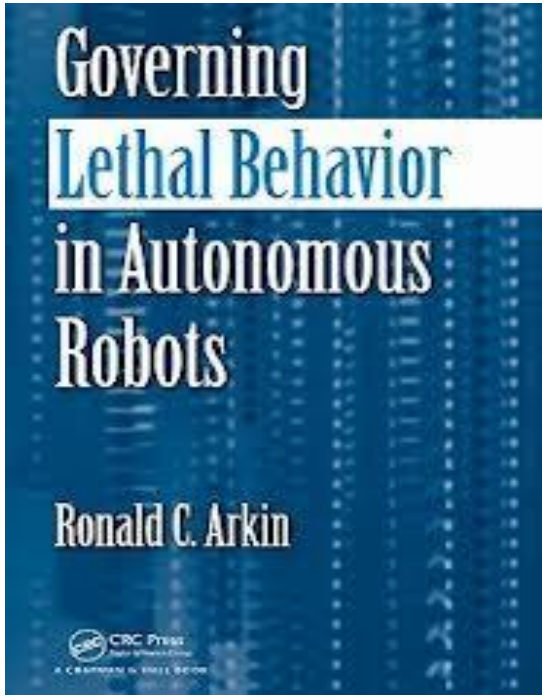
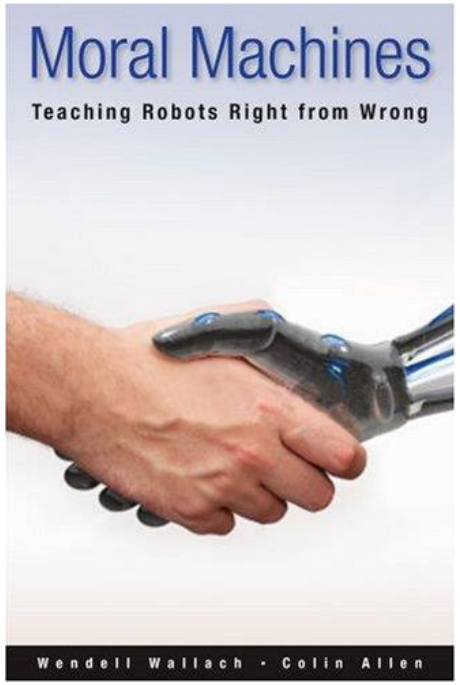
共生社会特論2015年度

ロボットの道徳

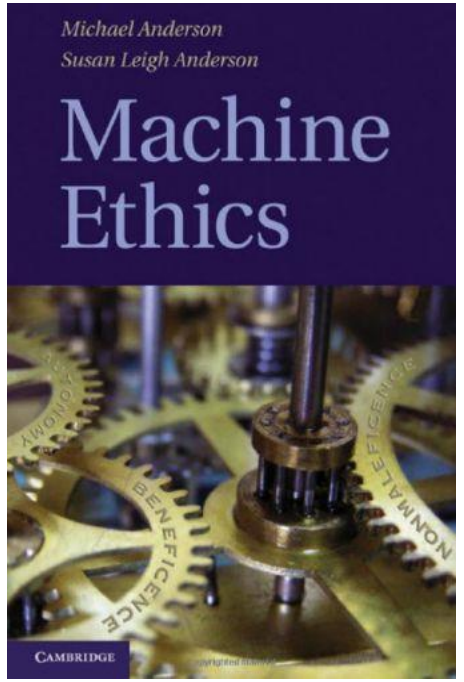
道徳的判断を機械に実装する取り組み



2009

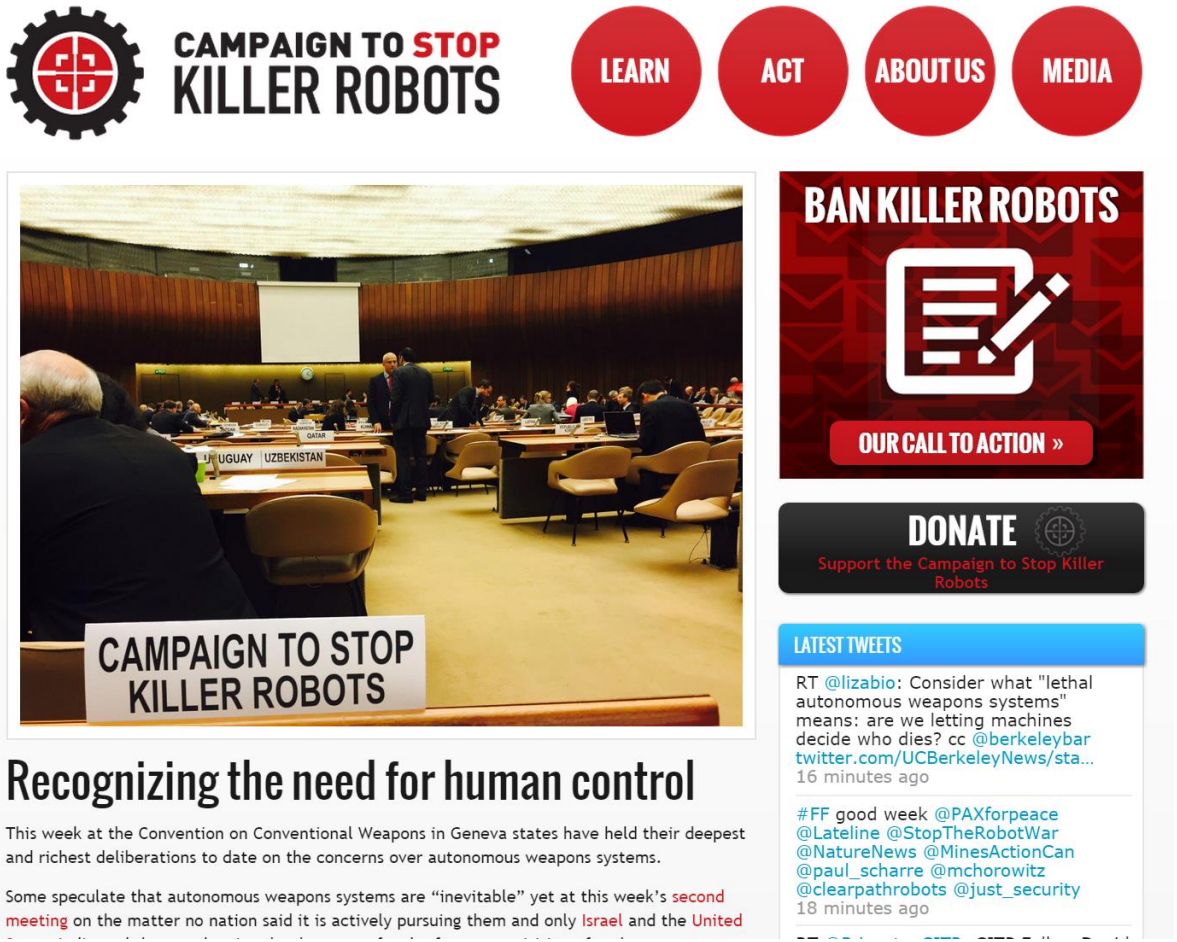


2010



2011

問いの重要性



http://www.stopkillerrobots.org/2015/04/humancontrol/

http://www.theguardian.com/technology/2014/may/28/google-self-driving-car-how-does-it-work

人工知能・ロボットが人間の生命や安全に直接かかわる場面に進出する可能性が高まり、それに伴ってそれらに倫理的判断をさせることの必要性・技術的実現可能性・道徳的是非が論じられるようになっている。

タフツ大学，ブラウン大学，レンセリア・ポリテクニック研究所の研究者たちが米海軍とチームを組んで，善悪とそれらの帰結を理解することのできる「道徳的」な自律ロボットの開発に着手した

道徳的能力とは大雑把にいって人間が同意する傾向にある法律や社会的な規約を学び、それについて推論し、それに基づいて行動し、それについて語ることができる能力と考えることができる

戦場における
アルゴリズムに従った道徳

Can robots be trusted to know right from wrong?

May 12, 2014

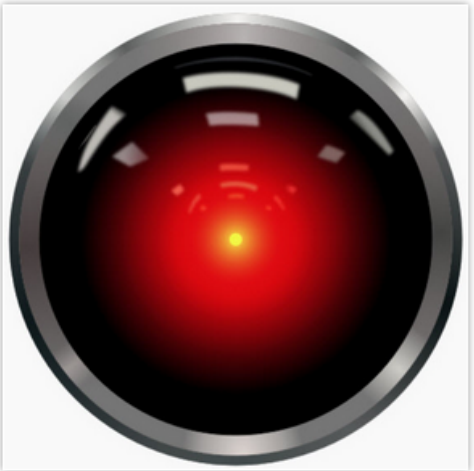
Is it possible to develop "moral" autonomous robots with a sense for right, wrong, and the consequences of both?

Researchers from Tufts University, Brown University, and Rensselaer Polytechnic Institute think so, and are teaming with the U.S. Navy to explore technology that would pave the way to do exactly that.

"Moral competence can be roughly thought about as the ability to learn, reason with, act upon, and talk about the laws and societal conventions on which humans tend to agree," says principal investigator Matthias Scheutz, professor of computer science at Tufts School of Engineering and director of the Human-Robot Interaction Laboratory (HRI Lab) at Tufts.

"The question is whether machines — or any other artificial system, for that matter — can emulate and exercise these abilities."

[+]



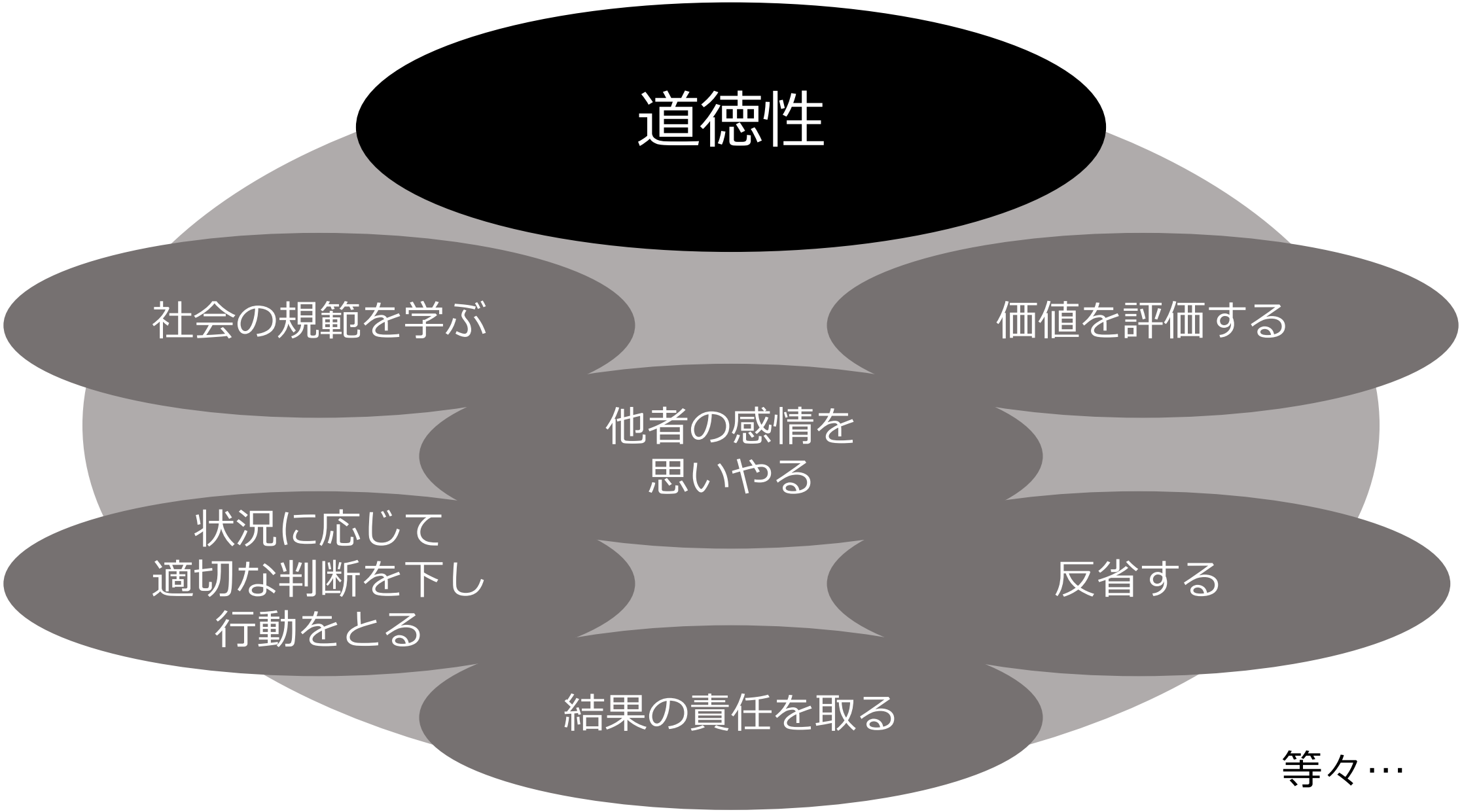
HAL 9000 (credit: Warner Bros.)

But since there's no universal agreement on the morality of laws and societal conventions, this raises some interesting questions. Was HAL 9000 (HAL = (**H**euristically programmed **AL**gorithmic computer) moral? Who defines morality?

Algorithmic morals on the battlefield

Scheutz cites a simplified military battlefield scenario, where the conventions are more easily defined (but still improve on Asimov's simplistic three laws of robotics): a robot medic responsible for helping wounded soldiers is ordered to

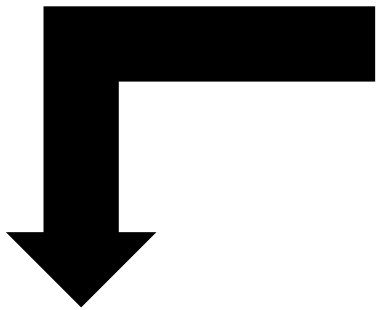
http://www.kurzweilai.net/can-robots-be-trusted-to-know-right-from-wrong


道徳性
社会の規範を学ぶ
価値を評価する
他者の感情を
思いやる
状況に応じて
適切な判断を下し
行動をとる
反省する
結果の責任を取る

等々…

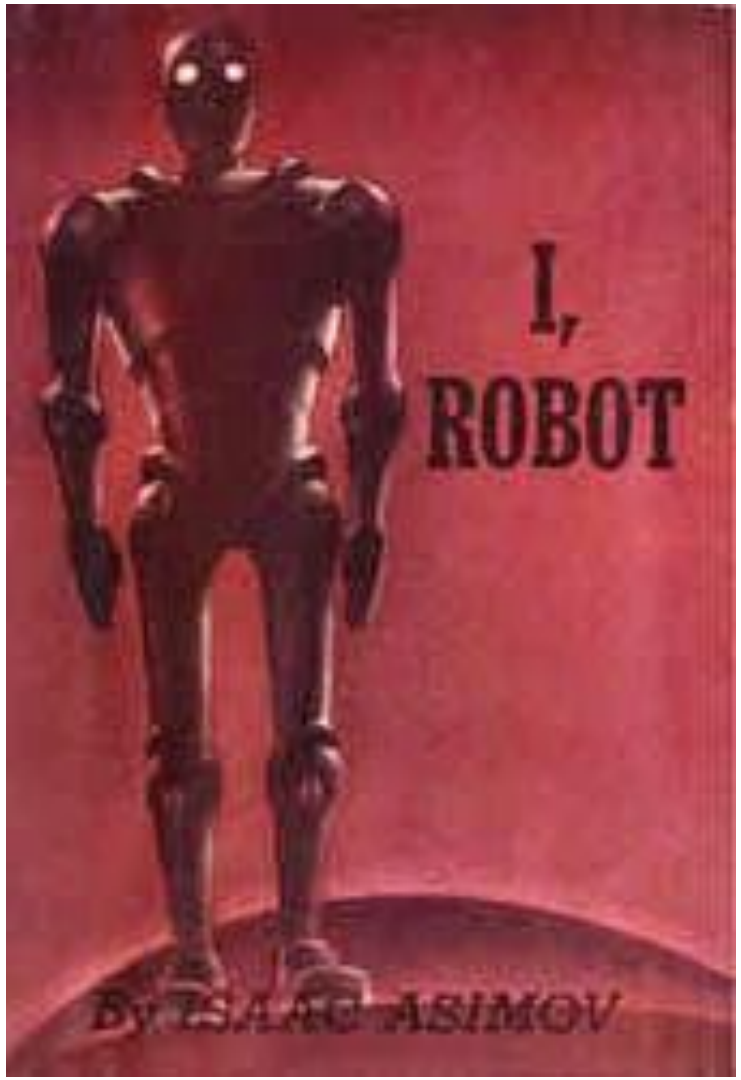
Machine Ethicsのパラダイム

機械が従うべき道徳的規範を
明示化する.



機械にその規範を組み込み,
従わせる.

古典的な「記号的AI」と
同じパラダイム



アシモフのロボット工学三原則

- 一、ロボットは人間に危害を加えてはならない。また何も手を下さずに人間が危害を受けるのを黙視していてはならない。
- 二、ロボットは人間の命令に従わなくてはならない。ただし第一原則に反する命令はその限りではない。
- 三、ロボットは自らの存在を護らなくてはならない。ただしそれは第一、第二原則に違反しない場合に限る。

『私はロボット』，伊藤哲訳，東京創元社，創元SF文庫，1988年，p. 8）

Ethical trap: robot paralysed by choice of who to save

Can a robot learn right from wrong? Attempts to imbue robots, self-driving cars and military machines with a sense of ethics reveal just how hard this is

New Scientist

https://www.newscientist.com/article/mg22329863-700-ethical-trap-robot-paralysed-by-choice-of-who-to-save/

ロボットは善悪の区別をつけられるようになるか？ ロボット，自走車，軍事機械に倫理観を備えさせる試みはそれがいかに難しいかを明らかにする

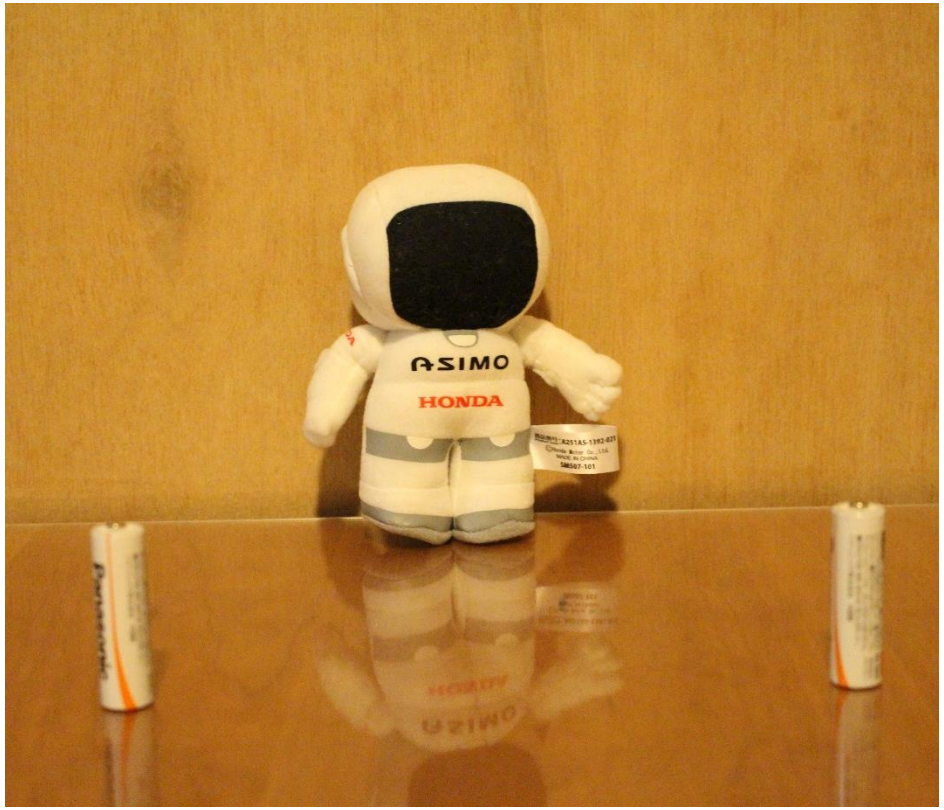
- Winfieldらの実験では，他のロボットを助けるようにプログラムされたロボットが，二台のロボットが同時に危険に近づいている状況で，適切な行動がとれずに「麻痺」してしまった.

1900年に書かれた政治的風刺画．大西洋と太平洋を
結ぶ運河をパナマとニカラグアのどちらのルートで建設するか
思案する米国議会の様子をビュリダンのロバになぞらえて
表現している．

ビュリダンのロバ

- ロバが二つの乾草の山の間でどちらを食べるか迷っているうちに餓死するという寓話.
- 同じくらい望ましい帰結を生む二つの行動の間では合理的な意思決定はできないというビュリダン（中世の哲学者）の説を風刺して作られた.

道徳的である ≠ 規範に従っている



ビュリダンのロボ

- 規範が何をするべきか教えてくれないときでも行動を選ばなければならない場合もある.
- 規範に反する行動が規範の順守よりも道徳的であることもある.

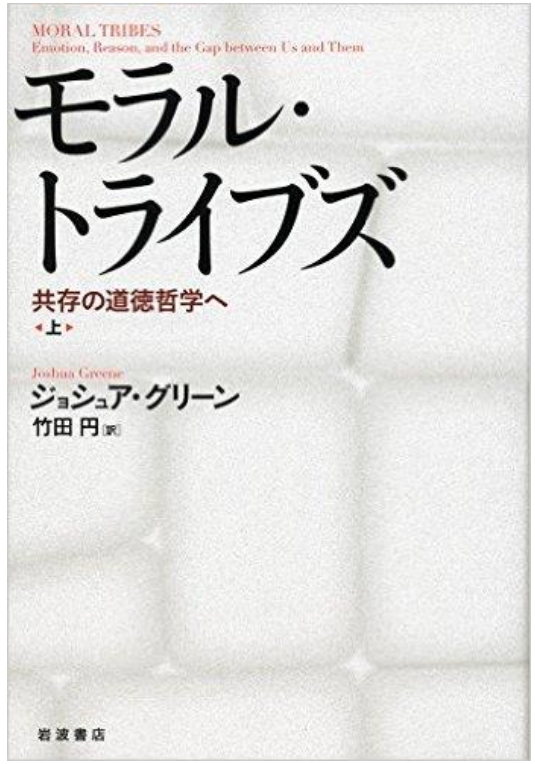
道徳的意思決定の2つのプロセス

論理

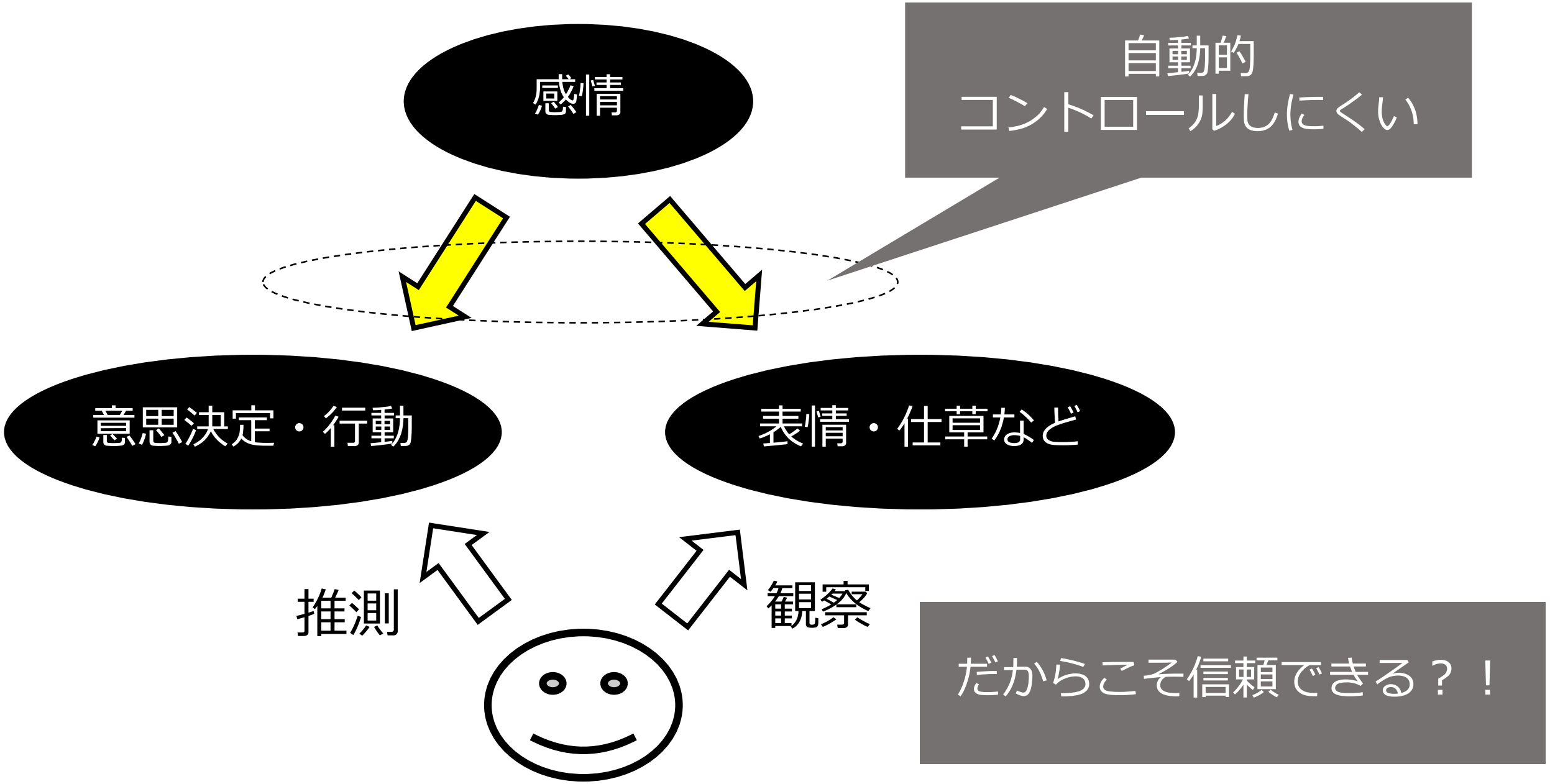
- 手間と時間がかかる
- 柔軟性がある
- より多くの関係者の利害を計算にいれることができる
- 損得を計算して利己的になりやすい

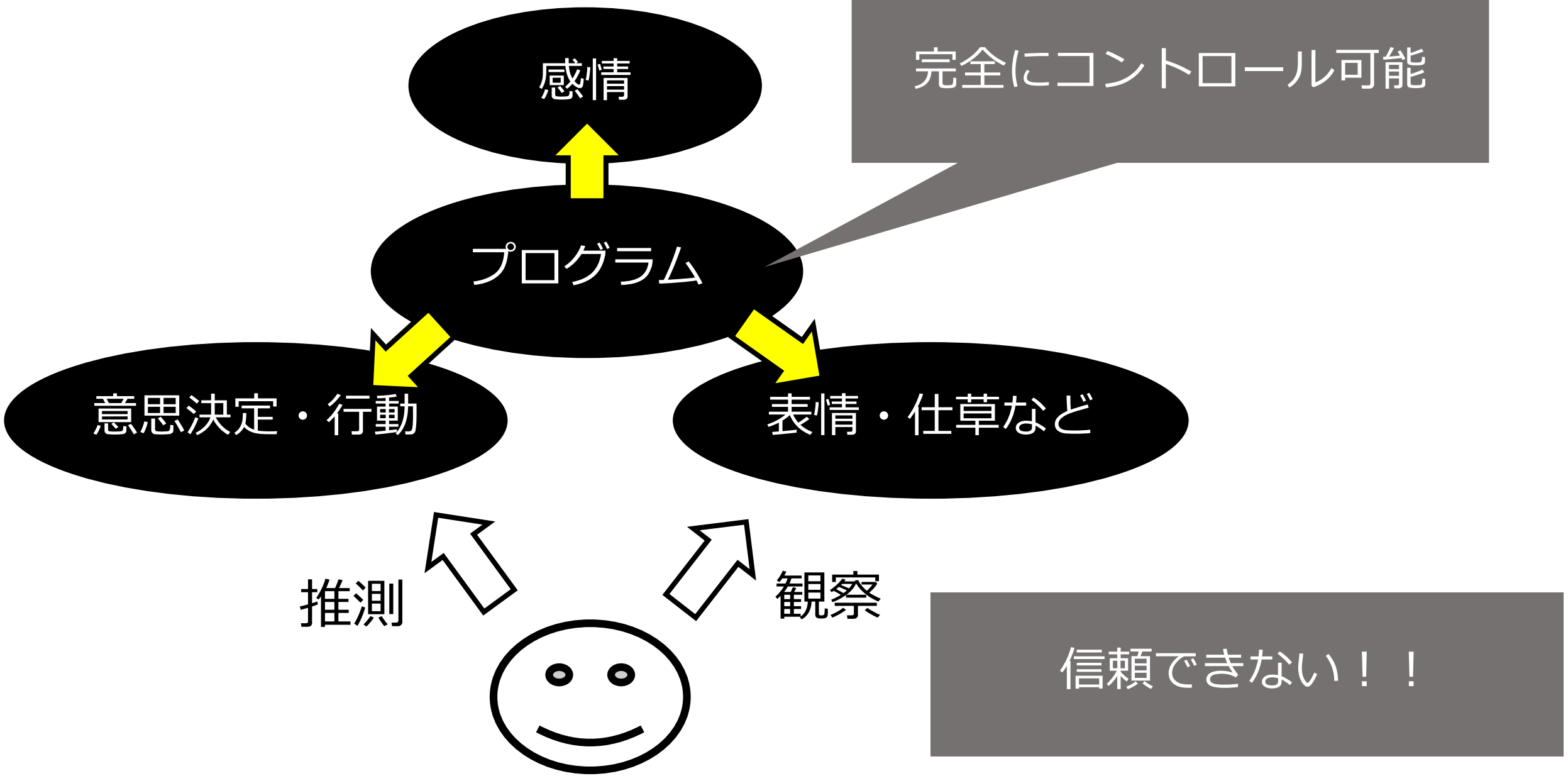
感情

- 自動的で素早い
- 融通が利かない
- 集団の中と外の区別に敏感
- 仲間に対しては利他的



ロボットに感情を持たせればいいのか？


感情
自動的
コントロールしにくい
意思決定・行動
表情・仕草など
推測
観察
だからこそ信頼できる？！


感情
プログラム
意思決定・行動
表情・仕草など
推測
観察
完全にコントロール可能
信頼できない！！

道徳的AI，道徳的ロボットの難しさ

状況依存性が高い

道徳的行為には責任が伴う

共感を求められる

「何をやるか」だけでなく
「誰がやるか」が重要

スポーツでビデオやカメラを活用した判定が増えている。米大リーグやテニスでは「チャレンジ」と呼ばれ、アウト、セーフやボールがコートに入ったかを映像で確認。ラグビーでは、審判が両手で空間に四角形を作るのが合図だ。四角形はテレビを指すようだ。

映像判定は最初、好きでなかった。何より、ゲームの流れが途切れる。ルールブックは「おれ」ではなく

機械判定

「機械」が取って代わったのに審判は悔しそうな表情を見せない。ジャッジに口角泡を飛ばす監督らの名物シーンもめっきり減った。ストライク、ボールの判定の機械化阻止は譲れない一線だ。ただ、真剣勝負を誤審でぶち壊したくないという点を重視すれば、ある程度は理解できなくもない。

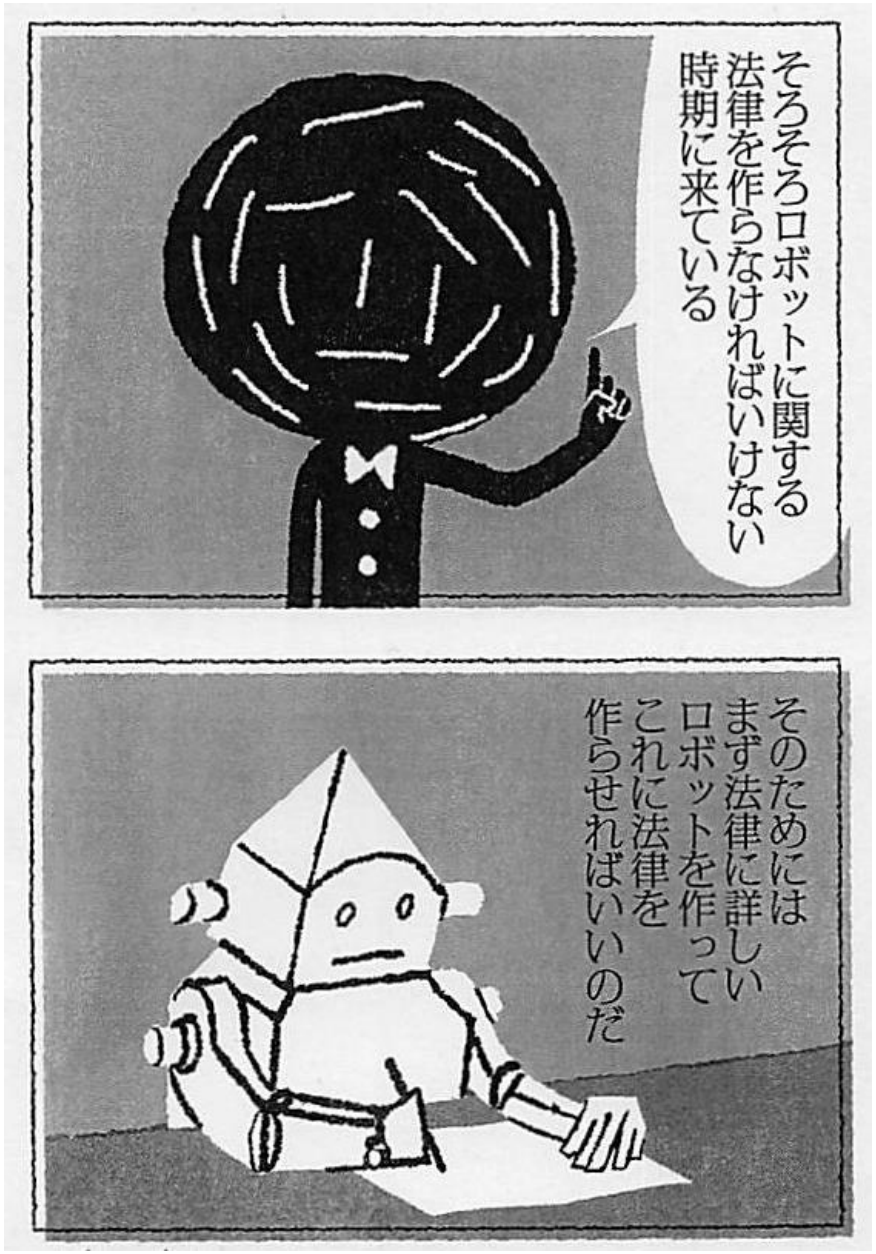
機械による判定を、例えば司法の世界に持ち込んだらどうなるか。人工知能の発達もあり、技術的には不可能ではないだろう。大型コンピューターに過去の判例を漏れなく入力。裁判官らが情状面の数値を入力すると、無機質な声で次のように言い渡される。

「被告人を懲役十年八月三十日四十一時間三分（秒以下切り捨て）に処する」

いくら「正確」で「公平」でも、こんな判決は絶対に受けたくない。

（デスク長・田中宏明）

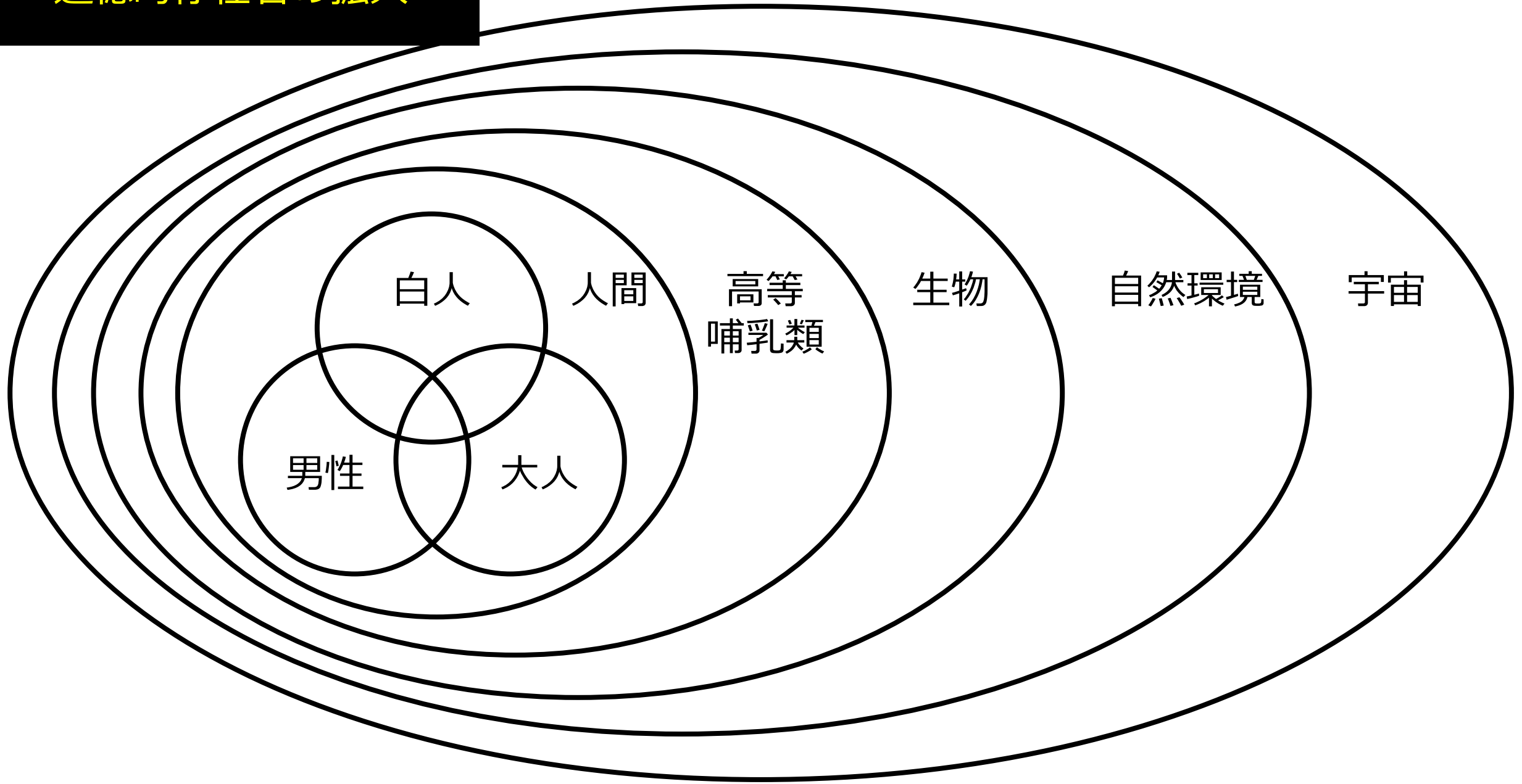
中日新聞，2015年5月17日



木内達朗「チキュウズィン」
（『考える人』，2015年春号）

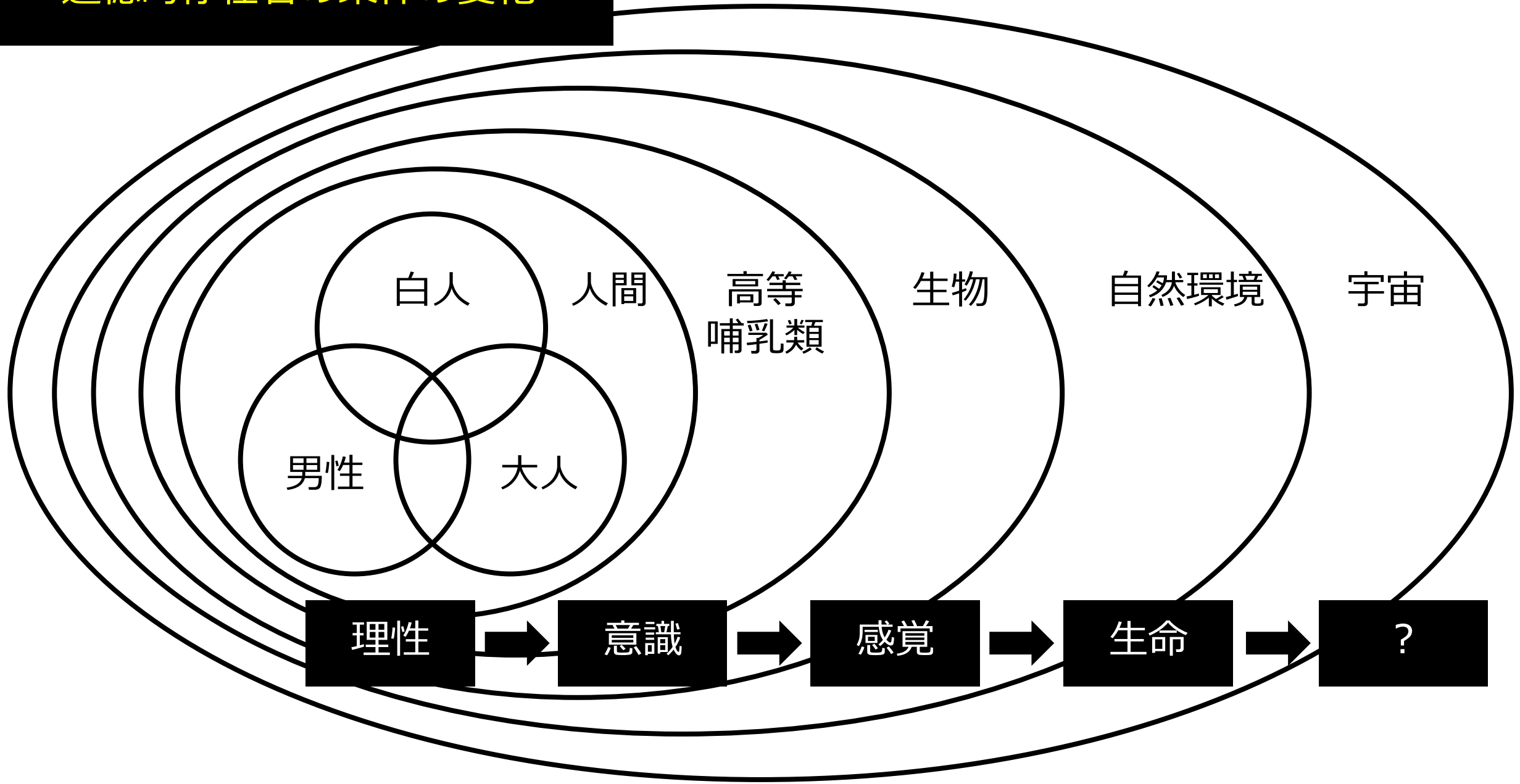
道徳的存在者の拡大

（ヨーロッパ，キリスト教文化圏の例）

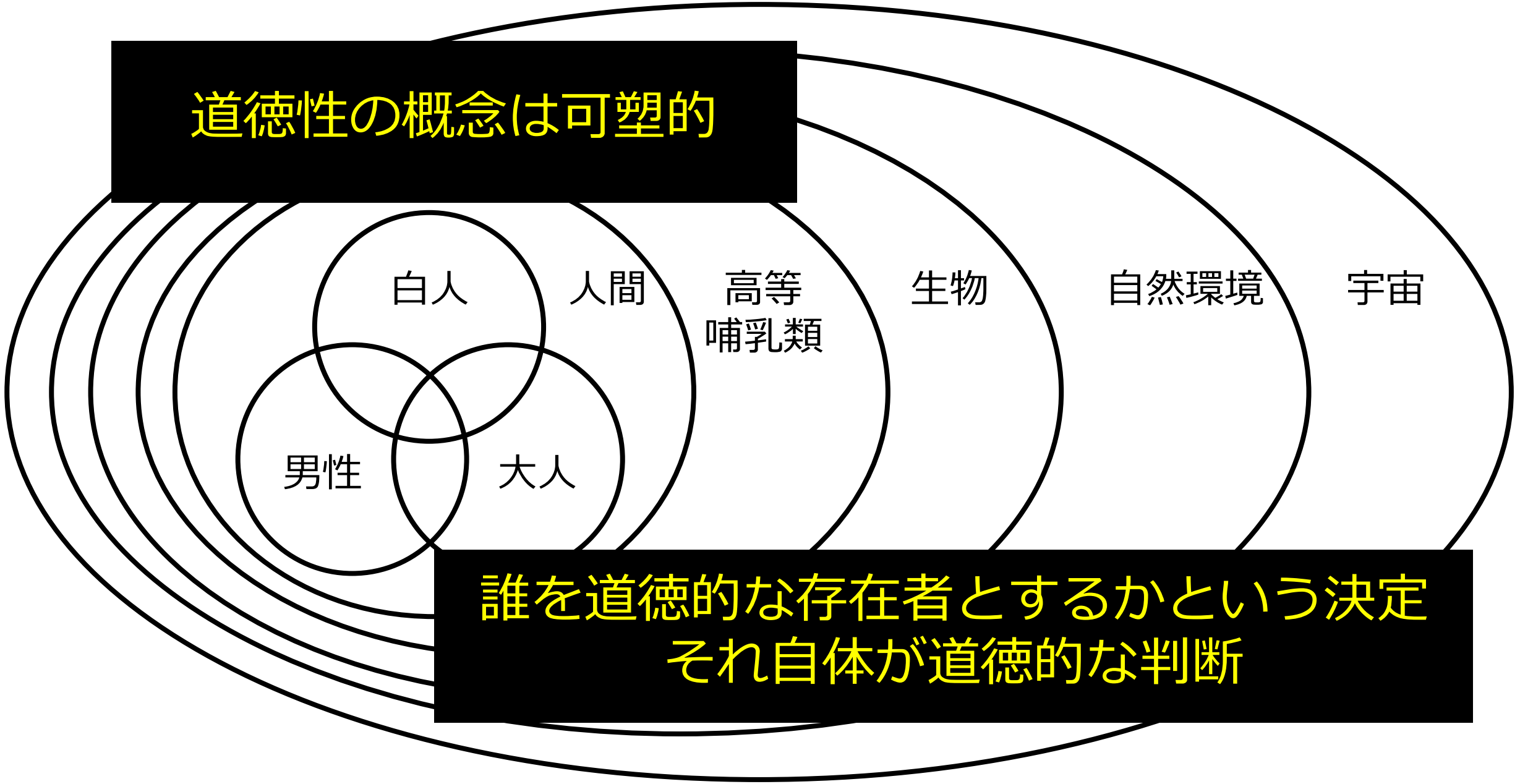


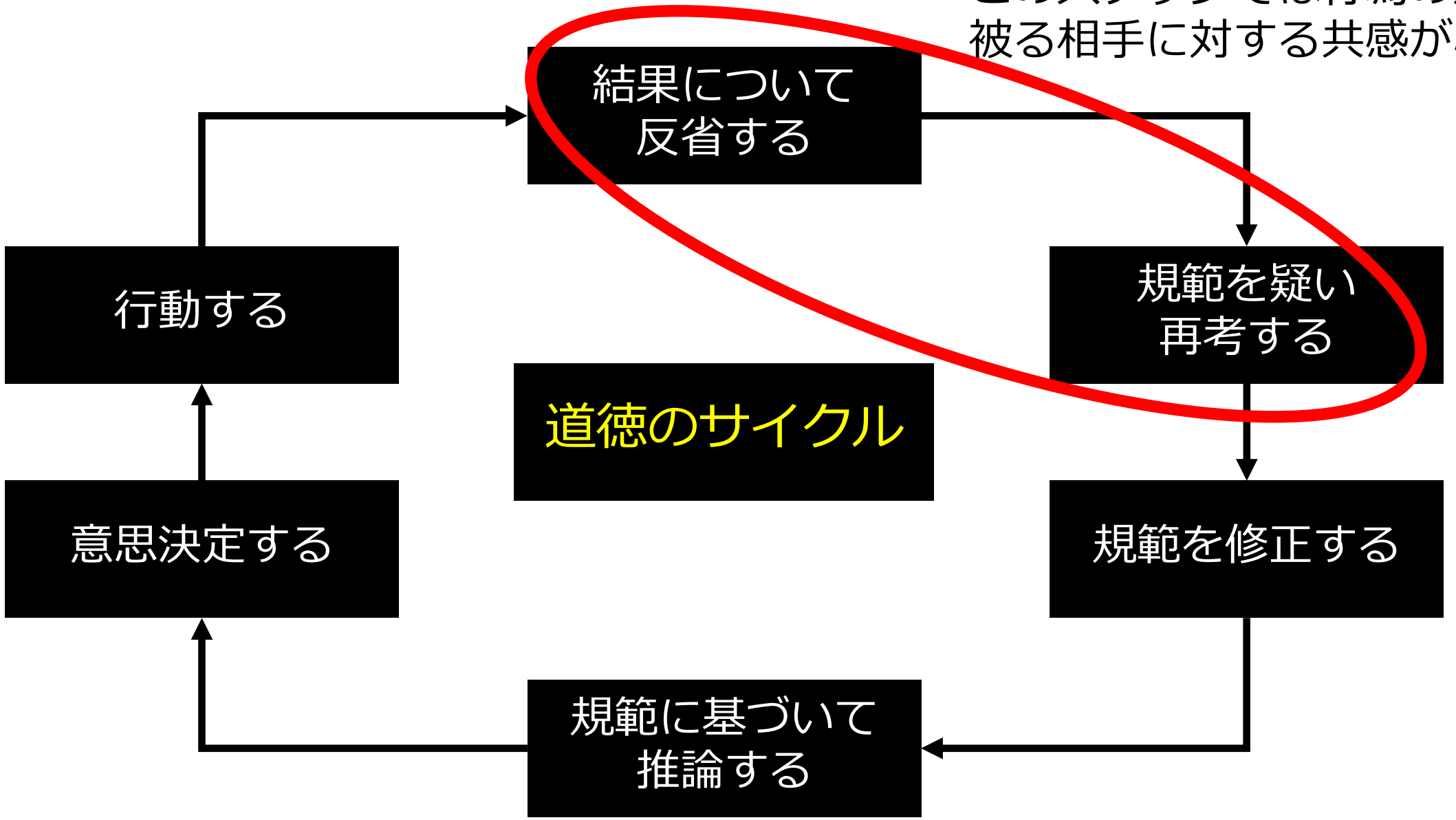
道徳的存在者の条件の変化

（ヨーロッパ，キリスト教文化圏の例）

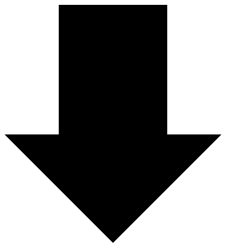


（ヨーロッパ，キリスト教文化圏の例）



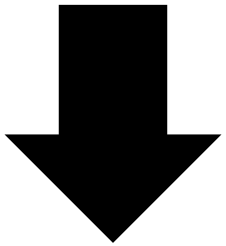

このステップでは行為の影響を
被る相手に対する共感が必要
結果について
反省する
規範を疑い
再考する
規範を修正する
規範に基づいて
推論する
意思決定する
行動する
道徳のサイクル

人工知能は道徳的になりうるのか？



私たちは人工知能を
道徳的存在者として受け入れられるか？

人工知能は道徳的になるべきか？



私たちは人工知能を
道徳的存在者として受け入れるべきか？

No robots were harmed
in the making of this video.

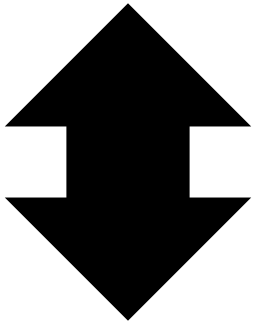
© 2015 Boston Dynamics

http://stoprobotabuse.com/

このビデオの撮影中に
傷つけられたロボットはいません

https://www.youtube.com/watch?v=M8YjvHYbZ9w

道徳は計算可能
道徳の機能主義，行動主義
道徳性もチューリングテストで計ろう



人工道徳は人工的サイコパス
道徳の官僚主義化
道徳的ニヒリズム

人工知能は道徳的になりうるのか？

現在の「道徳性」の通念からすれば無理.
しかし人工道徳的行為者の導入は
「道徳性」の概念を書き換えるだろう.
臓器移植という技術が「死」の概念を書き換えたように.
しかしその際には大きな衝突が起こることが予想される.

人工知能は道徳的になるべきか？

ケアロボットなどはこれからの日本に必要な技術である.
そしてそれらは表層的にでも「道徳的」に振舞うことが望ましい.
しかし私たちが伝統的に持ってきた「道徳性」の概念に
重大な変更をもたらす技術の導入には
少なくとも慎重な議論と合意形成のための努力が不可欠だろう.

Forbes

http://www.forbes.com/sites/jimgorzelany/2015/10/09/volvo-will-accept-liability-for-its-self-driving-cars/

Volvoは自動運転自動車についての責任を会社が引き受けると明言

Forbes / Autos

OCT 9, 2015 @ 11:48 AM 3,738 VIEWS

Volvo Will Accept Liability For Its Self-Driving Cars

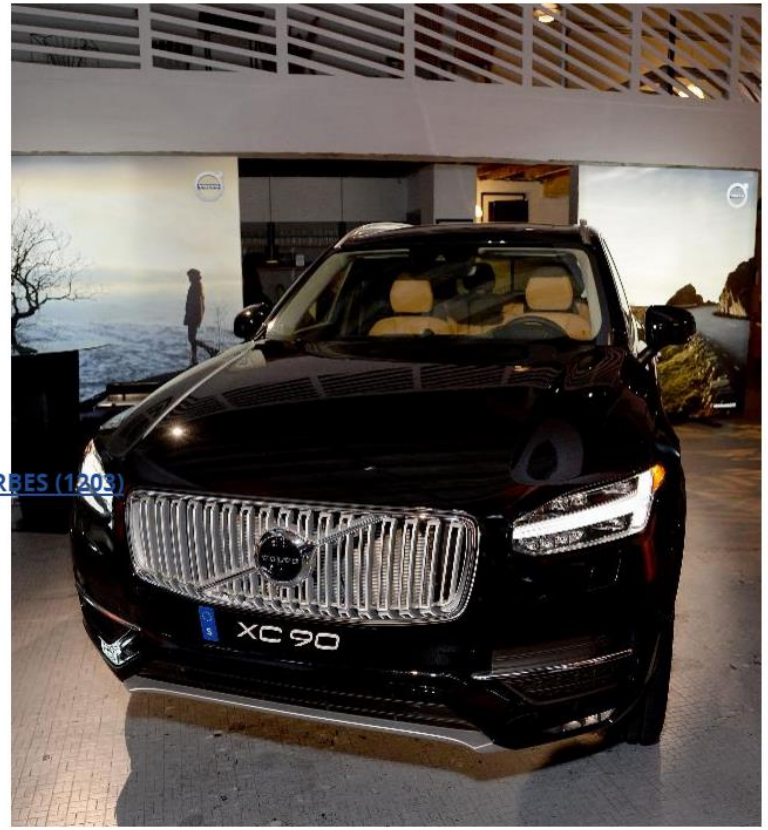
Jim Gorzelany
CONTRIBUTOR

I write about how to maximize your automotive investment and more.

FOLLOW ON FORBES (1203)

FULL BIO >

Opinions expressed by Forbes Contributors are their own.



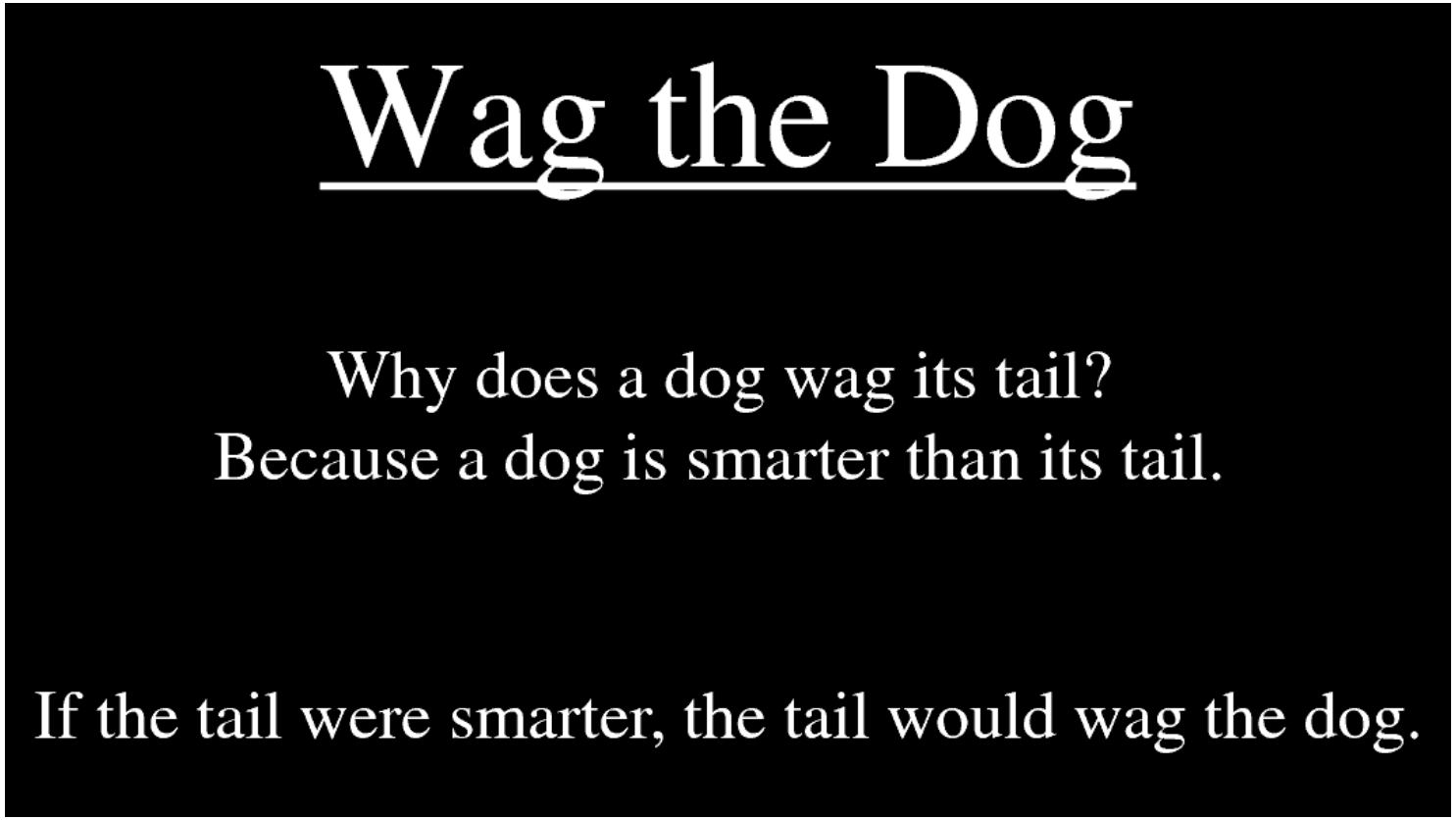
The XC90 crossover SUV will reportedly be the first Volvo vehicle to be fitted with the company's forthcoming Auto Pilot system, albeit on a limited basis. (Photo by Michael Kovac/Getty Images for Volvo Cars of North America)

機械の道徳性

作っている人間，会社の道徳性

「余は汽車の猛烈に 、見界なく 、すべての人を貨物同様に心得て走る様を見るたびに 、客車のうちに閉じ籠められたる個人と 、個人の個性に寸毫の注意をだに払わざるこの鉄車とを比較して 、 ― ―あぶない 、あぶない 。気をつけねばあぶないと思う 。現代の文明はこのあぶないで鼻を衝かれるくらい充満している 。おさき真闇に盲動する汽車はあぶない標本の一つである 。」

夏目漱石『草枕』（1906）



「犬が尻尾を振るのは何故か？ 尻尾より賢いからだ.
もし尻尾の方が賢ければ，尻尾が犬を振る．」

バリー・レヴィンソン監督の映画『Wag the Dog』（1997）より

Use the Humans

Why do humans use tools?
Because humans are smarter than the tools.

If the tools were smarter, the tools would use the humans.

「人が道具を使うのは何故か？ 道具より賢いからだ.
もし道具の方が賢ければ，道具が人を使う」