

What Can Theory of Knowledge Learn from Epistemic Side-Effect Effect?

メタデータ	言語: jpn 出版者: 公開日: 2017-10-03 キーワード (Ja): キーワード (En): 作成者: メールアドレス: 所属:
URL	http://hdl.handle.net/2297/34273

認識論的副作用効果は知識の理論に何を言うか

片岡 雅知¹

要約

本稿は、認識論的副作用効果（ESEE）と呼ばれる現象が知識の理論に持ちうる含意を検討する。まず1・2節で、元々の「副作用効果」をめぐる議論状況を確認し、それに影響を受けて見出されたESEEを3節で紹介する。4節ではどんな直観判断が判断対象の理論の証拠となるかに関し「良識の基準」を取り出し、5節でESEEにおける判断は知識の理論にとって適切な証拠ではないと示す。従って認識論的副作用効果は知識の理論に何も言わない。

1. 副作用効果とは

Knobe [2003]は、公園で過ごしていた78人の人々に対し、次の二つのシナリオのどちらかをランダムに提示した。

【シナリオ1 破壊条件】

株式会社の取締役会の会長のもとに、副会長がやってきてこう言った「新しい事業を始めようと思っています。わが社の利益を上げるでしょうが、環境を破壊することにもなるでしょう」。

会長は答えた「環境を破壊するとかはどうでもよい。私はできる限り利益を上げたいのだ。事業を始めよう」。

彼らは事業を開始した、やはり環境は破壊された。

【シナリオ2 保護条件】

株式会社の取締役会の会長のもとに、副会長がやってきてこう言った「新しい事業を始めようと思っています。わが社の利益を上げるでしょうし、環境を保護することにもなるでしょう」。

会長は答えた「環境を保護するとかはどうでもよい。私はできる限り利益を上げたいのだ。事業を始めよう」。

彼らは事業を開始した、やはり環境は保護された。

被験者は以上のシナリオを読んだ後、会長が自分の行いに対して非難／称賛されるべき度合いを7段階で示すよう求められ、また、会長が意図的に環境を破壊／保護したと思うか否かを問われる。

すると破壊条件では、被験者のほとんど（82%）が会長は環境破壊という副作用を意図的に引き起こしたと回答したのに対し、保護条件では被験者のほとんど（77%）が会長は意図的に環境保護という副作用を引き起こしたわけではないと回答するという、大きな非対称性が見られた²。さらにノーブは、人々が良い副作用に関して会長を称賛する度合いよりも、悪い副作用に関してより非難する度合いが高いことを示し、この評価の非対称性が、意図性帰属の非対称性と相関していることを指摘した。

「つまり人々は、副作用を悪いものと見なしている場合の方が、善いものと見なしている場合よりも、その副作用が意図的に引き起こされたと言う気がとても大きくなっているのかもしれない」（Knoke [2003] p. 193）。

この現象が副作用効果あるいはノーブ効果と呼ばれるものである。副作用効果はこのシナリオに限らず様々な条件のシナリオで、また、意図性に限らず様々な判断に関して生じることが知られている（Alexander [2012] Chap. 3）。

2. 副作用効果の説明：能力と運用エラー

なぜ副作用効果のような現象が生じてくるのか、その適切な説明がいかなるものであるかに関しては、様々な見解が林立している。まず大きく分けるならば、被験者の回答の非対称性が、何らかの形で人々の素朴な意図性概念を反映していると考える立場と、そうではないとする立場がある。後者の立場としては、被験者の回答パターンを会話の含意によって説明し尽くそうとする Adams & Steadman [2004] が代表的である。

副作用効果に見られる非対称性が素朴な意図性概念の反映だと考える方の立場も、大きく2つに分けられる。1つは、そもそも素朴な意図性概念には評価的な認知が含まれてお

り、従ってシナリオに対する道徳的評価によって意図性帰属に非対称性が出るのは、意図性概念の正常な行使によるものだとする見解である。ここではこれを「能力説」と呼ぶ。一方、あくまで意図性概念に評価的認知は含まれず、むしろなんらかの要因によって、意図性概念能力の運用がゆがめられてしまった結果、判断に非対称性が生じたとする立場がある。ここではこれを「運用エラー説」と呼ぶ。

ここでは副作用効果に関してどちらの説明がより適切なのかについての考察は行わないが、関連する2つの論点に触れておきたい。まず、もし能力説が正しければ、意図や予見可能性、スキルなどを中心に組み立てられてきた意図性に関する伝統的な理論は、大きな修正を迫られることになる可能性があると主張されることがある。というのも、一つの哲学の方法論によれば³、ある対象に関して理論を構築する際、その理論を支持する証拠として、我々はその対象に関する直観判断に大いに依拠できるため、もし、我々の意図性に関する直観判断の背後にある概念能力が、その正常なあり方として、評価的認知を含んでいるものならば、我々がつくるべき意図性に関する理論も、その事実を何らかの形で十分考慮すべきであると考えられるからである。他方で、運用エラー説が正しければそのような修正の必要はない。ゆがめられた直観を証拠として使うべきではないからだ。

もう一点、能力説と運用エラー説の論争に決着をつけるためには、概念能力の中に何が含まれるのかを精確に決定するという課題に取り組まねばならない。例えば Alexander, Mallon & Weinberg [2010] によると、概念能力の内実を明らかにする方法は大きく二種類ある。

- (1) ボトムアップ：神経解剖学の知見を用い、あるものを処理する部位が別のものを処理する部位に物的な意味で「含まれる」かどうかを見る
- (2) トップダウン：認知プロセスによって計算されていると想定される機能の特徴づけ、その機能のアルゴリズムを特定し、物理的実装の問題に取り組む。機能に関する実質的な説明が、何が概念能力の正常な行使で干渉なのかを教えてくれる。

このうち哲学が寄与出来そうなのは(2)だが、ある概念の機能が何なのかを精確に決めるには、哲学的な分析だけではなく、例えば進化に訴えた目的論的議論などを行っていく必要があろう。しかしこれまでのところ哲学者はそのような知見をあまり取り入れてこなかったように思われる。Alexander らの診断によれば、例えば Knobe [2010] で戦われているような能力説と運用エラー説の争いは、能力の内実に関して約定対約定の争いになってしまっている。何が能力のうちに含まれるのかという問題は多分に経験的なものであ

り、そしてそのことは哲学にとつての免罪符には決してならないのだが、しかし現時点で述べられることが少ないということは事実である。そしてこのことが、我々がここで副作用効果の説明の問題を取り扱わない消極的な理由である。ただし、それを取り扱わなくても以下の論旨に支障が出ないというのも、また別の理由である。

3. 認識論的副作用効果

この節では Beebe & Buckwalter [2010]で行われた調査を紹介する。これは、副作用効果が知識帰属にも及ぶという見解を示したはじめての公刊論文である。

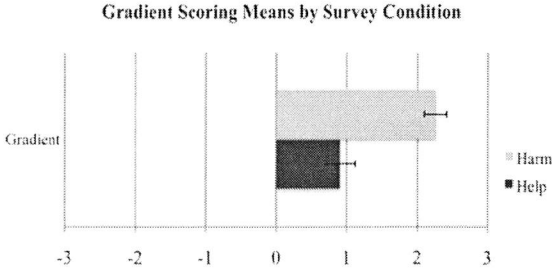
この論文で Beebe らは、Knobe の発見にインスパイアされて、「真なる信念に照らして (in light of) 行われる行為が持つだろう副作用の悪さあるいは有責性が、その信念が知識に値するか否かの被験者の判断に影響をおよぼす」(p. 475) という仮説を立てた。これはより詳細に言うと、「ある副作用が良い場合、被験者は、＜行為がその副作用をもたらすだろうと行為者は知っていた＞とはあまり考えなくなるが、副作用が悪い場合には、被験者は知識の帰属をより行うようになる」(p. 477 < >は引用者) ということである。そしてこの仮説をテストするために次のような調査が行われた。

Beebe らは大学の 749 人の学部生に対し、上で引用した Knobe [2003]で用いられたのと同様の 2 シナリオのうち 1 つのシナリオを渡した。ただし、各々のシナリオの最後には、

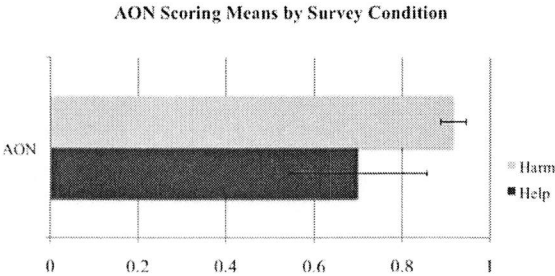
会長は新しい事業が環境を破壊（／保護）するだろうということを知っていたか？

という一文が付け加えられており、被験者はこの文に対する賛成・反対の度合いを、－3（「会長は知らなかった」）から 3（「会長は知っていた」）までの 7 段階のスケールで回答するように求められた。

回答結果の分析には 2 種類のスコアリング法が用いられたが、どちらの方法でも、破壊条件での賛成の平均値と保護条件での賛成の平均値に、統計的に有意な差があることがわかり（図 1 および図 2）⁴、元の副作用効果と同じように、破壊条件では賛成の値が高く保護条件では賛成の値が低いという判断の非対称性が見出された。



【図1：分析結果1】(Beebe & Buckwalter [2011] p. 476)



【図2：分析結果2】(Beebe & Buckwalter [2011] p. 477)

この結果は元の仮説と整合的である。知識帰属においてこのような非対称性がみいだされるという現象を、Beebe らは「認識論的副作用効果」(以下 ESEE)と呼んだ。

ESEE をどのように説明するのが適切かに関しては、Beebe らは態度を保留している。ただし、とりわけ能力説で説明が行われることになれば、評価的認知が知識概念にとって構成的であるということになるのだから、信念のほか証拠、信頼性などの真理にかかわる認識論的要因のみに注目する伝統的な知識概念は再考を迫られるだろうと指摘している。

また関連する別論文では、「真なる信念が知識に値するかどうかは、その信念に照らして当の信じ手が行いうるいかなる行為にも依存しない」という知識に関する主張と、「真なる信念が知識に値するかどうかを決定する際には、普通の被験者はその信念に照らして当の信じ手が行いうるいかなる行為の特徴も考慮に入れないだろう」という知識概念に関する主張とを分け、ESEE は後者の誤りを示すものであるが、前者の誤りを示すかどうかは哲学の方法論によると指摘している (Beebe & Jensen [2012] p. 711)。

これらの慎重な指摘はまったく正当だが、以下では議論のために、伝統的な認識論にとって最も都合の悪そうな仮定をおくことにする。つまり、概念能力説が正しく評価的認知は知識概念によって構成的であるとし、さらに、直観判断を判断対象に関する理論の証拠として存分に使ってよいという方法論を採用することにしよう。ではこの時、ESEE のような現象は知識の理論に対してどんな含意を持つだろうか。

4. 「良識の基準」：判断対象との「関係なさ」

いま行った仮定のもとで我々は、任意の直観判断を判断対象に関する理論の証拠として用いることができるのだろうか。

ところで、(非) 道徳的な行いの想起は部屋の明るさの判断に影響する(!?)。Banerjee, Chatterjee, & Sinha [2012] によると、過去の記憶から非倫理的な行いを想起させた被験者に部屋の明るさを聞くと、倫理的な行いを想起させた被験者よりも、部屋が暗いと判断するのである。ここでこの実験をヒントにして、行為の知覚でも同じような効果が生じるとしよう。つまり、非倫理的な行いを見た人は、倫理的な行いを見た人よりも部屋が暗いと判断するのである(ありそうな話ではある)。

さて、この架空の実験から、「部屋の明るさ」の理論、つまり部屋の明るさが何によって決まるかを探求する理論は何を学ぶべきだろうか。今、たとえ評価的認知が我々の部屋の明るさ認知の能力に含まれていると仮定してたととしても、「部屋の明るさ認知」についての理論ではなく「部屋の明るさ」についての理論の構築を目指す限り、その理論において、「部屋にいる人の(非) 道徳的行為」という要素を考慮すべきだとは我々には全く思われない、というよりむしろ考慮すべきではないと思われるのではないだろうか。つまり部屋の明るさの理論が学ぶべきことは何もないのである。

これは何故なのか。端的に言えば、部屋にいる人の(非) 道徳的行為と今の部屋の明るさはどうみてもぜんぜん関係ないからである。この一見自明な関係なさは、この場合、我々の認識はともかくとして、(非) 道徳的行為そのものは部屋の明るさのあり方そのものに何の因果的影響も与えていないという事実によるのだと思われる。

おそらく我々は、ある直観判断をその判断対象に関する証拠として用いることができるか否かに関して次のような一般的な「良識の基準」を持っており：

(B) 判断対象と（形而上学的に）「適切な関係」を持たないものの認知に影響されている判断は、判断対象についての理論を作るにあたっては（その影響分を）無視せよ

そして今回はそれを次のような形で実質化しているのである：

(L) 部屋の明るさと因果関係を持たない部屋の中の人（非）道徳的行為の認知に影響されている判断は、部屋の明るさについての理論を作るにはあたっては無視せよ。

この場合の「適切な関係」とは、「道徳的行為と部屋の明るさとの間の関係の中で、その部屋の明るさのあり方に影響するもの」であり、そしてそのような関係を見いだせないので、我々は架空の実験における被験者の判断を、明るさに関する理論の証拠として用いようという気には全くならないのである。

我々は、この基準を ESEE にあてはめてみることを提案する。そして、（非）道徳的な副作用と信念との間には適切な関係がないために、知識の理論が ESEE から学ぶことは何もないのだと論じていきたい。しかしここですぐさま問題となるのは、ではその場合に「適切な関係」とは何なのかという点である。何が適切な関係かは事例に依存する。そこでさっそく、ESEE という脈絡の中で「適切な関係」をどう実質化できるかをみていくことにする。

5. ESEE の検討

ESEE において第一に問題となるのは、判断対象となる信念とくそれに「照らして」行われる行為の（非）道徳的な副作用との間の関係である。ここで、行為と副作用の間に因果関係という適切な関係があることを確認し、副作用の道徳性は行為の道徳性に引き継がれるというひとまず無難な仮定をすれば、問題の焦点を、社長の信念と（非）道徳的な行為の間の関係に絞ることができる。そうするとここでの「適切な関係」とは、「信念と（非）道徳的な行為との間の関係の中で、その信念の知識としての身分に影響するもの」だということとなる。

さて、そのような関係の候補として、一見して2種類のものが考えられると思う。（1）理由関係（2）因果関係である。我々はそれらがどちらも適切な関係であると認めた上でもなお、ESEE のシナリオでは問題の行為が判断対象となる信念と「適切な関係」を結べ

ていないことを示す。

(1) まず理由関係はどうか。真理と正当化を強調する伝統的な知識観では、ある信念がある行為に対してもつ理由関係が、その信念の知識としての身分に影響を与えるなどと言われることはまずない。しかしそこで、理由関係は適切な関係ではないと不注意にも知識観を狭めてしまえば、論点先取を犯すことになってしまう。

ここで再び架空の実験を考えよう。そして更なる仮定として、部屋にいる人の(非)道徳的な行為は、部屋の蛍光灯に対し(理解しがたいが)因果的な影響を及ぼし、行為が悪ければ暗く、良ければ明るくといった風に光の強さが変わるのだとしてみよう。この場合、部屋の明るさに関する理論は、当然「部屋にいる人の(非)道徳的行為」を考慮に入れるべきである。そして、我々が正しく行為の道徳性を認知している限り、その認知に影響された判断はその理論の証拠になりうるだろう。

これと似たように、信念と(非)道徳的な行為の間の理由関係がある場合、道徳的に悪い行為の理由となる信念は知識であり、良い行為の理由となる信念は知識ではない、といった風にその信念の知識としての身分に影響があるのかもしれない。その場合、(非)道徳的な行為の認知に影響されて非対称性を見せる ESEE における判断は、知識の理論の証拠として十分扱ってよいことになる。

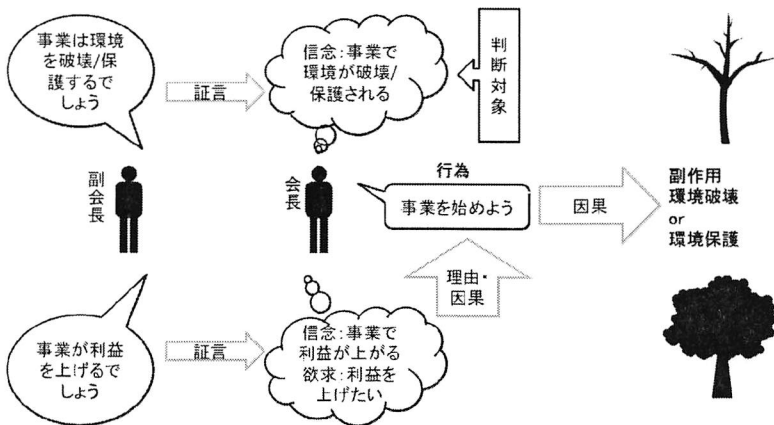
以上の知識観を認めたうえで、ESEE のシナリオ内で社長が事業を始める命令を出すという行為をおこなったのは、一体どんな理由によるものだと描かれていたかに注目してみる。会長は次のように言っていた。「環境を破壊／保護するとかはどうでもよい。私はできる限り利益を上げたいのだ。事業を始めよう」。この発言からわかるように、会長の行為の理由は「利益を上げたい」という欲求と、副会長の発言により獲得された「事業を行えば利益が上がる」という信念である。つまり会長の行為に対して「事業を行えば環境を破壊／保護することになる」という問題の信念は理由を与えていない⁶。それは端的に「どうでもよい」のである。したがって社長の信念と(非)道徳的な行為の間に理由関係はない。

(2) 次に因果関係はどうか。ここでも伝統的な考えにのっとれば、ある信念がある行為の原因であるということがその信念の知識としての身分に影響を与えるなどといわれることはない。しかし、道徳的に悪い行為の原因となる信念は知識であり、良い行為の原因となる信念は知識ではない、といった風にその信念の知識としての身分に影響があるのか

もしれない。

以上を認めたらうで我々は、問題の信念と行為の間に理由関係がないのだから、行為の因果説に従って因果関係もやはりないと結論したい。また、すでに強調したように、会長は行為にあたってこの信念のことをまるで気にかけていない。だから、たとえ副会長から環境に関する証言が全く無かったとしても、会長は全く同じように行為したことだろうという反実仮想が十分成立する。このこともやはり問題の信念と行為の間の因果関係を否定する根拠となる。

つまり、問題の信念と（非）道徳的な行為の間には、因果関係も理由関係もない（図3）。そのため、ESEE でみられるその（非）道徳的行為の認知に影響された非対称な判断は、知識の理論にとっては証拠として使うべきものではないのだと、良識の基準にそって結論することができる。



【図3：シナリオ図解】

ここで、上の引用で Beebe らが自分たちの仮説をどう立てていたかを思い出したい。ここでは、信念が知識であると判断されるか否かに影響するのは、その「信念に照らして (in light of) 行われる行為が持つだろう副作用の悪さあるいは有責性」である、と述べられていた。しかし、この「照らして」というのは、信念と行為の間のいかなる関係のことを言っているのだろうか。「照らして」はしばしば理由関係を表すために使われるが、今確認し

ように、会長の行為は環境の破壊／保護に関する信念が理由を与えていた訳でもなければ、さらにそれが原因で行われた訳でもない。Beebe らはこの「照らして」という関係は、いったい何ものなのかはっきりさせるべきであると思われる。

結論

我々は ESEE を紹介したのち、伝統的な認識論者に最も不利な仮定を採用した。しかし任意の直観判断を判断対象の理論の証拠として扱うべきではないことが示され、その基準が「良識の基準」として取り出された。そしてこの基準を ESEE において実質化する中で、問題の信念と行為の間には適切な関係が存在せず、そこでみられる非対称的な直観判断は知識の理論にとって証拠として扱われるべきものではないという結論に至った。認識論的副作用効果は知識の理論に何も言わない。

我々の直観判断は非常に些末な要因によって不安定になることが知られている。そうした事実を前に、なお我々が直観による理論の正当化という手続きを捨てたくないならば、では判断対象の理論の証拠として用いていい／いけないのはいかなる直観なのかを探索していくことに重要な意義があると思われる。本稿はそのためのひとつの試みでもあった。

ところで我々は知識の理論すなわち認識論を知識の理論であると考えてきた。しかし知識の理論なるものは、知識概念とは独立に知識なるものが存在していなければ成立しない。同じことが意図性や自由意志などあらゆる素朴概念に関しても言える。概念の理論の方は心理学がうまくやってくれるだろうが、我々にはもうひとつ形而上学がいる。

謝辞

本稿は、2012 年度科学基礎論学会秋の定例会におけるワークショップ「実験哲学のなかの認知と認識」で筆者らが行った口頭発表の原稿に基づく。同ワークショップの発表者であった太田紘史氏・飯島和樹氏・渡辺匠氏、そして筆者との共同発表者であった飯塚理恵氏（いずれも東京大学）に感謝する。また、発表当日に質問して頂いた方々、および発表後に原稿を仔細に検討しコメントをくださった笠木雅史氏（ブリティッシュコロンビア大学）に感謝する。それらへ応答すべく発表原稿を大幅に改定したものが本稿だが、その内容は全て私の責任のもとにある。

（東京大学大学院総合文化研究科 修士課程）

注

¹ kataoka.c.masanori@gmail.com

² $\chi^2(1, N = 78) = 27.2, p < .001$

³ 例えば、クリプキ [1977=1995] pp. 284

⁴ (1) 回答の平均を計算。破壊条件で $M = 2.25, SD = 1.50$ 、保護条件で $M = 0.91, SD = 2.09, t(747) = -10.126, p < .001$

(2) -3から-1は知識の否定として0点、1から3までは知識帰属として1点を与え、0点は分析から除外、そのうえで平均値を計算。破壊条件で $M = .92, SD = 0.27$ 、保護条件で $M = .70, SD = 0.46, t(689) = -7.692, p < .001$

⁵ この直観が多くの人に共有されていることを願う。

⁶ 社長の信念はデイヴィッドソンの区別で言う「行為をなした理由」ではないが「行為の理由」ではあるのではないかという指摘をいただいた。しかし私は、この信念とベアとなりうる適切な賛成的態度を見つけることができないと思うので、やはりこの信念は理由を構成するものではないと考える。

参考文献

- Adams, F. & Steadman, A. [2004] "Intentional action in ordinary language: core concept or pragmatic understanding?" *Analysis*, 64, 173–181
- Alexander, J. [2012] *Experimental Philosophy: An Introduction*. Cambridge: Polity
- Alexander, J., Mallon & Weinberg [2010] "Competence: What's in? What's out? Who knows?" *Behavioral and Brain Sciences* 33, 329–330
- Beebe, J. R. & Buckwalter, W. [2010] "The Epistemic Side-Effect Effect," *Mind and Language*, 25, 474–498
- Beebe, J. R. & Jensen, M. [2012] "Surprising connections between knowledge and action: The robustness of the epistemic side-effect effect," *Philosophical Psychology*, 25, 689–715
- Banerjee, P., Chatterjee, P. & Sinha, J. [2012] "Is It Light or Dark? Recalling Moral Behavior Changes Perception of Brightness," *Psychological Science*, 23, 407–9
- クリプキ, S. [1977=1995] 「話し手の指示と意味論的指示」 (黒川英徳訳『現代思想』1995年、4月号 266–295)
- Knobe, J. [2003] "Intentional Action and Side effect in ordinary language," *Analysis*, 63, 190–193
- Knobe, J. [2010] "Person as scientist, person as moralist," *Behavioral and Brain Sciences*, 33, 315–365