

人と人工知能を繋げる ヒューマン・エージェント インタラクション (HAI)

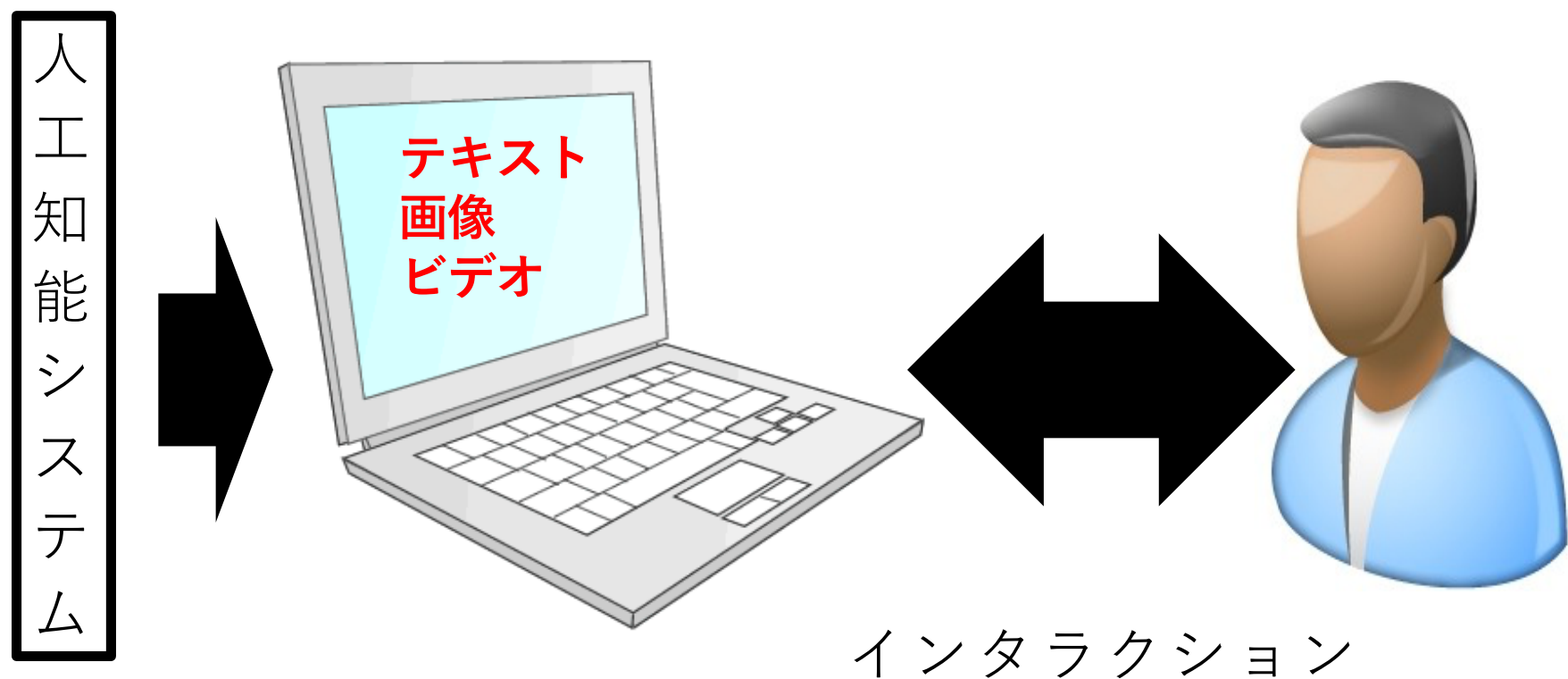
慶應義塾大学 理工学部

今井 倫太

自己紹介

- 1988-1994 慶應義塾大学 理工学部出身
 - 安西 祐一郎 研究室
- 1994- NTT ヒューマンインタフェース研究所
- 1997- ATR 知能映像通信研究所
- 2002- 慶應義塾大学 理工学部
ATR 知能ロボティクス研究所
- 2009-2010 シカゴ大学 訪問研究員
- 2015-2017人工知能学会理事
- 現在、理工学部情報工学科 教授, ATR 知能ロボティクス研究所

人と人工知能を繋げる

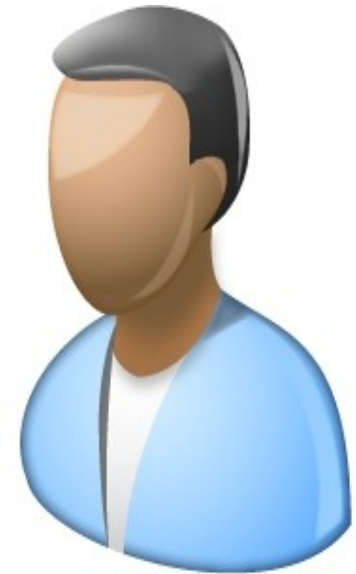
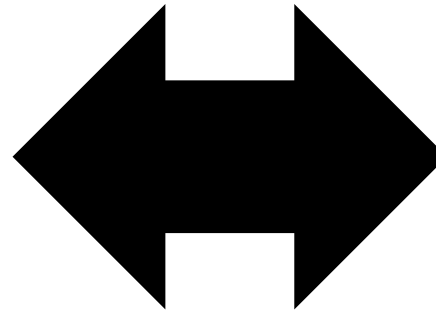


人と人工知能を繋げる存在

人工
知能
シス
テム



エージェント



ヒューマン・エージェント
インタラクション

本日の目標

- (ロボットを始めとする) エージェントシステムの存在意義
 - ただ、人目をひくからロボットなの？
 - 可愛いかったり愛着が湧くから？
- エージェントシステムに必要な設計論
 - 存在意義を主張するに足るエージェントを作る上で何が必要か？

概略

- AI全般・脅威・想定範囲内
- 自律性の向上 vs 今までの道具・機械・コンピュータ
- 人にすると理解が進む
- 人は人が好き？
- エージェントの設計論
- エージェントによる人と人工知能のインタラクション
- エージェントに必要なインタラクションデザイン
- そして未来へ

AI全般・脅威・想定範囲内

- 知的な情報処理のうち人にできて、コンピュータでできないことを実現する学問
- かなりナイーブなアプローチをとって来た
- 知識表現、推論、学習、探索、発見、制約充足、命題論理、エキスパートシステム、ゲーム、対話、画像認識・理解、音声認識、常識推論、事例ベース、ベイジアンネットワーク、ニューラルネットワーク
- トイ・プロブレム (積み木の世界の問題)

人工知能に奪われる・奪われない職種700位

• Michael A. Osborne オックスフォード大学

• 奪われる職種

データ分析および結果を利用する仕事

基準が明確で厳密性を問われる判定

定型的な作業や、定型的なやり取り業務

- 銀行融資の担当者
- スポーツの審判
- 不動産ブローカー
- レストランの案内係
- 保険の審査担当者
- 動物のブリーダー
- 電話オペレーター
- 給与・福利厚生担当者
- レジ係
- 簿記、会計、監査事務員
- etc

現状のAI研究の方向性で確実に実現されていく

人工知能に奪われる・奪われない職種

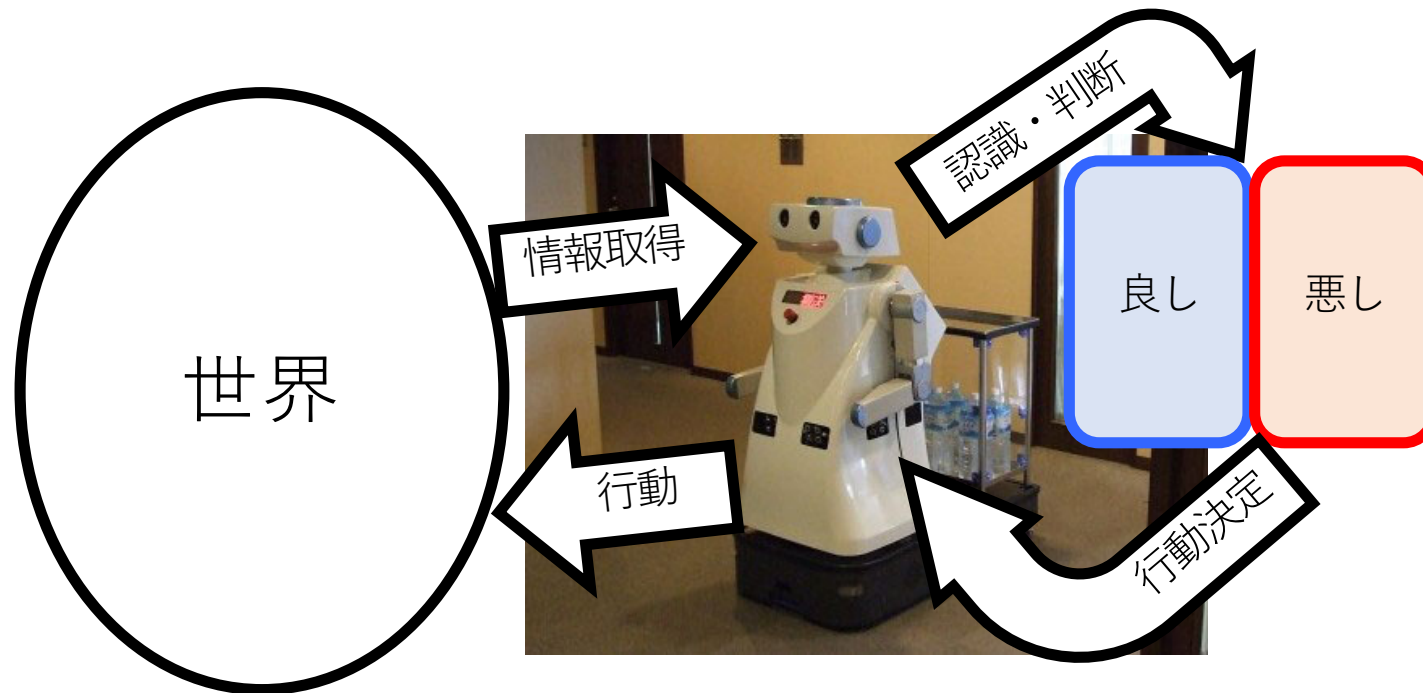
• 奪われない職種

インタラクションスキルが必要な仕事

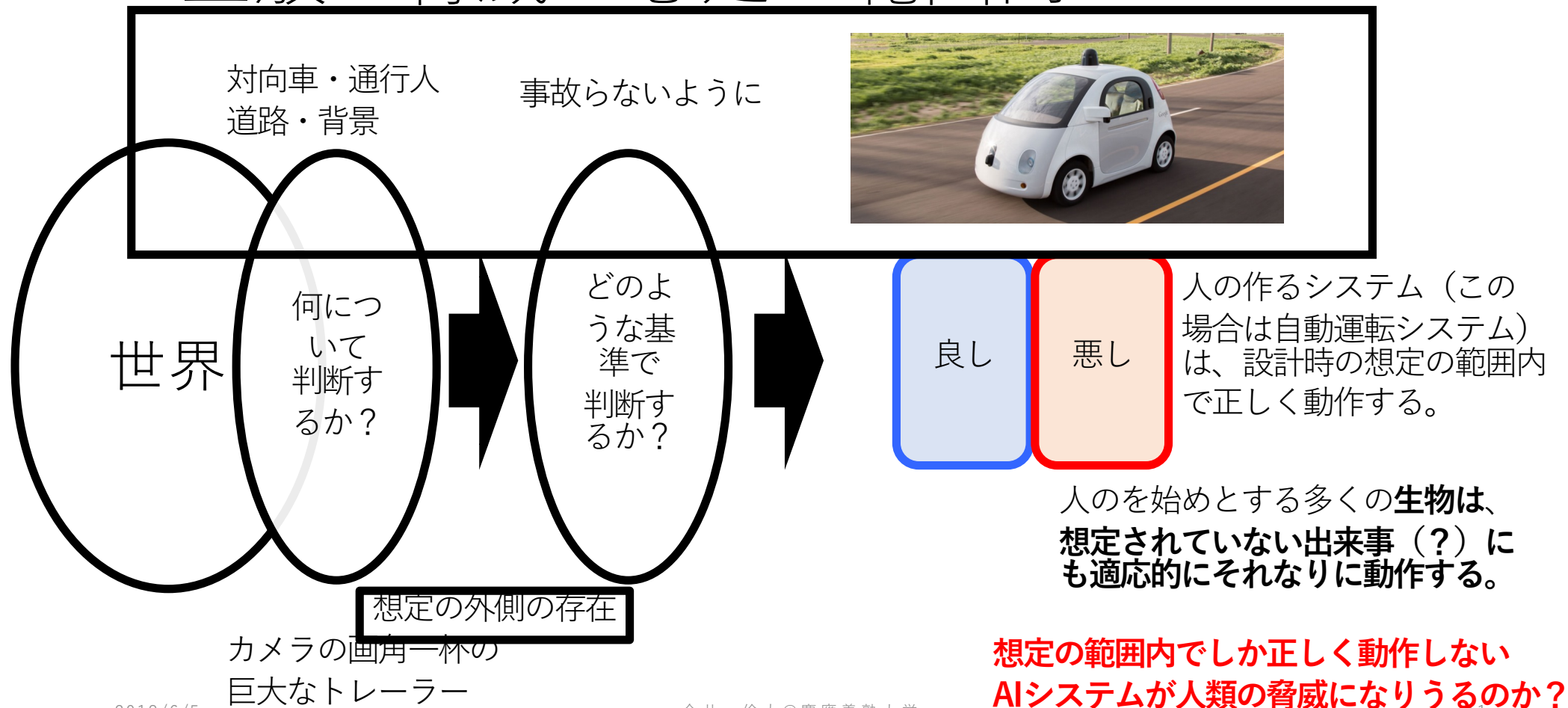
現状のAI研究に加えさらに、
インタラクションの研究をして
初めて支援方法とかが見えてくる領域

- レクレーションセラピスト
- 最前線のメカニック、修理工
- 緊急事態の管理監督者
- メンタルヘルスと薬物利用者サポート
- 聴覚医療従事者
- 作業両方師
- 義肢装具師
- ヘルスケアソーシャルワーカー
- 口腔外科
- 消防監督者
- 栄養士
- 施設管理者
- 振り付け師
- セールスエンジニア
- 内科医と外科医
- 教育コーディネーター
- 心理学者
- 警察と探偵
- 歯科医師
- 小学校教員

AI全般・脅威・想定の範囲内



AI全般・脅威・想定外の範囲内



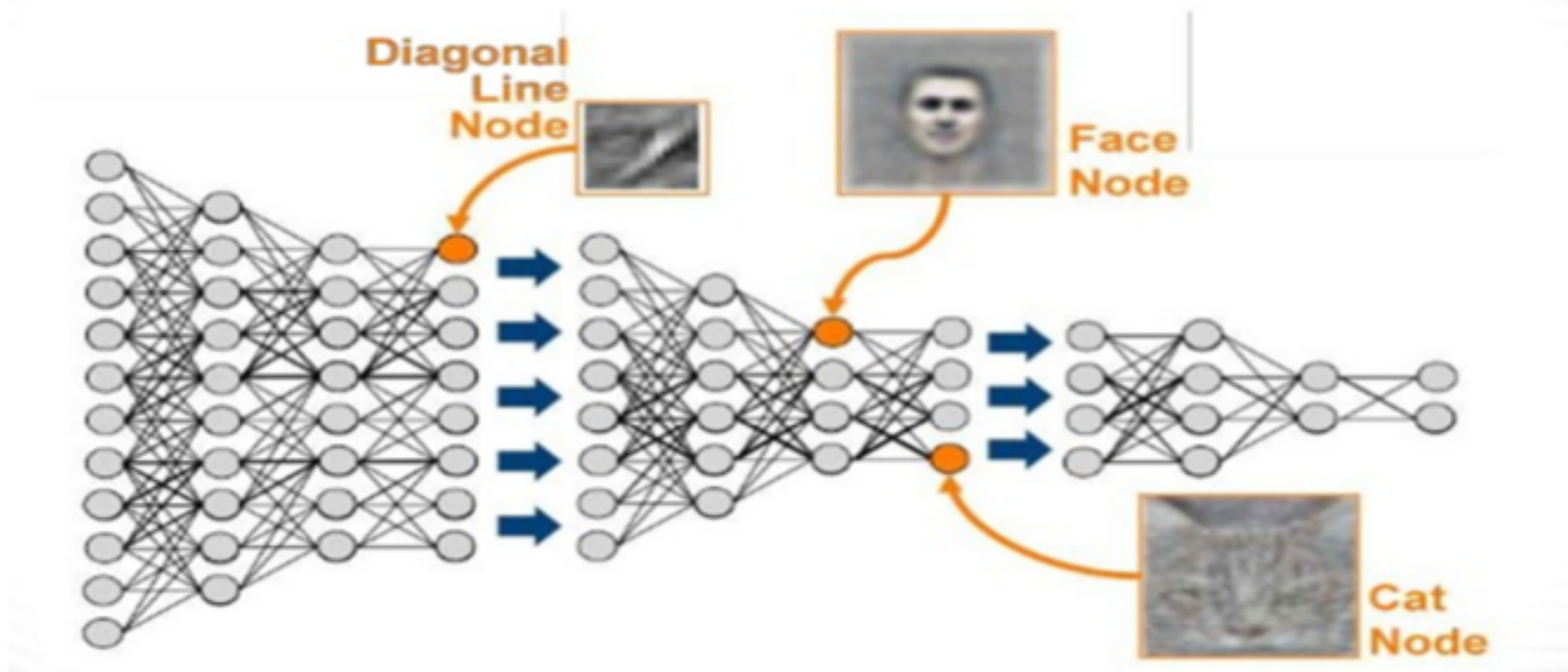
インタラクシ ョ ンとは

What is a plan for?



AI全般・脅威・想定範囲内

深層学習 (Deep learning)



AI全般・脅威・想定範囲内 Watson

- IBMの質問応答システム「Watson」が、米国の人気クイズ番組「Jeopardy!」に挑戦し、人間のチャンピオンを打ち破って最高金額を獲得（2011年2月）
- 100万冊の本を読むのに相当する自然言語で書かれた情報の断片を分析
 - 非構造データ分析（≡機械学習+ビッグデータ）
 - 自然言語処理

AI全般・脅威・想定範囲内

AIは想定範囲内を抜け出したのか？

No

- ・データの用意
- ・何を学習させるのかの選定
- ・ネットワーク構造の選定やパラメタの設定
- ・目標やモチベーション、価値は、人が与えている

機械が自律して活動できることが増えた。

自律性につながる前準備（環境認識能力など）が格段に向上した。

脅威といより、**人にとって捉えづらい**今までになかった**機械**が誕生しつつある

自律性の向上
コンピュータ

vs 今までの道具・機械・

今まで



掃除する



乗る



話す
調べる

人に、道具・機械を使うモチベーションがもともと存在する。

自律性の向上 vs 今までの道具・機械・ コンピュータ

登場しつつある機械



自動的に掃除してくれる

使う人にモチベーションがなくても、
機械の側から自律的に働きかけてくる。
さらに、自律的に行動する。

自律して目的地へ連れて
行ってくれる



向こうから話しかけられる
教えてくる

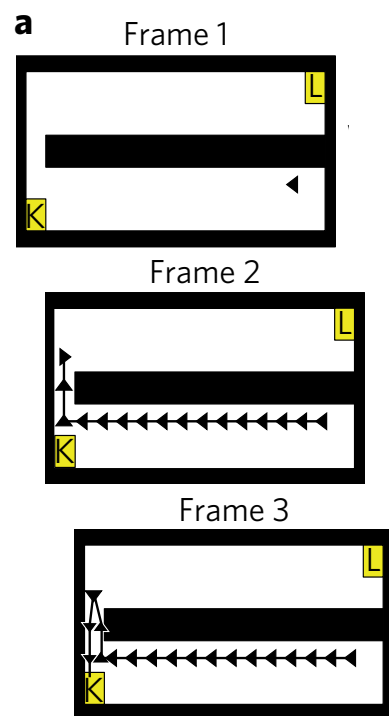
自律性の向上 vs 今までの道具・機械・ コンピュータ

**使う人にモチベーションがなくても、
機械の側から自律的に働きかけてくる。
さらに、自律的に行動する。**

人と機械の関係は、単純に、操作したら動く、聞いたら返事してくれるといった、人が主導型の使い方にとってかわって、機械の側から、情報を提示してくる、何かをお願いすると、複雑な過程を経て時間をかけて結果を出すといった、人と機械が対等もしくは、機械が主導的に何かをする存在にかわった。

**AIによって、能動的で複雑な事をする存在となった機械を人は
理解して使いこなせるのか？**

人になると理解が進む

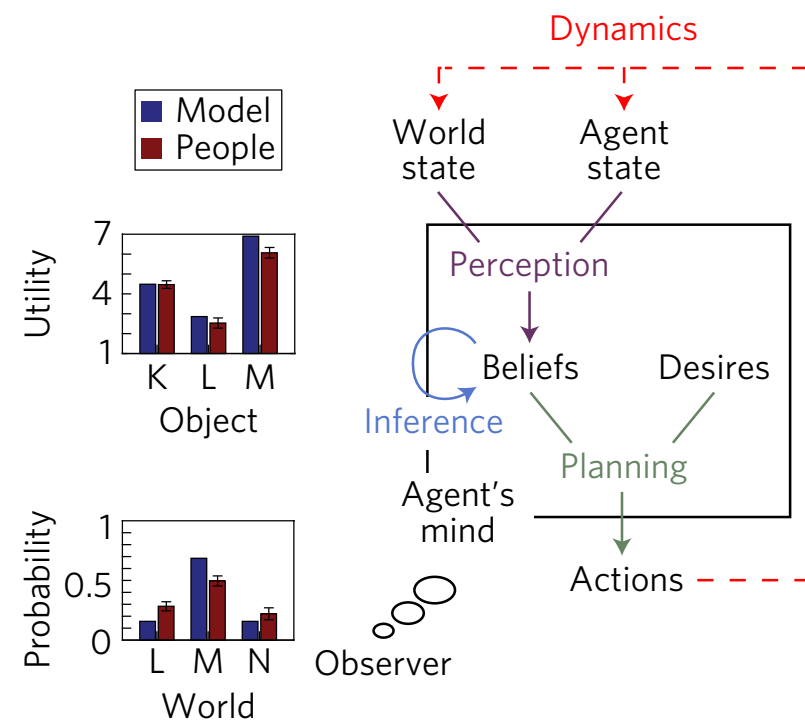


人は、自律的に行動する
相手を、認知的な内部
状態を仮定して理解する。

自律性が高まったAIシステム
を、人に理解させるひとつの
デザイン戦略となりうるか？

なぜロボットなのか？
なぜCGキャラクターなのか？

b

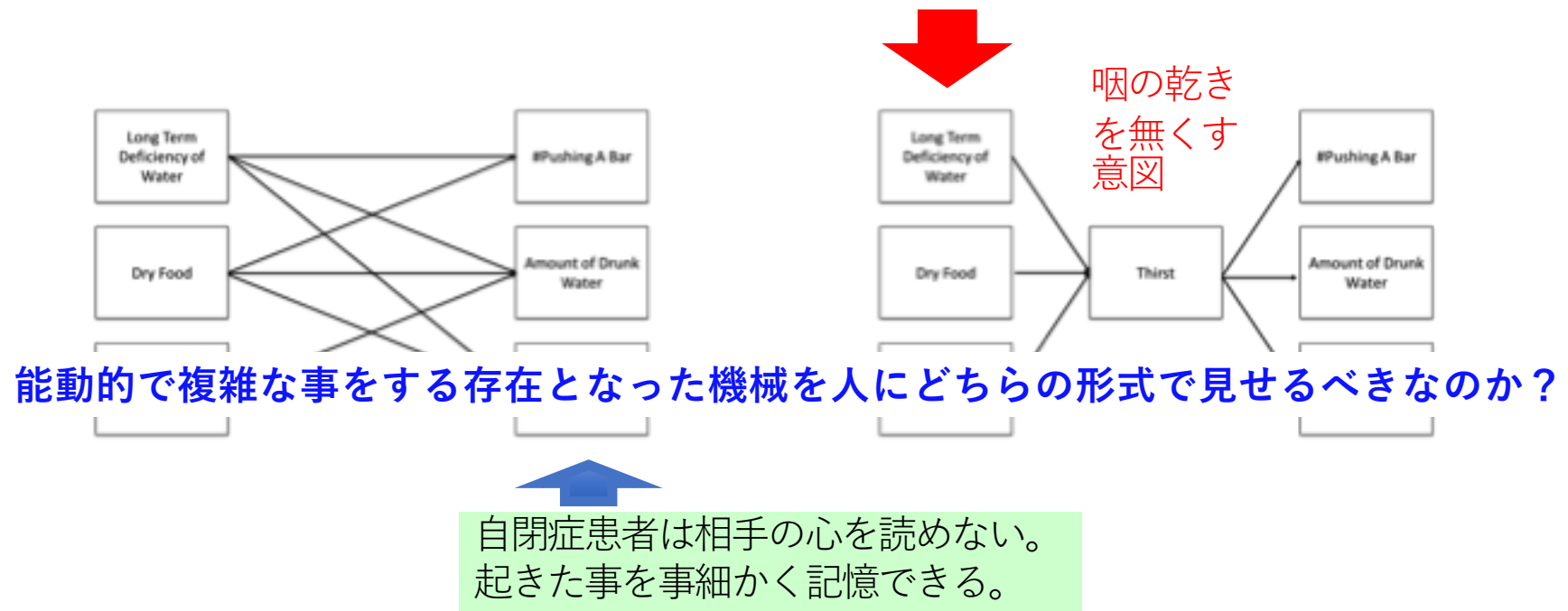


[1] Chris L. Baker, Julian Jara-Ettinger, Rebecca Saxe and Joshua B. Tenenbaum, "Rational quantitative attribution of beliefs, desires and percepts in human mentalizing," Nature Human Behavior, Vol.1, No. 64, pp.1-10, 2017

Behavior-Reading vs. Mind-Reading

振る舞いを読む vs. 心を読む

効率よく相手（エージェント）の事を理解できる。



Whiten 1996, "When does smart behaviour-reading become mind-reading?"

Theories of theories of mind, Cambridge University Press, 1996, 277-292

人になると理解が進む



[2] Tadakazu Kashiwabara, Hirotaka Osawa, Kazuhiko Shinozawa, Michita Imai, "TEROOS: a wearable avatar to enhance joint activities," Proceedings of ACM the SIGCHI Conference on Human Factors in Computing Systems, pp. 2001-2004, 2012

2018/07/5

人の社会的な反応



2018/6/5

22

人の社会的な反応

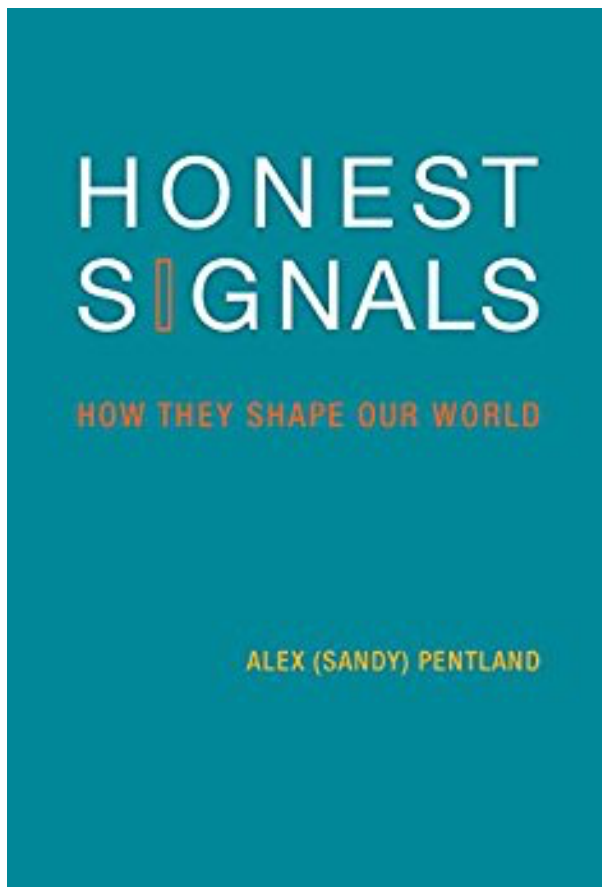


2018/6/5

今井 倫太@慶應義塾大学

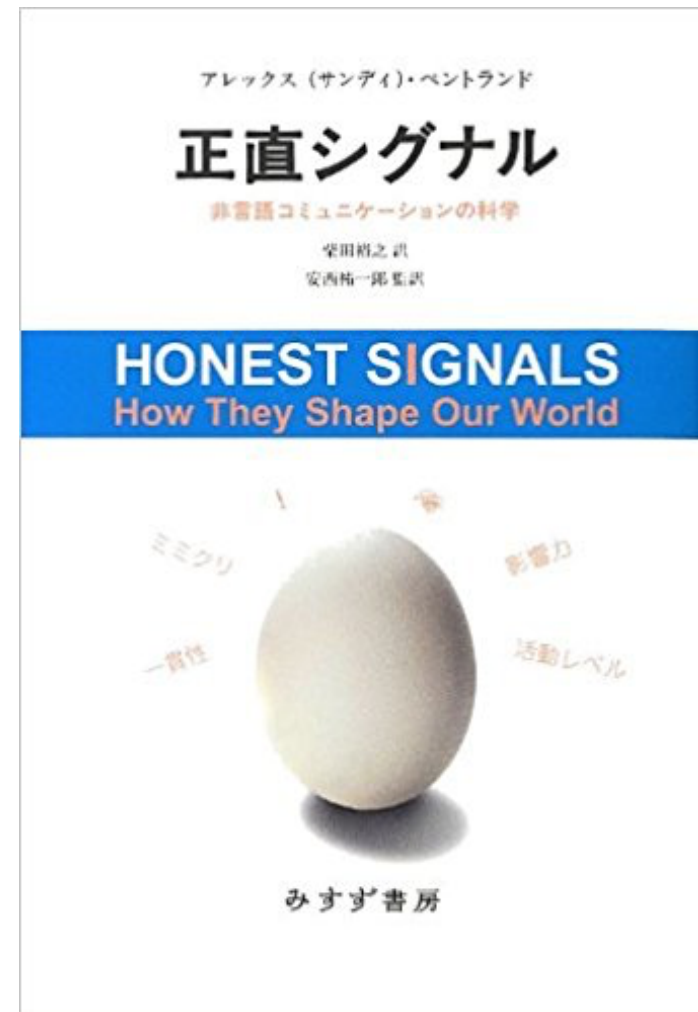
23

人は人が好き？



2018/6/5

今井 倫太@慶應義塾大学



24

人は人が好き？



なぜ、アナウンサーが登場するのか？

エージェントの設計論

- 擬人化の3要件
 - 1. 人間らしさを感じさせるデザイン要素 (elicited agent knowledge)
 - 2. 他のエージェントの振る舞いを理解・説明するモチベーション (effectance motivation)
 - 3. 社会的な接触の欲求 (sociality motivation)

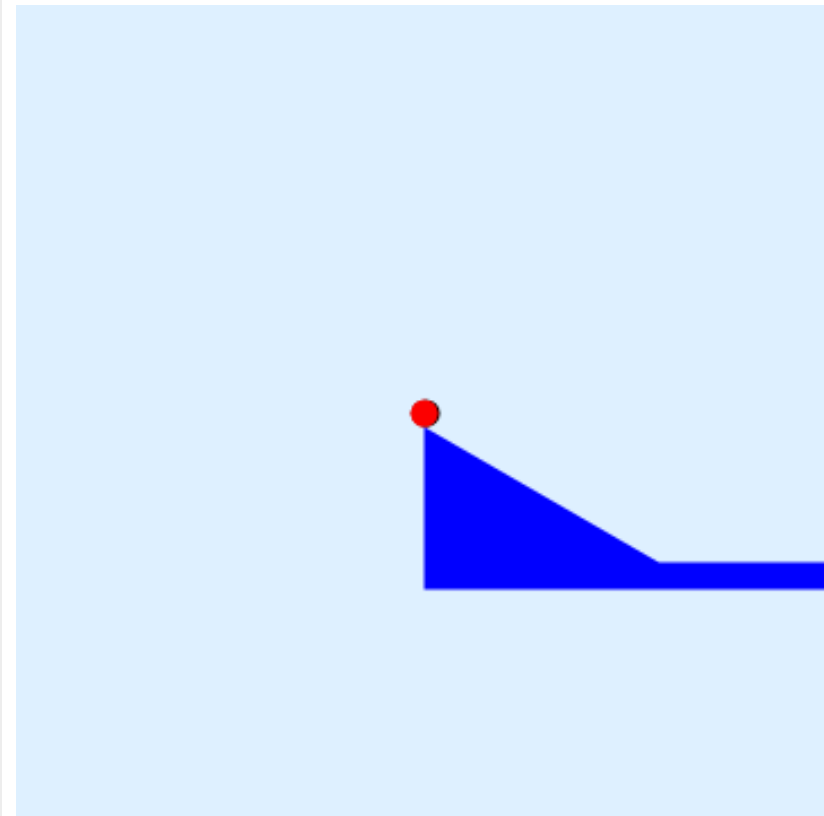
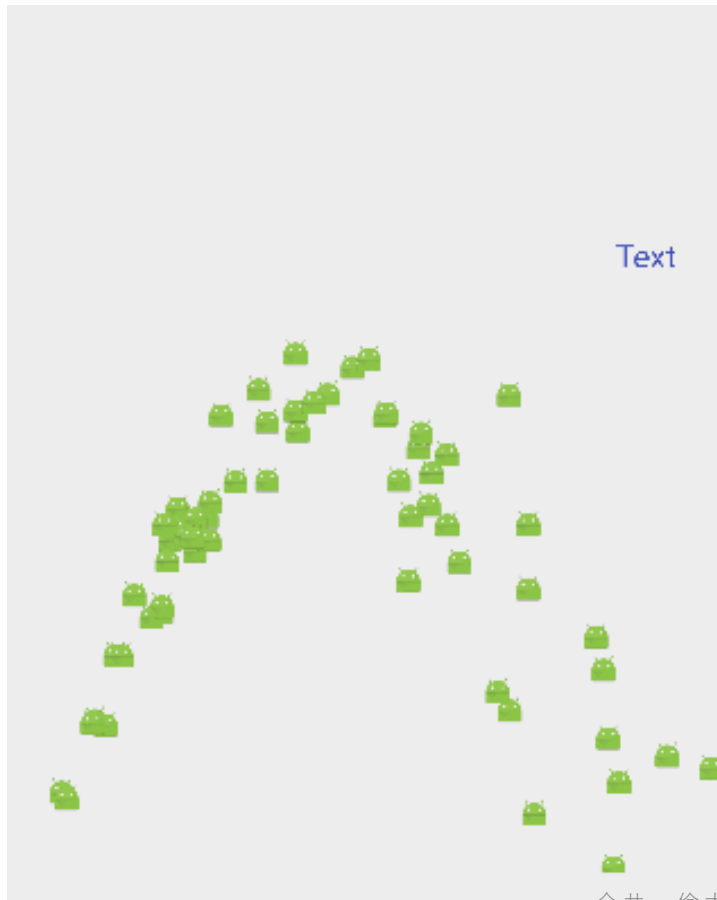
[3] Nicholas Epley, Adam Waytz, and John T. Cacioppo, “On Seeing Human: A Three-Factor Theory of Anthropomorphism,” *Psychological Review*, Vol.114, No.4, pp.884-886, 2007.

デネットのスタンス

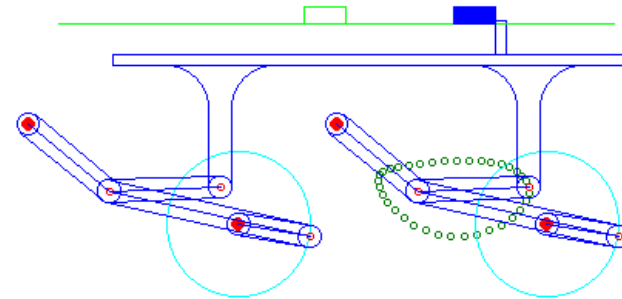
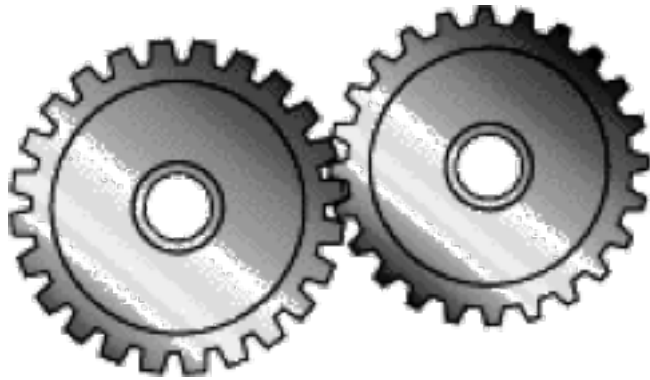
Daniel Dennett, "Intentional Systems," Journal of Philosophy 68: 87-106, 1971

- 人は動いているものをどのように理解するのか？
 - 物理スタンス
 - 設計スタンス
 - 意図スタンス

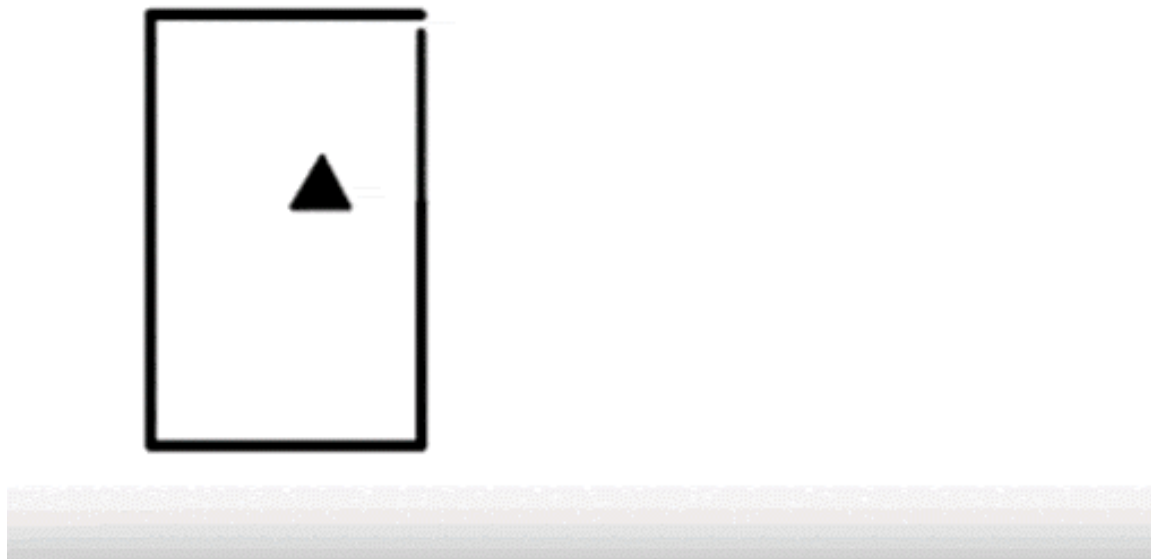
物理スタンス



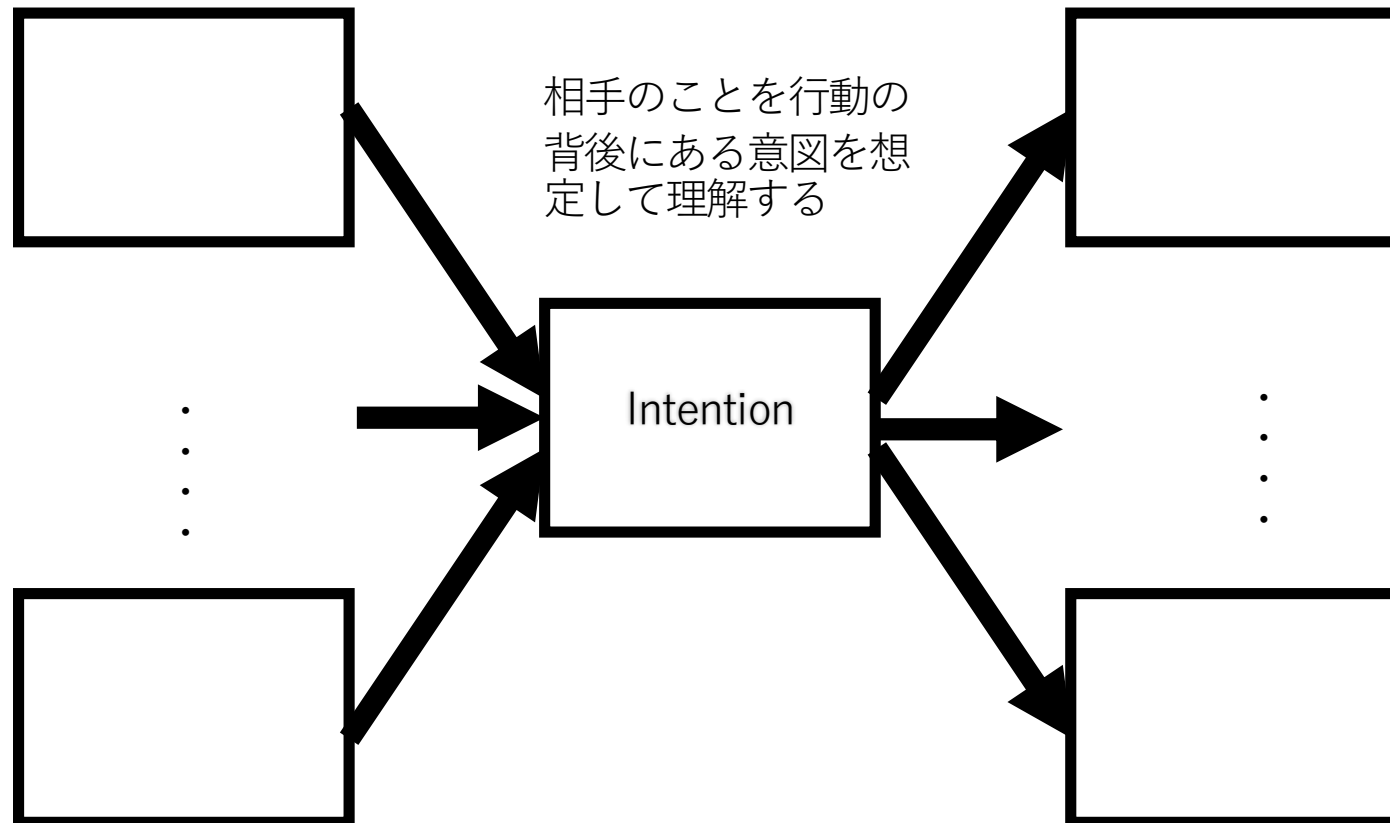
設計スタンス



意図スタンス



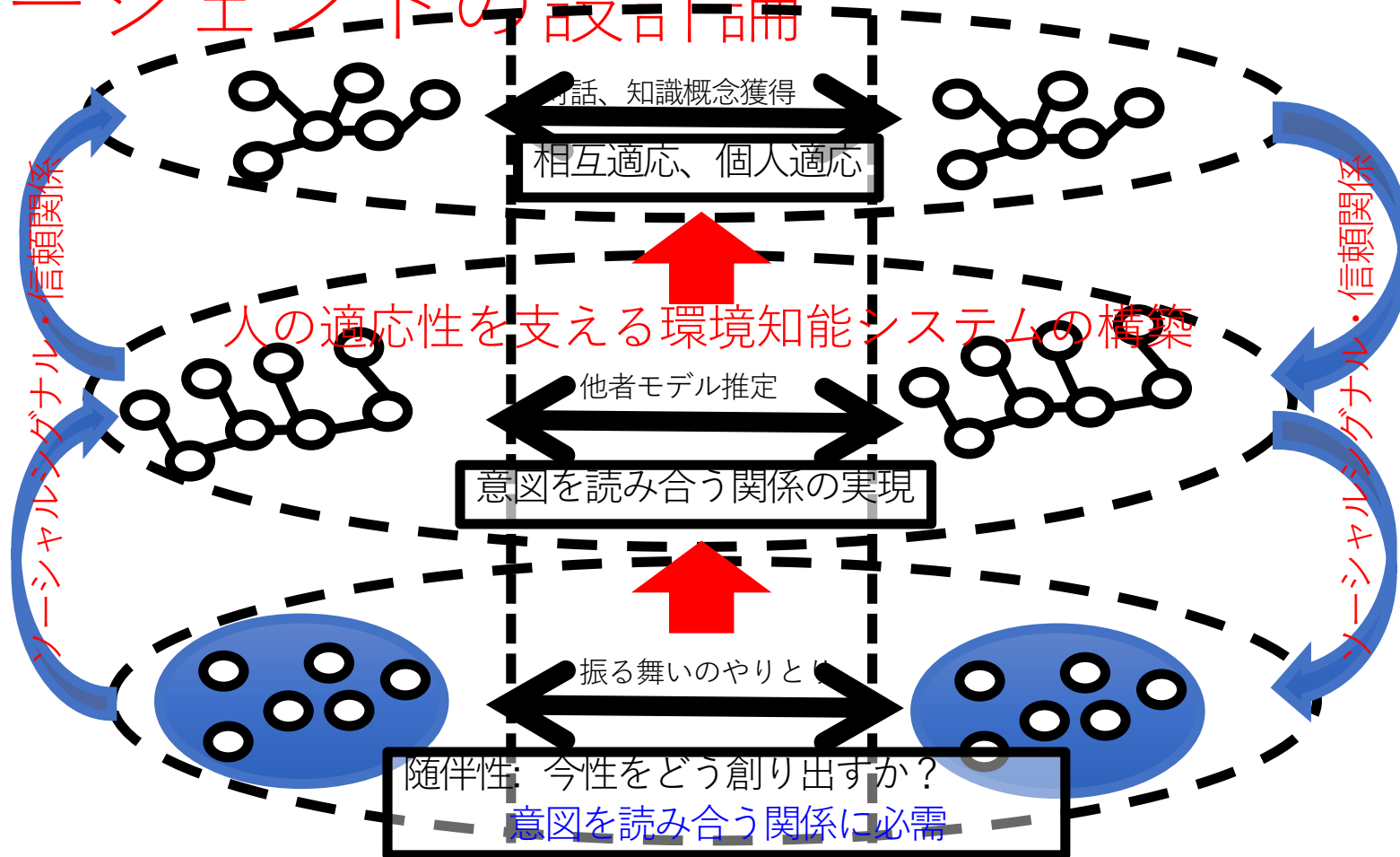
意図スタンス



設計スタンスと意図スタンス

- 設計スタンス
 - ある目的があって、それを達成するように誰かに設計されている。
- 意図スタンス
 - 複数想定される実行可能な行動がある中で、エージェントが、他の事をやらずにある一つの事をあえて選択して行っているように見える。

エージェントの設計論

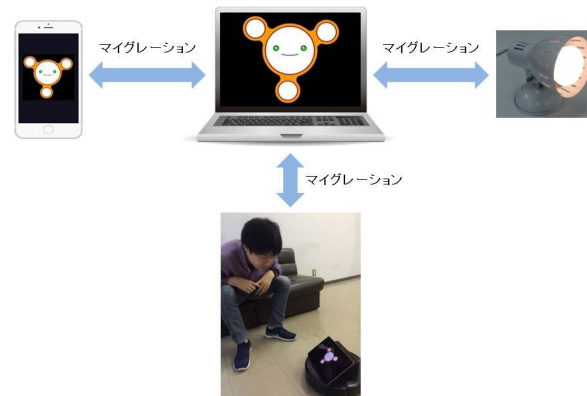


構築事例

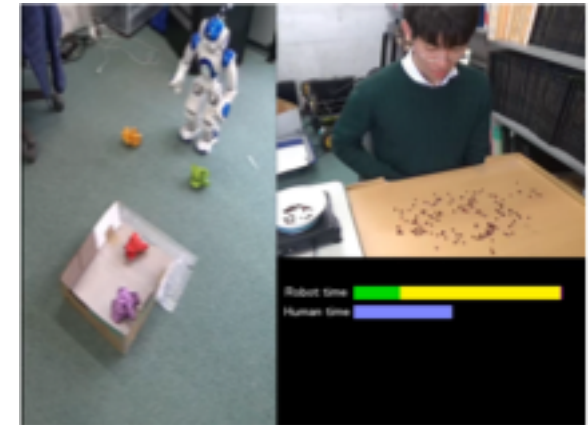
- 今性を示すロボット
- 機器のメンタライジングと人の受容性
- 時間ボリュームのあるタスクがメンタライジングのアプリケーションエリア
- 人からどう思われているかをロボットが知ること
- ロボットが人を思い計ると人はコミュニケーションに没入する



2018/6/5



今井 倫太@慶應義塾大学



34

今性を示すロボット

2017



Yusuke Takimoto, Komei Hasegawa, Taichi Sono, Michita Imai,
"A Simple Bi-Layered Architecture to Enhance the Liveness of a Robot",
The IEEE/RSJ International Conference on Intelligent Robots and Systems,
IROS2017, 2017

今性を示すロボット

2017



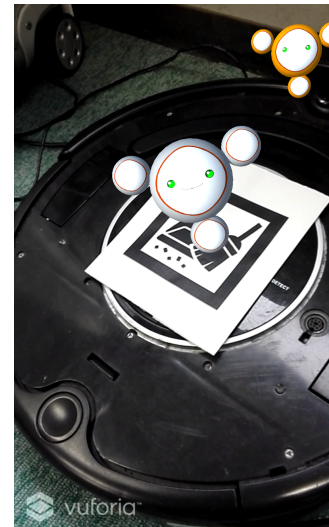
Yusuke Takimoto, Komei Hasegawa, Taichi Sono, Michita Imai,
"A Simple Bi-Layered Architecture to Enhance the Liveness of a Robot",
The IEEE/RSJ International Conference on Intelligent Robots and Systems,
IROS2017, 2017

機器のメンタライジングと人の受容性

- プッシュ型サービスに、メンタルイメージを利用する
- 自律的に振る舞うAIシステムの側から、ユーザに能動的にお知らせ（プッシュ型のお知らせ）をするにはどうしたらいいか？

機器のメンタライジングと人の受容性

- IoT機器にやどるSeirei Agent
- IoT機器毎に専用エージェントを持たせることで、**ユーザは、機器のメンタルイメージを想像でき、プッシュ型サービスに違和感なく対応できる。**
- 自律性のあるサービスには、メンタルイメージの付与が重要である。



評価実験

– 目的：

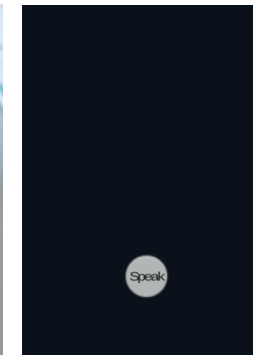
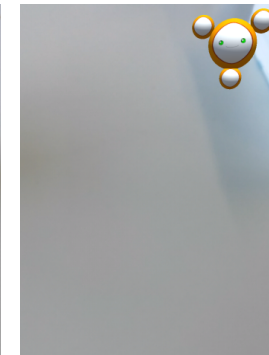
複数デバイス環境でプッシュ型インタラクションが発生する状況における SEIREI-Agent の有効性の検証

– 概要：

3種類のエージェントを使用し, 3つのデバイスに関するプッシュ通知を順次発生させる (電球, 冷蔵庫, ルンバ)

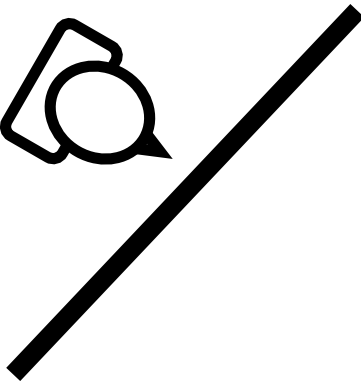
1. SEIREI-Agent
2. 単一のCGエージェント
3. 音声エージェント

– 実験参加者：22名



評価実験

実験者



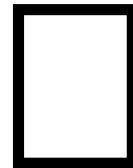
冷蔵庫

ルンバ

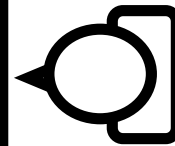
電球

タブレット

机

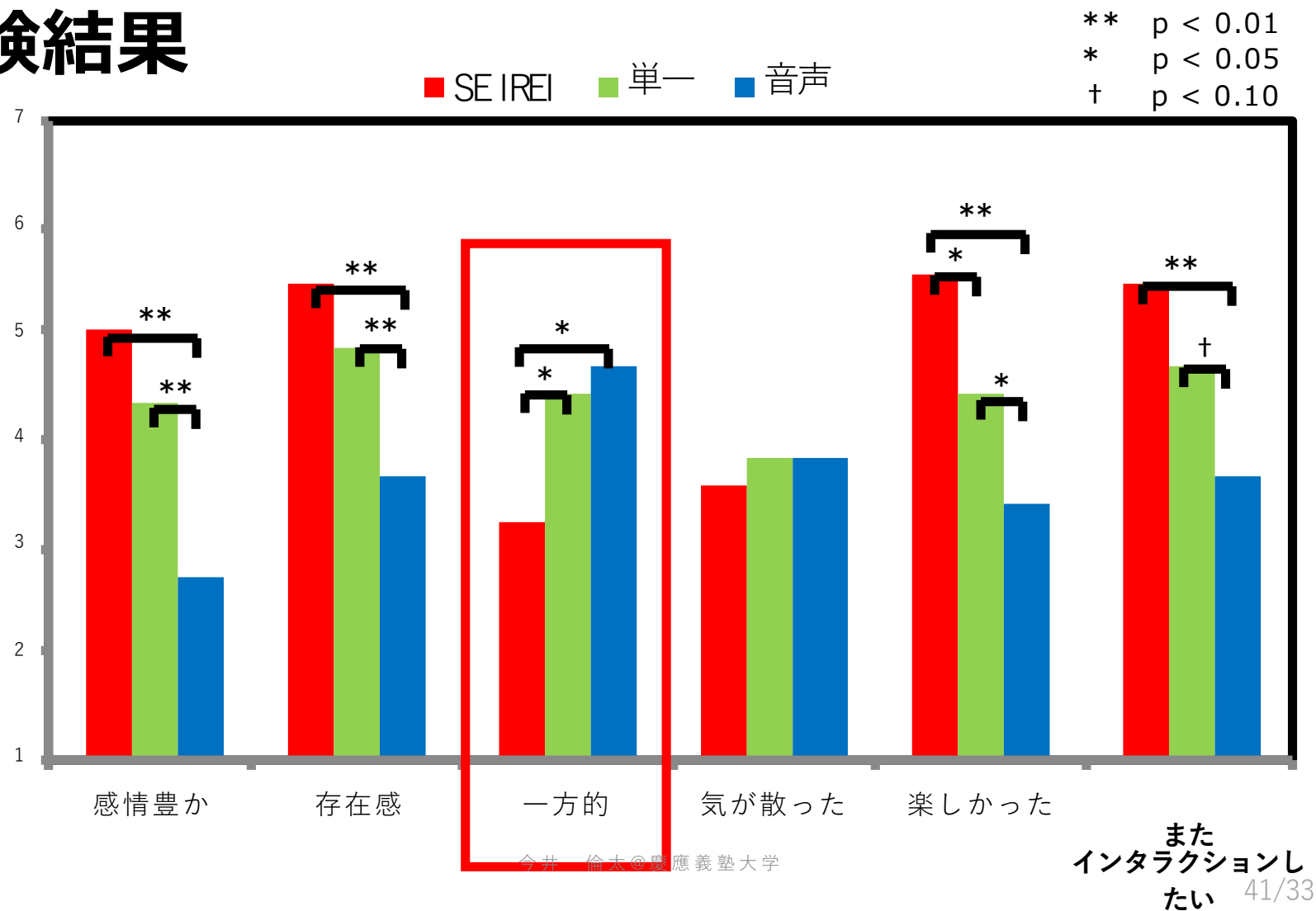


写経タスク

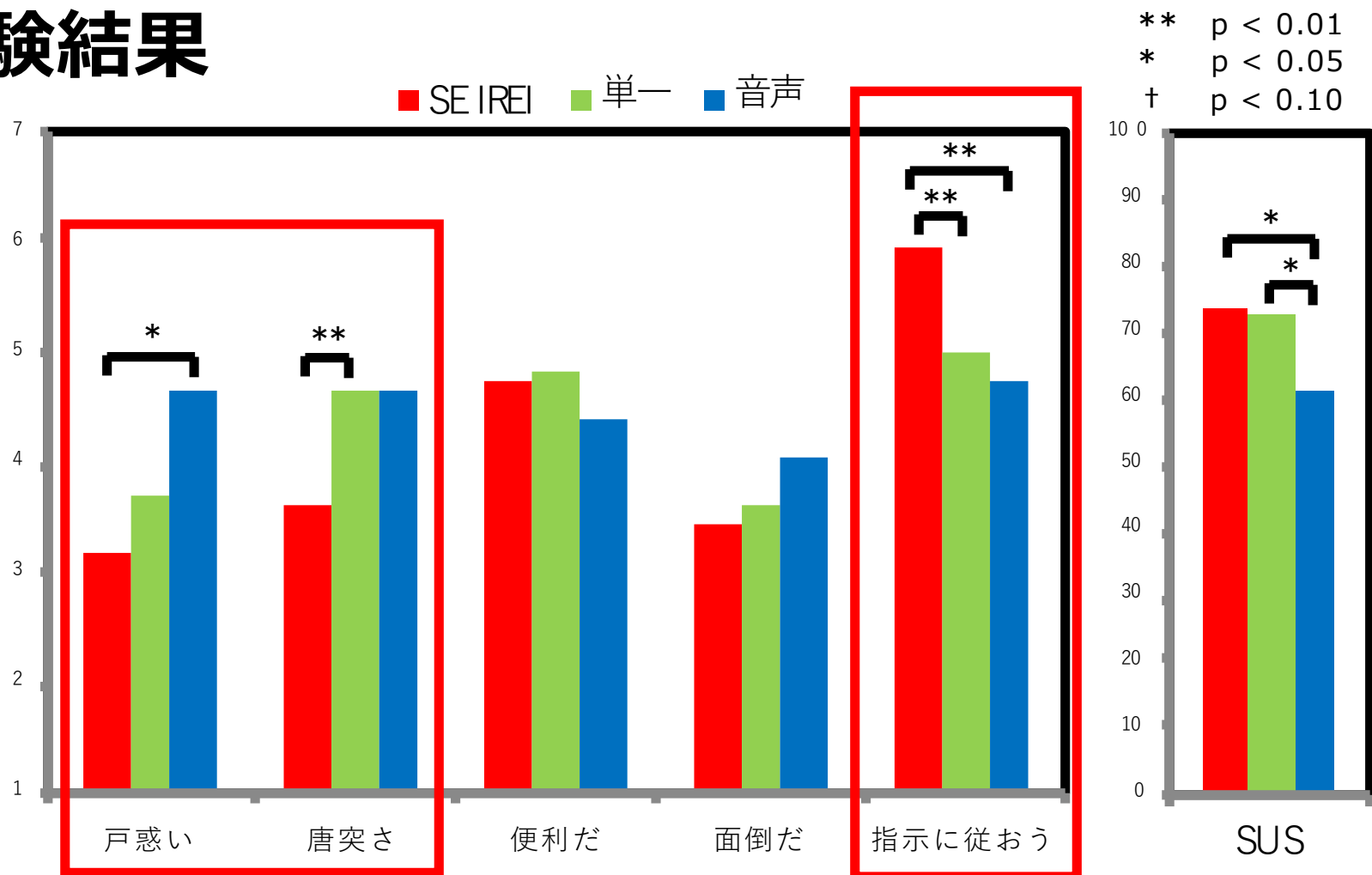


参加者

実験結果



実験結果



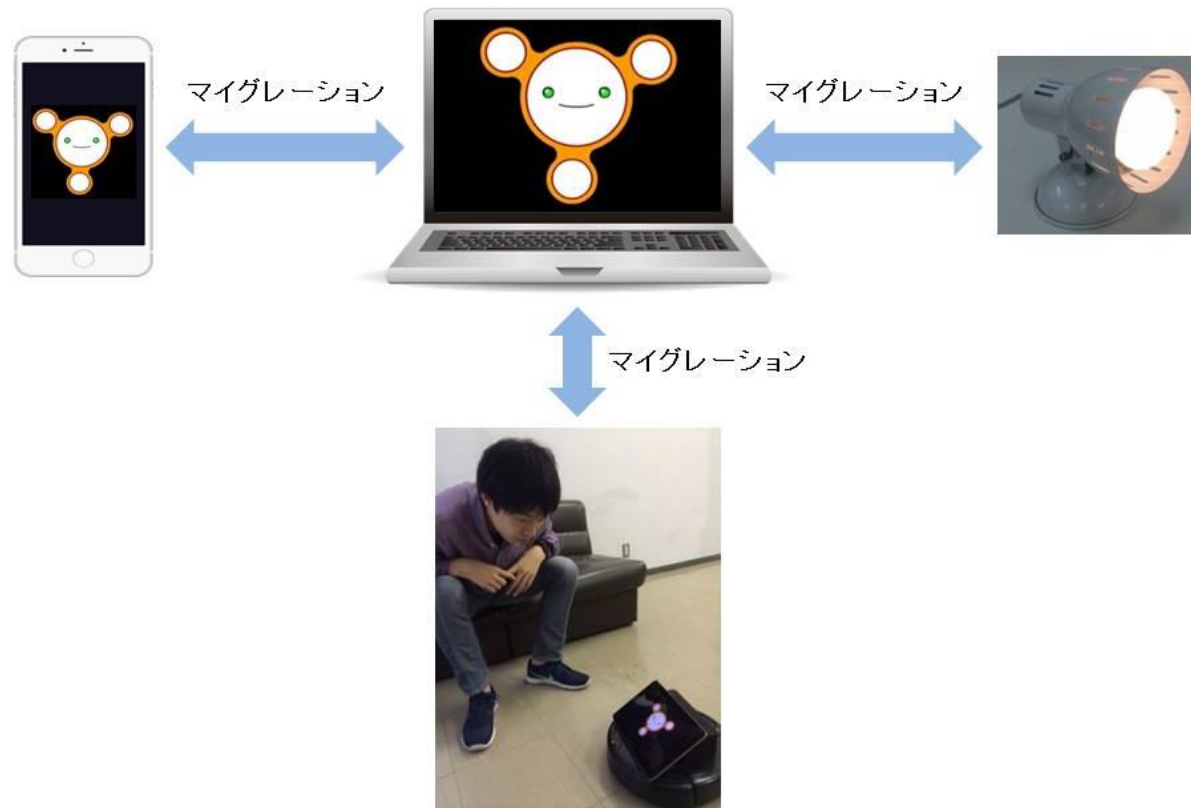
時間ボリュームのあるタスクが
メンタライジングのアプリケーションエリア

コンピュータや携帯端末で事足りることごとに
エージェントを導入するのは不毛？

道案内や情報検索、商品の購入、整理券の発券など

**時間ボリューム(時間展開)のある仕事を自律的に行う
人工知能システムが、HAIの真のターゲット**

時間ボリュームのあるタスクが メンタライジングのアプリケーションエリア



[4] Mamoru Yamanouchi, Taichi Sono, Michita Imai, "The Use of The BDI Model As Design Principle for A Migratable Agent", The 4th International Conference on Human-Agent Interaction, pp.115-122, 2016

時間ボリュームのあるタスクが メンタライジングのアプリケーションエリア

目的：時間をかけて結果を出す必要のあるタスクに向けたUIとしてのエージェントインタフェースを実現する。

アプローチ：意図的行動：複数ある未来の可能性の中から、与えられた目標が実現できることを確約できる範囲内で、目標の実現されている未来を意識的に選ぶことである。 [5]

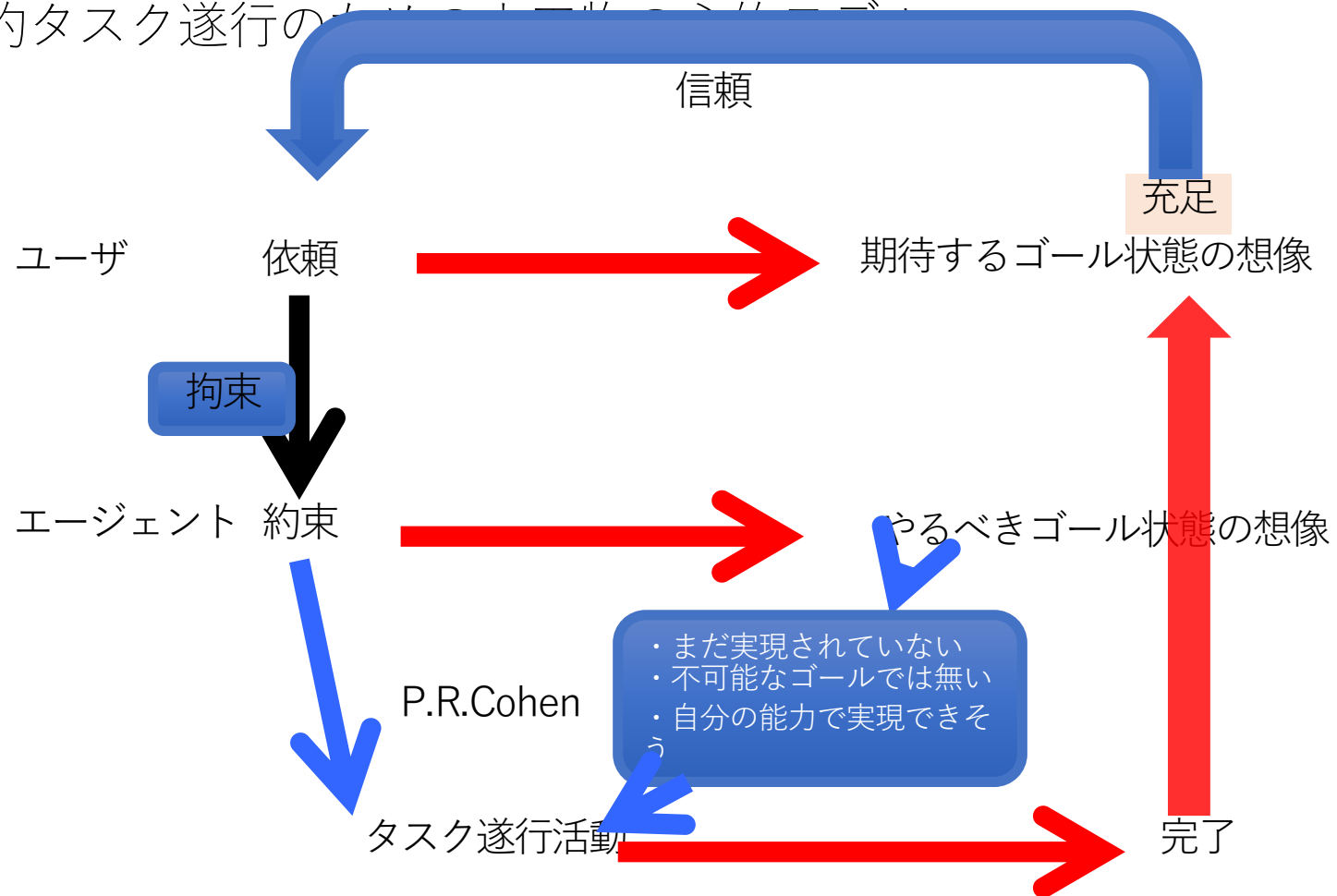
課題：ユーザが心に抱く未来（ゴール）と、エージェントが意図的に行動する過程の中で想定する未来の差異を考慮に入れたインタラクションデザインをすることで、協調や共生に向けた信頼関係を築く。

[5] P.R.Cohen, H.J.Levesque, "Intention is Choice with Commitment," Artificial Intelligence, Vol.42, issues 2-3, pp. 213-261, 1990.

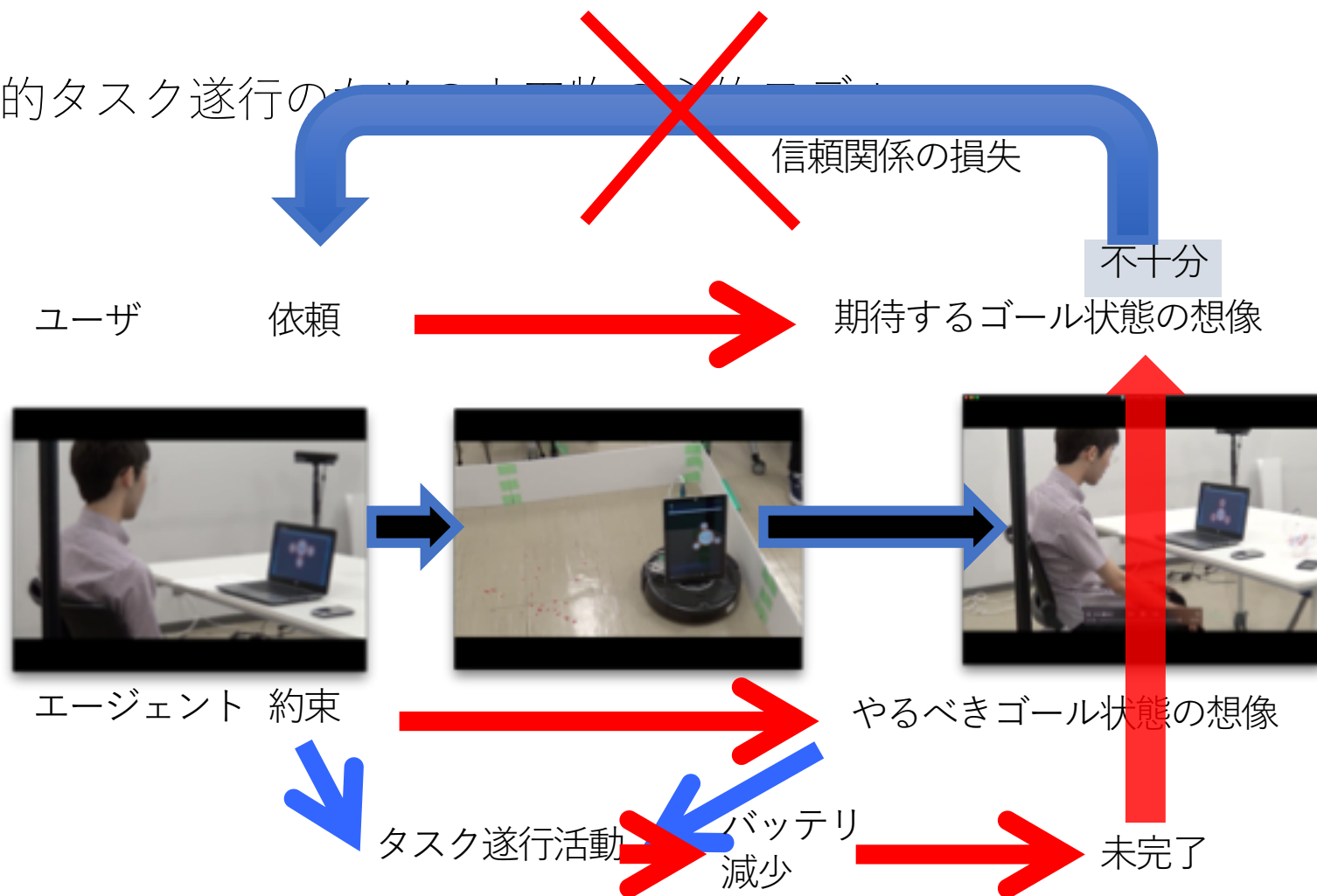
適応的タスク遂行のメタ認知モデル



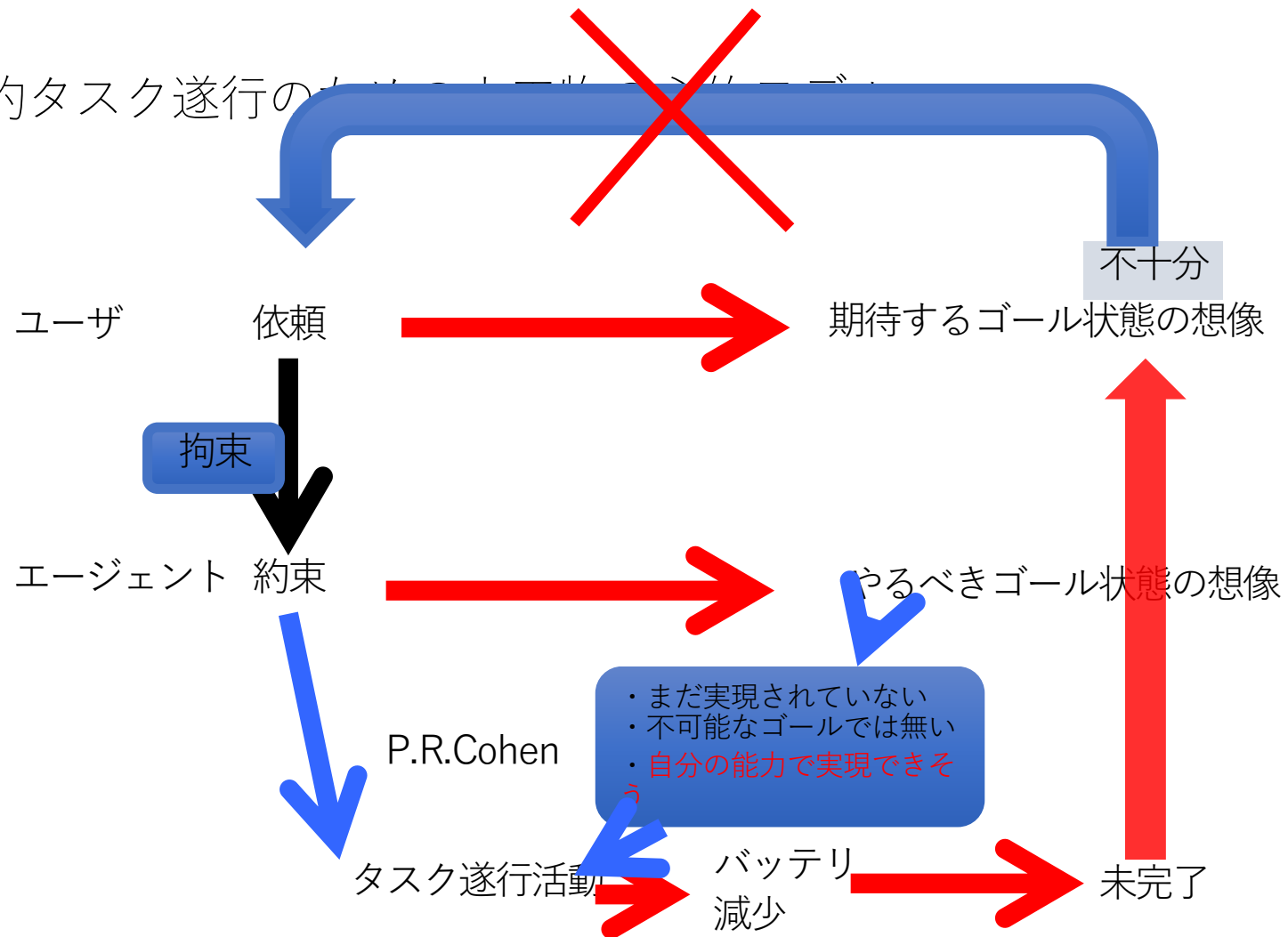
適応的タスク遂行のメタ認知モデル



適応的タスク遂行の失敗モード



適応的タスク遂行のモデル

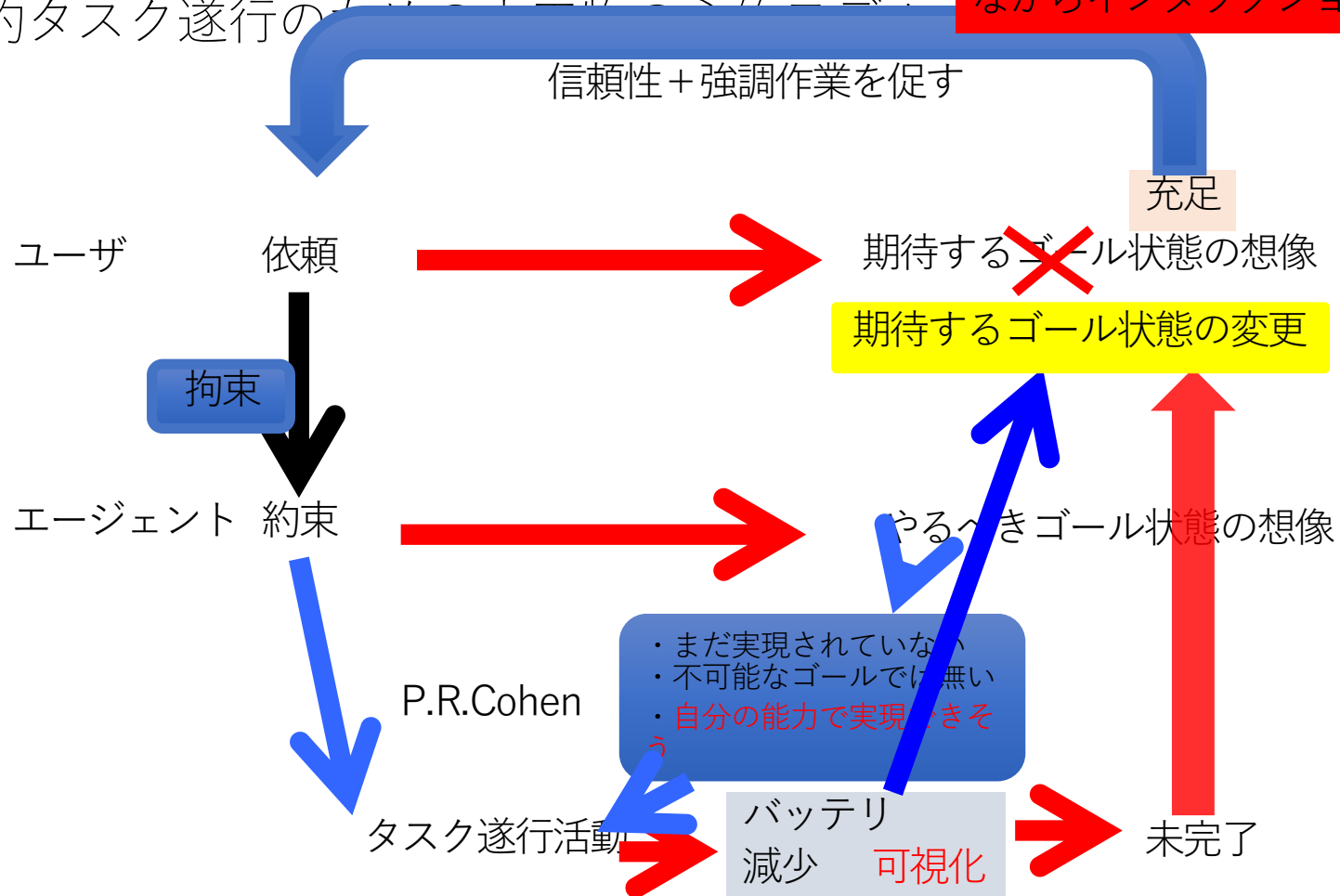


適応的タスク遂行のためのヒューマン・ロボット・システム



適応的タスク遂行のメタ認知メカニズム

ユーザが抱く世界とエージェントが想定する世界をすり合わせながらインタラクションする。



人からどう思われているかをロボットが 知ること

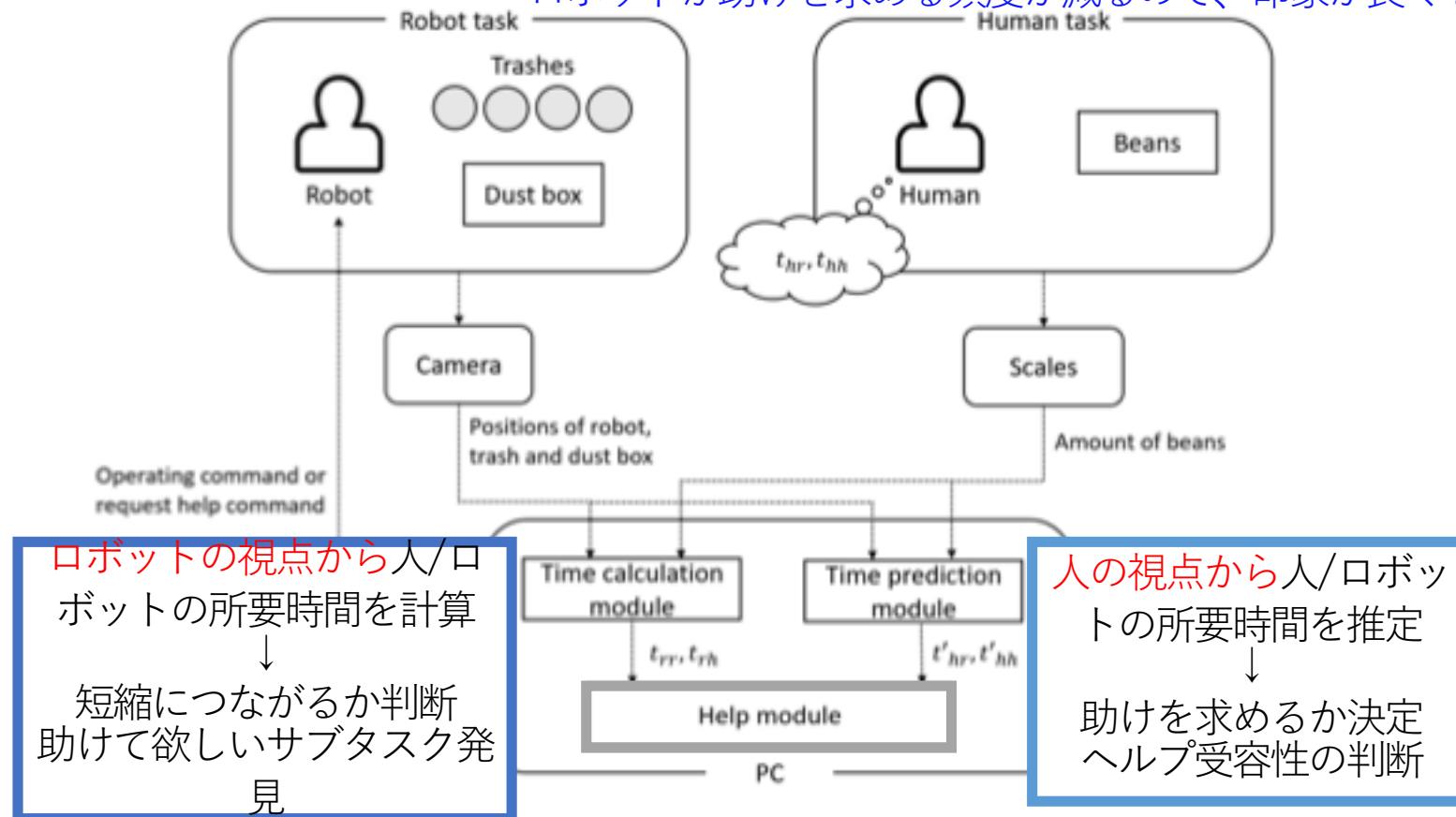
- 人工知能システムと人が助け合って作業するには、どのように
認知的内部状態を扱ったらいいか？
 - ロボットが助けを求める場面において

Help-Estimator

ロボットの視点から助けてもらいたい仕事を選んで、
人の視点から助けを言うべきか（人が助けてくれそうか）判断する。

ロボットは助けてもらいやすくなる。

ロボットが助けを求める頻度が減るので、印象が良くなる。

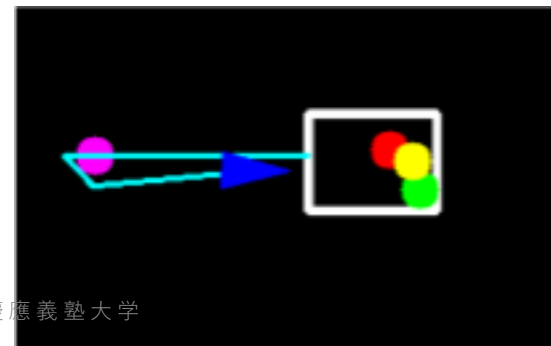
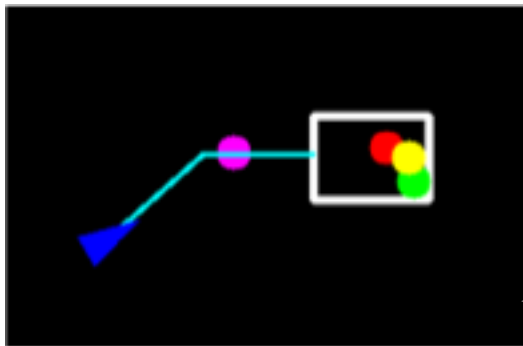


人の所要時間 t_{rh} の計算(ロボット視点)

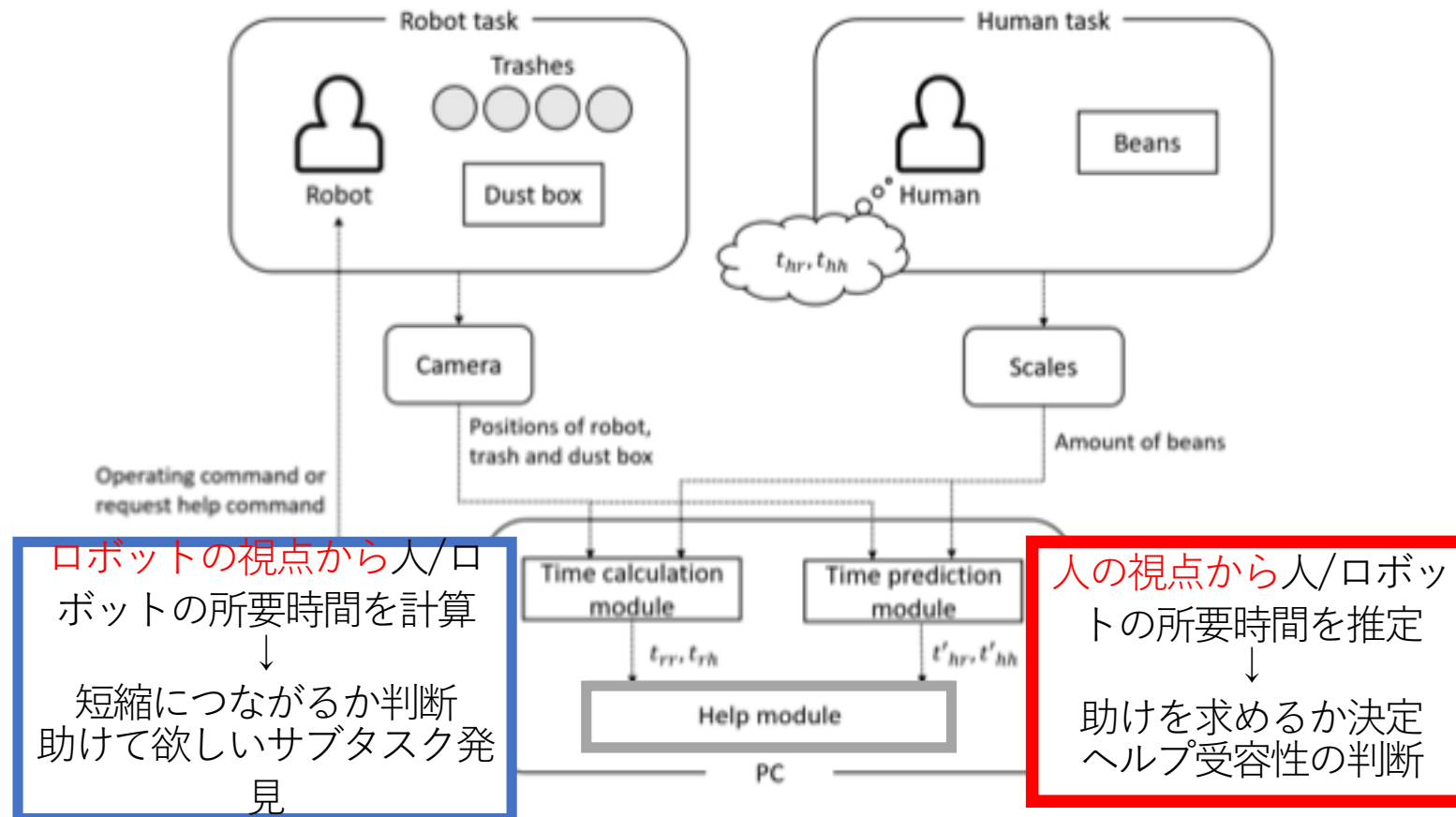
- 小豆を拾った個数 n と小豆を拾う速度 $pace$ で計算
- t 秒間で n 個の小豆を拾ったとすると
 - $pace = n/t$
- 小豆の総量を N とすると、残りの所要時間は
 - $t_{rh} = (N - n)/pace$

ロボットの所要時間 t_{rr} の計算（ロボット視点）

- 紙ゴミまでの経路を決定
 - ロボットとゴミの位置から独自のアルゴリズムで決定
- 経路の総距離 D ，回転角 Φ を元に所要時間を計算
 - $t_{rr} = D/v + \Phi/\omega$
 - (v ：ロボットの速度， ω ：ロボットの角速度)



Help-Estimator

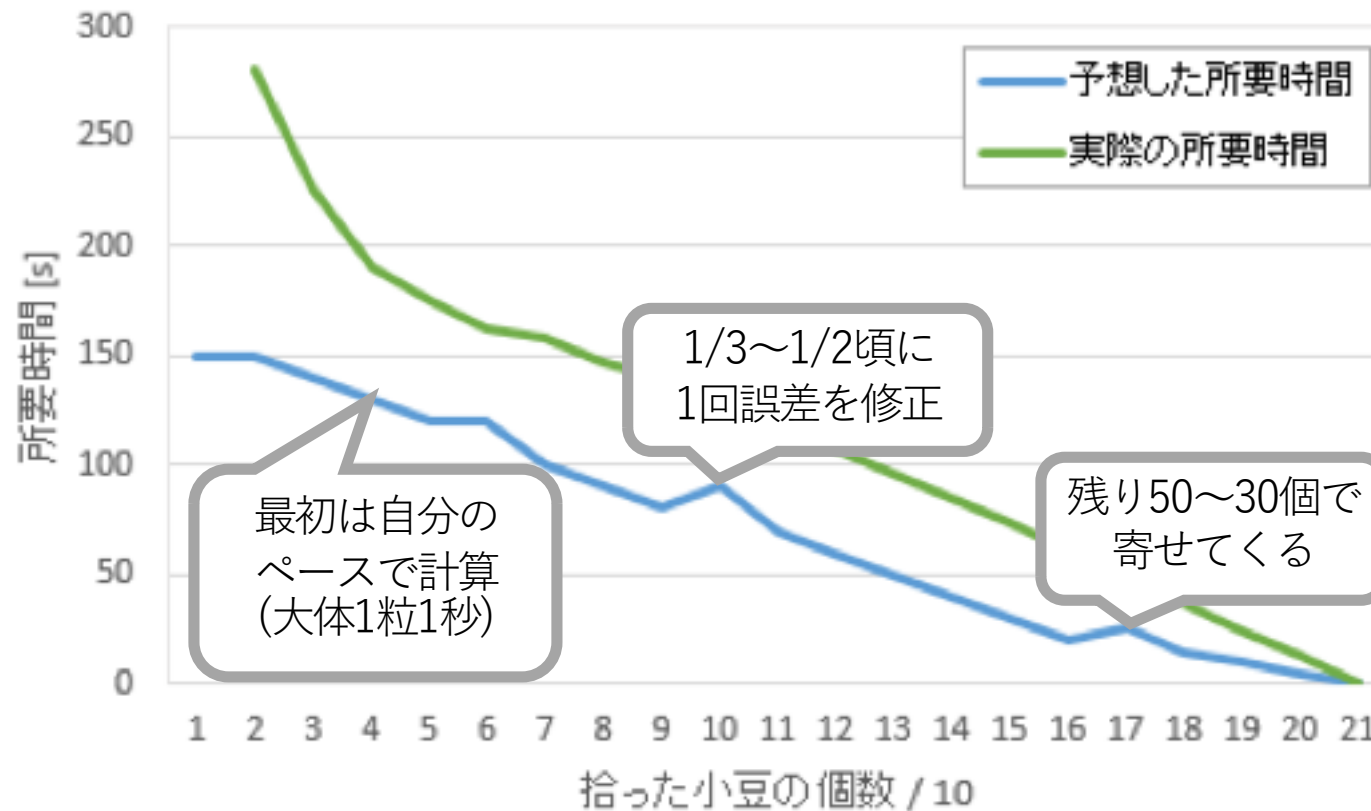


人のタスクの所要時間の予想(人の視点)

- 小豆を200粒拾ってもらう
 - タスク開始前に所要時間について予想してもらう
 - 1粒ずつ拾い，10粒拾うごとに残りの所要時間を予想
- 8人の実験参加者に対して実施



人のタスクの所要時間の予想(人の視点)



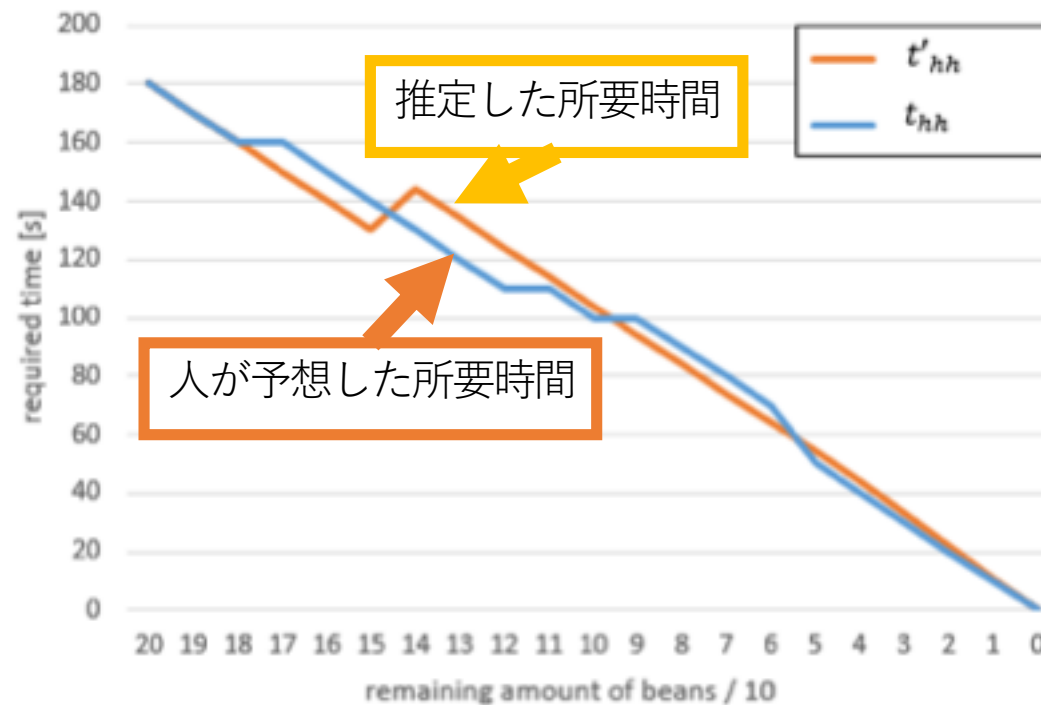
人のタスクの所要時間の予想(人の視点)

- 知見をもとに関数を作成

$$\bullet \text{ } pred_{t_{hh}}(n) = \begin{cases} t_H - v_h(N - n) & n > p \\ \frac{t_H - v_h(N - p) - t_{rh}}{2} + t_{rh} & n = p \\ pred_{t_{hh}}(p) - v_h(p - n) & 30 < n < p \\ n \cdot pred_{t_{hh}}(30)/30 & n < 30 \end{cases}$$

- t_H : 最初に人に予想してもらった所要時間
- v_h : 小豆を1粒を拾うのにかかる時間(今回は1秒/粒)
- p : 修正を加えるタイミング(今回は1/3の時点)

人のタスクの所要時間の予想(人の視点)



誤差：±25.57秒，相関：0.947 で推定

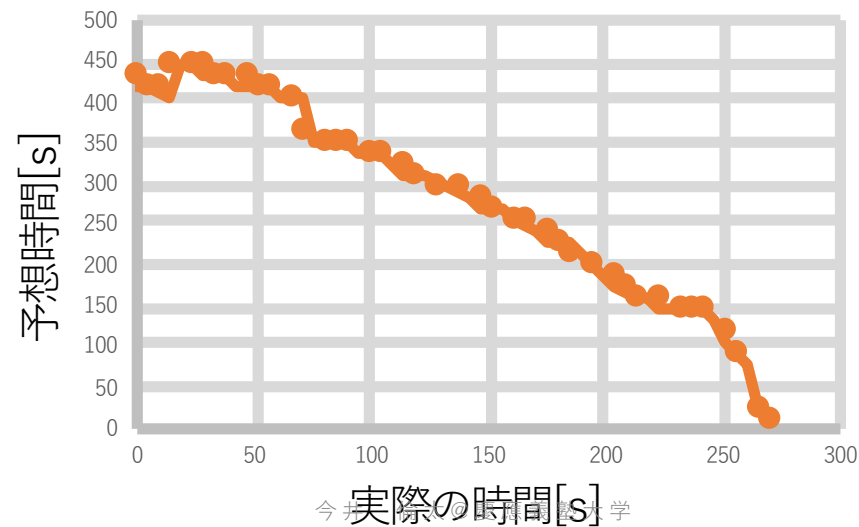
ロボットのタスクの所要時間の予想(人の視点)

- ロボットが紙ゴミをゴミ箱に入れる時間を予想
 - ロボットはWoZで動かし、4歩で1回停止(1step)
 - 1stepごとに、残りの所要時間について予想
- 5人の実験参加者に
- 対して実施



ロボットのタスクの所要時間の予想(人の視点)

- 全部で何stepかかるかで予想
 - 順調なら1stepあたりの所要時間を引いていく
 - ゴミが変なところに転がるなどした場合は増やす

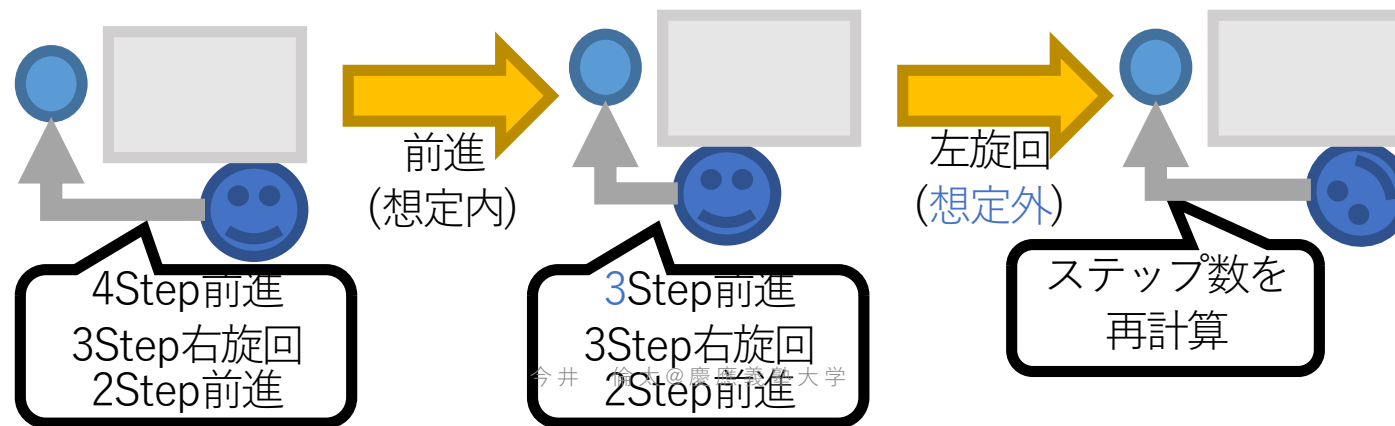


ロボットのタスクの所要時間の予想(人の視点)

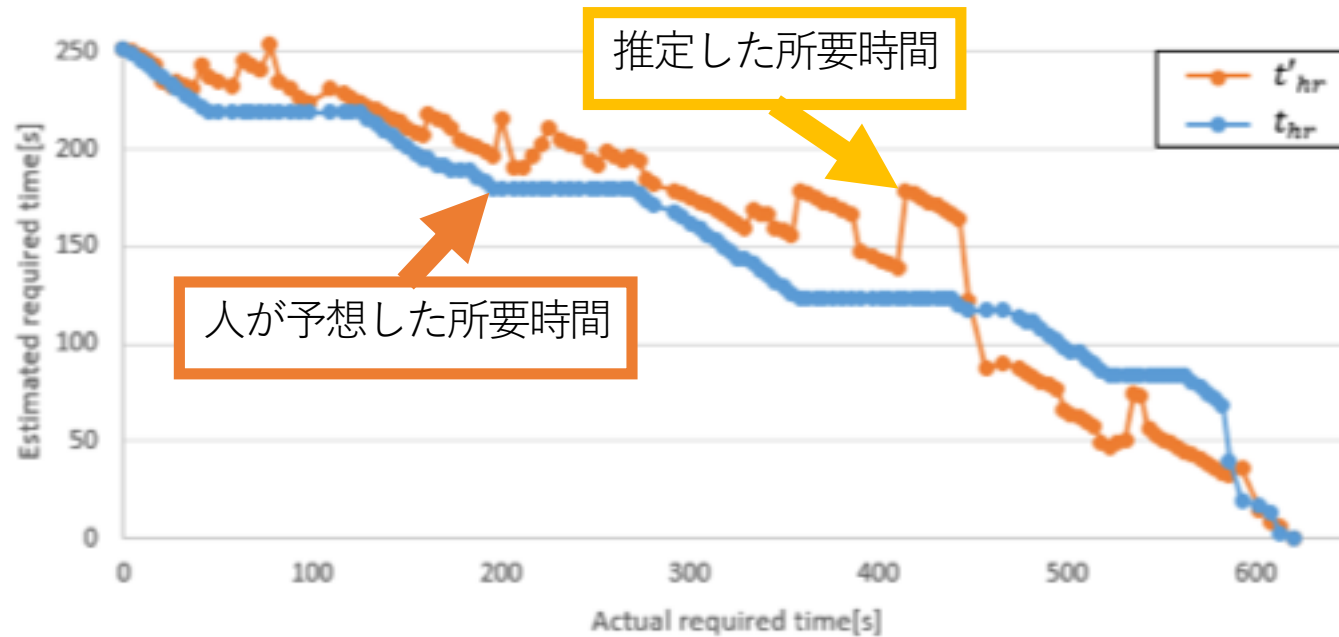
- 経路をもとにステップ数 $Step$ を計算
 - 想定内のコマンドが入力→ $Step$ をデクリメント
 - 想定外のコマンドが入力→ステップ数を再計算

$$\bullet t'_{hr} = \frac{\text{最初の予想時間}}{\text{タスク開始時のステップ数}} \times Step$$

主観的時間見積もり



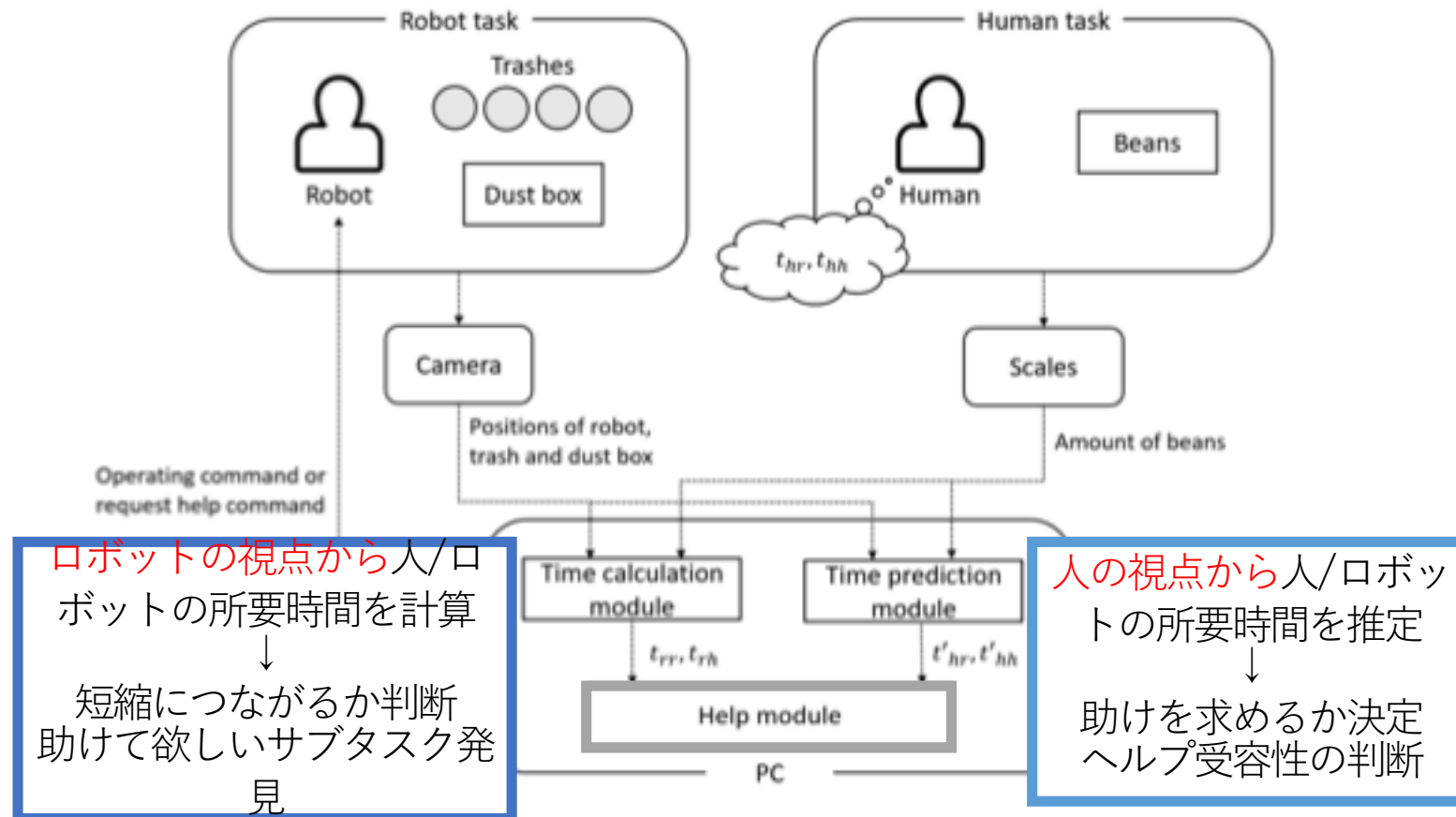
ロボットのタスクの所要時間の予想(人の視点)



誤差： ± 22.22 秒，相関：0.942 で推定

Help-Estimator

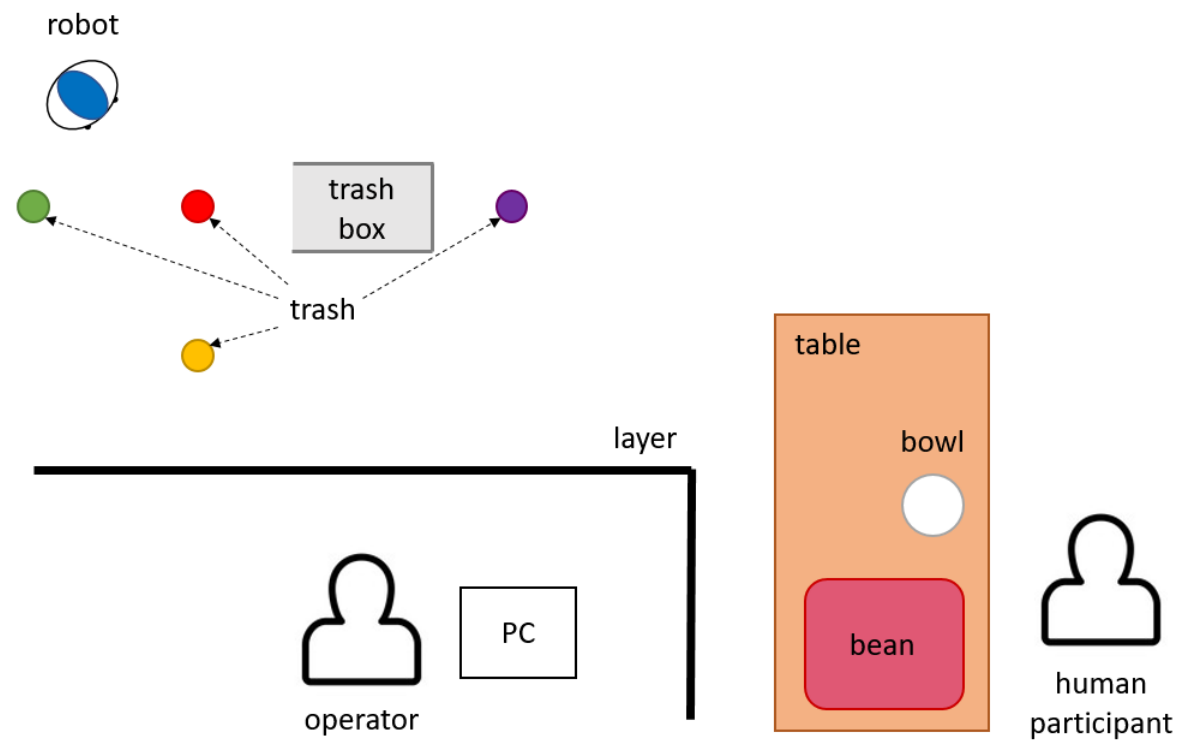
ロボットの視点から助けてもらいたい仕事を選んで、
人の視点から助けを言うべきか（人が助けてくれそうか）判断する。



実験条件

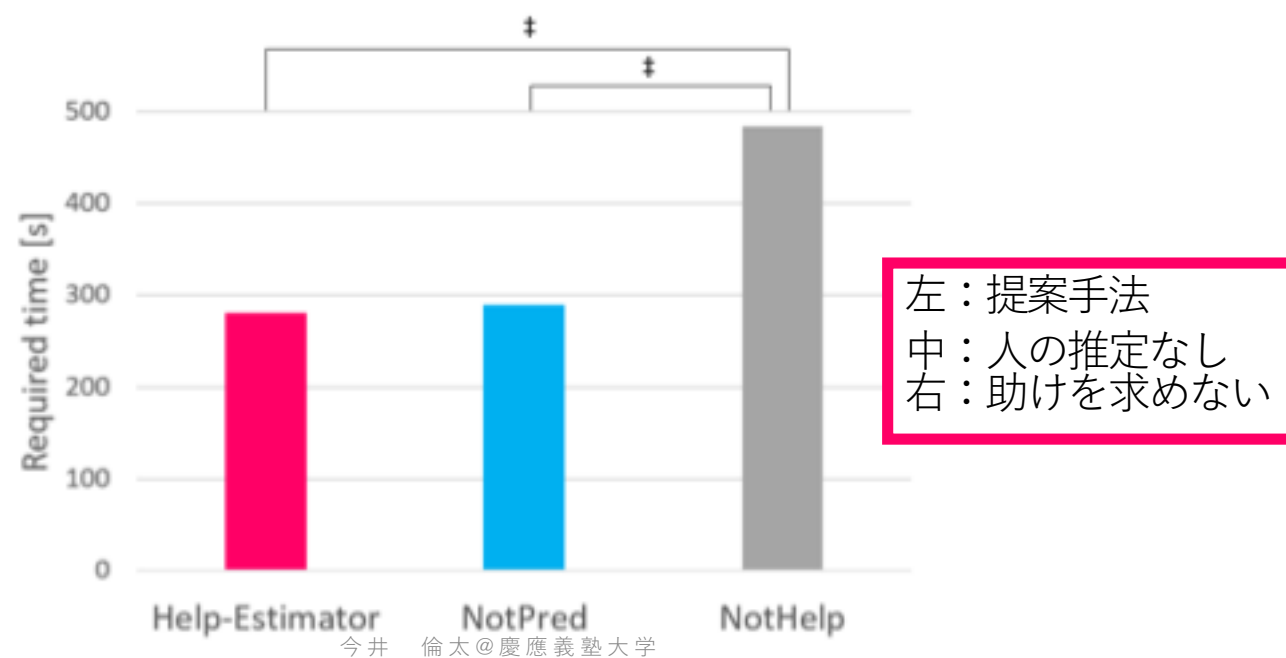
- Help-Estimator(提案手法)
 - NotPred(所要時間が短縮されるかどうかでHelp)
 - NotHelp(助けを求めない)
-
- 目的
 - ロボットのHelpにより，所要時間が短縮されるか
 - Help-EstimatorとNotHelpを比較
 - 人の予想所要時間を推定した方が助けてもらえるか
 - Help-EstimatorとNotPredを比較

実験



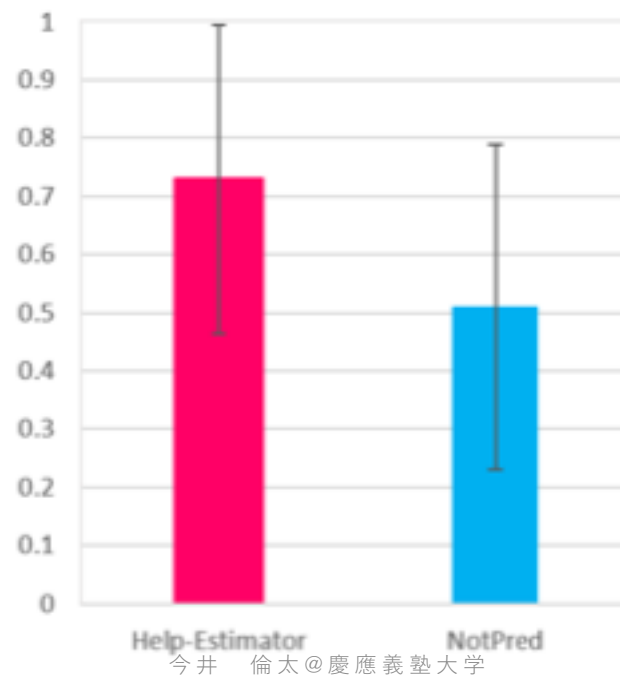
協調タスクの所要時間

- 助けを求めることで**所要時間が短縮**
 - ANOVAおよびTukey-Kramer法で検定($p < 0.005$)



助けてもらった割合

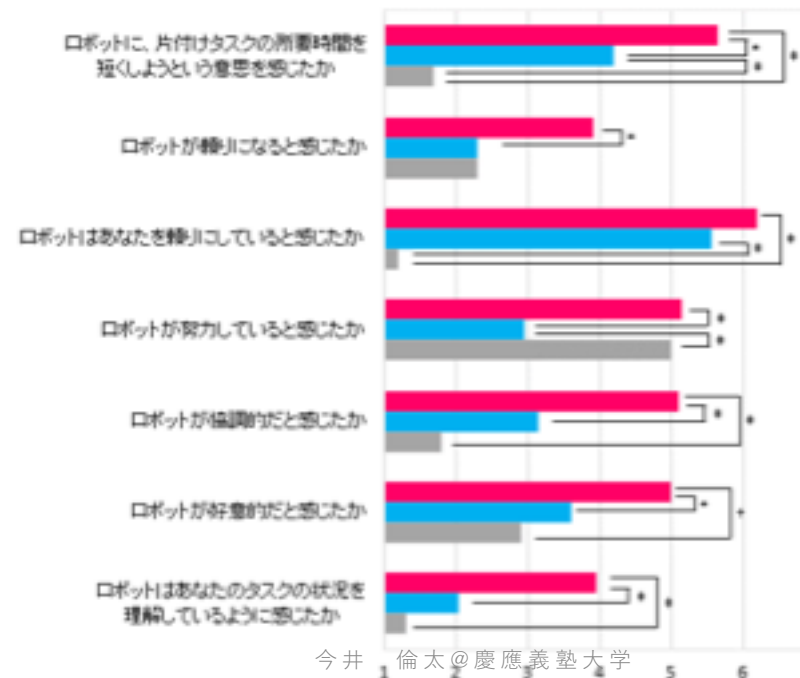
- 人が予想した所要時間を推定することで人から助けてもらえる割合が改善
- 有意差($p < 0.05$)



左：提案手法
右：人の推定なし

アンケート評価

- Help-Estimatorが総じて高評価
 - ANOVAおよびTukey-Kramer法で検定



上：提案手法
中：人の推定なし
下：助けを求めない

ロボットが人を思い計ると人は
ロボットとのコミュニケーションに没入する

- 人間の五感を利用した演出

- 同じ状況に対して同じ感覚を持たせる

- 二つの演出ステップ

ロボットが人を思い計ると人は
ロボットとのコミュニケーションに没入する

• 人間の五感を利用した演出

- 同じ状況に対して同じ感覚を持たせる

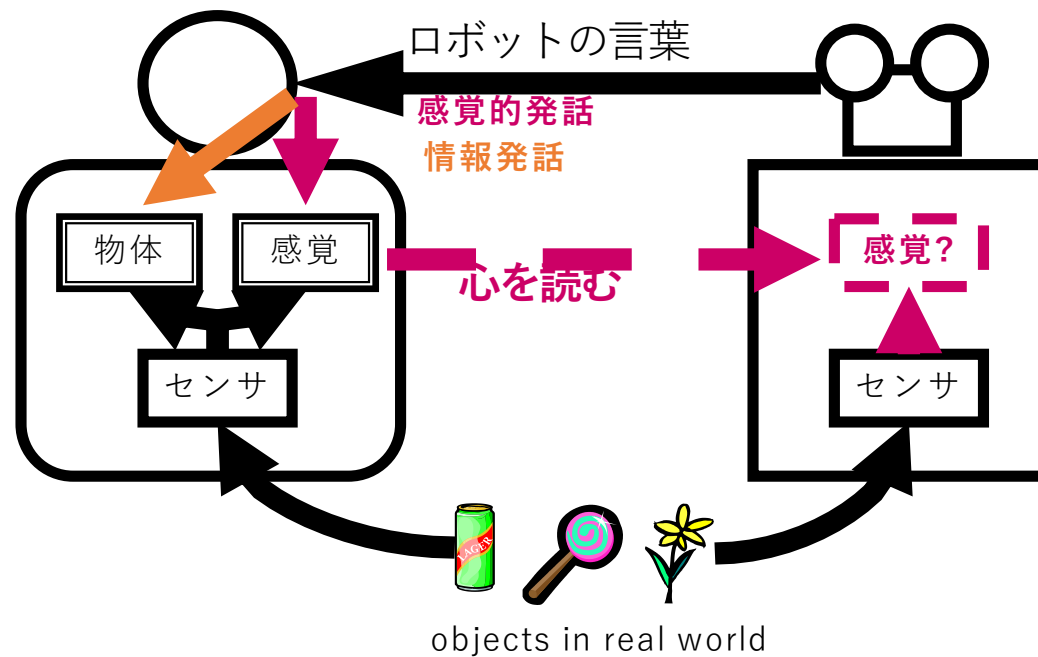


ロボットが人を思い計ると人は
ロボットとのコミュニケーションに没入する

- 比較実験



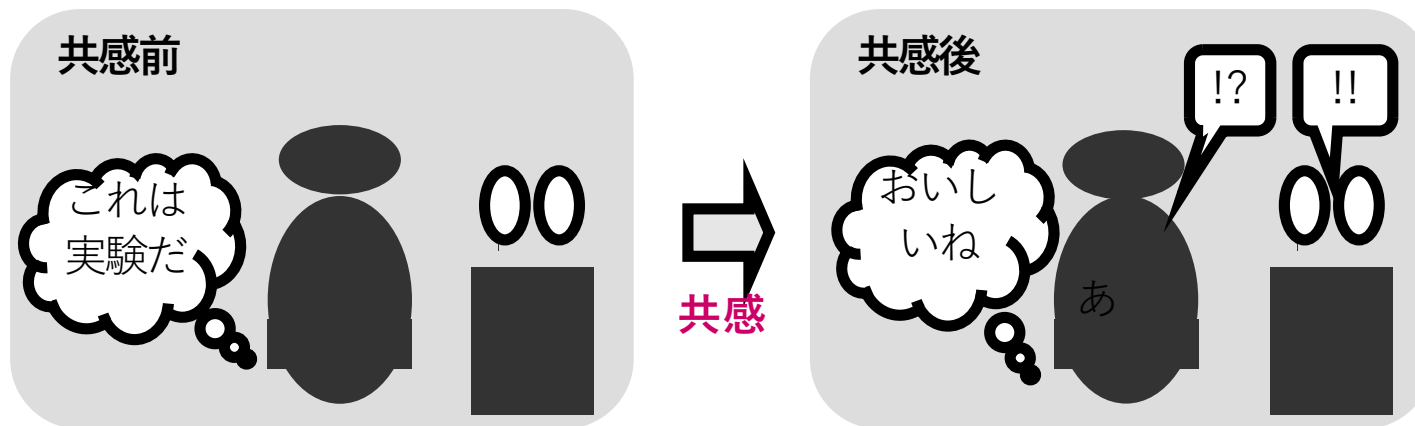
ロボットが人を思い計ると人は
ロボットとのコミュニケーションに没入する
• 感覚的発話による共感のモデル



ロボットが人を思い計ると人は
ロボットとのコミュニケーションに没入する

・実験仮説

ロボットからの感覚的発話によって「ロボットと共感した」と感じた
被験者は、**ロボットとのコミュニケーションを作為的にとらず
コミュニケーション自体に集中できる**



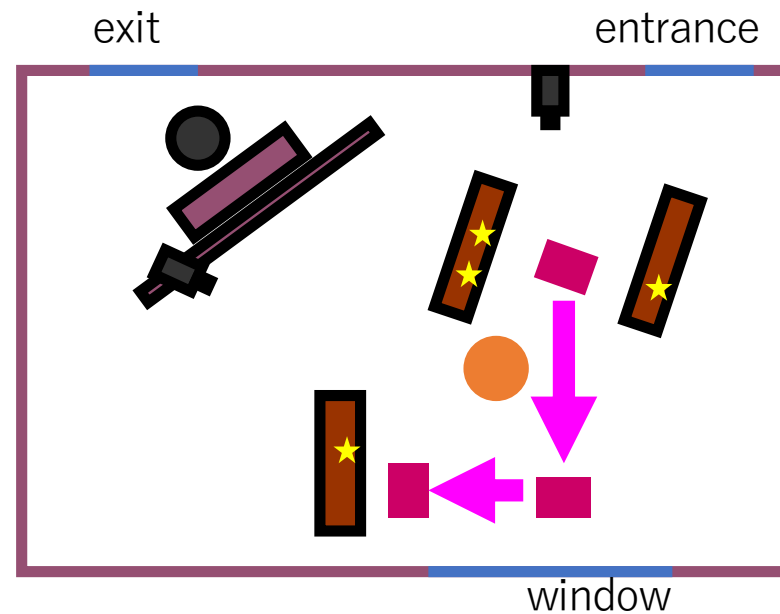
ロボットが人を思い計ると人は

ロボットとのコミュニケーションに没入する

- 被験者：男女の大学生21名を、実験群11名・対照群10名に

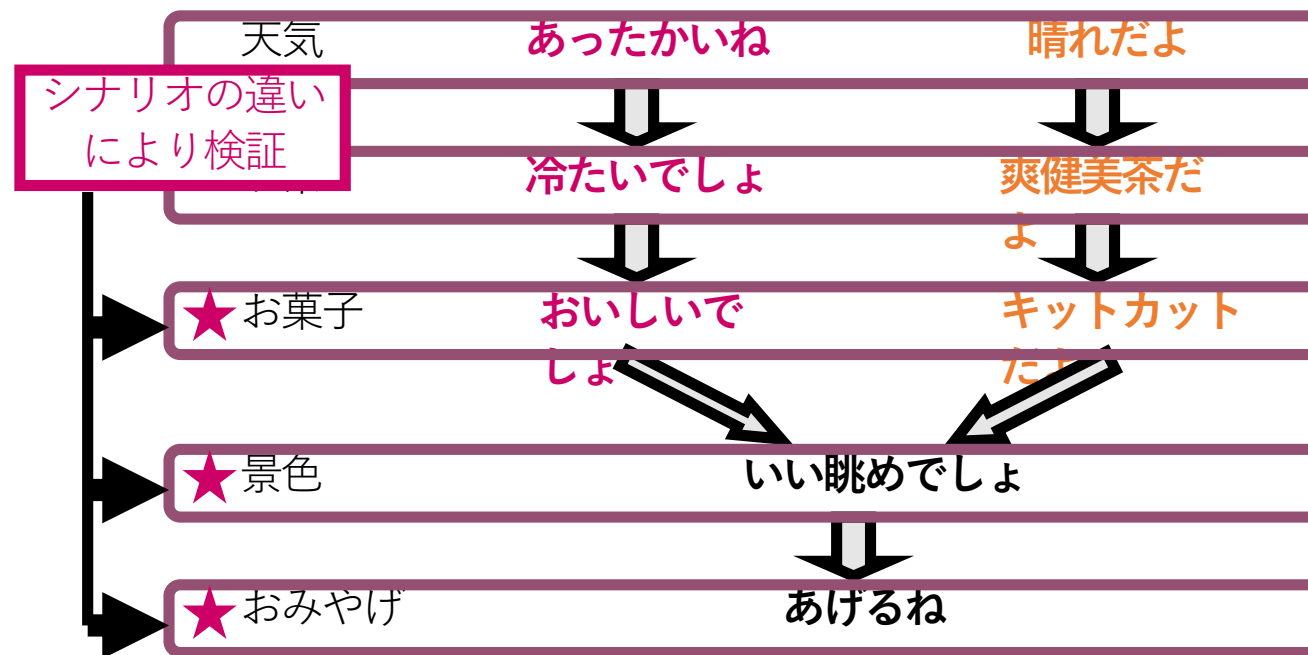
ランダムに配分

- 実験環境：大学の教室内にロボットと小道具などを配置



ロボットが人を思い計ると人は ロボットとのコミュニケーションに没入する

- シナリオ 実験群(感覚的) 対照群(情報伝達)



ロボットが人を思い計ると人は
ロボットとのコミュニケーションに没入する

- お菓子を食べる実験群の被験者



R: これ、あげるね
H: どうも
ありがとう
(お菓子を取る)
R: 食べてみてね
H: はい
(食べる)
R: おいしいでしょ
H: うん、おいしい
R: よかった

ロボットが人を思い計ると人は ロボットとのコミュニケーションに没入する

- お菓子を食べない対照群の被験者



H: (お菓子を取る)
R: これ、あげるね
H: …ありがとう
R: 食べてみてね
H: …
 (お菓子を見る)
R: キットカットだよ
H: …
 (お菓子を見る)

ロボットが人を思い計ると人は
ロボットとのコミュニケーションに没入する

- 景色に関する発話に返答する実験群の被験者



R: (窓に近づく)
H: (窓に近づく)
(外を見る)
R: わあ
いい眺めでしょ
H: そうですね

ロボットが人を思い計ると人は ロボットとのコミュニケーションに没入する

- 景色に関する発話に返答しない対照群の被験者



R: (窓に近づく)
H: (Robovieの
後ろから
ついていく)
R: わあ
H: (外を見る)
R: いい眺めでしょ
H: (外を見る)
...

ロボットが人を思い計ると人は ロボットとのコミュニケーションに没入する

• おみやげを持って帰る実験群の被験者



R: (回転して)
あ、そうだ
おみやげがあるんだ
H: おみやげがあるの？
R: どっちな好きな方を
取ってね
H: ほんとに？
R: Robovieは赤が
好きなんだ
H: じゃあ私も赤が好き
だから赤をもらっても
いい？(赤を取る)
なんだろう
R: 今日は来てくれて
ありがとう
H: こちらこそ、ありがとう
R: あっちが出口だよ
H: あ、ほんとに？
ありがとう、
ごちそうさまでした
R: ばいばい！
H: ばいばい！ Robovie
ばいばい！(退出)

ロボットが人を思い計ると人は ロボットとのコミュニケーションに没入する

- おみやげを持って帰らない対照群の被験者



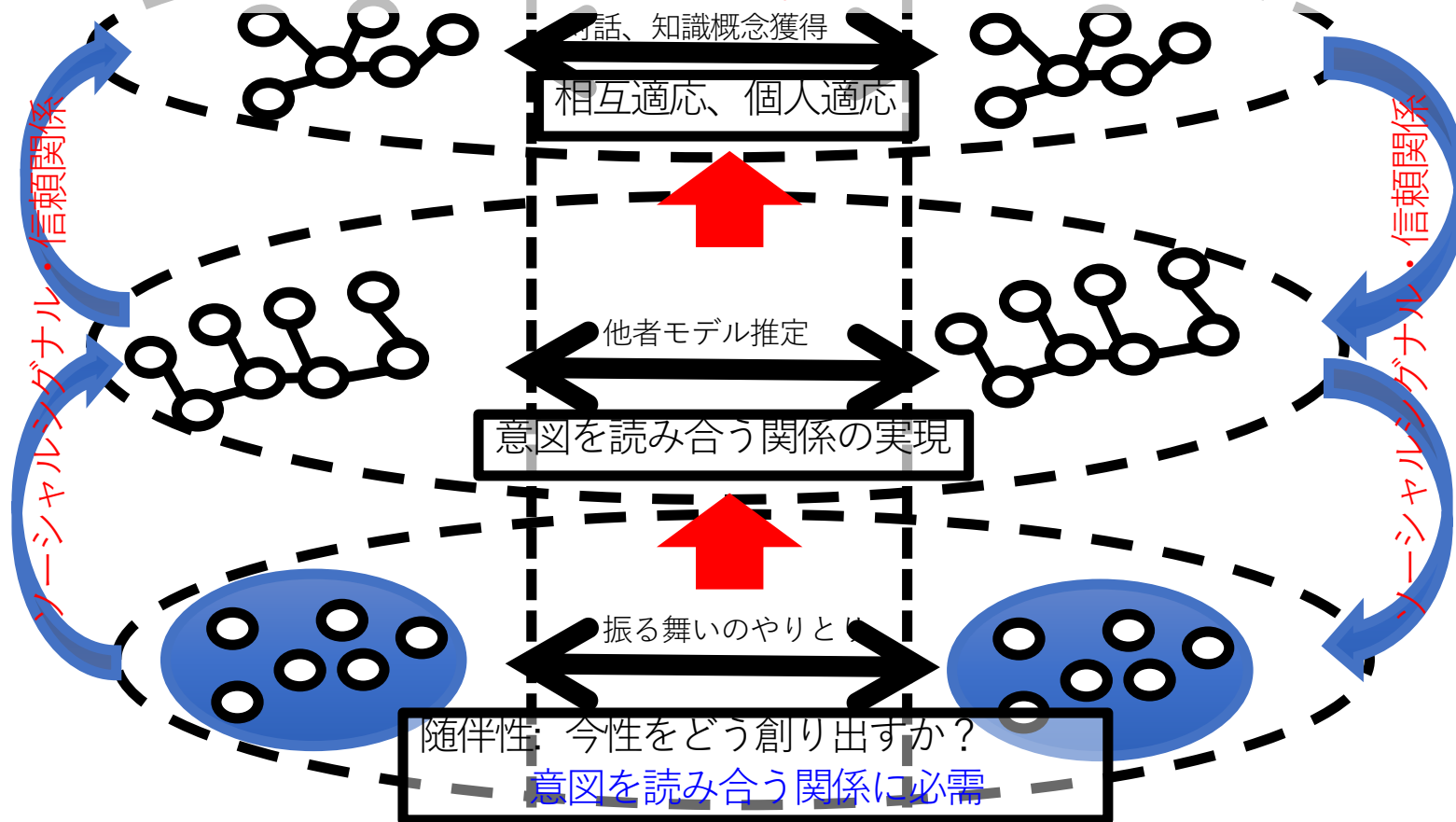
R: (回転して)
あ、そうだ
おみやげがあるんだ
H: (青を手にする)
R: どっちか好きな方を
取ってね
H: ...
(青を観察)
R: Robovieは赤が
好きなんだ
H: じゃあ俺は青で
R: 今日は来てくれて
ありがとう
あっちが出口だよ
H: ...
R: ばいばい！
H: ばいばい
(青を机に戻す)
(退出)

ロボットが人を思い計ると人は ロボットとのコミュニケーションに没入する

- 設問： あなたはおみやげを持って部屋を出ましたか？ (yes / no)
また、そうした理由を教えてください。
- おみやげを持って帰らなかった被験者の理由
 - 「実験だから」
 - 「持って帰っていいのかわからなかったから」
- おみやげを持って帰った被験者の理由
 - 「せっかくもらったので」
 - 「おみやげだから」
- おみやげを持って帰らなかった被験者は、「実験である」ことを意識していた

エージェントに必要なインタラクションデザイン

目標：人とインタラクションする自律した人工物のデザイン要件を
このモデルの下で発見・整理する。



そして未来へ

- 僕らは、自律的に振る舞い、意図を持つように見えるものを擬人化して、相手の認知状態を想定しながら理解する。
- 自律的に振る舞い人工知能システムを人と繋げるためには、擬人化可能なエージェントをインタフェースとして使うのが望ましい。
 - かならずしも、擬人化＝人のような姿 である必要はない。
- 子供やお年寄りに対しては、物理的身体を持つロボットはとても有効であるが、成人に対しては、かならずしも物理的身体は必要じゃない可能性が大きい。
- 以上を加味した、人工知能システムのためのユーザインタフェースデザインを確立していくことが重要である。

本日の目標

- (ロボットを始めとする) エージェントシステムの存在意義
 - 人に**意図スタンス**を取らせることができる
 - エージェントの意図に基づく行動と解釈させることで、AIシステムからのサービスを人に効率良く把握させる。
 - 感情、注意、関係性、社会性、他者モデル、認知バイアスが影で働いている
- エージェントシステムに必要な設計論
 - 今性
 - 擬人化とメンタライジング
 - 意図を読み合うメカニズム