計においても重要である。

システムが人間に許容されるには、人間の信頼 (trust) を得ることが不可欠である。trust は、主に社会学の立場から研究されてきた概念である [2, 69, 126]。Muir[59] は、人間による自動化システムの利用に trust が重要な影響を与えているとし、自動化システムへの trust に関する基礎理論を構築している。Lee と Moray[49, 50] は、システムの誤作動が trust を失う原因であることを実験的に明らかにし、trust の動的なモデルを定量的に構築した。さらに、Lee と Moray[51] は、オペレータ自身の能力への自信 (self-confidence) を導入し、自動化システムへの依存性が、trust と self-confidence の差によって定まることを明らかにした。



人間の信頼には、能力への信頼と意図への信頼との二つの側面がある [126]。本章では、SAA の信頼性のみならず、SAA 意図をオペレータがいかに関与するかにも注目して、SAA への trust や、self-confidence の性質を解明する。

4.2 節では、本章で実験に利用するマイクロワールド (SCARLETT) を説明する。4.3 節では、実験の計画と方法を述べる。4.4 節では、タスクを遂行するにあたり被験者の特性を示し、本実験で用いた SAA の能力との関係を示す。4.5 節では、システム安全を確保するためにレベル 6 以上の自動化が有効に機能すること、および完全自動化をすべきではないことを、実験結果の解析を通じて明らかにする。4.6 節では、trust と self-confidence を、静的・動的観点からモデル化し、それらの性質を明らかにする。4.7 節では、高レベルの自動化に対する trust の振舞を解析し、オペレータのモード認識を妨げず、trust の低下を防ぐための新しい自動化レベルを提案する。

## 4.2 マイクロワールド 2

本章では、3章で用いたマイクロワールドを改良した、SCARLETT (Supervisory Control And Responses to LEaks: Tara at Tsukuba) [37, 38, 56] を用いて認知工学的実験を行うこととした。なお、SCARLETT は、Gensym 社の G2<sup>®</sup> 上で構築されている。

### 4.2.1 プラントの記述

SCARLETT は、集合住宅 (AP) の室温調節のためのセントラルヒーティングシステムである (Fig. 4.1)。このプラントは低温系としての集合住宅、熱を供給する高温系、熱交換器 (HX) からなる。高温系には、2つの予熱槽 (R1, R2) と各々につけられたヒーター (H1, H2) ・水供給ポンプ (P1, P2)、加熱槽 (R3) とそのヒーター (H3)、R3 の出力を決定するポンプ (P3)、バルブ (V1 - V7)、パイプがある。パイプの一部は、L1 ~ L6 と名前がつけられている。スイッチングバルブ (SV) は、熱交換器を出た水を、L4 を通る系と

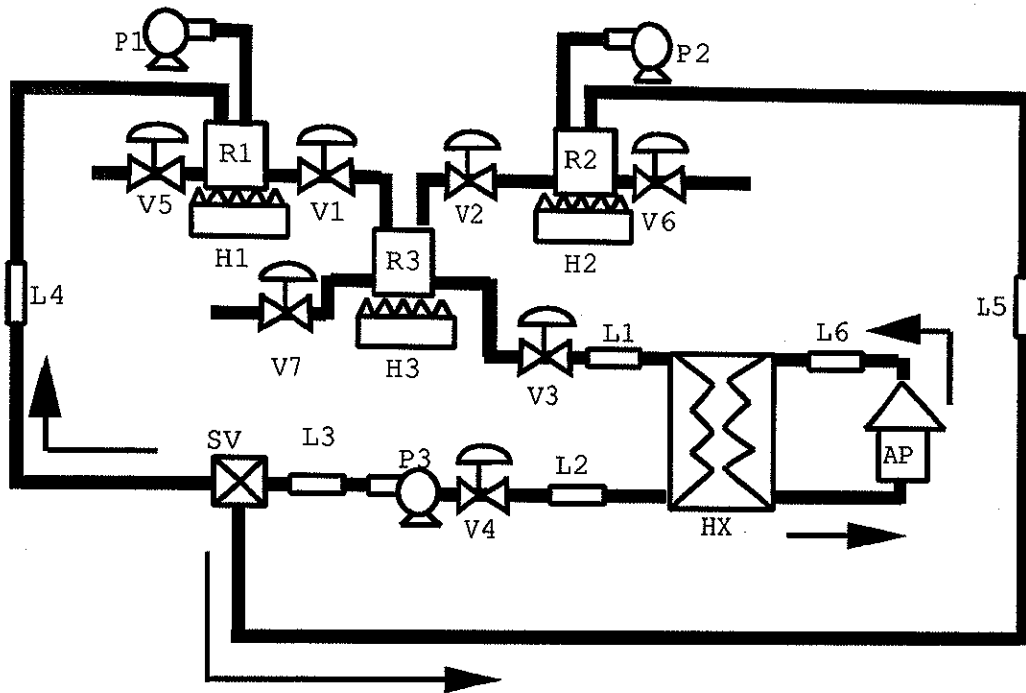


Figure 4.1: Microworld 2(SCARLETT)

L5 を通る系とに分岐するためのものである。

タンク R1, R2 からの流出量は重力のみに依存する, すなわち水位に比例して決まるが, R3 からの流出量はポンプ P3 の設定値により決定される. なお, 各タンクには, 水を廃棄するためのバルブ (V5, V6, V7) があるが, これらのバルブを開いた場合, いずれのタンクでも水位に比例した量が放出される。

#### 4.2.2 ヒューマンインタフェース

オペレータは, Fig. 4.2 の制御パネルを通じてプロセスの監視や各機器の操作をする. タスクの遂行に必要な情報, 例えば, 顧客からの要求温度やタンクの水位・温度, パイプ中の流速等の情報は, 数値として各アイコン付近に表示される。

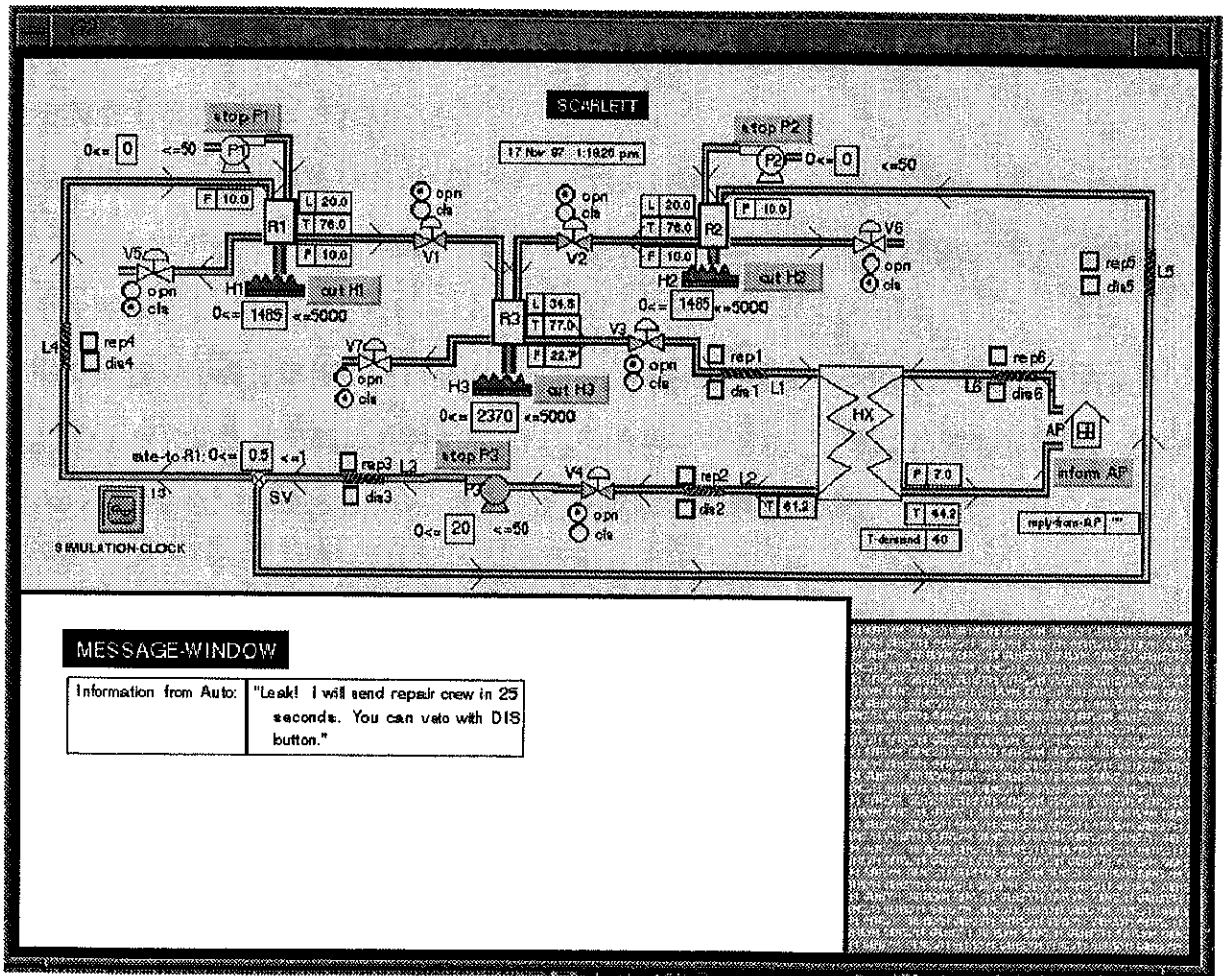


Figure 4.2: Interface for SCARLETT

### 4.2.3 オペレータのタスク

本プラントにおいて、ヒューマンオペレータがなすべきタスクは二つある。

#### 集合住宅の温度制御

オペレータは、集合住宅 (AP) に住む顧客からの要求を満たすように、AP 側の水の温度を調整することを課されている。

集合住宅から熱交換器 (HX) に流入する水は、流量と温度が一定で固定されているものとする。したがって、顧客の要求を満足するためには、オペレータは高温系の水の温度と流量を調節しなければならない。

熱交換器の集合住宅側の出口温度 ( $T_{out2}$ ) を所望の温度にするための基本的な方法を、 $T_{out2}$  を上げたい場合を例に説明する。 $T_{out2}$  を上昇させるには、熱交換器の高温側の流入エネルギーを増加させればよい。次の二通りの方法が可能である。

1. 熱交換器の高温側に流入する水の温度を上げる。
2. 熱交換器の高温側に流入する水の流量を増やす。

前者の場合、R3 を中心として各タンク (R1,R2,R3) におけるヒーター (H1,H2,H3) の加熱量を増すことによって実現可能である。ただし、タンク内の水の温度は直ちには上昇しないため、熱交換器の高温側流入エネルギーを目的の値にするまでには時間がかかる。

後者の、水の流量を増やす方法は、循環ポンプ P3 の設定値を上げることによって実現可能である。この方法では、熱交換器の高温側流入エネルギーを、直接増やすことができる。したがって、前者の方法に比べ、集合住宅側の温度をより早く上昇させることができる。ただし、P3 の流量が増加することにより、R1, R2 のタンクの流量が増え、その結果各々のタンクの入出量のバランスが崩れ、各タンク内の水位が変化する。ヒーターの加熱量が一定と仮定すると、タンク内の水位変化によってタンク内の水の温度が変化する。したがって、全体的に流量と温度のバランスが崩れ、不安定になる。この場合、オペレー

タは、システムを安定させるためにきめ細かくポンプやヒーターを操作しなければならない。

## プラントの安全確保

パイプをはじめとして、システム内の各構成要素には故障などの異常事象が発生し得るものとする。異常事象には、プラントの稼動を継続しながら修復できる類のものと、プラントを停止しなければ安全性を確保できないものがありうる。したがって、本プラントでは、オペレータは異常事象の性質を見抜き、プラントの稼動を継続できるか、もしくは停止すべきかの判断を下さなければならない。また、異常事象が生じたもとで稼動を継続するにしても、システムを正常な状態に戻すためには必要な処置を講じなければならない。

### 4.2.4 異常事象と事故

本章での実験では、パイプ L1 ~ L6 において異常が発生し得るものとする。パイプの異常には2種類ある。

#### 1. パイプ漏洩

パイプにひびが生じ、そこから水が流出する。単位時間あたりの漏出量は一定である。パイプ漏洩に対しては、修理チームを派遣することによって、プラント稼働中にパイプを修復できる。

#### 2. パイプ破断

パイプが完全に断裂し、その破断箇所を通過する流体がすべて失われる。パイプ破断が発生した場合、迅速にプラントを停止する必要がある。ただし、発生箇所によっては、パイプ漏洩とパイプ破断とが類似の現象を起こすことがあり、事象の種類を特定することが困難な場合もある。破断の起きていないプラントを停止することは許され

ないので、オペレータは事象を正確に同定する必要がある。

異常事象の生起などをきっかけとして、次の事象のうち少なくとも一つが発生した場合、事故が発生したとみなす。

1. 空だきによるタンク破壊.
2. 空気の混入によるポンプ破壊.
3. アパートへの熱湯もしくは冷たすぎる水の供給.
4. 大量の水漏れによるプラント浸水.

パイプ破断が生じることにより、直ちに事故が発生する場所もある（たとえばL1での破断の場合、10秒程度で事故にいたる）。しかし、多くの場合オペレータが適切な処置を迅速に実施することによって防ぐことができる。すなわち、パイプの異常が発生しただけでは事故に至るとは限らない。

#### 4.2.5 安全制御システム

異常事象に対処するために、このプラントでは「安全制御システム」が備えられている。安全制御システムは、異常の検知・診断を行う異常診断機構、オペレータへ異常発生を通報する警報生成機構、安全確保のために制御操作を実行する安全確保機構からなる。

安全制御システムの異常診断機構が異常の発生を検知した場合、異常が生じた旨のメッセージを、Fig. 4.2におけるMESSAGE-WINDOWに提示し、異常発生個所を赤色点滅させる。ただし、メッセージの内容やメッセージ提示後の行動は、次に示す安全制御モードに依存して異なる。

安全制御システムは、次の2つのいずれかの安全制御モードで作動する。

1. 手動制御モード
2. 状況適応自動制御モード

パイプ漏洩、パイプ破断それぞれの場合において、安全制御システムがいかに機能するかを示す。

- パイプ漏洩の場合

1. 手動制御モード：異常診断機構がパイプ漏洩を検知すると、警報生成機構が「パイプ漏洩が発生しました。修理チームを派遣してください」とのアラームメッセージと、異常発生箇所の情報を提示する。オペレータは、通報された診断が正しいと判断した場合修理チームを派遣しなければならないが、その診断が正しくないとは判断したときはアラームをリセットすればよい（修理チームは派遣されない）。
2. 状況適応自動制御モード：パイプ漏洩が発生していると認められる場合、「パイプ漏洩が発生しました。25秒以内に修理チームを派遣します」との提案と、異常発生箇所が通知される。パイプ修理・アラームリセットについては、手動制御モードと同様にオペレータが自分で行う。ただし、メッセージ提示後25秒後までにその提案が拒否されない場合、安全確保機構が修理チームを自動的に派遣する（自動化レベル6）。

- パイプ破断の場合

1. 手動制御モード：異常診断機構がパイプ破断の発生を検出した場合、警報生成システムが起動し、プラント停止手順と異常発生箇所とを表示する。安全制御システム自体はプラントを停止しない。すなわち、オペレータが手動でプラント停止手順を実行する必要がある。ただし、診断情報は正しくないとはオペレータが判断した場合、アラームをリセットすれば、通常運転を継続できる。
2. 状況適応自動制御モード：パイプ破断が発生していることが検出されると、「パイプ破断が発生しました。\*秒後にプラントを停止します」とのメッセージと異常発生箇所が提示される。なお、\*に当てはまる数値は事故に至るまでの時間余裕に依存して決定される。この通報に対してオペレータがとるべき行動は

手動モードと同じである。ただし、一定時間内にアラームがリセットされなければ、安全確保機構がプラントを停止する（自動化レベル6）点が手動モードと異なっている。ところで、L1でのパイプ破断の場合、事故に至るまでの時間が極めて短い。そこで、L1破断の場合に限り安全確保機構がプラントを緊急停止し、その後「事故の恐れがあったため、プラントを停止しました」と通知する（自動化レベル7）。以上のように、状況適応自動制御モードでは、自動化レベルは事象の種類や発生個所によって異なる。

## 4.3 実験の方法

### 4.3.1 実験計画

本章では、4つの要因による要因実験を実施する。本実験は、 $2 \times 2 \times 2 \times 3$ の混合計画である。各要因のそれぞれの水準を説明する。

**第1要因** 異常事象の種類。パイプ漏洩とパイプ破断のいずれかが起こる。なお、異常発生場所や時刻はランダムに決定される。

**第2要因** 安全制御モード。本実験では、安全制御システムの動作について、手動制御（M）モード、および状況適応自動制御（SAA）モードの二つを各被験者に体験させる。

**第3要因** モード体験順序。被験者を、Mモードでの試行を先に実施する（M→SAA）群と、SAAモードでの試行を先に実施する（SAA→M）群とに無作為に分ける。

**第4要因** 安全制御システムの信頼性。安全制御システムの3つの機構のうち、異常診断機構が誤りを犯すことがある。診断の信頼性のレベルとして、次の3つの場合を考える。すなわち、(1) 完全な信頼性の場合（診断は100%正しい）、(2) 「高信頼性」の場合（被験者に提示される診断情報の信頼性が90%、すなわち10回のうち9回は診断が正しい）、(3) 「低信頼性」の場合（70%の診断が正しい）、である。ただし、



異常の有無は正しく検知できるが、パイプ漏洩をパイプ破断と診断する（またはその逆）ことがあるものとする。各被験者には、(1)～(3)のいずれかの信頼性を体験させる。

#### 4.3.2 手順

実験には、情報科学を専攻する学部生および大学院生 30 名が被験者として参加した。各被験者につき、実験は 1 日約 1 時間で 3 日連続して行い、報酬として一律 3,000 円を最終日に支払った。

実験の実施手順を以下に示す。

##### 1 日目

1. 教示: 実験の目的, SCARLETT の物理的特性, 遂行すべきタスク, 制御パネルの操作方法, について書面で説明する。
2. トレーニング: (1) プラント制御技術を身につけさせるために, 異常事象が発生しない条件で 7 回の試行を行う。(2) パイプ漏洩発生時の, プラントの挙動と提示されるメッセージを理解させるために 5 回の試行を観察させる。さらに, 各制御モードでの安全制御をそれぞれ体験させる。(3) 異常事象の種類をパイプ破断に変え, (2) と同じ手順を実施する。

##### 2 日目

1. プリテスト: 制御技術の習熟度を確認するために, 異常が発生しない条件で 3 回の試行を実施する。
2. データ収集: 被験者に割り当てたモード体験順序に応じて, いずれかのモードで 20 回 (パイプ漏洩, パイプ破断各 10 回) の試行を実施する。なお, 10 回の試行が終了した時点で, 約 5 分の休憩を入れる。

### 3日目

1. データ収集：2日目1. と異なる安全制御モードで、20回の試行を実施する。
2. ポストテスト：実験中の学習効果を調べるために、プリテストと同じ要領で試行を実施する。

#### 4.3.3 評価指標

本実験は、次の8つを評価指標とし、データを収集した。すなわち、

1. アパートへの供給水の実現温度と要求温度との2乗平均誤差(RMSE)
2. 異常発生から安全制御行動を人間が開始するまでの時間
3. 手動プラント停止の完遂に要する時間
4. パイプ異常による循環水漏出量
5. 誤停止率
6. 事故率
7. 安全制御システムへの信頼 (trust)
8. 被験者自身の技量への自信 (self-confidence)

1, 2, 3 や事故・誤停止の有無は、各試行終了後画面に表示することによって被験者に報告する。なお、7, 8 は主観評価である。

#### 4.4 タスク遂行に関する被験者の特性

本章のデータ解析にあたっては、実験の第四要因である信頼性(%)が[0, 100]における任意の実数値をとることができることから、この要因を無作為要因(random factor)として扱う。被験者間要因が無作為要因である分散分析において、無作為要因を含まない効果の有意性の検定は、「その効果と無作為要因との交互作用」を誤差項として行われる [94](pp.

323-326). 得られたデータを分散分析する際, 第四要因を含まない効果, たとえば, 制御モードの主効果に対する誤差変動の自由度は  $(3 - 1) \times (2 - 1) = 2$  になる. 本実験では, 固定要因 (fixed factor) の水準数はいずれも 2 であるため, この誤差変動の自由度は常に 2 になる.

なお, 本章におけるデータは, Statsoft 社の STATISTICA<sup>®</sup> (Windows 版, Ver. 5) を使用して解析した.

#### 4.4.1 プラント運転成績の学習効果

被験者が SCARLETT における制御タスクにどの程度習熟したか, またその学習が実験を通じていかに進んだかの解析を通じて, 温度制御タスクの難しさを示す. ここでは, 「データ収集」の前後に行われたプリテスト, ポストテストの「顧客からの要求温度と実現温度との 2 乗平均誤差 (Root Mean Square Error: RMSE)」を評価する.

Table 4.1 に, RMSE の測定結果を表す. プリテストよりもポストテストの方が RMSE の値は小さい. 実際, 「データ収集」の前と後との間での学習を要因に追加した, 信頼性, モード体験順序, 学習の三元配置分散分析によると, プリテストとポストテストとの差は有意であった ( $F(1,2)=30.33, p=0.031$ ). なお, プリテスト・ポストテストにおいては, 異常事象が発生しないことから, 異常事象の種類と制御モードを要因からはずしてある.

RMSE に学習効果が見られた. すなわち, 実験のデータ収集の段階を通じて, 温度制御のスキルは向上していた. このことは, 温度制御のタスクが, そのスキルを容易には習得できないものであることを示している. これは, SCARLETT と同様に, マイクロワールドとして認知工学的実験研究にしばしば利用されている Pasteurizer[49, 50, 51, 60]での結果とよく似ている. たとえば, Muir と Moray[60] は, Pasteurizer において, 被験者のスキルの学習を完了させるまでに 7 時間以上費やしている.

なお, 以上の結果は, トレーニングが不十分であったことを示すものではない. 実際, プリテストと同じシナリオで, 被験者が一切の操作をしない場合 (NO OPERATOR) の

Table 4.1: Root Mean Square Error (RMSE) Scores of the Central Heating Plant

RMSE	mean	s.d.
PRETEST	2.07	0.829
POSTEST	1.64	0.337
NO OPERATOR	3.64	0.194

RMSE のデータを Table 4.1 に示す（なお，s.d. は標準偏差を表す）．プリテストと NO OPERATOR との間で RMSE を比較すると，前者の RMSE の方が後者のそれより小さいことが高度に有意である ( $t=11.7$ ,  $df=38$ ,  $p<0.001$ )．被験者が不慣れである場合には，プラントを不安定にしてしまうことにより，NO OPERATOR の RMSE を上回る状況，すなわち成績を悪化させる場合も観察されたことから，プリテストを実施する段階で，被験者のスキルは一定のレベルに達していたといえる．

#### 4.4.2 異常診断に要する時間と自動化の介入時刻

警報が与えられたもとで被験者が異常診断に要する時間や，安全制御行動の遂行に要する時間を評価し，状況適応的な自動化が適切に設計されていたか検討する．

SAA モードでは，異常診断機構がパイプ異常を検知すると，安全制御システムの安全制御機構は Table 4.2 に示す時間（秒）だけ人間からの拒否を受け付ける．なお，Table 4.2 の値は，異常発生から事故に至るまでの時間の長さを中心に反映した設定である．

人間の信頼性解析手法 [11] に基づき，パイプ漏洩に対する反応時間を，対数正規確率紙にプロットした (Fig. 4.3)．この時間は，アラームが提示された時刻から，修理クルーの派遣やアラームのキャンセルなど，何らかの行動を人間が実施するまでの時間として測定した．図の各点は，各被験者の反応時間の，全試行にわたる平均値である．なお，被験者毎に反応時間をプロットしたところ，各被験者の反応時間のデータが対数正規分布によ

Table 4.2: Time Delay to Allow Veto

Fault	L1	L2	L3	L4	L5	L6
Leak	25	25	25	25	25	25
Break	0	10	18	18	18	10

く従うことも確かめられた。

Fig. 4.3 から、ほとんどの被験者が警報提示時刻から約 15 秒の間に反応できていることがわかる。本実験では、(1) パイプ漏洩から事故は発生しないこと、(2) 漏洩しているパイプを修理しないと水が漏れつづけることによる損失を被ること、(3) 25 秒たつまでにほぼすべての被験者が対処行動を終了させていることから、SAA が人間による拒否を受け付ける時間の長さ (25 秒) は、遅すぎも早すぎもせず、妥当な設定であるといえる。

パイプ破断の場合を考える。M モードでパイプ破断が発生するケースについて、人間がシステム停止を始めるまでの時間を測定した。パイプ破断の場合、異常発生個所によっては異常発生後直ちに事故に至ることがあり、このとき人間の認知推論プロセスは打ち切られてしまう。このようなデータに対しては、パイプ漏洩の場合と同様な解析をすることは困難であり、生存分析 [48] が適している。Fig. 4.4 は、「生存曲線」(累積未反応確率曲線) を示している。また、パイプ破断への反応時間 (秒) の四分位数を Table 4.3 に示す。いずれの個所での破断でも、人間の反応時間の中央値は 10 ~ 15 秒程度であり、L1 での破断がもたらす事故を防ぐことはほとんど不可能である。また、プラント停止操作を完了するのにさらに十数秒かかる [132] ことから、L2, L6 でも、手動操作で事故を防ぐのが困難となっている。結局、SAA は、異常の発生から事故に至るまでの時間が短い L1, L2, L6 では人間の平均的な反応時間よりも早くにプラント停止に介入するが、それ以外のケースでは、人間の平均的な反応時間よりもやや遅れてプラント停止を開始する。以上のことから、本実験での SAA は適切に設計されているといえる。

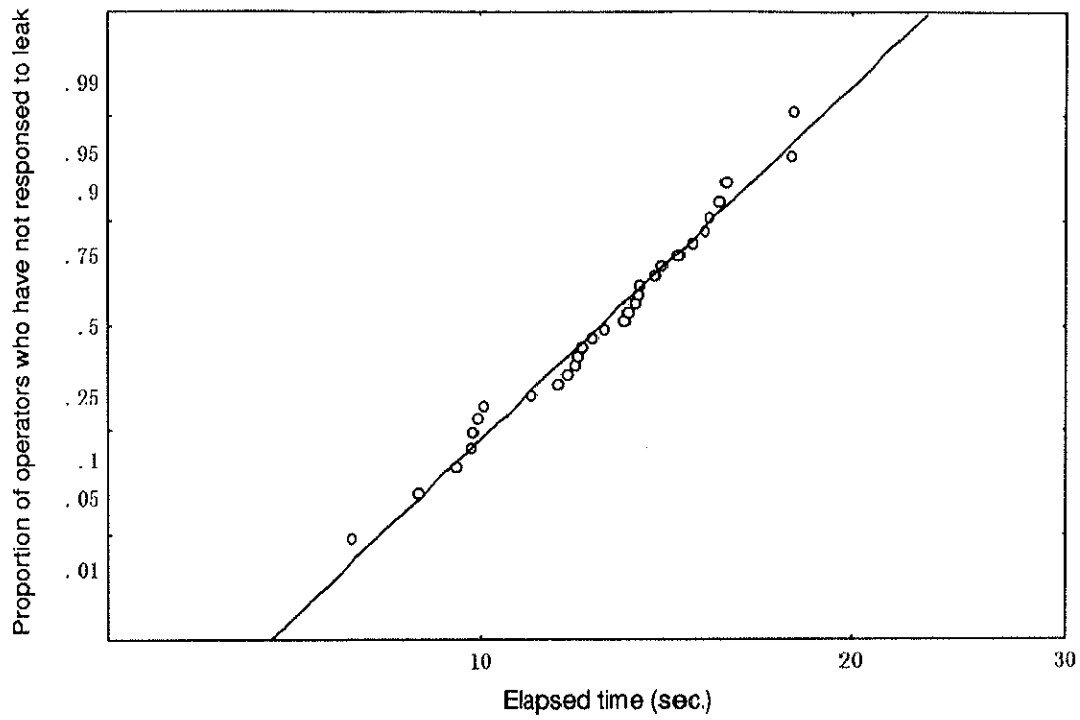


Figure 4.3: Time-Reliability Curve in Response to Leaks

Table 4.3: Elapsed Time before Initiating Shutdown Procedure

quartile	L1	L2	L3	L4	L5	L6
25%	10	8	8	10	14	8
50%	11	11	9	14	15	11
75%	16	-	14	17	23	12

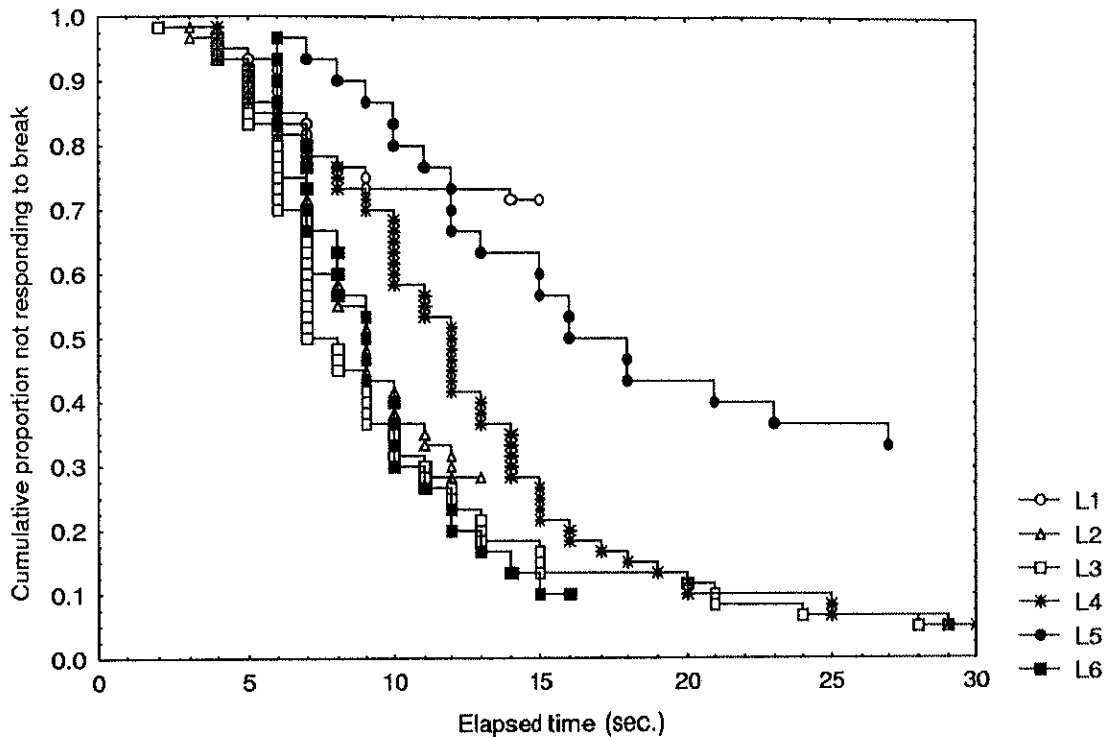


Figure 4.4: Survival Function Curves in Response to Breaks

## 4.5 状況適応自動化による安全確保

### 4.5.1 事故防止への自動化の貢献

レベル6以上の自動化(SAAモード)が、事故を防止するのにいかに役立つか解析する。

**事故率** 事故が発生したのはほとんどパイプ破断の場合であったので、パイプ破断が発生した試行でのみ事故率を計算する。Table 4.4に、事故率の集計結果を示す。異常事象の種類を除いた三元配置の分散分析によると、事故の発生した頻度に関しては、つぎの2点が明らかになった。

1. 安全制御システムの信頼性の主効果が有意であった ( $F(2,24)=9.825, p<0.001$ )。このことは、異常診断機構の信頼性が下がると事故率が増加することを意味する。
2. 制御モードの主効果が有意であり ( $F(1,2)=438.86, p=0.0023$ )、SAAモードの事故

Table 4.4: Proportion of Accidents

Reliability	Mode	mean	s.d.
100%	SAA	0.02	0.42
100%	M	0.36	0.14
90%	SAA	0.09	0.16
90%	M	0.42	0.22
70%	SAA	0.26	0.12
70%	M	0.55	0.14

率がMモードよりも低い (Fig. 4.5). このことは、事故を防ぐためにはSAAモードが有効であることを示している。

制御モードと安全制御システムの信頼性との間には、交互作用が見られなかった。このことは、安全制御システムの信頼性が低下すると、各モードで同じように事故率が高まっていくことを示している。

#### 4.5.2 ヒューマンオペレータの存在意義

レベル5を超える自動化が事故防止に貢献することが明らかとなった。しかし、安全確保のための意思決定とその実行を完全に自動化してよいわけではない。ここでは、その理由を、誤停止率、循環水漏出量の解析を通じて明らかにする。

**誤停止率** Table 4.5は、誤停止の頻度の集計結果である。パイプ破断がおきた場合にはプラントを停止しなければならないとしているので、誤停止は、パイプ漏洩が発生した場合もしくは異常事象が何も生起していない状況におけるプラント停止として定義している。



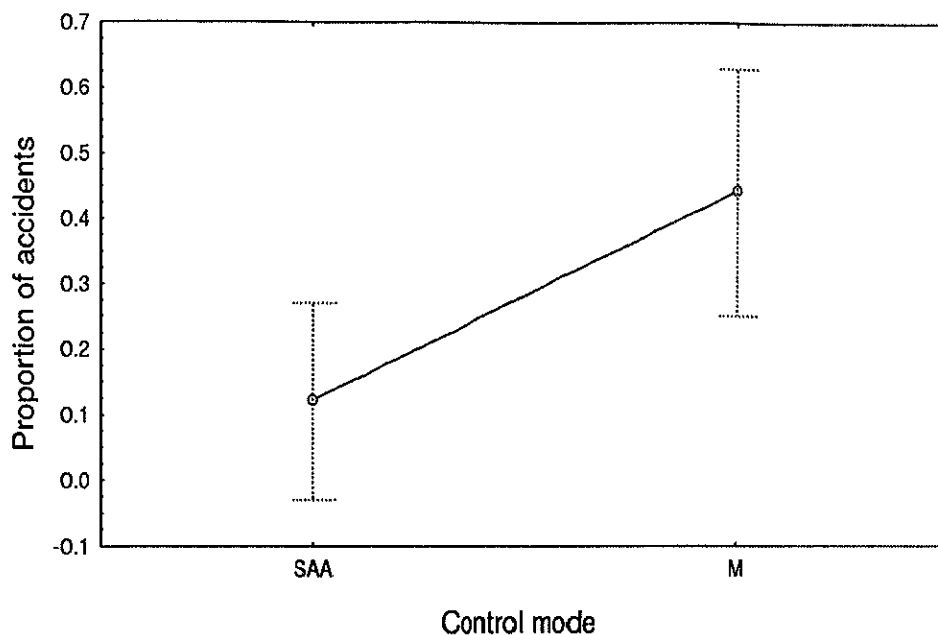


Figure 4.5: Effect of Control Mode on Proportion of Accidents

異常事象の種類を除いた三元配置の分散分析の結果，誤停止率に関してつぎのことが明らかになった。

1. 安全制御システムの信頼性の主効果が有意であった ( $F(2,24)=25.56$ ,  $p<0.0001$ )。これは，異常診断機構の信頼性が低下するにつれて誤停止が増えることと解釈できる。
2. 安全制御システムの信頼性と制御モードとの1次の交互作用が有意であり ( $F(2,24)=4.8$ ,  $p=0.018$ )。これは，コンピュータの信頼性が高いときには制御モード間での誤停止率に差がないが，コンピュータの信頼性が下がってくるとSAAモードでの誤停止率がMモードでの誤停止率よりも悪くなる (Fig. 4.6) ことと解釈できる。

この結果から，誤停止を防ぐために，人間の判断の余地を完全に排除すべきではないこと，すなわち完全な自動化をすべきではないことがわかる。

**循環水漏出量** パイプ漏洩・破断による循環水の喪失量の集計結果を，Table 4.6 に示す。分散分析の結果，つぎの2点が明らかになった。

Table 4.5: Probability of False Shutdowns

Reliability	Mode	mean	s.d.
100%	SAA	0.03	0.067
100%	M	0.03	0.048
90%	SAA	0.11	0.102
90%	M	0.11	0.088
70%	SAA	0.32	0.091
70%	M	0.20	0.094

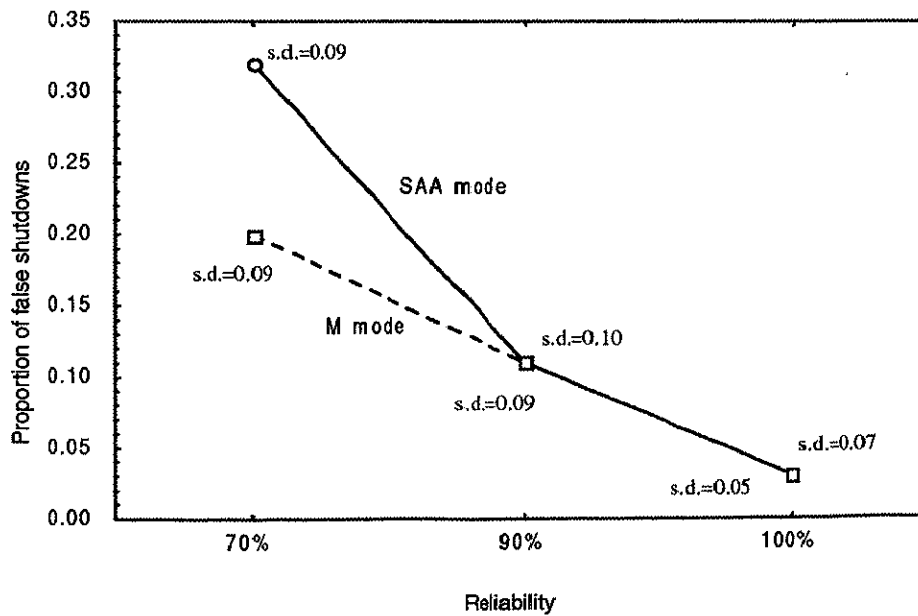


Figure 4.6: Effect of Control Mode on Probability of False Shutdowns

Table 4.6: Fluid Loss

Reliability	SAA				M			
	Leak		Break		Leak		Break	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
100%	56.22	22.43	189.37	38.59	43.95	17.76	306.69	70.04
90%	106.05	79.73	284.08	114.74	78.63	63.66	323.87	77.62
70%	76.65	29.85	266.10	56.02	90.00	40.45	305.16	105.88

1. 異常事象の種類の主効果が有意であり ( $F(1,2)=2160.22, p<0.001$ ), パイプ破断による水の漏出量が, パイプ漏洩による漏出量よりも多い。
2. 異常事象の種類, 安全制御システムの信頼性, および安全制御モードの2次の交互作用が有意であった ( $F(2,23)=4.630, p=0.020$ ) (Fig. 4.7). このことは, 次のように解釈される. パイプ漏洩の場合, 異常診断機構の信頼性が高い場合にはMモードの方がSAAモードよりも失う水の量は少ないが, 信頼性が低下するとMモードの方が少ないとはいえなくなる. パイプ破断の場合, 全体としてSAAモードの方がMモードより喪失量が少なく, Mモードではコンピュータの信頼性によらずほぼ一定の水を失うが, SAAモードではコンピュータの信頼性が下がると失う水の量が多くなる。

循環水漏出量の解析結果は, つぎのことを示している. すなわち, 異常の規模が大きく, システムの信頼性が十分に高い場合には自動化の方が人間よりも有効であるものの, 安全制御システムの信頼性が低下したり, 異常の規模が小さい場合には, 自動化と人間との差が小さい. したがって, 安全制御システムが完璧でない限り, 人間を排除することは積極的には支持されない。

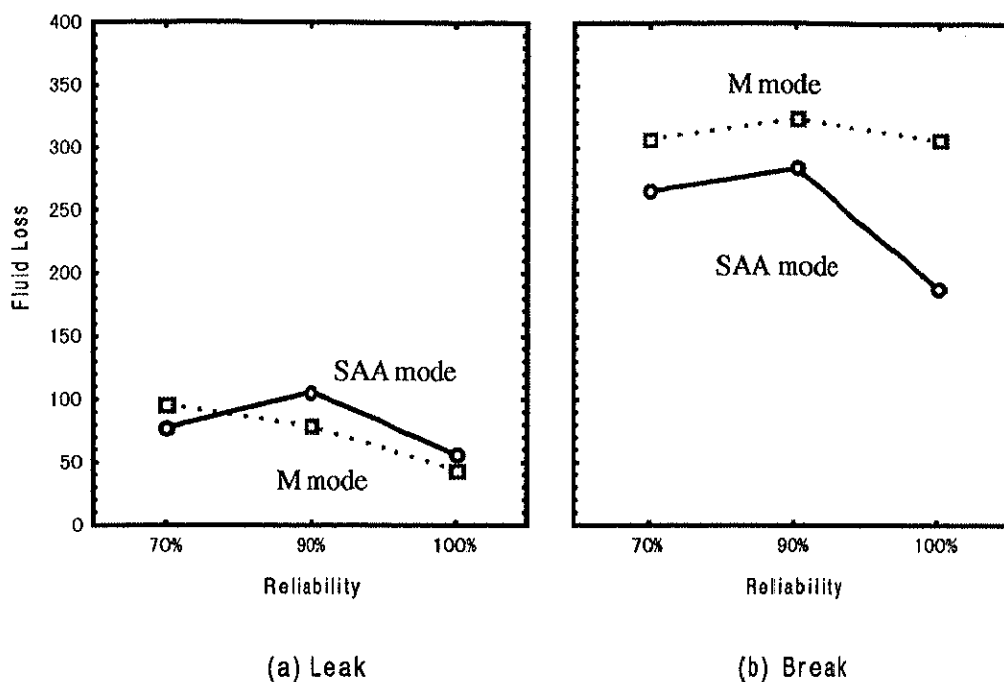


Figure 4.7: Fluid Loss in Three-Way Interaction among Reliability, Mode, and Fault

## 4.6 trust, self-confidence のモデル

状況適応的な自動化がシステム安全の確保に有効であることが示されたが、SAAの実現のためには、人間の自動化システムへの依存性・許容性を考慮する必要がある。すなわち、状況適応自動化機構を持つ自動化システムを人間が受け入れ、利用しようと思えるものにならなければならない。Lee と Moray[51]によると、自動化システムを利用しようとするか否かは、自動化への信頼感と自分に対する自信とに影響される。そこで、ここでは、信頼 (trust) と自信 (self-confidence) のモデルを静的と動的の二つの観点から構築する。

### 4.6.1 分散分析に基づく静的モデル

「データ収集」における各試行終了後、自動化システムに対する被験者の trust, 被験者自身の能力に対する自信 (self-confidence) を、0 から 10 までの 11 段階で被験者に答えさせた。

Table 4.7: Subjective Ratings of Trust in Automation

Reliability	SAA				M			
	Leak		Break		Leak		Break	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
100%	8.23	0.94	8.48	0.81	8.51	0.86	8.74	0.87
90%	7.64	1.19	7.52	1.53	7.62	1.06	7.37	0.98
70%	4.87	1.23	4.67	1.47	6.09	1.47	5.56	1.22

#### trust

trust の評価値の集計結果を Table 4.7 に示す。分散分析の結果、以下の 3 点が明らかになった。

1. 安全制御システムの信頼性の主効果が高度に有意であった ( $F(2,24)=25.803, p<0.0001$ ). 多重比較によると、安全制御システムの信頼性が高い (低い) と、trust も高く (低く) なるといえる。
2. 安全制御システムの信頼性と異常事象の種類との交互作用が高度に有意であった ( $F(2,24)=6.830, p=0.004$ ). ただし、この交互作用がもたらす影響は、有意ではあるものの絶対値としては小さく、重要ではない [56].
3. 安全制御システムの信頼性と制御モードとの交互作用に有意傾向が見られた ( $F(2,24)=3.138, p=0.062$ ). Fig. 4.8 を見ると、安全制御システムの信頼性が低いとき、SAA モードにおいて trust がより低くなる傾向があることがわかる。

3 番目の結果は、次のように解釈できる。すなわち、SAA モードでは人間ではなく自動化システムが安全制御を実行する分、実行した内容が誤った場合に安全制御システムへの信頼感がより低下しやすい。これに対し、M モードでは、意思決定し、これを実行する

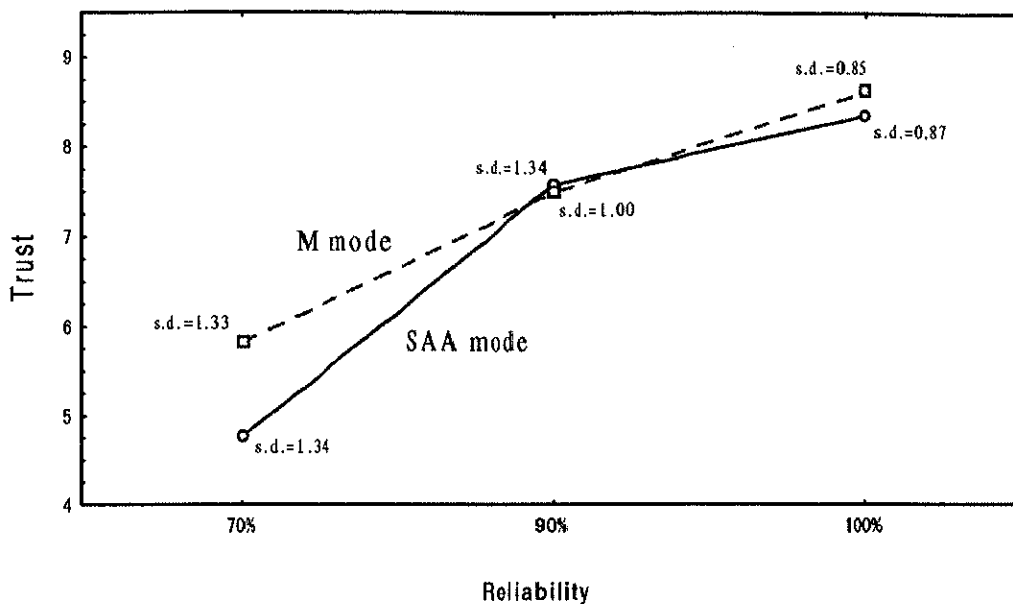


Figure 4.8: Interaction of Reliability of Automation and Mode of Control on Trust

のはあくまでも被験者本人である。安全制御システムから与えられた情報が誤っていたとしても、SAAモードの場合と比べ信頼感は低下しにくい。

### self-confidence

Table 4.8 に self-confidence の集計結果を示す。分散分析を行ったところ、モード体験順序の主効果が有意であった ( $F(1,2)=19.988, p=0.0466$ )。しかし、「モード体験順序」という要因は現実世界にはなく、本論文の興味の対象ではない。

制御モードの主効果 ( $F(1,2)=14.112, p=0.064$ ) と異常事象の種類の主効果 ( $F(1,2)=11.581, p=0.0766$ ) に有意傾向が見られた。しかし、これらの効果は、0 から 10 のスケールにおいて 0.5 程度の差をもたらすに過ぎず、重要なものではない [56]。

興味深い点は、安全制御システムの信頼性の主効果が有意でなかったことである。このことは、self-confidence が、集合住宅の温度制御のスキルと異常事象への対処の能力とを総合的に評価した結果与えられるものと考えることによって説明できる。本実験では、温度制御はすべて手動で行わせたため、温度制御に関しては安全制御システムの信頼性に依

Table 4.8: Subjective Ratings of Self-Confidence

Reliability	SAA				M			
	Leak		Break		Leak		Break	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
100%	4.64	1.99	4.22	2.28	5.31	2.24	4.74	2.59
90%	4.55	2.66	3.91	2.46	5.00	2.13	4.46	2.15
70%	4.88	1.44	4.81	1.51	5.20	1.29	4.91	1.42

存しないものと考えられる。これに対し、もし安全制御システムから誤った診断情報を提示され、これに従って正しくない安全制御操作を実行すれば、その結果「異常事象への対処」のスキルへの自信は低下するものと考えられる。すなわち、システムの信頼性が低下するに従い、self-confidence も低下するはずである。

self-confidence が、弱いながらも安全制御システムの信頼性に影響を受けることは、次節の解析によって明らかにされる。

#### 4.6.2 時間遅れを反映した動的モデル

trust は時間的な側面を持つことが知られている [49, 50, 51, 59, 60, 69]。self-confidence も同様に、時間的な変動を伴う。trust は時間的な側面を反映したモデル化が提案されている [49, 50] が、self-confidence 自体はそのようなモデル化は試みられたことがない。本節では、SAA に対する trust, self-confidence 両者の動的モデルを構築し、その性質を明らかにする。

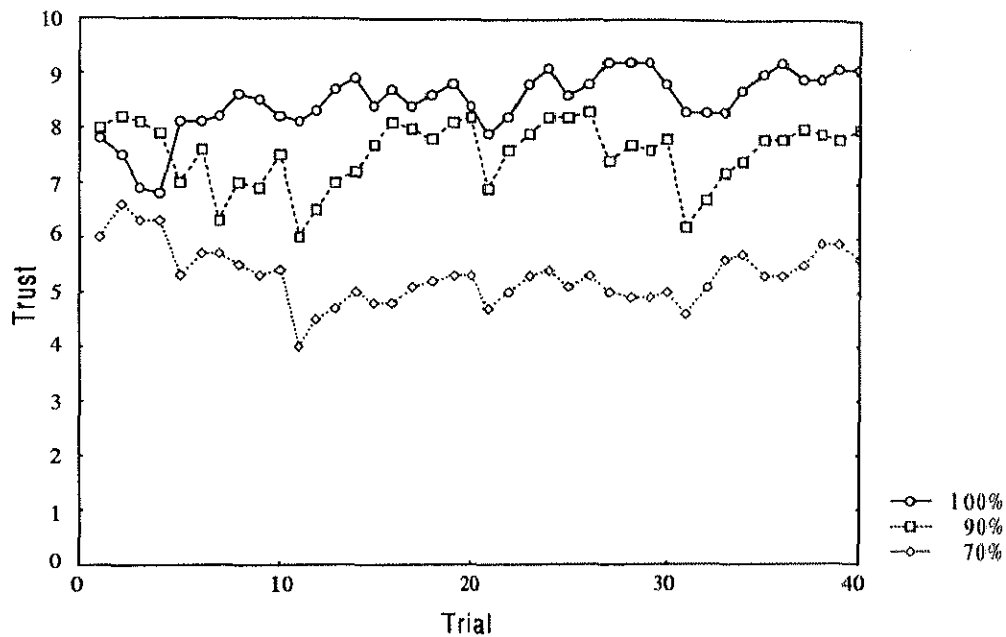


Figure 4.9: Dynamic Behavior of Trust

## trust

trust の値は、安全制御システムの信頼性以外にも、様々な要因に影響を受ける。ここでは、trust の値に影響を及ぼす要因を明らかにし、trust の変動モデルを構築する。Fig. 4.9 に、安全制御システムの信頼性の各群における、各試行の trust の平均値を示す。なお、SAA → M 群では、90%の信頼性のもとで trials 5, 11, 27, 31 に、70%の信頼性のもとで trials 1, 5, 9, 11, 15, 16, 21, 27, 28, 31, 34, 40 に誤報があった。M → SAA 群では、90%, 70%の信頼性のもとでそれぞれ trials 7, 11, 25, 31, trials 1, 7, 8, 11, 14, 20, 21, 25, 29, 31, 35, 36 に誤報が発生した。誤報のある試行では、一試行前の値と比べ trust が低下するケースが多い。このことは、以下の解析によって支持される。

trust の変動について、文献 [49] と同様、一次遅れ系としてモデル化した。まず、trust の変動に影響を及ぼす要因を抽出するための回帰分析を行った。以下の 13 変数が独立変数の候補である。

### 1. 安全制御システムの信頼性 (Reliability)



2. モード体験順序
3. 制御モード (Mode)
4. 異常事象の種類
5. 誤報 (FalseAlarm) の有無
6. 異常事象発生場所
7. 自動パイプ修理の有無
8. 自動プラント停止の有無
9. 手動パイプ修理の有無
10. 手動プラント停止手順開始の有無
11. 警報のキャンセル (Disagree) の有無
12. 事故 (Accident) の有無
13. 誤停止の有無

変数減少法によるステップワイズ式の回帰分析の結果, Reliability (100%: 100, 90%: 90, 70%: 70), Mode (Mモード: 1, SAAモード: 0), FalseAlarm (有: 1, 無: 0), Disagree (有: 1, 無: 0) が採用された. このときの決定係数は  $R^2 = 0.444$  と低い. なお, 文献 [49] での結果と同様, Durbin-Watson 検定によると残差系列が無相関であると言えず, 回帰式の妥当性は保証できない. したがって, このモデルには改善の余地がある.

そこで, 採用された4つの変数のうち, 試行ごとに変化し得るもの, すなわち FalseAlarm および Disagree について, 一試行前の値 (FalseAlarm(n-1), Disagree(n-1)) を独立変数として追加し (n は試行番号), trust の一試行前の値も独立変数として加えた. これらの変数に対して, 変数減少法によって回帰分析を再度実行すると, Mode が独立変数から除去され, 次式を得た.

$$T(n) = \frac{2.78}{100} Reliability + 0.69T(n-1) - 1.01 FalseAlarm(n) - 0.46 Disagree(n)$$

$$-0.33FalseAlarm(n-1) - 0.32Disagree(n-1) \quad (4.1)$$

ただし、 $n$  回目の試行における trust を  $T(n)$  と表す。

このときの決定係数は  $R^2 = 0.751$  となり単純な回帰分析に比べあてはまりがよく、残差系列も無相関であった。したがって、trust の変動を記述する一つのモデルとして妥当なものとなっている。

この trust のモデルから次の 3 点がいえる。

1. 安全制御システムの信頼性が増せば trust は上昇する (4.6 節の解析結果に合致する)
2. 安全制御システムの誤診断が人間の trust を低下させる
3. 安全制御システムの診断が正しい場合でも、人間の見解との間に相違がある場合には、trust が低下する

本論文では、trust のモデルを構築するにあたり、一試行前の値のみを用いた。2 試行以前のデータを取り入れていないのは、2 試行前以前のデータを用いたモデルで解析してみても、ほとんど決定係数が改善されなかったためである。このことは、つぎの self-confidence の解析においても同様であった。

### self-confidence

Fig. 4.10 は、self-confidence が試行の進行に伴いいかに変動したかを示している。図の各点は、安全制御システムの信頼性の各群における平均値である。

trust と同様なプロセスを通じて、self-confidence の動的モデルを構築したところ、次式を得た。

$$SC(n) = \frac{0.69}{100}Reliability + 0.88SC(n-1) - 0.92Accident(n) + 0.72Accident(n-1) \quad (4.2)$$

ここで、 $n$  回目の試行における self-confidence の値を  $SC(n)$  で表す。Accident は、事故の

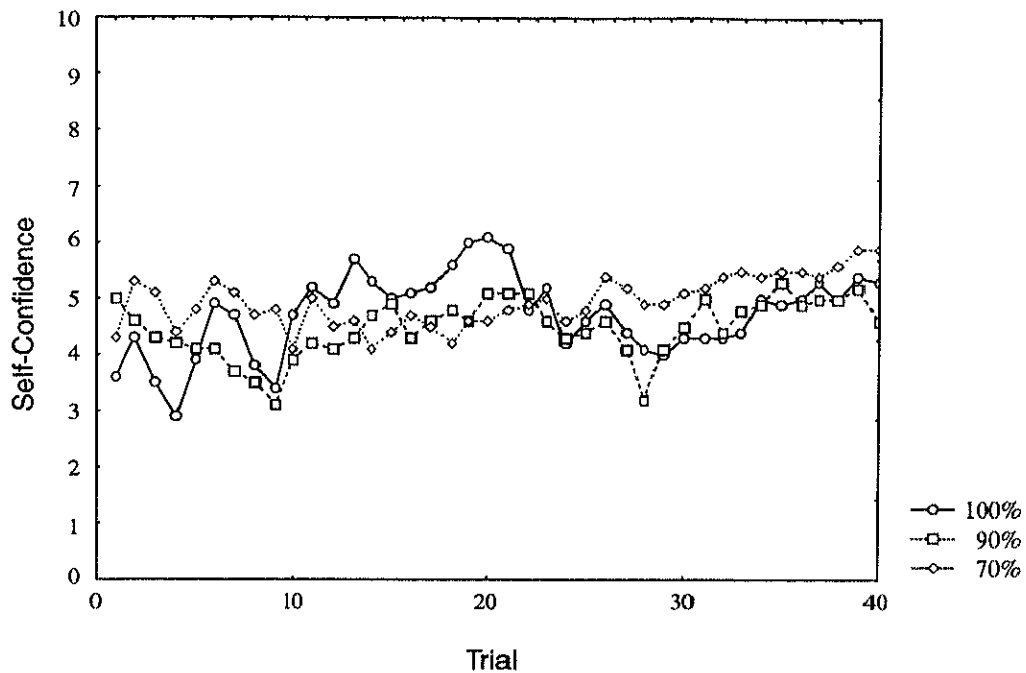


Figure 4.10: Dynamic Behavior of Self-Confidence

生起の有無を表す変数であり，1 ならば事故の発生を，0 ならば事故の非発生を意味する．

(4.2) 式の決定係数は  $R^2 = 0.73$  であり，残差系列も無相関であった．この self-confidence のモデルから，次のことがいえる．

1. self-confidence は，弱いながらも安全制御システムの信頼性に依存する．
2. 現時刻における事故の発生は self-confidence のレベルを 1 程度下げるが，直ちに回復しやすい．すなわち，失敗を経験しても早く立ち直る．

#### trust と self-confidence との比較

(4.1)，(4.2) 式において，各項の係数の値は実験の条件や対象とするシステムによって異なりうるので，それらの値自体を論ずることには意味がない．しかし，式に現れる変数，およびその係数の符号と正規化した係数の大小関係を比較することは有意義であると考えられる．(4.1)，(4.2) 式を比較すると，次のことがいえる．

1. 両者とも、一試行前の自身の値に強く依存する。
2. 安全制御システムの信頼性は、trust へは大きく影響するのに対し、self-confidence への影響は小さい。
3. self-confidence は、事故の生起を含め被験者自身の心理的狀態によってほぼ説明される。これに対し、trust の振舞を説明するには、安全制御システムとの相互作用を考慮に入れることが不可欠である。
4. trust は警報が誤ると低下し、直ちにはもとの値へ回復しない (FalseAlarm の項の係数が、試行  $n$ ,  $n-1$  の場合とも負) のに対し、self-confidence は事故が起きると低下するがすぐにもとの値へ回復しようとする (Accident( $n-1$ ) の係数が正)。

ただし、実験の様子を観察したところによると、自動化システムが権限を持ってプラントを停止した場合に、self-confidence が低下することがまれに見られた。たとえば、プラント停止手順をオペレータが実行している際に、コンピュータが介入してプラントを停止したとき、被験者が「間に合わなかった」と感じた場合が見られた。このことから、self-confidence が、完全に被験者内部の要因だけにより定まるとは限らない。

ところで、trust を Reliability と trust 自身だけで記述してみよう。すなわち、すべての試行において FalseAlarm, Disagree が生起しないものとする。この場合、trust の値の更新を繰り返すと、trust の初期値にはよらずに、大体  $0.09 \text{Reliability}$  に収束する。たとえば、100%の信頼性に対しては trust の値が 9 に収束する。このことは、100%の群の被験者が安全制御システムの信頼性をよく見抜いていることを示している。

これに対し、self-confidence では、収束先の値は、大体  $0.06 \text{Reliability}$  である。実験で被験者が体験したのは 70 % の信頼性までであることから、self-confidence の値は 4 から 6 程度の値に落ち着くことがこのモデルからいえる。実際、Fig. 4.10 では、各群の self-confidence の平均値がほぼ 4 から 6 の間の値を取っている。self-confidence が被験者自身に強く依存することから、被験者に偏りがなければ self-confidence の平均は 5 になると考えられる。ここで得られた結果は、被験者が適切に各群に配置された、すなわち実験が適

Table 4.9: Mean Values of Trust as a Function of Site of Break

Reliability	L1		L2		L3		L4		L5		L6	
	SAA	M	SAA	M	SAA	M	SAA	M	SAA	M	SAA	M
100%	8.6	8.5	8.3	9.0	8.7	8.8	8.0	9.0	8.55	8.3	8.2	8.6
90%	7.0	7.15	7.9	7.4	7.55	7.2	7.3	7.9	7.45	6.1	8.0	8.3
70%	4.45	5.8	4.6	5.75	4.85	5.65	3.9	5.6	5.05	4.5	4.75	5.5

切に実施されたことを裏付けるものである。

## 4.7 意思決定の自動化への許容

自動化システムへの依存性，すなわち自動化システムを利用しようとするか否かは， $\text{trust}(T)$ ， $\text{self-confidence}(SC)$ の値によって定まる [51]。すなわち， $T - SC$ が正ならば自動化に頼ろうとし，負ならば手動操作を選択する。しかし，本実験では，自動化システムの利用を被験者が積極的に選択できるようには設定していない。ここでは， $T - SC$ を評価することによる自動化システムへの依存性の議論はしない。

Muir と Moray[60]の実験では， $\text{trust}$ が低下したときには手動操作を選択することを見出している。レベル7の自動化の場合など，本実験では手動操作へ切り替えることができない状況もあるので，自動化システムに頼ろうとしたか被験者の選択を解析することはできないが，どのような場合に  $\text{trust}$ が下がるのか明らかにすることはできる。ここでは，高いレベルの自動化の実現可能性，すなわち，パイプ破断に対する自動化のあり方に注目する。

Table 4.9 に示す異常事象の発生個所別の  $\text{trust}$ の平均値を見ると，場所毎，すなわち自動化レベルの違いによって  $\text{trust}$ の「値」に大きな違いは見られない。

Table 4.9 は、ある場所で破断が発生した場合の trust について、ほかの場所での破断の場合との「高低」を比較しているが、「増減」を評価するものではない。trust は時間的な側面を持つ [49, 50] ことから、高いレベルの自動制御が行われた場合などに trust がどのような振る舞いをするかを、動的な観点から評価する必要がある。

#### 4.7.1 trust の動的変化

4.6.2 節では、人間の意見と安全制御システムの意見との不一致が、trust の低下につながることを明らかにした。本節以降では、仮に人間と安全制御システムとの意見の不一致があっても、人間の意思が制御に反映されない場合、すなわち、自動化レベルが 7 である場合に注目する。本実験では、レベル 7 の自動化を必要とする異常事象は LI 破断のみであり、その発生頻度は比較的少ない。したがって、trust の「値」についての定量的な解析は困難である。そこで、連続する 2 つの試行における trust の差分の符号を調べることとし、本節では、実験全体におけるその符号の頻度を解析する。

被験者は Fig. 4.11 に示された尺度に基づいて trust のレベルを回答する。現試行での trust の値を被験者が入力しようとするとき、前の試行での値がこのウィンドウに示されている。すなわち、被験者は、一試行の自分の trust を参考にして、現在の trust を決定できる。したがって、trust の値の変化は偶然ではなく、被験者の意思に基づくものであると言える。

Table 4.10 に、trust の上昇・低下の頻度分布を示す。なお，“+”は一試行前と比べ trust が上昇すること，“0”は同じ値に留まること，“-”は低下することを表し、括弧内の数値は残差を表す（残差の右肩についている\*,\*\*については後述する）。実験は 10 試行で 1 つのまとめ（セット）をなし、各セット最初の試行の終了時には、Fig. 4 のポインタ（▼）は初期値として 5 を指している。そこで、全 40 回の試行のうち、1, 11, 21, 31 番目の試行では trust の変動を考えないこととする。

Table 4.10 を見ると、全体として trust は同じ値に留まりやすく、trust の上昇の頻度は

**SUBJECTIVE-RATING**

The trial is over. Drag the pointers to give your opinion.

1. On the basis of my experience until now, I trust the automation

not at all 0 
4
10
 completely

2. On the basis of my experience until now, I am confident that I can control the plant by myself without the automation

not at all 0 
8
10
 completely

If you want to change your mind, you can do it again.  
When you have finished, click "done"

done

When "done" is clicked, all data are stored and next run begins.

Figure 4.11: Rating Scale for Subjective Judgments of Trust and Self-Confidence

Table 4.10: Difference of Trust

Reliability	+	0	-	total
100%	65(-.230*)	264(3.51**)	31(-2.60**)	360
90%	91(3.48**)	226(-2.30*)	43(0.91)	360
70%	70(-1.18)	244(-0.08)	46(-0.91)	360

低下のそれよりもやや多い。このことは、「基本的には、trust は経験を積むにつれて徐々に上昇する」ことを表していると考えられる。この結果は、Muir と Moray[60] が指摘した、「自動化への trust を高めるには、自動化を使わせる必要がある」ことに関連する。

安全制御システムの信頼度ごとの度数分布には偏りがある。実際、trust の差分の符号の度数について  $\chi^2$  検定を行ったところ、安全制御システムの信頼性間に有意な違いが見られた ( $\chi^2(4)=11.157, p<0.05$ )。残差を分析したところ、期待度数と実際の度数との差が、\* のものは危険率 5% で、\*\* のものは危険率 1% で有意であった。したがって、次のことがいえる。

1. 100%の信頼性のもとでは、他の信頼性の場合と比べて trust は同じ値に留まる傾向があり、上昇や低下の頻度が少ない。
2. 90%の信頼性のもとでは、他と比べ trust は同じ値には留まりにくく、上昇する頻度が多い。
3. 70%の信頼性のもとでは、trust の変動の分布は全体の分布にほぼ一致する。

#### 4.7.2 高レベルの自動化への trust

レベル7の自動化による介入に対して信頼が増しているケースがある。このことを裏づけるものとして、実験後のインタビューで「十分に信頼できる」とする意見が得られた。これに対し、何名かの被験者が「一体何が起きたのか分からなかった」と答えているが、このことは「オートメーションサプライズ」[77]が発生していることを示唆している。さらに、レベル7の自動化による介入への不快感を示した被験者も見られた。

Table 4.11 はレベル7の自動化による介入のもとでの、trust の変化に関する頻度を抜き出したものである。ただし、100%、90%の信頼性のもとでは各被験者につき2試行あるが、70%のもとでは1試行が診断を誤るシナリオであるため1試行で集計した。レベル7の自動化は、信頼度の群ごとに異なる影響を与えている。Table 4.11において、100%のも



Table 4.11: Changes of Trust unde Level 7 Autonomy

Reliability	+	0	-	total
100%	4	16	0	20
90%	3	8	9	20
70%	2	6	2	10

とでは全く trust が低下しないのに対して、90%のもとでは20 ケースのうち9回 trust が低下している。trust の変動を、低下・非低下のカテゴリに分けて2項検定 [86] を行った結果、100%および70%の信頼性のもとでは trust の低下数は Table 4.10 の各信頼性のもとでの分布に比べて違いは見られなかった。これに対し、90%の信頼性のもとでは Table 4.10 の90%での分布に比べ trust の低下数には有意な差があった ( $p=0.000233$ )。本節の trust に関する解析結果から、次の結論を得る。

1. 100%の信頼性のもとでは、被験者は過去の経験から安全制御システムに対して高い信頼感を抱いている。たとえ、自動化レベル7による安全制御への介入を体験しても、信頼感は損なわれることがない。
2. 90%の信頼性のもとでは、一見100%の信頼性と同じ程度に被験者は安全制御システムを信頼している。しかし、誤診断を経験していることからその信頼は安定しておらず、レベル7の自動化による介入に対し快く思わないことをきっかけとして信頼感が低下する傾向が見られる。
3. 70%の信頼性のもとでは、誤報が多く発生する結果、被験者は安全制御システムを十分には信用しなくなる。しかも、システムの挙動に対する被験者の感受性が次第に鈍り、信頼感は低い値で安定するようになる。したがって、自動化レベル7による介入があっても信頼感のさらなる低下は起きにくい。

### 4.7.3 モード認識を阻害しない自動化レベル

信頼性が90%のときに、レベル7の自動化に対して trust が低下することが顕著であったが、安全制御システムが人間の信頼を失うのは、モード認識を損ねさせることがあるからであると考えられる。システムの信頼性を完全に、100%にすることは現実的ではない。そこで、本実験での90%の信頼性に象徴されるような「高信頼性」のもとで、人間の trust を低下させない、人に許容されるしきみを確立しなければならない。

一つの案として、レベル7の自動化を要する緊急事態において、次の自動化レベルを提案する [37]。

**定義 6** コンピュータがすべてを実行する。ただし、実行する前に、これから何を実行するか人間に報告する。

この自動化のレベルは、6よりも高く、7よりも低いと考えられることから、6.5のレベルであるということができる。

このレベル6.5の自動化によって、人間のモード認識を阻害せず、人間から自動化システムへ権限の委譲がスムーズに達成されるものと期待できる。

## 4.8 むすび

本章では、認知実験を通じて、レベル5を超える自動化が安全確保に有効であること、安全制御に関連して、人間を完全に排除すべきではないことを確認した。すなわち、システム安全の確保に SAA が有効であることを明らかにした。

本章で得られた結果は、SAA がシステムの安全確保に貢献することを示している。すなわち、「人間が最終決定権を持つ」べきとする人間中心の原理は、人間の状況判断に十分な時間が確保でき、人間が責任を全うできるという条件付きで受け入れるべきものであるといえる。このことは、数理モデルの解析 [30] ですでに指摘されていたが、実験によって確認した点に本研究の意義がある。

本章では、実験データに基づいて、trust と self-confidence を静的、動的側面からのモデル化も行った。その結果、事故の発生が self-confidence を低下させるが比較的早くもとの値へ回復するのに対し、コンピュータの誤診断が trust を低下させ、その場合 trust は直ちにはもとの値へ回復しようとしなかったことがわかった。trust がシステムの信頼性に強く依存し、人間がその信頼性をよく見抜いていることも明らかになった。また、人間とシステムとの意見の相違が trust を低下させることも示された。さらに、レベル7の自動化は安全確保には有効であるが、システムが「高信頼度」のときに人間の信頼を失わせるきっかけともなりうることを解明した。この結果に関連して、モード認識を阻害しない新しいタイプの自動化レベル、すなわち6.5のレベルの自動化を提案した。

本章の最後で提案した新たな自動化レベルを現実のシステムに応用することにより、人間とコンピュータとの信頼関係がより円滑に構築できるようになるものと期待できる。

## 第5章 結論

### 5.1 本論文の結果と意義

本論文では、監視制御系におけるオペレータの活動、とくに認知、判断、実行の支援に関して、システムの大規模性に伴う状況の不確実性と、システムの動特性がもたらす意思決定の時間制約、の二つの観点から論じた。

本論文は、大きく二つの部分に分かれる。すなわち、不確実状況下での意思決定支援のために、証拠理論における信念の更新機構を整理・拡張し、その性質を解明した第2章と、時間制約のもとで意思決定を自動化しうる状況適応自動化の必要性や有効性などを明らかにした第3、4章である。

第2章では、証拠理論における情報処理の方法として情報統合と情報更新とがあることを示した。また、情報更新について、従来提案されている規則の問題点を指摘し、そうした問題点を解消する新たな情報更新規則を、従来の方法を含む一般的な形で与えた。さらに、プラント安全制御の問題を例にとり、期待損失を最小化する意味で適用規則を最適化する観点から情報統合と情報更新との比較をし、両者を使い分ける指針を示した。

第2章の主要な結果は、以下の通りである。

1. 情報更新の一部をなす条件付けに、本論文で新たに提案した規則を用いることにより、情報更新があらゆる状況で適用できるようになった。
2. 本論文で提案した信念の条件付けの規則が非可算無限個存在しうることから、情報更新に柔軟さを持たせられる。すなわち、新たに獲得した情報をどの程度受け入れるかを、条件付け規則の選び方として表現できる。

3. 情報統合は安全制御方策の選び方に応じて最適規則が定まる。これに対し、情報更新はどの安全制御方策が選ばれているかには無関係に最適規則が定まる意味で、オペレータとは独立に規則を選択できる。
4. 情報統合の場合、最適規則は時刻に依存するのに対し、情報更新の場合、最適規則は時刻に依存しない。
5. システム状態が時間の推移に伴って変化しない場合のように、静的な環境に適用する信念更新手法としては情報統合がふさわしい。これに対し、時々刻々状況が変化する動的な場合には情報更新が適している。

本章で得た結果の意義は、合理的な情報の解釈や判断を提案する意思決定支援に、証拠理論の情報更新を応用できるようになった点にある。とくに、動的な環境では情報更新を採用すべきである。さらに、情報更新は、規範的意思決定に適用できるだけでなく、その柔軟さにより人間の決定過程を記述できる可能性がある。このことから、情報の合理的解釈と人間による解釈とのずれを評価しそのずれを人間に認知させることにより、バイアスにかかりにくくするタイプの支援ができると考えられる。

本章の成果の具体的な応用例としては、安全監視システムや警報システムの設計がある。すでに、Tanaka[90, 91, 92, 118]により、証拠理論的安全監視システムの研究が静的な問題設定のもとで進められているが、その研究を動的な状況における安全監視システムの設計へ拡張できると期待される。

本論文の後半において、第3章では、状況適応的な意思決定の自動化の必要性を確かめる実験を行った。第4章では、状況適応自動化について、認知実験により有効性を明らかにし、その実現可能性を trust と self-confidence の観点から検討した。

第3, 4章で得られた主要な結果は以下の通りである。

1. 状況適応自動化が、システム安全を確保するためには必要であり有効である。すなわち、緊急度が高い場合にはコンピュータが意思決定の権限を持つ方がよいが、緊急

度が低い場合などには人間が自動化システムの誤作動による影響を打ち消すのに役立つ。

2. 事故発生といった、オペレータ自身にも原因がある失敗を経験した場合、self-confidence は低下するものの比較的早く回復する。これに対し、コンピュータの失敗（誤診断）を経験した場合 trust は低下し、すぐには回復しない。
3. trust がコンピュータの信頼性に強く依存している。また、人間はコンピュータの信頼性をよく見抜いている。
4. コンピュータの判断が正しくとも、人間の判断とコンピュータの判断とが食い違う場合に trust は低下する。
5. 自動化が高いレベル（7）になったときに trust は失われることがある。ただし、このことはコンピュータの信頼性に依存し、コンピュータが「高信頼度」なときにその傾向が顕著になる。

以上のことから、安全確保のための状況適応自動化を実現するには、つぎのことが要請される。

1. コンピュータの信頼性は可能な限り高くなければならない。
2. ただし、コンピュータの意図を明確に人間に伝えられなければならない。

第3,4章で得られた結果の意義は次の点にある。すなわち、緊急時にはコンピュータが決定権を持つべきであるが、その場合コンピュータはオペレータに説明する義務も負う。これに関し、本論文では、コンピュータが説明責任を果たすための方法として、コンピュータの意図を明示させる新しい自動化（レベル6.5）を提案した。

状況適応自動化は、航空の分野における緊急時のシステム安全の確保などに応用可能であると考えられる。たとえば、旅客機の離陸時にエンジン故障が発生した場合の離陸中断・継続に関する意思決定では、状況適応的な自動化が必要かつ有効であることを示す結果が報告されつつある [31, 36].

## 5.2 今後の研究課題

2章で得られた結果を受け、今後の研究課題や展開として、以下のことが考えられる。

1. 確率論に基づく意思決定論では、情報を逐次獲得する状況をモデル化した意思決定が、逐次決定過程 [93, 119] として知られている。証拠理論に基づく逐次決定過程を定式化し、動的な意思決定を最適化する枠組みを確立することが必要である。
2. 数学モデルは、抽象化の段階で現実世界における様々な要因を無視してしまうことから、数理的解析によって得られた結果が実際に成立するかは自明ではない。そこで、証拠理論的な意思決定支援システムを実際に構築し、認知実験によってその有効性を検討しなければならない。
3. 現在、エキスパートの意思決定の過程を記述する考え方として、Naturalistic Decision Making(NDM)[101]が注目されつつある。NDMとは、現実の場面において人が意思決定する際に、過去の経験をいかに利用するかを記述する試みである [44](p. 4)。Klein [44, 45] は、「エキスパートは必ずしもすべての選択肢を比較検討して意思決定しているわけではない」ことを示している。これに対し証拠理論では考察の対象の「粗さ」を必要に応じて変更できることから、考察の範囲をエキスパートがいかに設定しているかを証拠理論により説明できるようになると考えられる。また、情報更新が非可算無限個の規則を持つという多様性により、エキスパートの状況把握の過程を情報更新によりモデル化できると考えられ、この点について今後検討する必要がある。
4. 本論文では、trust, self-confidenceといった人間の主観を、いわば確率的に定量化した。たとえば、trustが1であることは、ほとんど信頼していないことを意味し、trustが5であることは信頼感と不信感が拮抗していることを意味する。この方法では、経験が浅い段階での信頼感と、経験を十分に積んだのちに形成された信頼感とを区別できないという問題もある。たとえば、闇雲に信じ込む「盲信」と、過去の経験から「過信」することとが同じものとして扱われる。trustなど人間の主観を証拠理論

の測度で表現し、その動的な変動を証拠理論の情報更新機構で記述することにより、trust と self-confidence のより適切なモデルが構築できるようになると考えられる。

また、第3、4章の研究に関連して、以下の事項が今後の研究課題である。

1. 本論文では、状況に応じた自動化レベルの変更が安全確保に有効であることを確かめた。しかし、個々の状況において自動化をどのレベルに設定すべきかは自明ではない。状況適応自動化を実現するには、時間余裕やリスクと最適な自動化レベルとの関係を明確にして、自動化システムの設計方法を確立する必要がある。
2. コンピュータの意図を人間にわからせることが、人間とコンピュータとの間で適切な信頼関係を築くために必要である。そのためには、人間とコンピュータとの対話が円滑にできなければならない。人間とコンピュータとの間での信頼関係を確立できるための、ヒューマンマシンインタフェースの設計原理を構築しなければならない。
3. 本研究では、一人のオペレータと一つのコンピュータとからなるヒューマンマシンシステムを解析の対象としたが、現実には、オペレーションは複数の人間により共同で遂行される。チーム行動の観点からヒューマンマシンシステムを考察し、人間（群）とコンピュータとの役割分担を最適化するための方策を構築することが望まれる。
4. 認知実験に対しては、被験者や状況設定の妥当性が批判の対象となる。たとえば、本論文では実験の都合上一回の試行は6分半や2分半であったが、現実には一人のオペレータが数時間続けて勤務する。そこで、現実に存在するシステムのシミュレータを用い、現実的な状況を想定して、実際のオペレータを被験者とする認知実験を行い、状況適応自動化の有効性を検証する必要がある。



## 謝辞

7年間にわたりご指導を賜りました，筑波大学電子・情報工学系教授 稲垣 敏之先生に深く感謝いたします。稲垣先生には，本研究に着手するきっかけを与えていただいただけでなく，私が筑波大学 助手として，筑波大学先端学際領域研究 (TARA) センター マルチメディア情報研究アспект研究プロジェクトに参加し，研究に専念する機会をくださいました。

TARA マルチメディア情報研究アспект，元アспект教授 (現 筑波大学電子・情報工学系長) 板野 肯三先生，現アспект教授 寅市 和男先生には，TARA 研究プロジェクトの推進に様々な形で便宜を図っていただいたことに謝意を表します。また，寅市先生には，励ましの言葉を何度も頂いたばかりでなく，学位論文の審査委員として，貴重なご意見を賜りました。

筑波大学電子・情報工学系 教授 大田 友一先生，同 平井 有三先生，同 宮本定明先生には，博士論文審査委員として多くの有益なご意見を賜りましたことに感謝いたします。

筑波大学 電子・情報工学系 教授 西原 清一先生，助教授 和田 耕一先生には，TARA 研究員として実験の遂行にご協力いただくなど，多くのご支援を賜りましたことに謝意を表します。

TARA 研究プロジェクト客員教授 門田 一雄先生 ((財) 原子力発電機構ヒューマンファクターセンター所長) には，先生が筑波大学に来学された折りに頻繁に議論させていただき，有益な数多くのご示唆をいただきました。ここに謝意を表します。

英国サリー大学心理学科教授 Neville P. MORAY 先生には，TARA 研究プロジェクト

を通じて、公私共にお世話になったことに深く感謝いたします。Moray 先生は、研究者としてあるべき姿を身を持って教えてくださいました。また、2度にわたり Moray 先生の私宅に滞在させていただき、議論を交わさせていただいたことは、その後の研究に役立ったばかりでなく、一生の思い出となりました。

科学技術庁航空宇宙技術研究所飛行実験部人間機械研究室長 田中 敬司氏、東京電力(株)技術開発本部原子力研究所ヒューマンファクターグループ主管研究員 河野 龍太郎氏には、TARA 研究プロジェクトの一員として、研究の遂行にご協力いただきましたことに感謝いたします。とくに河野氏には、認知実験を遂行するにあたり多くの貴重なご助言を賜りました。

運輸省船舶技術研究所システム技術部松岡猛氏、福戸淳司氏には、SCARLETT の開発にあたりご協力をいただきましたことに対してお礼申し上げます。

電気通信大学大学院情報システム学研究科情報システム運用学専攻社会情報システム学講座教授 太田敏澄先生、同助教授 田中健次先生、同助手 石田 和成先生には、私が電気通信大学において学位論文をまとめることを快く了承していただき、様々な便宜を図って下さいましたことに感謝いたします。また田中先生には、筑波大学大学院在学中より証拠理論の研究に関してしばしば討論させていただき、有益なご示唆を賜りました。

筑波大学電子・情報工学系認知システム科学研究室の皆様には、本研究を進めるにあたり有意義な討論を何度もさせていただきましたことに謝意を表します。

最後に、今日に至るまでさまざまな形で支えてくださった、父敬三、母かづ江および親族に深く感謝します。

## 参考文献

- [1] L. Bainbridge. Ironies of automation. In J. Rasmussen, et al., editors, *New Technology and Human Error*, pp. 271–283. Wiley, 1987.
- [2] B. Barber. *The Logic and Limits of Trust*. Rutger University Press, 1983.
- [3] J. M. Bernado and A. F. M. Smith. *Bayesian Theory*. Wiley, 1994.
- [4] C. E. Billings. Human centered aircraft automation: A concept and guidelines. Technical Report TM-103885, NASA, 1991.
- [5] C. E. Billings. Toward a human-centered aircraft automation philosophy. *The International Journal of Aviation Psychology*, Vol. 1, No. 4, pp. 261–270, 1991.
- [6] C. E. Billings. *Aviation Automation*. LEA, 1997.
- [7] Boeing. *Statistical Summary of Commercial Jet Aircraft Accidents*. 1995.
- [8] B. G. Buchanan and E. H. Shortliffe. *Rule-Based Expert Systems - The MYCIN Experiments of the Stanford Heuristic Programming Project*. Addison-Wesley, 1984.
- [9] L. M. de Campos, M. T. Lamata, and S. Moral. The concept of conditional fuzzy measure. *International Journal of Intelligent Systems*, Vol. 5, pp. 237–246, 1990.
- [10] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, Vol. 38, pp. 325–339, 1967.

- [11] E. M. Dougherty and J. R. Fragola. *Human Reliability Analysis*. Wiley, 1988.
- [12] D. Dubois and H. Prade. Combination and propagation of uncertainty with belief functions. In *Proceedings of International Joint Conference on Artificial Intelligence*, Vol. 1, pp. 111–113, 1985.
- [13] D. Dubois and H. Prade. On the unicity of Dempster rule of combination. *International Journal of Intelligent Systems*, Vol. 1, pp. 133–142, 1986.
- [14] D. Dubois and H. Prade. Conditioning, non-monotonic logic, and non-standard uncertainty models. In H. T. Nguyen I. R. Goodman, M. M. Gupta and G. S. Rogers, editors, *Conditional Logic in Expert Systems*, pp. 115–158. Elsevier Science Publishers B. V., 1991.
- [15] D. Dubois and H. Prade. Updating with belief functions, ordinal conditional functions and possibility measures. In *Uncertainty in Artificial Intelligence*, Vol. 6, pp. 311–329. Elsevier Science Publishers B. V., 1991.
- [16] D. Dubois and H. Prade. Focusing versus updating in belief function theory. In R. R. Yager, et al., editors, *Advances in the Dempster-Shafer Theory of Evidence*, pp. 71–95. Wiley, 1994.
- [17] D. Dubois and H. Prade. A survey of belief revision and updating rules in various uncertainty models. *International Journal of Intelligent Systems*, Vol. 9, pp. 61–100, 1994.
- [18] D. Dubois and H. Prade. A synthetic view of belief revision with uncertain inputs in the framework of possibility theory. *International Journal of Approximate Reasoning*, Vol. 17, pp. 295–324, 1997.

- [19] W. Edwards. Conservatism in human information processing. In B. Kleinmuntz, editor, *Formal Representation of Human Judgement*. Wiley, 1968.
- [20] M. R. Endsley. Measurement of situation awareness in dynamic systems. *Human Factors*, Vol. 37, No. 1, pp. 64–84, 1995.
- [21] M. R. Endsley. Toward a theory of situation awareness in dynamic systems. *Human Factors*, Vol. 37, No. 1, pp. 32–64, 1995.
- [22] FAA. *Takeoff Safety Training Aid*, 1992.
- [23] I. R. Goodman, M. M. Gupta, H. T. Nguyen, and G. S. Rogers, editors. *Conditional Logic in Expert Systems*. North-Holland, 1991.
- [24] J. Guan and D. A. Bell. *Evidence Theory and Its Applications*, Vol. 1. North-Holland, 1991.
- [25] J. Y. Halpern and R. Fagin. A new approach to updating beliefs. In *Uncertainty in Artificial Intelligence*, Vol. 6, pp. 347–373. Elsevier Science Publishers B. V., 1991.
- [26] H. Ichihashi and H. Tanaka. Jeffrey-like rules of conditioning for the Dempster-Shafer theory of evidence. *International Journal of Approximate Reasoning*, Vol. 3, pp. 143–156, 1989.
- [27] T. Inagaki. Interdependence between safety-control policy and multiple-sensor schemes via Dempster-Shafer theory. *IEEE Transactions on Reliability*, Vol. 40, No. 2, pp. 182–188, 1991.
- [28] T. Inagaki. Dempster-Shafer theory and its applications. In K. B. Misra, editor, *New Trends in System Reliability Evaluation*, Vol. 16 of *Fundamental Studies in Engineering*, chapter 15, pp. 587–624. Elsevier Science Publishers B. V., 1993.

- [29] T. Inagaki. Situation-adaptive degree of automation for system safety. In *Proceedings of IEEE International Workshop on Robot and Human Communication*, Vol. 2, pp. 231–236, 1993.
- [30] T. Inagaki. Situation-adaptive responsibility allocation for human-centered automation. *Transactions of the Society of Instrument and Control Engineers*, Vol. 31, No. 3, pp. 292–298, 1995.
- [31] T. Inagaki. To go or not to go: Decision under time criticality and situation-adaptive autonomy for takeoff safety. In *Proceedings of IASTED International Conference on Applied Modeling and Simulation*, pp. 144–147, 1997.
- [32] T. Inagaki and K. Fujioka. Decision support for multiple-fault diagnosis under vague information. In *Proceedings of 12th European Annual Conference on Human Decision Making and Manual Control*, 1993.
- [33] T. Inagaki and S. Hasegawa. Safety-control of large-complex systems: Intelligent information processing and decision making. In *Proceedings of 12th European Annual Conference on Human Decision Making and Manual Control*, 1993.
- [34] T. Inagaki and Y. Ikebe. A mathematical analysis of human-machine interface configurations for a safety monitoring system. *IEEE Transactions on Reliability*, Vol. 37, pp. 55–40, 1988.
- [35] T. Inagaki and M. Itoh. Trust, autonomy, and authority in human-machine systems: Situation-adaptive coordination for systems safety. In *Proceedings of Cognitive Systems Engineering for Process Control*, pp. 176–183, 1996.

- [36] T. Inagaki and M. Itoh. Situation-adaptive autonomy: The potential for improving takeoff safety. In *Proc. IEEE International Workshop on Robot and Human Communication*, pp. 302–307, 1997.
- [37] T. Inagaki, N. Moray, and M. Itoh. Trust and time-criticality: Their effects on the situation-adaptive autonomy. In *Proceedings of International Symposium on Artificial Intelligence, Robotics, and Intellectual Human Activity Support for Nuclear Applications*, pp. 93–104, 1997.
- [38] T. Inagaki, N. Moray, and M. Itoh. Trust, self-confidence and authority in human-machine systems. In *Proceedings of 7th IFAC/IFIP/IFORS/IEA Conference on Analysis, Design and Evaluation of Man-Machine Systems*, pp. 491–496, 1998.
- [39] M. Itoh and T. Inagaki. Design of human-interface for situation awareness. In *Proceedings of IEEE International Workshop on Robot and Human Communication*, pp. 478–483, 1996.
- [40] M. Itoh and T. Inagaki. Combination and updating for belief revision in the theory of evidence. In G. Grahne, editor, *Proceedings of the 6th Scandinavian Conference on Artificial Intelligence*, Vol. 40 of *Frontiers in Artificial Intelligence and Applications*, pp. 71–82. IOS Press, 1997.
- [41] J. Jaffray. Dynamic decision making with belief functions. In R. R. Yager, M. Fedrizzi, and J. Kacprzyk, editors, *Advances in the Dempster-Shafer Theory of Evidence*, chapter 15, pp. 331–352. Wiley, 1994.
- [42] R. C. Jeffrey. *The Logic of Decision*. University of Chicago Press, 1983.
- [43] B. H. Kantowitz and R. D. Sorkin. Allocation of functions. In G. Salvendy, editor, *Handbook of Human Factors*, chapter 3.3, pp. 355–369. Wiley, 1st edition, 1987.

- [44] G. Klein. The recognition-primed decision (RPD) model: Looking back, looking forward. In C. E. Zsombok and G. Klein, editors, *Naturalistic Decision Making*, chapter 27, pp. 285–292. LEA, 1997.
- [45] G. A. Klein. Recognition-primed decisions. *Advances in Man-Machine Systems Research*, Vol. 5, pp. 47–93, 1989.
- [46] T. Kletz. Some incidents that have occurred, mainly in computer-controlled process plants. In *Computer Control and Human Error*. Institution of Chemical Engineers, 1995.
- [47] G. J. Klir. Measures of uncertainty in the Dempster-Shafer theory of evidence. In R. R. Yager, et al., editors, *Advances in the Dempster-Shafer Theory of Evidence*, pp. 35–49. Wiley, 1994.
- [48] C. T. Le. *Applied Survival Analysis*. Wiley, 1997.
- [49] J. Lee. *Trust, Self-Confidence, and Operator's Adaptation to Automation*. PhD thesis, University of Illinois, 1992.
- [50] J. Lee and N. Moray. Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, Vol. 35, No. 10, pp. 1243–1270, 1992.
- [51] J. Lee and N. Moray. Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies*, Vol. 40, pp. 153–184, 1994.
- [52] F. Lévy. A survey of belief revision and updating in classical logic. In L. Sombé, editor, *Revision and Updating in Knowledge Base*, pp. 29–59. Wiley, 1994.
- [53] D. K. Lewis. Probabilities of conditionals and conditional probabilities. *The Philosophical Review*, Vol. 85, pp. 297–315, 1976.



- [54] J. Lovell and J. Kluger. *APOLLO 13*. Pocket Books, 1994.
- [55] R. Molloy and R. Parasuraman. Monitoring an automated system for a single failure: Vigilance and task complexity effects. *Human Factors*, Vol. 38, No. 2, pp. 311–322, 1996.
- [56] N. Moray, T. Inagaki, and M. Itoh. Situation adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology: Applied*. (to appear).
- [57] M. Mouloua and R. Parasuraman, editors. *Human Performance in Automated Systems: Current Research and Trends*. LEA, 1994.
- [58] B. M. Muir. Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies*, Vol. 27, pp. 527–539, 1987.
- [59] B. M. Muir. Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. Vol. 37, No. 11, pp. 1905–1922, 1994.
- [60] B. M. Muir and N. Moray. Trust in automation. part II. Experimental studies of trust and human intervention in a process control simulation. Vol. 39, No. 3, pp. 429–460, 1996.
- [61] T. Murai, M. Miyakoshi, and M. Shimbo. Measure-based semantics for modal logics. In R. Lowen, et al., editors, *Fuzzy Logic: State of the Art*, pp. 395–405. Kluwer, 1993.
- [62] T. Murai, M. Miyakoshi, and M. Shimbo. Soundness and completeness theorems between the Dempster-Shafer theory and logic of belief. In *Proc. 3rd FUZZ-IEEE*, pp. 855–858, 1994.

- [63] R. Parasuraman, R. Molloy, and I. L. Singh. Performance consequences of automation-induced “complacency”. *International Journal of Aviation Psychology*, Vol. 3, No. 1, pp. 1–23, 1993.
- [64] R. Parasuraman and M. Mouloua, editors. *Automation and Human Performance*. LEA, 1996.
- [65] R. Parasuraman and V. Riley. Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, Vol. 39, No. 2, pp. 230–252, 1997.
- [66] F. P. Ramsey. Truth and probability. In R. B. Braithwaite, editor, *Foundations of Mathematics and Other Logical Essays*. Routledge and Kegan Paul, 1931.
- [67] J. Rasmussen. Skill, rules and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 13, No. 3, 1983.
- [68] J. Reason. *Managing the Risks of Organizational Accidents*. Ashgate, 1997.
- [69] J. K. Rempel, J. G. Holmes, and M. P. Zanna. Trust in close relationships. *Journal of Personality and Social Psychology*, Vol. 49, No. 1, pp. 95–112, 1985.
- [70] W. B. Rouse. Adaptive allocation of decision making responsibility between supervisor and computer. In T. B. Sheridan and G. Johanssen, editors, *Monitoring Behavior and Supervisory Control*, pp. 295–306. Plenum Press, 1976.
- [71] W. B. Rouse. *Design for Success*. Wiley, 1991.
- [72] W. B. Rouse. Twenty years of adaptive aiding: Origins of the concept and lessons learned. In M. Mouloua and R. Parasuraman, editors, *Human Performance in Automated Systems: Current research and trends*, pp. 28–32. LEA, 1994.

- [73] S. Russell and P. Norvig. *Artificial Intelligence*. Prentice Hall, 1995.
- [74] N. B. Sarter and D. D. Woods. Pilot interaction with cockpit automation. *International Journal of Aviation Psychology*, Vol. 2, No. 4, pp. 303–321, 1992.
- [75] N. B. Sarter and D. D. Woods. Pilot interaction with cockpit automation II. *International Journal of Aviation Psychology*, Vol. 4, No. 1, pp. 1–28, 1994.
- [76] N. B. Sarter and D. D. Woods. How in the world did we ever get into that mode? mode error and awareness in supervisory control. *Human Factors*, Vol. 37, No. 1, pp. 5–19, 1995.
- [77] N. B. Sarter and D. D. Woods. Strong, silent, and ‘out-of-the-loop’: Properties of advanced (cockpit) automation and their impact on human-automation interaction. Technical Report CSEL report 95-TR-1, Ohio State University, 1995.
- [78] M. W. Scerbo. Theoretical perspectives on adaptive automation. In Mouloua Parasuraman, editor, *Automation and Human Performance*, pp. 37–63. LEA, 1996.
- [79] W. B. Scott. New research identifies causes of CFIT. *Aviation Week and Space Technology*, pp. 70–71, June 17 1996.
- [80] J. W. Senders and N. P. Moray, editors. *Human Error: Cause, Prediction, and Reduction*. LEA, 1991.
- [81] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
- [82] J. Sharit. Allocation of functions. In G. Salvendy, editor, *Handbook of Human Factors and Ergonomics*, chapter 11, pp. 301–339. Wiley, 2nd edition, 1997.
- [83] T. B. Sheridan. *Telerobotics, Automation, and Human Supervisory Control*. MIT Press, 1992.

- [84] T. B. Sheridan. Human centered automation: Oxymoron or common sense? In *Keynote address of IEEE Systems, Man and Cybernetics Conference*, 1995.
- [85] T. B. Sheridan. Supervisory control. In G. Salvendy, editor, *Handbook of Human Factors and Ergonomics*, chapter 39, pp. 1295–1327. Wiley, 2nd edition, 1997.
- [86] S. Siegel. *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill Kogakusha, 1956.
- [87] P. Smets. Belief function. In *Nonstandard Logics for Automated Reasoning*, pp. 253–286. Academic Press, 1988.
- [88] P. Smets. The transferable belief model for expert judgements and reliability problems. *Reliability Engineering and System Safety*, Vol. 38, pp. 59–66, 1992.
- [89] G. Stix. Along for the ride? *Scientific American*, pp. 76–85, July 1991.
- [90] K. Tanaka. A design method for incorporating human judgement into monitoring systems. In *Proceedings of 6th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design and Evaluation of Man-Machine Systems*, pp. 717–722, 1995.
- [91] K. Tanaka. Human-machine cooperative judgements in non-homogeneous safety monitoring systems. In *Proceedings of 5th IEEE International Workshop on Robot and Human Communication*, pp. 484–488, 1996.
- [92] K. Tanaka. Dempster-Shafer theoretical approach to reliable monitoring design under uncertainty. *Advances in Systems Science and Applications*, pp. 201–207, 1997.
- [93] A. Wald. *Sequential Analysis*. Dover Publications, 1947.

- [94] B. J. Winer, D. R. Brown, and K. M. Mitchels. *Statistical Principles in Experimental Design*. MacGraw-Hill, 1991.
- [95] D. D. Woods. The effects of automation on human's role: Experience from non-aviation industries. In Orlyady Norman, editor, *Flight Deck Automation: Promises and Realities*, No. CP-10036, pp. 61-85. NASA, 1989.
- [96] R. R. Yager. On the Dempster-Shafer framework and new combination rules. *Information Science*, Vol. 41, pp. 93-137, 1987.
- [97] R. R. Yager, J. Kacprzyk, and M. Fedrizzzi, editors. *Advances in the Dempster-Shafer Theory of Evidence*. Wiley, 1995.
- [98] J. Yen. GERTIS: A Dempster-Shafer approach to diagnosing hierarchical hypothesis. *Communications of the ACM*, Vol. 32, No. 5, pp. 573-585, 1989.
- [99] J. Yen. Implementing evidential reasoning in expert systems. In L. N. Kanal, T. S. Levitt, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, Vol. 3, pp. 333-344. Elsevier Science Publishers B. V., 1989.
- [100] L. A. Zadeh. Review of Shafer's: A mathematical theory of evidence. *AI Magazine*, Vol. 5, No. 3, pp. 81-83, 1984.
- [101] C. E. Zsombok and G. Klein, editors. *Naturalistic Decision Making*. LEA, 1997.
- [102] K. ヒンティッカ. 認識と信念. 紀伊国屋書店, 1975. (永井, 内田 訳).
- [103] 網谷りょういち. 信楽高原鐵道事故. 日本經濟評論社, 1997.
- [104] J. J. ギブソン. 生態学的視覚論. サイエンス社, 1985. (古崎 他 訳).
- [105] I. ピーターソン. 殺人バグを追え. 日経 BP 社, 1997. (伊豆原 弓 訳).

- [106] R. ブラウン. 地球白書. ダイヤモンド社, 1988. (本田 幸雄 監訳).
- [107] 伊藤, 稲垣. 証拠理論における確信関数の更新と新たな条件付けの提案. 計測自動制御学会論文集, Vol. 21, No. 12, pp. 2011-2017, 1995.
- [108] 伊藤, 稲垣. 大規模複雑プラント安全制御のための情報更新規則の最適性について. システム制御情報学会論文誌, Vol. 10, No. 1, pp. 26-32, 1997.
- [109] 伊藤, 稲垣. 混合戦略によるプラント安全制御のための証拠理論的な情報統合と更新. 計測自動制御学会論文集, Vol. 34, No. 5, pp. 415-421, 1998.
- [110] 伊藤, 稲垣. マイクロワールドアプローチに基づく状況認識の解析. ヒューマンインタフェース学会論文誌, Vol. 1, No. 1, 1999. (掲載予定).
- [111] 伊藤, 稲垣, N. Moray. システム安全制御の状況適応的自動化と人間の信頼. (投稿中).
- [112] 井上, 幸田, 熊本, 高見. 安全監視システムの最適論理構成. 計測と制御, Vol. 24, No. 2, pp. 142-154, 1985.
- [113] 守一雄. 認知心理学. 岩波書店, 1995.
- [114] 稲垣, 伊藤. 証拠理論. 日本ファジィ学会誌, Vol. 10, No. 3, pp. 445-450, 1998.
- [115] 王, 稲垣. ヒューマン・インタフェースと安全制御方策の組み合わせを考慮した警報システムの構成と確率的解析. システム制御情報学会論文誌, Vol. 6, No. 3, pp. 156-164, 1993.
- [116] 田中敬司. 人間と機械の協調と調和 —航空機を例に—. 計測と制御, Vol. 32, No. 3, pp. 199-204, 1993.
- [117] 田中健次. 安全監視と危険監視の最適組み合わせによる高信頼度安全監視システム. 計測自動制御学会論文集, Vol. 28, No. 6, pp. 725-732, 1992.

- [118] 田中健次. 不確実性を考慮した安全監視システムの数理モデル. 応用数理, Vol. 5, No. 3, pp. 2-13, 1995.
- [119] 宮沢光一. 情報・決定論序説. 岩波書店, 1971.
- [120] 村田厚生. 認知科学. 朝倉書店, 1997.
- [121] 航空事故調査委員会. 航空事故調査報告書 中華航空公司所属エアバス・インダストリー式 A300B4 - 622R 型 B1816 名古屋航空. 日本航空技術協会, 1996.
- [122] 航空事故調査委員会. 航空事故調査報告書 ガルーダ・インドネシア航空所属 ダグラス式 DC-10-30 型 PK-GIE 福岡空港. 日本航空技術協会, 1997.
- [123] 佐々木正人. アフォーダンス. 岩波書店, 1994.
- [124] 資格試験研究会. 数的推理. 実務教育出版, 1996.
- [125] 寺野寿郎. システム工学入門. 共立出版, 1985.
- [126] 山岸俊男. 信頼の構造. 東京大学出版会, 1998.
- [127] 岸田純之助. 巨大技術の安全性. 電力新報社, 1987.
- [128] 菅野, 室伏. ファジィ測度, 第3巻. 日刊工業新聞社, 1993.
- [129] 石淵, 市橋, 田中. あいまいな症状に対するセット・カバー論的故障診断. システム制御情報学会論文誌, Vol. 2, No. 2, pp. 54-61, 1989.
- [130] 前田, 市橋. ベイズ定理の証拠理論への拡張について. システム制御情報学会論文誌, Vol. 5, No. 12, pp. 481-490, 1992.
- [131] 村井, 深海. 様相論理 (2) . 日本ファジィ学会誌, Vol. 7, No. 2, pp. 222-238, 1995.

- [132] 渡辺泰夫. ヒューマン・マシンシステムの異常時における人間の対応と自動化システムへの依存性. 筑波大学第三学群情報学類卒業論文, 1998.
- [133] 長谷川, 稲垣. Dempster-Shafer 理論に基づくシステム安全制御のための意思決定支援. 計測自動制御学会論文集, Vol. 7, No. 5, pp. 160-167, 1994.
- [134] 榎木哲夫. 資源制約下での人間-人工物の協調系のデザイン. 人工知能学会誌, Vol. 13, No. 3, pp. 371-378, 1998.
- [135] 田村, 小林, 鳩野. Dempster-Shafer の確率理論を導入した不確実下の価値関数. システム制御情報学会論文誌, Vol. 4, No. 4, pp. 133-139, 1991.
- [136] 日本損害保険協会. 世界の重大産業災害. 1993.
- [137] 松原望. 意思決定の基礎. 朝倉書店, 1985.
- [138] 大田友一. トップダウン的画像理解における仮説の妥当性評価に関する一考察. 情報処理学会論文誌, Vol. 34, No. 10, pp. 2071-2074, 1993.
- [139] 松山隆司. Dempster-Shafer の確率モデルに基づく evidential reasoning の論理的意味に関する考察. 人工知能学会誌, Vol. 4, No. 3, pp. 340-350, 1989.



筑波大学附属図書館



1 00990 12464 6

本学関係