主　論　文　要　旨

No.1

報告番号	甲　㊙　第　　号	氏　名	安西　祐一郎

主論文題名：

The Epistemology of Learning and Interaction: A Goal-Directed Adaptive Agent is an Epistemic Agent

（学習と相互作用の認識論：目標指向の適応的行為主体は認識的行為主体である）

（内容の要旨）

本論文は、「The Epistemology of Learning and Interaction: A Goal-Directed Adaptive Agent is an Epistemic Agent（学習と相互作用の認識論：目標指向の適応的行為主体は認識的行為主体である）」と題する英文の論文であり、序章（Introduction）、第１～３章（Chapters 1～3）、結言と認識論への貢献（Concluding Remarks and Contributions to Epistemology）から成る。

本論文の目的は、認識論哲学における議論を認知科学の成果に基づいて新たな方法で問い直すことにより、長年にわたって多くの見解が積み重ねられてきた認識論の諸問題、特に、内的表象（internal representation）に関する諸議論、実在論（realism）と反実在論（anti-realism）の論争、いかにしての知識（knowledge-how）といかにしての信念（belief-how）の諸議論、行為の因果理論（causal theories of action）、認識論の社会的側面（social aspects of epistemology）、認識論の自然化（naturalization of epistemology）等の問題に新たな知見をもたらすことにある。特に、認識論において深く議論されてきた「認識的行為主体（epistemic agent）」（（認識論の意味での）知識あるいは真理を追究する行為主体）の概念を、認知科学において多くの研究が進められてきた「目標指向の適応的行為主体（goal-directed adaptive agent）」（環境の変化に適応しつつ目標を生み出し、またその目標の達成に向けて行為を準備し実行する行為主体）の概念に関係づけるとともに、「目標指向の適応的行為主体は認識的行為主体である」というテーゼに肯定的に答えることにある。

本論文では、目的の達成に向け、対象領域を人間の基本的活動である学習（learning）と相互作用（interaction）に限定することによって、複雑な理論構築を可能にする。特に、知識（knowledge）、信念（belief）などのような認識論の基本概念を認知科学の側からもあらためて定義するとともに、認識論において多用されてきた真理（truth）、認知科学においてよく使われてきた目標（goal）、両方の分野で用いられてきた内的表象（internal representation）、因果性（causality）、行為（action）などの概念を、認識論における新たな立場に立ってあらためて検討している。

この新たな立場として、本論文では特に、行為主体に内在する情報源と外在する情報源の両方から行為主体が能動的に内的表象を構成し（construct）、制御し（control）、調整する（regulate）過程に焦点を当てるとともに、いくつかの基準的規範（norm）を導入し、これらの規範のもとで、実在論と反実在論の諸議論を中心に従来の多様な議論を整理することを通して、「過程指向構成論（process-oriented constructivism）」を提唱している。特に、この立場に立って内的表象の構造を分析するとともに、多様な内的表象に対してそれらが表象する対象の実在性を議論し、過程指向構成論を実在論の新たな立場として位置づけている。また、この複雑な作業を通して、「目標指向の適応的行為主体」に関わる経験的な認知科学研究の成果を「認識的行為主体」という認識論の規範的な概念に結びつけ、「目標指向の適応的行為主体は認識的行為主体である」ことを主張している。

まず、序章では、本論文の目的、論文作成の動機、および第 1～3 章の概要を述べるとともに、さらに認識論への本論文の貢献について予めまとめている。

本論文における議論の本体は第１～３章であるが、特に本論文の主題は第３章で述べる認識論的議論にある。これに対して第１章と第２章は、第３章の認識論的議論で用いられる認知科学的諸研究について、概要をまとめたものである。認識論と認知科学の関係づけを通して認識論に新たな貢献をするという本論文の目的に沿って、第１章と第２章では、特に学習と相互作用の認知科学の諸研究の成果が、心理学、神経科学、進化の研究、人類学、言語学、コンピュータ科学、その他の関連分野を含

めて広くまとめられており、第３章での議論を支える土台となっている。第１～２章における議論の多くの部分は本論文の著者自身の長年にわたる認知科学的研究に関連しているが、本論文の主題は認識論への新たな貢献であり、第１～２章に述べた、著者の研究を含む認知科学的研究の概要は、本論文の主題ではなく、第 3 章における認識論の議論を支えるためのものである。

第１章は「学習の理論とモデル（Theory and Models of Learning）」と題して８節から成る。1.1～1.3 節で認知科学における学習の研究を広く概観した後、1.4～1.5 節において、著者らの長年の研究成果である「することによる学習の理論（The theory of learning by doing）」とその意味について概要を述べている。この理論の特徴の一つは「手続き的理論（procedural theory）」だという点にあり、この点が本論文の認識論的議論における新たな観点の一つを提供することになる。また、この理論は、以前に学習された問題解決方略を用いて問題解決の新しい認知方略が発見されていく学習の過程を明らかにするとともに、方略の学習順序が問題解決の領域や問題解決の行為主体に依存しないことを示唆している。これらの成果は、第 3 章において信念が知識に変化していく過程の議論の経験科学的な基盤を提供している。この 1.4～1.5 節の概要を受け、1.6～1.8 節では、学習の認知科学的研究における近年の成果についてさらに幅広くまとめている。

第２章は「相互作用の理論とモデル（Theory and Models of Interaction）」と題して 11 節から成る。まず、2.1～2.4 節において、認知科学における相互作用の研究を、特に人間とロボットの相互作用に関する著者らの新しい方法論とその成果を含めて広く概観している。次に、2.5～2.6 節で、著者の研究成果である「情報共有による相互作用の理論（The theory of interaction by information sharing）」およびその意味について概要を述べ、著者らの人間・ロボット間相互作用の研究を中心にした実証研究の概要も併せて述べている。この理論も、第 1 章で概要を述べた理論と同様に「手続き的理論」である点に特徴がある。また、理論の中に、外在する対象の内的表象、あるいは自己、第二人称他者、第三人称他者が構成する内的表象の内的表象、さらにはそれらを統合した複雑な表象等について、それらを構成する手続き、対象あるいは内的表象の「相似性（similarity）」を発見する手続き等を含んでいる点にも特徴がある。相似性の概念は認識論の問題として第 3 章で扱われ、推論可能性の基準的規範として表現される。さらに、この理論は、行為主体の推論能力が第一人称の推論に限定されることを仮定している。この仮定も、第 3 章における推論可能性の基準的規範として示される。理論はしたがって、第 3 章における認識論の諸主張の重要な経験的基盤となる。

また、特に 2.5 節では、第 3 章の認識論的議論の基礎となる、行為主体（agent）、文脈（context）、環境（environment）、ビュー（view）（内的表象の内的表象を指す）、その他多くの概念を定義する。また、2.6 節において、知識、信念、意図、行為、欲求、目標等の概念を議論するとともに、内的表象の第３章の議論で広範に用いられる知識および信念の概念について、認知科学の立場からの定義として、知識の 4 条件（活用可能性 utilizability、頑健性 robustness、適応可能性 adaptability、許可可能性 admittability）を提示している。この定義の特徴の一つは、認識論において一般に知識の概念と強く関連している真理の概念を陽に導入していない点にある。

さらに、2.7～2.8 節で相互作用における最近の認知科学的研究に言及し、2.9 節では、認知神経科学の新しい潮流を踏まえ、相互作用研究の神経科学的な基礎となる「機能的にネットワーク化された神経基盤（functionally networked neural platforms）」について述べている。

これら第１章と第２章の概論を受け、第３章において認識論の面から広範な議論が展開される。

第３章は、「過程指向構成論者の立場からの表象（Representations from Process-Oriented Constructivist's Standpoint）」と題し、15 節から成る。まず、3.1 節（表象と過程指向構成論 Representations and Process-Oriented Constructivism）において、表象とは何か、過程指向構成論とは何かについて述べるとともに、第１章と第２章に述べた概要の中で認識論から見て重要な論点等を整理する。また、特に、規範的な認識論と記述的な認知科学の境界を成す基準的規範として、6 つの規範（動機規範 motivatedness norm、構造化可能性規範 constructability norm、過程可能性規範 processability norm、推論可能性規範 inferrability norm、知識規範 knowledge norm、世界規範 world norm）を導入し、議論する。次に、3.2 節（三項関係としての表象 Representations as Triadic Relations）で、表象の構造として表象（representation）、表象される対象（represented entity）、表象主体（representer）の三項関係を導入し、特に表象主体の概念について、Dretske, Giere, Millikan らの議論を引用しつつ、その必要性を主張する。

3.3 節（表象の内容と構造の枠組みおよび要因 Frameworks and Factors for Representational Contents and Structures）では、表象の構造についてさらに構造形式（form）および表示形式（format）

の 2 つの概念を導入し、本論文の一つの論点である内的表象について、その構造を詳細に検討する。さらに、3.4 節（表象構造のさらなる要因 : 知識の構造化可能性、認知方略、構造化された世界 More Factors for Representational Structures: Knowledge Structuralizability, Cognitive Strategies, and the Structured World）において、表象の構成における重要な要因である、内的表象としての知識の構造化に関する制約、表象主体が構造形式や表示形式を決める際の認知方略、世界の構造と内的表象の構造の関係等について議論している。また、3.5 節（表象のための表示形式の分類 Classifications of Formats for Representations）では、多様な表示形式が存在すること、表象主体による表示形式の選択が内的表象の構成過程において重要であること等を述べている。

3.6 節（目標指向・方略駆動によって表現すること：我々の認知研究からの例 Goal-Directed Strategy-Driven Representing: Examples from Our Cognitive Studies）では、それまでに述べた三項関係、表象主体、構造形式、表示形式、表象構成の認知方略等についての議論の例として、問題解決（problem solving）、連想記憶（associative memory）、神経科学における統計的方法（statistical methods in neuroscience）の 3 つの分野に関わる著者らの研究を取り上げ、研究目標に沿って表象の表示形式を選択する方略に依存して科学者の研究方法が大幅に変わることを示唆している。本節は、3.1～3.5 節における内的表象の構造および表象主体による表象の構成に関する議論を受けて、科学哲学において議論が十分なされているとは言い難い、科学者の研究方法と表象の関係についての具体的な例示となっている。

第 3 章 3.7 節～3.14 節では、内的表象に関する 3.1～3.6 節の議論を受け、第 1～2 章の経験科学的研究に言及しつつ、本論文で提唱している過程指向構成論の立場と認識論における従来からの議論との関係を、科学哲学を中心とする認識論の多様な主張に対比して議論している。

まず、3.7 節（認識論の中に過程指向構成論を位置づける：はじめに Placing Process-Oriented Constructivism in Epistemology: An Introduction）において、認識論としての過程指向構成論の特徴を従来からの認識論と対比しつつ概観する。

次に、3.8 節（認識的行為主体を彫琢する : 能動的行為主体はなぜ認識的でなければならないのか？ Sculpting an Epistemic Agent: Why Must an Active Agent be Epistemic?）（「能動的行為主体」は「目標指向の適応的行為主体」と同意）では、2.6 節に述べた知識の認知科学的定義および知識の条件について、認識論における真理の概念および真理条件との関係づけを行っている。特に、第 2 章に述べた人間・ロボット間相互作用の研究に言及し、人間・ロボット間相互作用におけるロボットにとっての真理、知識とは何か、という問題を議論するとともに、真理の対応理論（correspondence theories of truth）、真理の一貫性理論（coherence theories of truth）、知識の正当化理論（the justification theory of knowledge）、知識の信頼性理論（the reliability theory of knowledge）、知識の因果性理論（the causal theory of knowledge）等、伝統的な真理と知識の議論と過程指向構成論の関係を検討している。そのうえで、認識論における真理および知識の扱いと認知科学における扱いの橋渡しを 2.6 節の知識条件を介して行いつつ、「（認知科学的定義による）知識条件を満たす信念は（認識論的定義による標準的な）真理条件を満たす」ことを示している。このことは、本論文の目的の一つである「目標指向の適応的行為主体は認識的行為主体である」というテーゼの主張の基盤になる。 なお、3.7 節では、感情（emotion）の概念についても触れ、過程指向構成論においては感情の概念を 3.1～3.6 節で述べた内的表象の構造概念の中に自然な形で取り込めることを示唆している。

3.9 節（目標指向の適応的行為主体にとって表象は何を意味するか？ 表象される対象の存在 What Do Representations Mean to Goal-Directed Adaptive Agents? Existence of Represented Entities）では、認知科学における諸研究を認識論の立場から再考するため、まず、第 1 章で述べた「することによる学習の理論」と第 2 章で述べた「情報共有による相互作用の理論」を科学哲学における科学理論の面から見たとき、ともに科学理論であると言えることを主張する。他方で、これらの理論が、従来の科学哲学では十分扱ってきたとは言えない「手続き的な科学理論」であることを指摘する。

3.9 節では、次に、さまざまな構造を持つ内的表象について、表象の対象の実在に関する議論を展開する。まず、問題解決方略（problem-solving strategy）について、方略をアルゴリズムとみなせること、ただしそのアルゴリズムは行為主体が自分で構成した内的表象であり、コンピュータで実行されるのではなく行為主体が自分で実行すること、したがってアルゴリズムとその実行結果には行為主体から見て因果的な関係があることを指摘する。また、新しい方略が学習される過程が、（2.6 節で定義した意味での）信念が知識に変換される過程であることを指摘する。そのうえで、学習過程で初心者（novice）から熟達者（expert）に変化する行為主体について、熟達者の問題解決方略の実行結

果が実在することを、収束的実在論（convergent realism）、特に最良の説明への推論（the inference to the best explanation）の議論を援用し、「最良の」についてはコンピュータ科学における計算量の考え方を導入して主張する。

さらに、科学哲学における反実在論をリードしてきた議論の一つである構成的経験論（constructive empiricism）と過程指向構成論の違いを述べ、構成的経験論を批判するとともに過程指向構成論を擁護する。その際、収束的実在論への批判についても言及し、特に悲観的帰納法（pessimistic induction）について、行為主体の因果的認識と 2.6 節の知識条件を土台とするここでの議論にはあてはまらないことを主張する。また、第 1 章で概要を述べた、問題解決方略の学習順序が領域や行為主体に依存せずに定まるという「することによる学習の理論」の経験科学的主張を取り上げ、この学習順序が方略を自ら手続きを学習する領域・行為主体非依存のアルゴリズムの実行結果の構造にほかならないことを指摘するとともに、その構造の実在を、収束的実在論における最良の説明への推論の方法について「最良の」を「（2.6 節の知識条件に照らして）妥当な」に弱めた議論によって主張している。

次に、内的表象の第 2 の例として、問題解決のための内的モデル（internal model）（問題を解くために外的環境（定義は 2.5 節）を推論することによって行為主体が構成する、問題の内的表象）を取り上げ、まず、科学哲学における議論と共通する論点を指摘する。そのうえで、1.5 節に概要を述べた「同一の行為主体であっても、多くの問題を繰り返し解いている間に、より問題を解きやすい内的モデルを構成できるようになる」という経験科学的研究の成果に言及し、熟達者が構成する内的モデルが科学哲学における科学理論の範疇に入ることを指摘する。これらの議論は、第 1 章の認知科学的研究を含めて、初等物理学の問題解決を例として用いている。そのうえで、収束的実在論と最良の説明への推論の議論を援用して、熟達者の内的モデルの対象が実在することを主張する。また、内的モデルの問題は科学哲学における科学理論の議論と密接な関係があるため、本論文ではこの箇所で、実在論として、実体実在論（entity realism）、認識的構造実在論（epistemic structural realism）、存在的構造実在論（ontic structural realism）、半実在論（semi-realism）、その他、また反実在論として、構成的経験論としての認識的構成論（epistemic structuralism）、悲観的帰納法、プラグマティック経験論（pragmatic empiricism）、ラディカル構成論（radical constructivism）、社会的構成論（social constructivism）、その他の多様な論について、過程指向構成論との対比で議論し、反実在論の主張を退けるとともに多くの実在論についてもそれらの弱点を指摘し、過程指向構成論の長所を主張している。また、理論の対象を本質的には自然科学的対象に限定せず、しかも理論と対象の関係を断ち切らない認識的構造実在論の長所を認めている。さらに、実在論に立つ過程指向構成論において、内的表象の構成過程に注目することによって反実在論の一部を取り込むことができることを指摘し、実在論と反実在論の二項対立的議論を超える議論が可能になることを示唆している。

内的表象の第 3 の例として、3.9 節では「ビューのビュー（a view on a view）」（対象が他者あるいは自己が構成するビュー（内的表象の内的表象）であるようなビュー）を取り上げている。このような表象は、それ自体複雑な構造を持つばかりでなく、自然科学の認識論で主に扱われる、観察可能と考えられる対象とは異なり、対象が観察可能ではないと考えられる点に特徴がある。また、構造の複雑さを超えて議論するには 3.1～3.6 節に述べた内的表象の構造に関する規範的な議論が不可欠である。そこで、ここではまず、（「ビューのビュー」を構成することについての）熟達者が構成する「ビューのビュー」は観察不可能と考えられる対象についての理論とみなせること、およびそのような「ビューのビュー」は 2.6 節で定義した知識条件を満たすことを指摘する。そのうえで、「ビューのビュー」が表象する「ビュー」は実在することを、収束的実在論と妥当な説明への推論の方法を援用して主張する。さらに、構成的経験論をはじめとする反実在論への批判を述べる。

さらに、3.9 節では、第 4 の例として「共有（される）ビュー（shared view）」を取り上げる。相互作用に参加（participate）している各行為主体が、すべての行為主体が相似的な内的表象を構成していると（第一人称推論によって）推論しているとき、それらの内的表象を総合して（当該行為主体によって推論され、内的表象として構成されている）共有ビューと呼ぶ。この定義によれば、相互作用に参加しているすべての行為主体が相似的な内的表象を推論によって構成しているとき、またそのときに限って、その相似的な内的表象を共有ビューと呼ぶ。

ここでは特に、認識論の社会的側面を議論するため、相互作用の中で「社会的相互作用（social interaction）」（「社会性（sociality）」を含む目標を持つ行為主体同士による相互作用）について議論している。特に、第 2 章に述べた情報共有による相互作用の理論、進化に関する諸研究、特に社会脳仮説（the Social Brain Hypothesis）および進化生物学における社会的シグナル（social signal）の概

念、および収束的実在論の議論を併用することで、共有ビューの実在を主張する議論を行っている。また、3.13 節で取り上げる、認識論の社会的側面の議論に向けて、過程指向構成論の考え方が、共有ビューが内的表象として構成されているように各行為主体が社会的相互作用の中で推論する過程を説明し得ることを主張する。

以上のように、3.9 節では、多様な内的表象について、それらが表象する対象が実在することをさまざまな方法によって主張している。これらの主張は、実在論としての過程指向構成論の基底となるものである。

次に、3.10 節（複雑な内的表現の構造と内容 Structures and Contents of Complex Internal Representations）では、それまでの議論を踏まえ、内的表象とその構成過程の複雑さを許容しつつ実在論の立場を取る過程指向構成論においてさらに議論しておくべき、いくつかの論点について述べている。まず、「内的表象」と「内的表象の内的表象」を同一視すべきかどうかについて、いわゆる KK 仮説（KK 原理）をもとに議論し、2.5 節で述べた内的表象の定義との一貫性を保つには、KK 仮説の成立を肯定すべきことを主張している。また、図やグラフを内的表象の外化とみなしたときに図の理解および描図の過程が複雑な内的表象の構成過程とどのように関係しているのか、相似性の推論が内的表象の構成過程にどう関係しているのかについても議論している。さらに、構成的経験論、ラディカル構成論、社会的構成論など、多様ないわゆる「構成論」について、さらにまとめて批判を述べている。また、過程指向構成論の立場から、認知科学の哲学を席巻した機能主義（functionalism）、計算主義（computationalism）についての批判を述べている。

次に、3.11 節（いかにしての知識といかにしての信念：我々の立場から Knowledge-How and Belief-How: From Our Standpoint）では、第 1 章で扱った「することによる学習の理論」と第 2 章で扱った「情報共有による相互作用の理論」がともに「手続き的理論」であること、また、特に学習の研究において、初心者が熟達者に変化する過程と信念が知識に変換される過程を（2.6 節の知識条件を通して）関係づけられることから、いかにしての信念がいかにしての知識に変換される過程について、従来の認識論の議論とは違った切り口からの議論を展開している。

特に、いかにしての信念を、記号による表示形式と記号列による構造形式（例えば行為の記号的表示と記号的表示の系列）によって対象を表象する内的表象としてのいかにしての信念（explicit-belief-how）とそのような形式では表象し難い対象（自転車の乗り方など）を表象するいかにしての信念（implicit-belief-how）を分け、Stanley と Williamson 等によるいわゆる知性主義（intellectualism）的な議論に基づいて、 前者についての実在論的な議論は 3.9 節に述べた議論の範疇で可能であることを論じている。また、前者については、いかにしての信念からいかにしての知識への変換についての議論が可能であることを指摘している。これに対して、後者（記号列の構造形式では表象し難い対象を表象するいかにしての信念）については、従来の知性主義的アプローチでは議論が困難だが、表象の構造形式と表示形式を前もって確定せずに表象主体が形式を方略によって決めるとする過程指向構成論の立場では、少なくとも議論の俎上に載り得ることを示唆している。

3.12 節（内的表象における目標状態と行為の因果的、帰属的、目的論的、合理的関係）Causal, Attributional, Teleological, and Rational Relations of Goal States and Actions in Internal Representations）では、認識論において長く議論されてきた原因、理由、行為、およびそれらの間の関係に関する議論について、新たな切り口を提供する。

すなわち、第 1 章で概要を述べたように、経験科学的研究によれば、同一の行為主体であっても、内的表象として構成する内的状態と行為の関係を、因果関係（causal relation）、帰属関係（attributional relation）（結果から原因への関係）、目的論的関係（teleological relation）、合理的関係（rational relation）（目標状態の達成についての何らかの評価のもとで最も効果的あるいは効率的と考えられる行為と状態の関係）等、多様な関係として解釈できることが示唆されている。また、やはり第 1 章に概要を述べたように、「することによる学習の理論」で主張されている問題解決方略の学習過程において、初心者が熟達者になるにつれて、内的状態と行為の関係の解釈が、因果関係⇒帰属関係⇒目的論的関係⇒合理的関係の順序で変化していくことが、経験科学の面から示唆されている。

本節では、これらの経験科学的結果を踏まえ、伝統的な行為論において、古くは Davidson と Anscombe の議論のように対立関係にあった、「原因としての行為」と「理由としての行為」を再検討し、同一の行為主体において一体として理解できること、また、熟達者になるにつれて内的状態と行為の関係の解釈が同一の行為主体において変化し得ることを主張している。これらの主張は、同一の行為主体が初心者から熟達者になる学習過程については行為論の中であまり議論されてこなかっ

たために、認識論の議論として本格的に俎上にのぼることがなかったと考えられる。

3.12 節ではさらに、複数の行為主体の行為が同時に生起したり、複数の行為主体の連携による行為が生起するといった場合を含め、社会的相互作用の中での状態と行為の因果関係・目的論的関係・合理的関係についての問題についても議論している。

3.13 節（認識論の社会的側面と我々の過程指向構成論 Social Aspects of Epistemology and Our Process-Oriented Constructivism）では、これまで議論してきた個々の行為主体についての認識論を離れ、社会的相互作用の認識論について議論している。特に、第 2 章に述べた「情報共有による相互作用の理論」とその人間・ロボット間相互作用への応用に関する経験科学的研究に言及しつつ、過程指向構成論の面から、「社会的行為（social action）」の概念について、2.6 節に述べた知識条件との関係を含め、認識論的な分析を行っている。また、社会的信頼性主義（social reliabilism）と過程指向構成論の関係、相対主義（relativism）と過程指向構成論の両立可能性を論じている。

さらに、「知識の大域的および局所的安定性（global and local stability of knowledge）」の概念を導入するとともに、知識の正当化理論、過程指向構成論、相対主義のいう知識を、それぞれ大域的に安定な知識、局所的に安定な知識、大域的に不安定な知識として整理し、関係づけている。

本節では、こうした議論を踏まえ、Bratman による shared agency の議論（集合的概念を導入しない）と Gilbert による joint commitment の議論（導入する）を引用しつつ、社会的相互作用に参加している行為主体が第一人称推論だけに限定した推論を行うという基準的規範のもとでは、社会的相互作用を説明するのに集合的概念は必要がないこと、ただし、行為主体の適応性を重視する過程指向構成論は、プランの概念を重視する Bratman の議論とは方向性がまったく異なることも指摘している。

3.14 節では、かつて Quine らによって大きく取り上げられた、認識論の自然化について議論している。特に、認識論において以前に起こった心理学的転回（The Psychological Turn）、プラグマティズム的転回（The Pragmatic Turn）等の議論について再考するとともに、第 1 章と第 2 章で概要を述べた認知科学的研究における特徴的な方法論のうちから、発話プロトコル分析、コンピュータシミュレーション、大規模データ解析と神経科学の組み合わせによる脳活動の分析、人間・ロボット間相互作用を利用した相互作用研究の合成的（synthetic）方法について、認識論の自然化と関連づけ、科学方法論としての妥当性を検討している。また、認識論の自然化の可能性を探る例題として、Davidson がかつて提案した「三角測量による説明（triangular explanation）」等を用い、認識論の中で行われてきた多くの議論の自然化が可能なこと、ただし完全な自然化、さらには消去主義的唯物論（eliminative materialism）の議論は成立しないことを主張している。そのうえで、認知科学、神経科学等の経験科学的研究によって従来の認識論の相当部分は自然化されるが、規範的認識論のすべてが自然化されることはないとする、中庸の（modest）自然化認識論が妥当である旨を述べている。

最後に 3.15 節において第 3 章の要約と総括を行い、それまでの議論を総合して、「目標指向の適応的行為主体は認識的行為主体である」こと、また、内的表象を構成し、制御し、調整できることは、目標の達成を指向し環境の変化に適応していく行為主体にとっても知識と真理を探究していく認識的行為主体にとっても必要な前提条件になるという、深い意味を持っていることを主張している。

このように、第 3 章では、本論文の目的に沿って、認識論の立場から認知科学における学習と相互作用の研究を再検討するとともに、過程指向構成論を従来の認識論の議論の中に位置づけ、目標指向の適応的行為主体が認識的行為主体であることを示している。そして、これらの検討と議論を通して、認識論に対していくつかの新しい主張を行っている。

最後に、第 3 章の後に「結言と認識論への貢献」を付し、第 1～3 章の議論を要約するとともに、特に第 3 章で述べた新しい主張を、以下のような 10 項目の貢献にまとめている。

貢献１：　「目標指向の適応的行為主体は認識的行為主体であること」を、多くの概念と理論の構築を通して主張していること。また、4 つの「知識条件」と 6 つの「基準的規範」に基づいて認知科学の経験的知見を規範的に説明するとともに、認識論の多様な議論との対比を通して、実在論に基づく「過程指向構成論」の立場を確立したこと。

貢献２：　問題解決の初心者が解決方略を学習して熟達者に変化する過程は信念が知識に変換される過程とみなすことができ、しかも方略の学習順序には問題解決領域および行為主体に独立な構造が実在することを、科学哲学における実在論を援用して主張したこと。

貢献３：　複雑な内的表象の構造を規定するとともに、その構成過程を認識論の中に明確に位置づけたこと。特に、内的表象を表象主体、表象、表象される対象の３つ組で構成されるとみなすとともに、

実在論の立場に立った独自の内的表象論を提唱したこと。
貢献４：　問題解決方略、内的モデル、ビューのビュー、共有ビューなどの内的表象によって表象される多様な対象が実在することを、主として収束的実在論を援用して示す論を立てて実在論の立場を明確にしたこと。また、構成的経験論（例えば van Fraassen の経験的構成論）等の反実在論、また認知科学の哲学の主流であった機能主義や計算主義を批判的に論じるとともに、従来の実在論と反実在論の論争を止揚し、過程指向構成論に基づく新たな議論を提示していること。また、実在論の立場を明確にするにあたっては、認識論の代表的な議論を挙げて過程指向構成論と対比し、過程指向構成論の立場を認識論の新たな立場として提唱していること。
貢献５：　方略のような手続きを学習する過程における、「いかにしての信念」から「いかにしての知識」への変換について、記号列として表象される「いかにしての信念」とそうでない「いかにしての信念」を区別して、前者については実在論の議論が基本的に成り立つことを明確にしたこと。また、後者についても、過程指向構成論における内的表象の議論に矛盾せず、議論に取り込めることを示唆したこと。
貢献６：　初心者が熟達者に変化する学習の過程において、行為主体としての学習者が、自ら内的表象として構成した状態と行為の間の関係について、因果関係、帰属関係、目的論的関係、合理的関係のように解釈を変化させていく過程に関して、因果性と非因果性に関わる行為の理論の議論にも言及しつつ、認識論の立場から新たな光を当てていること。
貢献７：　認識論の社会的側面に関する議論を検討しつつ、複数の「目標指向の適応的行為主体」による社会的相互作用についての認識論的議論を「集合的概念」の導入なしに行えることを示していること。また、認識論の社会的側面に関する議論の中で「集合的概念」を導入しない他の議論（例えば Bratman の shared agency）とも異なることを示していること。
貢献８：　Quine らを嚆矢とする「認識論の自然化」の議論を取り上げ、経験科学的研究によって認識論には従来に増して「自然化」の進行が起こるが、規範的な認識論が完全に消去されることはなく、規範的認識論と記述的科学が本論文で提示した「過程指向構成論」を媒介として「隣人として領域を分け合う」ことになると主張していること。
貢献９：　本論文で提示している過程指向構成論が、認識論において提起されてきた多くのいわゆる「構成主義」とは一線を画し、ここに述べている貢献にあるように、認識論に対して新たな知見を与える立場にあることを主張していること。
貢献１０：　認識論は、「真理とは何か」という命題を扱い、（個別には違った議論もあるが）総論としては「真理」が存在することを念頭に置いた議論を展開してきた。他方、認知科学では、「真理」の存在を明示的に前提とすることなく、目標指向の適応的行為主体が「いかにして」環境の変化に適応しながら目標を発見し達成するために行為を準備し実行しているのかを、経験科学の方法を用いて解明しようとしてきた。これらに対して、本論文のテーゼ「目標指向の適応的行為主体は認識的行為主体である」は、真理あるいは知識の探究を第一義とせずに環境の変化に適応しながら目標を発見し達成するために行為を準備し実行することを第一義とする、目標指向の適応的行為主体が、実は真理あるいは知識の探究を第一義とする認識的行為主体であることを主張する言明である。
　したがって、この言明が妥当であることを主張する本論文の内容は、真理に関する議論が中心にはない認知科学の経験科学的研究と成果を、真理についての議論に関連する認識論の中心的議論に引き込む役割を果たしている。特に、学習と相互作用に関わる人間の知についての経験的知見から知識の条件を提示し、また認識論の面から規準的規範を提示して、それらを哲学における真理および知識の条件に結びつけることにより、認識論に新たな地平を拓く議論を提供している。

　以上、本論文は、著者自身の認知科学的研究を含めて認知科学およびその関連分野の広範な研究を援用しつつ、学習と相互作用に関する新たな認識論的議論を展開して、過程指向構成論と呼ぶ新しい実在論の立場を提唱するものである。

Keio University

Thesis Abstract

No.

Registration Number:	□ "KOU" ◯ "OTSU" No. *Office use only	Name:	Yuichiro Anzai

Title of Thesis:

The Epistemology of Learning and Interaction: A Goal-Directed Adaptive Agent is an Epistemic Agent

Summary of Thesis:

This dissertation, entitled "The Epistemology of Learning and Interaction: A Goal-Directed Adaptive Agent is an Epistemic Agent", aims at providing new contributions to longstanding arguments in epistemology, such as what internal representations are, the realism/anti-realism debate, causal theories of action, social aspects of epistemology, and naturalization of epistemology, through rethinking those arguments in new ways and referring to empirical results from cognitive science. In particular, the dissertation relates the concept of an *epistemic agent* in epistemology to the concept of a *goal-directed adaptive agent* in cognitive science, and argues for the thesis that *a goal-directed adaptive agent is an epistemic agent*.

Towards achieving these aims, the dissertation restricts its attention to two principal human activities: learning and interaction. This makes easier and simpler the assembly of various new theoretical and conceptual constructs for discussions and arguments. Under this restriction, the dissertation redefines basic epistemic concepts, such as knowledge and belief, from the perspective of cognitive science. Also, it reconsiders representative concepts such as truth, internal representation, causality, and action, from a new standpoint to be proposed and properly positioned in conventional epistemology.

As this new standpoint, the dissertation proposes what it calls *process-oriented constructivism*. This standpoint focuses on processes of an agent for actively constructing, controlling, and regulating internal representations, using information originated in both internal and external sources. Also, it introduces several epistemic norms, and reconsiders various preceding arguments in epistemology, such as the realism/anti-realism debate as a typical example, analyzes structures of various kinds of internal representations, argues for the existence of entities represented by those internal representations, and positions itself in epistemology as a new standpoint based on realism. Further, through these complex discussions, the standpoint relates empirical results in cognitive research on goal-directed adaptive agents to the normative concept of an epistemic agent in epistemology, and contends that a goal-directed adaptive agent is an epistemic agent.

The dissertation consists of the following parts: the Introduction, Chapters 1-3, and the Concluding Remarks and Contributions to Epistemology. First the Introduction provides the aims, motivations, and sketches of Chapters 1-3, as well as a preview of its contributions to epistemology.

Then Chapters 1-3 follow, which constitute the body of the dissertation. Among those three chapters, Chapter 3 plays the principal role for achieving the aims of the dissertation by presenting the main epistemological arguments. Chapters 1 and 2, on the other hand, summarize representative research in

cognitive science, especially works on learning and interaction, respectively. The contents of those chapters are extensively used in discussions to be conducted in Chapter 3. The summaries compiled in Chapters 1-2 refer to a wide spectrum of literature in psychology, neuroscience, evolutionary studies, anthropology, linguistics, computer science, and other fields related to cognitive science. The central parts of those summaries are taken from the works of the author of this dissertation and his colleagues.

At the beginning of Chapter 1 titled Theory and Models of Learning, a survey of cognitive studies on learning is given in Sections 1.1-1.3.

Then, Sections 1.4-1.5 provide an extensive summary of the *theory of learning by doing* and its implications, which compile the longstanding contribution of the author with his colleagues to this topic. Notably, the theory is a *procedural theory* rather than a substantive theory, providing a breaking point for the epistemological arguments of this dissertation. Also, the theory, elucidating processes of a problem-solving agent for discovering and acquiring new problem-solving strategies by itself, empirically explains that the order of strategies learned in those processes is fixed and independent of problem domains and agents. This descriptive result provides empirical bases to epistemic arguments in Chapter 3 on internal processes of an agent in which a belief is turned into knowledge.

These summaries are also used for discussions in the subsequent Sections 1.6-1.8, where more recent results from cognitive science on learning are discussed.

Chapter 2, titled Theory and Models of Interaction, is devoted to providing comprehensive summaries of cognitive research on interaction. First, Sections 2.1-2.4 summarize cognitive studies on interaction, particularly including the research on human-robot interaction by the author and his colleagues.

Next, Sections 2.5-2.6 provide summaries of the *theory of interaction by information sharing* and its implications, taken from the author's own work. These sections also include summaries of studies on human-robot interaction conducted by the author and his colleagues. Those studies are sources of inductively producing the theory. The theory is a procedural theory, and includes procedures for compositing internal representations with very complex structures like a representation of another agent's internal representation and a representation of one's own representation (the infinite-regress argument can be skirted by introducing a constraint on the capacity of an agent's internal mechanisms). The theory also includes procedures for detecting similarities. Further, it restricts the inferential capability of an agent only to the first-person inference, including no capability for the second- or third-person inference. These two points on similarity and inference are discussed in Chapter 3 from the epistemological perspective. In this way, the theory provides an empirical basis of the epistemological arguments in Chapter 3. Further, in Section 2.5, many important concepts are given their definitions: agent, context, environment, view (an internal representation of an internal representation), the distinction of internal and external, and others. Section 2.6 proceeds to discuss key concepts like knowledge, belief, intention, action, desire, and goal, and also provides, from the cognitive perspective, the four conditions of knowledge: utilizability, robustness, adaptability, and admittability. One of the characteristics of those conditions is that no explicit concern is given to the concept of truth, which is strongly related to the concept of knowledge in epistemology.

Sections 2.7-2.8 compile the recent advances of cognitive science on interaction. Section 2.9 extensively discusses the recent trends of cognitive neuroscience, and presents an information processing model of brain activities that the author calls the functionally networked neural platforms.

Chapter 3 is titled Representations from Process-Oriented Constructivist's Standpoint, and presents discussions and arguments on a broad spectrum of topics in epistemology to properly position process-oriented constructivism in conventional epistemology and to present new contributions.

First, in Section 3.1 (titled Representations and Process-Oriented Constructivism), discussions are given for what representations are and what process-oriented constructivism is, and also a list is provided for enumerating salient results given in Chapters 1-2 which play significant roles in subsequent arguments. Also, Section 3.1 introduces and discusses six norms (the motivatedness norm, constructability norm, processability norm, inferrability norm, knowledge norm, and world norm), which play pivotal roles in drawing a clear boundary between normative epistemology and descriptive cognitive science.

Section 3.2 (Representations as Triadic Relations), then, discusses the structure of internal representations. In particular, it introduces the triadic relation of a representer, representation, and represented entity to the structure of a representation, and argues for its usefulness by referring especially to the relevant works of Dretske, Giere, and Millikan.

In Section 3.3 (Frameworks and Factors for Representational Contents and Structures), further frameworks, called *form* and *format*, are introduced, and the structure of an internal representation, which is a key concept in the dissertation, is delineated in detail. Section 3.4 (More Factors for Representational Structures: Knowledge Structuralizability, Cognitive Strategies, and the Structured World) discusses various factors essential for the internal construction of representations, including constraints on the structuralization of knowledge as internal representations, cognitive strategies of a representer in deciding forms and formats, relationships between the structure of a world and structures of internal representations. Section 3.5 (Classifications of Formats for Representations) states that there exist many kinds of formats, and further that the representer must find it important to choose an appropriate format when it constructs an internal representation.

In Section 3.6 (Goal-Directed Strategy-Driven Representing: Examples from Our Cognitive Studies), following the discussions on representational structures and various new concepts such as representer, three studies of the author and his colleagues, taken from the domains of problem solving, associative memory, and statistical methods in neuroscience, are used as examples to argue that the methodologies to be adopted by scientists drastically change depending on their strategies for selecting forms of representations to attain their scientific goals. This argument provides, following the discussions on representations and their structures in Sections 3.1-3.5, a concrete example for suggesting relationships of scientific methodologies of scientists and representations constructed by those scientists; this topic apparently is less discussed in philosophy of science.

Following the discussions on internal representations and their structures in Sections 3.1-3.6, Sections 3.7-3.14 provide a wide variety of arguments on the relationships between process-oriented

constructivism and conventional arguments in epistemology. Those arguments proceed through comparing with various claims given in epistemology, particularly in philosophy of science, and also referring to empirical studies summarized in Chapters 1-2.

The arguments begin with Section 3.7 (Placing Process-Oriented Constructivism in Epistemology: An Introduction), in which surveys are given for the characteristics of process-oriented constructivism, especially by associating them with those in conventional epistemology.

Following this introduction, in Section 3.8 (Sculpting an Epistemic Agent: Why Must an Active Agent be Epistemic?), the definition and conditions of knowledge given in Section 2.6 are related to the concept and conditions of truth. It also discusses the topic of what truth and knowledge mean to a robot in a human-robot interaction. With these discussions, the characteristics of process-oriented constructivism are compared with various theories of truth or knowledge in epistemology such as: correspondence theories of truth, coherence theories of truth, the justification theory of knowledge, reliability theory of knowledge, and causal theory of knowledge. Following these discussions, Section 3.8 provides the argument that *a belief that qualifies the knowledge conditions given in Section 2.6 satisfies truth conditions with the standard definition in epistemology*, while bridging the concepts of truth and knowledge in epistemology to those in cognitive science by the conditions of knowledge presented in Section 2.6. This provides a base in arguing for the thesis that a goal-directed adaptive agent is an epistemic agent; giving an affirmative answer to this thesis is one of the aims of this dissertation. Section 3.7 also touches on the concept of emotion, and suggests that process-oriented constructivism is able to naturally integrate this concept to the structure of internal representations given in Sections 3.1-3.6.

Section 3.9 (What Do Representations Mean to Goal-Directed Adaptive Agents? Existence of Represented Entities) is devoted to discussions on connecting a wide range of arguments in epistemology to relevant aspects of cognitive research. First, it claims that the theory of learning by doing and the theory of interaction by information sharing, which are summarized in Chapters 1-2 respectively, are both scientific theories, if they are evaluated by standard conditions of scientific theories argued in philosophy of science. On the other hand, it is also recounted in the section that both of them are procedural theories, apparently not sufficiently discussed in philosophy of science.

In Section 3.9, this claim is followed by discussions given for the existence of entities represented by internal representations with various kinds of structures. First, a problem-solving strategy, discussed in Chapter 1, can be regarded as an algorithm, with the reservation that the algorithm is an internal representation constructed by a problem-solving agent. The algorithm is executed by the agent, not by a computer, and thus an output trace and the algorithm must be causally related through the agent's actions. Also, it is pointed out that processes of an agent for learning new strategies can be regarded as processes of the agent to transform a belief to knowledge (in the sense of the knowledge condition in Section 2.6). Then the section provides an argument for the existence of output traces of problem-solving strategies acquired by an expert agent in a learning process. The argument is carried out by applying convergent realism and the inference to the best explanation argument, and defining 'best' in the best explanation as the best for the space complexity of the agent's internal processes.

Further, in Section 3.9, constructive empiricism, a leading school for epistemic anti-realism, is called to task with a comparison to process-oriented constructivism. Also, the discussion refers to criticisms of convergent realism, especially the pessimistic induction argument. In Section 3.9, the dissertation contends that pessimistic induction must be discarded from the present discussion, because it is based on the causal relation between an algorithm executed through actions of an agent and its output traces, and also it relies on the knowledge conditions. Further, for the domain- and agent-independent fixed learning order of problem-solving strategies described in Chapter 1, it is pointed out that this learning order is a structure embedded in an output trace of a domain- and agent-independent algorithm for learning new strategies. Also, the section argues for the existence of this structure by applying the inference to an appropriate explanation. There, the 'best' in the best explanation is replaced by 'appropriate' in the sense of being qualified by the conditions of knowledge given in Section 2.6.

The second example of internal representations that Section 3.9 takes is an internal model for problem solving (an internal model is an internal representation of a problem, constructed by an agent through inferring its external environment for solving the problem), using the domain of elementary physics, the same domain used in the empirical studies summarized in Section 1.5. The section points out that such models have many commonalities with models and theories argued in philosophy of science. Then, referring to the experimental result given in Section 1.5 that the same agent learns to construct internal models which can be used more effectively or efficiently in solving problems, the section claims that an internal model constructed by an expert agent can be regarded as a scientific theory in philosophy of science. Accepting this claim, the argument contends the existence of an entity represented by an internal model constructed by an expert, applying convergent algorithm and the inference to the best explanation.

Section 3.9 also provides extensive discussions for comparing process-oriented constructivism with other realism and anti-realism schools in epistemology: entity realism, epistemic structural realism, ontic structural realism, semi-realism, and others for realism arguments; and epistemic structuralism, pessimistic induction, pragmatic empiricism, radical constructivism, social constructivism, and others for anti-realism arguments. From those discussions, the dissertation claims against representative anti-realism arguments, and also points out weaknesses of many representative realism arguments. On the other hand, it partly recognizes the merits of epistemic structural realism because of its two characteristics. One is that this school does not restrict entities represented by theories to those in natural sciences. The other is that it keeps relations of entities with their theories. Moreover, the argument points out that process-oriented constructivism, which attends to processes for constructing internal representations, is able to integrate at least a part of anti-realism arguments into the realism argument, and makes possible discussions that go beyond the dichotomy caused by the realist/anti-realist debate.

As the third example of internal representations, 'a view on a view' (a view that represents a view constructed by another agent or the self) is taken in Section 3.9. A view on a view has a complex structure, and also a represented entity is supposedly unobservable or non-physical. Discussions on this

kind of complex representations might necessitate normative arguments, which actually are provided in Sections 3.1-3.6. Here, it is pointed out that a view on a view constructed by an expert agent (an expert in constructing views on views in a specific domain) can be regarded as a theory of supposedly unobservable entities, and also a view on a view of an expert satisfies the knowledge conditions given in Section 2.6. Following this argument, it is claimed that a view represented by a view exists, by applying convergent realism and the inference to an appropriate explanation. Further, the argument is followed by critics of anti-realism arguments such as constructive empiricism for this kind of representations.

The fourth example of internal representations taken in Section 3.9 is a shared view. Suppose that each agent participating in an interaction constructs as an internal representation a set of views on internal representations of all the participating agents, where all of those representations of participating agents are inferred by the agent as being similar. If all the participating agents construct internal representations of this kind, and if they *are* all similar, then such a set of views on representations of all the agents, inferred and constructed as an internal representation by each agent, is called a shared view. The theory of interaction by information sharing, if it is applied to the theory of interaction by view sharing, guarantees that all the sets constructed by participating agents autonomously become similar. The definition of a shared view says nothing about this autonomy for similarity. A simple but typical example is a composite goal state shared by multiple agents in an interaction.

In the discussions on shared views, interactions are restricted to social interactions. A social interaction is an interaction in which a goal state of each participating agent includes sociality. The argument here contends the existence of shared views in a social interaction, by applying the theory of interaction given in Chapter 2, various studies on evolutionary processes (particularly those on the Social Brain Hypothesis and social signals), and the convergent realism argument, in a combined fashion. Also, as noted above, the argument claims for process-oriented constructivism that it is able to explain processes of each agent in a social interaction by inferring that all the participating agents have similar shared views. This supports arguments in Section 3.13 for social aspects of epistemology.

Overall, Section 3.9 affirms the existence of entities represented by various different kinds of internal representations in a variety of ways. This claim provides the basis for the realism stance of process-oriented constructivism.

Following the arguments in preceding sections, Section 3.10 (Structures and Contents of Complex Internal Representation) discusses some important points for relations of internal representations, their structures, and processes for constructing them. First, the KK thesis (or the KK principle) is discussed in relation to the question of whether to identify an internal representation with an internal representation of an internal representation, raised in defining internal representations in Section 2.5. The argument here claims that the KK thesis must hold if the definition of internal representations keeps consistency. Also, in this section, discussions are provided for how processes for understanding diagrams (as external representations) and diagramming are related to processes for constructing internal representations, and how inferences on similarities between representations are related to those processes. Furthermore, the section provides criticisms of what are labeled as ‘constructivism’ and the like: for

example, epistemic structuralism, radical constructivism, social constructivism, and others. It also takes a critical look at functionalism and computationalism, which once covered a large portion of philosophy of cognitive science.

Section 3.11 (Knowledge-How and Belief-How: From Our Standpoint) turns its attention to procedural aspects of knowledge and belief. Being aware that the theories of learning and interaction in Chapters 1-2 are both procedural theories, and that learning processes of a novice to become an agent can be related to processes of an agent for transforming a belief to knowledge, the section argues this topic from the perspective different from conventional epistemology. In particular, the argument here distinguishes explicit-belief-how, which is an internal representation that represents an entity with a symbolic form and format, from implicit-belief-how, which represents an entity whose form and format are not symbolic. Then, it proceeds to argue, based on the intellectualism argument by Stanley and Williams as well as others, that similar contentions to ones provided for the existence of entities represented by problem-solving strategies, internal models, and others can be applied to explicit-belief-how. Further, it points out that process-oriented constructivism is available also for discussions on implicit-belief-how since the standpoint does not state any specific form or format. On the other hand, it is difficult for intellectualism arguments to cope with such discussions for implicit-belief-how.

Section 3.12 (Causal, Attributional, Teleological, and Rational Relations of Goal States and Actions in Internal Representations) provides new perspectives for causes, reasons, actions, and their relations, argued for years in epistemology. As summarized in Chapter 1, the same agent can interpret relations on elements in the same set of internal states and actions in different ways as the causal relation, attributional relation, teleological relation, and rational relation. Further, also the theory of learning in Chapter 1 empirically suggests, in the process of learning strategies to become an expert in problem solving in a specific domain, the same learner is generally able to transform interpretations for relations on elements in a set of states and actions from the causal to attributional, further to teleological, and yet further to rational relation. It is argued here that an action as a cause and as a reason can be the same action of the same agent. The argument provides new perspective to the longstanding debate in causal theories of action many years ago on whether an action is a cause or a reason, for example, by Davidson and Anscombe.

In Section 3.12, further discussions are given for causal, teleological, and rational relations constructed as internal representations by agents participating in a social interaction, where actions of multiple agents could be exerted at the same time, or such actions may be exerted collaboratively.

Then, Section 3.13 (Social Aspects of Epistemology and Our Process-Oriented Constructivism) is dedicated to discussions on social aspects of epistemology. There, referring to the theory of interaction and experiments on human-robot interaction, both summarized in Chapter 2, epistemic analyses are given for the concept of social action from the process-oriented constructivist's standpoint, particularly from the perspective of the knowledge conditions given in Section 2.6. Also, the section discusses relationships between social reliabilism and process-oriented constructivism, as well as possible

reconciliation of relativism and process-oriented constructivism. Further, the concept of the global and local stability of knowledge is introduced, and the relations among the justification theory of knowledge (globally stable), process-oriented constructivism (locally stable), and relativism (globally unstable) are argued.

In this section, the dissertation further contends that no collective concept is necessary for explaining social interaction. This is because an agent participating in a social interaction makes only the first-person inference, but still it is possible for all the agents to share similar information without the second- or third-person inferential capabilities as argued in Section 3.9. This argument is given by referring to arguments in epistemology that do not introduce collective concepts (such as Bratman's shared agency) and those that recruit such concepts (like Gilbert's joint commitment). It is argued further that process-oriented constructivism, which emphasizes the adaptability of agents, runs very differently from Bratman's argument that stresses plans and planning roles.

In Section 3.14, naturalization of epistemology, the issue once raised by Quine and others, is discussed by referring to relevant literature in both epistemology and cognitive science. In particular, the psychological turn and the pragmatic turn are reconsidered, and also some representative experimental methods in cognitive science are reexamined. Those methods include the think-aloud protocol method, computer simulations, analyses of brain activities by massive-data analytics and neuroscience, and the synthetic method by using human-robot interaction. These methods are described in Chapters 1-2, but in this section it is argued that they can be regarded as scientific methods for advancing specific aspects of cognitive science.

Also, to examine how epistemology can be naturalized, the section employs as an example exercise the triangular explanation that Davidson once presented for advocating externalism. This exercise suggests that epistemology can be, or was already, naturalized to a considerable extent. However, the argument also claims that, although a large part of epistemology will be naturalized, the normative part of it will remain, and not be eliminated. The dissertation opposes the complete naturalization of epistemology, and eliminative materialism, and advocates the modest naturalized epistemology.

Section 3.15, the final section of Chapter 3, provides a summary of the chapter. Also it affirmatively states, by integrating the arguments given in the chapter, that a goal-directed adaptive agent is an epistemic agent, and that the capability of constructing, controlling, and regulating internal representations is an important prerequisite for both a goal-directed adaptive agent and an epistemic agent.

Thus, Chapter 3, along with the aims of this dissertation, reconsiders empirical studies of learning and interaction from the perspective of epistemology, proposes process-oriented constructivism, relates it to conventional arguments in epistemology, claims that a goal-directed adaptive agent is an epistemic agent, creates new relationships between epistemology and cognitive science, and accordingly, through these arguments, provides new contributions to epistemology.

Following Chapter 3, the dissertation places the Concluding Remarks and Contributions to Epistemology. This last part summarizes Chapters 1-3, and provides the contributions produced from

the arguments in Chapter 3 as the list of ten items that follow; the dissertation:

Contribution I: Affirmed the thesis that a goal-directed adaptive agent is an epistemic agent, through the deliberate introduction of conceptual and theoretical constructs and extensive arguments on those constructs. Also, provided normative explanations to empirical results in cognitive research on learning and interaction by introducing four conditions of knowledge and six epistemic norms. Further, established the process-oriented constructivist's standpoint based on realism arguments, through comparisons with a broad range of arguments in epistemology.

Contribution II: Argued that processes of a novice in a specific domain to learn new strategies, internal models, or views on views for becoming an expert can be regarded as processes of an agent for transforming its belief to knowledge. Also, contended by applying realism arguments the existence of the domain- and agent-independent fixed learning order of problem-solving strategies.

Contribution III: Defined new frameworks for stating the structures of complex internal representations, including forms and formats, as well as the structure with the triadic relation of a representer, representation, and represented entity. Further, using those frameworks, properly positioned processes for constructing those representations in arguments in epistemology.

Contribution IV: Argued for the existence of entities represented by various kinds of internal representations like problem-solving strategies, internal models, views on views, and shared views by applying convergent realism and other arguments. Also, provided critical arguments on anti-realism such as constructive empiricism that has led the anti-realism argument in philosophy of science, as well as functionalism and computationalism which served as main players for some time in philosophy of cognitive science. Further, presented a new realism argument by bridging conventional realism arguments with anti-realism. All these arguments were conducted by comparing process-oriented constructivism with arguments proposed by representative schools of epistemic thought.

Contribution V: Argued that processes for learning new strategies to become an expert can be related to processes for an agent to transform a belief-how to knowledge-how. Further, proposed to distinguish explicit-belief-how and implicit-belief-how by differences of forms and formats for those representations (symbolic for the former, whereas distributed for the latter), and claimed that realism arguments similar to those applied to other parts in Chapter 3 can be implemented to explicit-belief-how, but not to implicit-belief-how. Suggested, however, that the frameworks of representations proposed in the arguments for process-oriented constructivism could be applied to both.

Contribution VI: Shed new light on causal theories of action by attending to the transformation of an agent's interpretation of relations between internal states and actions from causal to attributional, further to teleological, and to rational relations, while the agent learns to become an expert.

Contribution VII: Contended that epistemic arguments on social interaction of multiple goal-directed adaptive agents can be conducted without introducing any collective concept. Also argued that the argument is very different from other arguments that do not use such concepts (such as Bratman's shared agency).

Contribution VIII: Through discussing naturalization of epistemology, whose modern arguments were started by Quine and others, from the perspectives of both epistemology and cognitive science, argued for the modest naturalized epistemology. Contends that more portions of normative epistemology will be naturalized but no complete naturalization will occur; normative epistemology and descriptive cognitive science will stay as neighbors mediated by process-oriented constructivism.

Contribution IX: Argued that process-oriented constructivism is superior to various 'constructivisms' in many points related to epistemology such as the contributions listed here.

Contribution X: Declared the new statement that substantially connects epistemology, which deals with 'what is truth?' and hypothesizes the existence of truth, and cognitive science, which does not explicitly assume the existence of truth but works on how goal-direct adaptive agents try to prepare and exert actions to discover and attain their goals through adapting to new environments. The affirmatively answered thesis that a goal-directed adaptive agent is an epistemic agent (Contribution I) is indeed this statement, which contends that any such agent, like a human problem solver such as a student, a business person, or any other, pursues knowledge and truth in the sense of epistemology. The statement, and the arguments given in Chapter 3 as well, provide new meanings originated in process-oriented constructivism to the epistemic concepts of knowledge, belief, truth, and others, at least in the realms of learning and interaction that constitute the two principal activities of human beings.

Overall, this dissertation presents a new realism stance called process-oriented constructivism and offers new contributions to epistemology, through extensive arguments on learning and interaction from the epistemological perspective, referring to a broad spectrum of literature in epistemology and cognitive science including the author's own works.