Action understanding as inverse planning

Chris L. Baker, Rebecca Saxe, Joshua B. Tenenbaum

Cognition, Vol. 133, No. 3, pp. 329−349, 2009

1. Introduction

- 社会的なやりとりは，相手がある行動を行うに至った心的状態の理解・予測の上に成り立っている
  - 心理状態　信念 (beliefs)，欲求 (desires)，など
  - “欲求を最も効率的に達成できる行動を選択する”という合理性原則に基づく (Fig. 1a; Dennett, 1987; Gopnik & Meltzoff, 1997; Perner, 1991; Wellman, 1990)

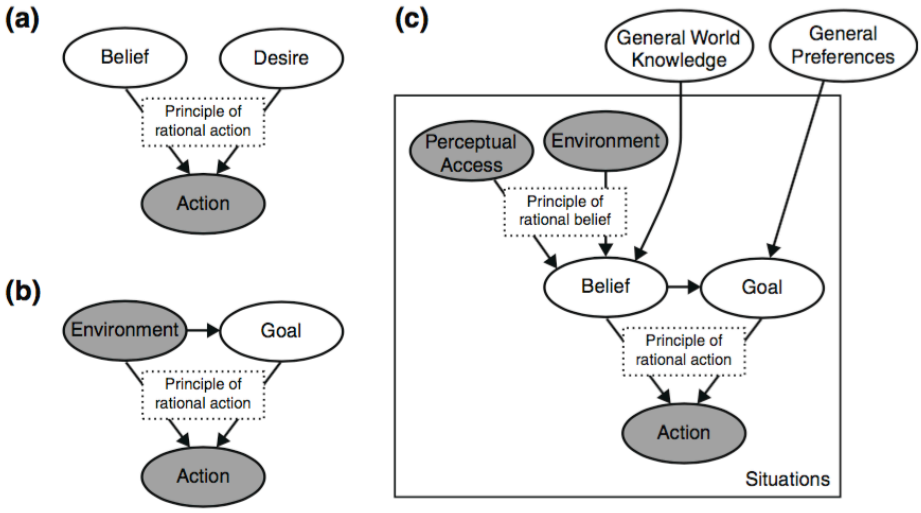


Fig. 1. Modeling intuitive theories of intentional action. Diagrams use causal graph notation. Shaded nodes represent observed variables; unshaded nodes represent latent variables whose values must be inferred. Dotted boxes indicate the causal relation between variables. In this example, the observations are an agent's Action in some Environment and the agent's Perceptual Access to the value of the Environment. Given this evidence, the agent's Belief and Goal must be inferred. (a) Non-formal "belief-desire psychology" account of folk theories of intentional action. Non-formal accounts of human action understanding typically assume some version of this causal structure, and define the causal relation between beliefs, desires and actions in terms of qualitative, context-specific commonsense rules (*e.g.* Fodor, 1992; Wellman and Bartsch, 1988). (b) A simplified version of (a) proposed by Gergely et al. (1995) and Csibra and Gergely (1997) as a model of infants' early-developing action understanding competency. We formalize this intuitive theory as Bayesian inverse planning. The functional form of the causal relation between Environment, Goal and Action is given by rational probabilistic planning in Markov decision problems, and goal inference is performed by Bayesian inversion of this model of planning. (c) A generalized framework for human action understanding. This relates the non-formal account in (a) to the formal model in (b) by sketching how agents' Beliefs and Desires (or Goals) depend on their Perceptual Access to the Environment, mediated by their General World Knowledge and General Preferences. The model in (b) is a limiting case of (c), in which agents are assumed to have complete Perceptual Access to the Environment, constraining their Beliefs to be equal to the Environment.

- 行動理解は，逆プランニング (inverse planning) や逆強化学習 (inverse reinforcement learning) の一種と見なされている (Ng & Russell, 2000)
  - エージェントの事後行動から事前状態を推論する不良設定問題
    - 単に合理性原則の逆をたどるだけでは解けない

- 生後 6 ヶ月にもなれば，行動と目標を与えられることで次の行動を予測できる (Woodward, 1998)
  - ここから，状況的制約である“環境”とそこに設けられた“目標”が，合理性原則にしたがって“行動”を決定する理論が提案された (Fig. 1b; Gergely et al., 1995)

- 実験では，参加者はエージェントが迷路を動く様子を観察し，その目標を推論する
  - エージェントの動きは，運動主体感 (sense of agency)・印象を強く喚起 (Heider & Simmel, 1944; Tremoulet & Feldman, 2000, 2006)

2. Computational framework

- 行動理解を，マルコフ決定問題 (MDPs) における確率的プランニングのモデルを逆ベイズ推定で定式化
  - MDPs　不確実な状況における連続的な意思決定モデルのフレームワーク (Dayan & Daw, 2008)
    - > 環境・エージェントに関連するすべてを,“状態”変数として記銘
    - > エージェントの行動によって生じる主観的利益・損失を表現

- MDPs をレストランの料理長に例える
  - 目標　資源活用の最大化，顧客満足度の最大化
  - 状態　スタッフの人数，材料，コンロ，調理器具，など
  - 行動　スタッフへの指示，資源の割り当て
    - > この因果モデルには，調理時間とパフォーマンスの不確実性を反映されている

- MDP モデルに目標指向プランニングの理論を当てはめた以下の方程式を定式化

$$P(Goal|Actions, Environment) \propto P(Actions|Goal, Environment)P(Goal|Environment) \quad (1)$$

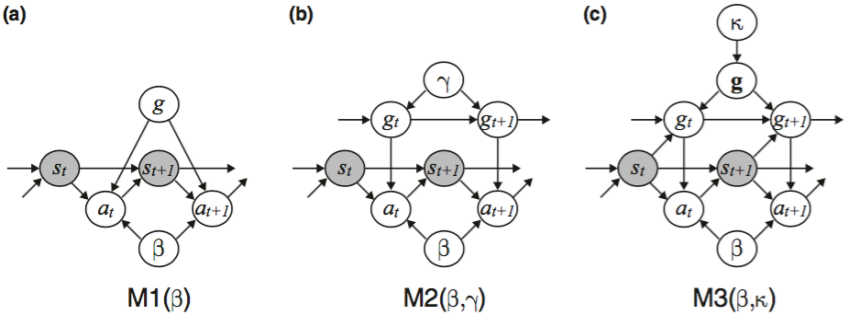
- また，過去の行動から新規状況における未来の行動 (Actions') を推論する方程式を定式化

$$P(Actions'|Actions, Environment) = \sum_{Goal} P(Actions'|Goal, Environment)P(Goal|Actions, Environment) \quad (2)$$

- エージェントの目標を人間がどう推論しているかを検証するため，参加者データとモデルの予測を比較
  - 実験 1　エージェントの現在の目標を推論
  - 実験 2　エージェントの過去の目標を遡って回想的に推論
  - 実験 3　エージェントの過去の行動から未来の目標を予測的に推論

- 3 つのモデルと 1 つのヒューリスティックス (下図参照)
  - M1: single underlying goal
    - > 目標は 1 つで，“決定パラメータ β に応じてエージェントが気まぐれな行動を起こす”と想定
      - * $\beta = 0$　完全にランダムな行動を取る

        - $0 < \beta \leq 5$　値が大きくなるにつれて最適な行動 (最短経路) を取る頻度が高まる
- M2: changing goals
    - β のほか, “目標変更パラメータ γ に応じてエージェントの目標が変更される”と想定
        - $\gamma = 0$　目標は変更されない (M1 と同等)
        - $0 < \gamma < 1$　値が大きくなるにつれて目標変更の頻度が高まる
        - $\gamma = 1$　行動する度に目標が変更される
- M3: complex goals
    - β のほか, “下位目標パラメータ κ に応じてエージェントに下位目標が設定される”と想定
        - $\kappa = 0$　下位目標は設定されない (M1 と同等)
        - $0 < \kappa \leq 1$　値が大きくなるにつれて複雑な目標が設定される
- H: heuristics alternative
    - “直前の行動のみからエージェントは目標を決定する”と想定
        - 行動する度に目標が任意に変更される (M2 の $\gamma = 1$，かつ履歴を無視した場合と同等)



2.1. Related work

- 従来モデルの問題点
    - 行動の配分が事前に付与 (Bui et al., 2002; Charniak & Goldman, 1991)
    - 行動の決定に過去の膨大なデータが必要 (Liao et al., 2004).
    - 目標表象の抽出が不可 (Kautz & Allen, 1986; Ng & Russell, 2000; Verma et al., 2006)

- 本論文では，人間の行動理解をモデル化するため，エージェントに対する目標表象と合理的プランニングの確率論モデルを統合する

3. Experiment 1

- 目標・障害物・エージェントの経路を変化させ，その時点 (現在) のエージェントの目標を推論する
    - M1・M2・M3・H が参加者の推論をどのくらい正確に説明するか？

3.1. Method

3.1.1. Participants

- MIT 参加者プールから 16 名 (男性 7 名，女性 9 名)

3.1.2. Stimuli

- エージェントが迷路を進む様子を俯瞰視点で提示 (Fig. 2)
  - エージェントの行動　東西南北 8 方位 (障害物がある場合は移動不可)

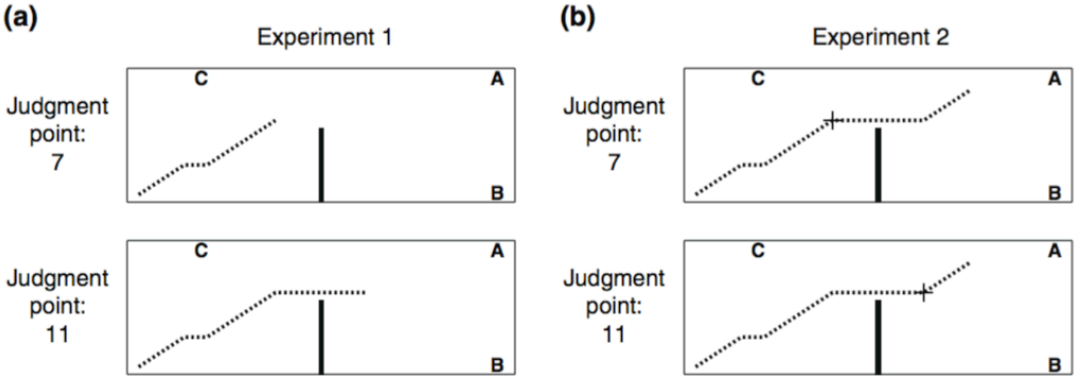


Fig. 2. Stimulus paradigm for Experiments 1 and 2. Each stimulus presented an animation of an agent's path (marked by a dashed line) ending at a judgment point: a pause in the animation that allowed participants to report their online inferences of the agent's goal at that point. (a) Experiment 1: online goal inference task. Subjects rated how likely each marked goal was at the judgment point. (b) Experiment 2: retrospective goal inference task. Subjects rated how likely each marked goal was at an earlier judgment point, given by the "+" along the dashed line.

3.1.3. Design

- 全 36 条件の刺激を Fig. 3 に示す
  - 4 (目標 C の位置: 1 / 2 / 3 / 4)　　A・B の位置は固定
    - × 3 (目標: A / B / C)　　エージェントの最終目標
    - × 3 (障害物の切れ間と経路)　　切れ間なし (Solid)
      切れ間を迂回 (Gap (around))，切れ間を通過 (Gap (through))

3.1.4. Procedure

- カバーストーリー"科学者が捕まえた知能を持つ未知の生物が，どの目標にたどり着くかを推論"
- エージェントの動きが止まったポイントで，エージェントの目標を推論
  - A・B・C のいずれかを選択
  - 選択しなかった 2 つの目標の可能性を 9 件法で評定
    - > "選択した目標と同程度に起こりそう" 〜 "まったく起こりそうにない"

3.1.5. Modeling

- 方程式(1)を用いた M1・M2・M3・H で比較

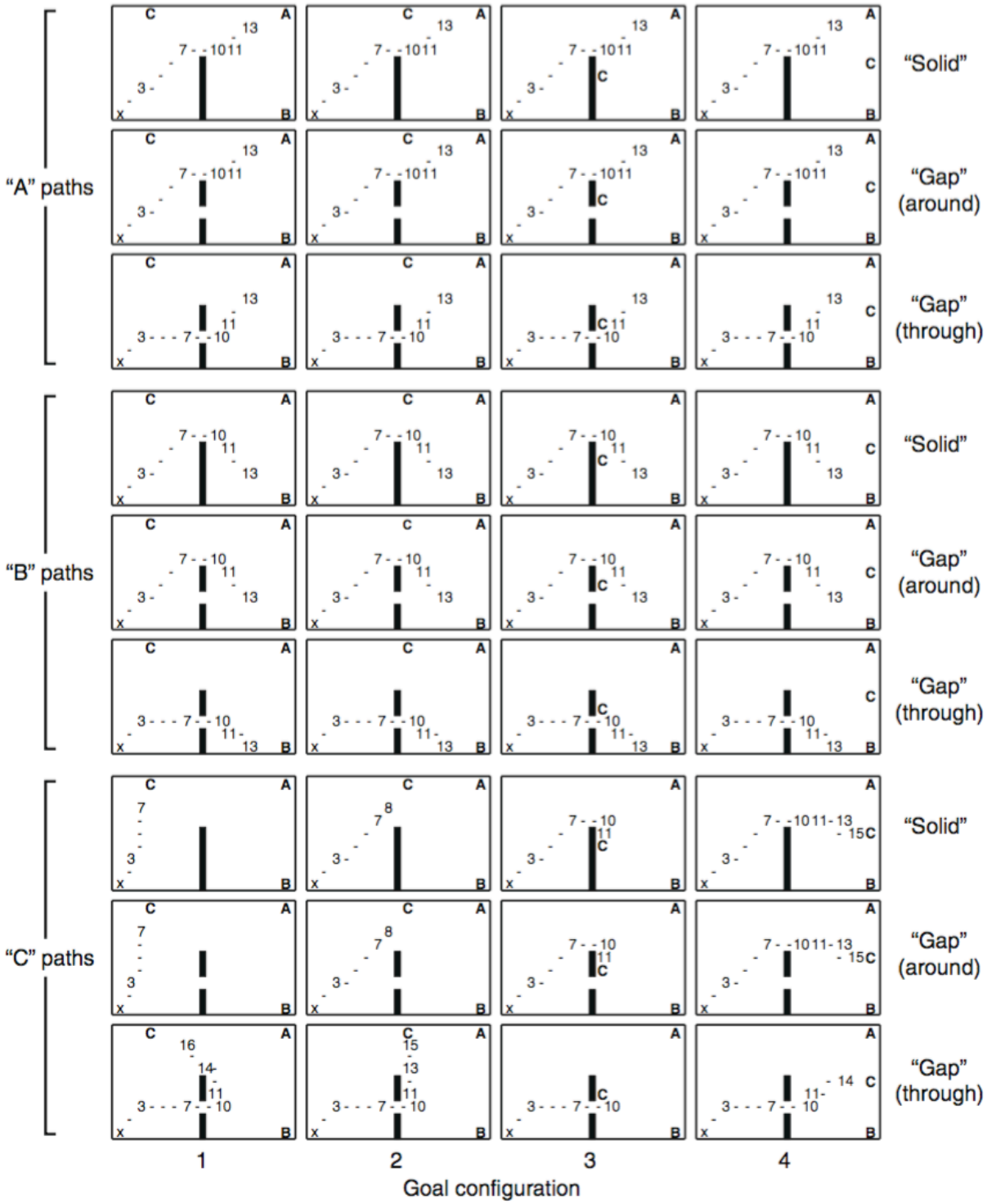


Fig. 3. All stimuli from Experiment 1. We varied three factors: goal configuration, obstacle shape and agent path. There were four goal configurations, displayed in columns 1–4. Path conditions are grouped as "A" paths, "B" paths and "C" paths. There were two obstacle shape conditions: "Solid" and "Gap". There were 36 conditions, and 99 unique stimuli in total.

3.2. Results

- 参加者の目標推論データと，最も適合度の高かった M2 ($\beta = 2.0$, $\gamma = 0.25$) の結果 (Fig. 4)
  - 初期段階では目標は不明確だが，経路が観察されるにつれて推定が確かになってくる
    - 条件 1・2　　ひとつの目標に絞れた地点で，A と B の不確実性が解消
    - 条件 3・4　　障害物の進み方から B / A の可能性が低下するが，急旋回で再上昇

– 刺激特徴が変わることで，多かれ少なかれ目標推定は影響を受ける

> 条件 5・6　　C が A と B の中間にあるため，相対的に A・B の可能性が低下

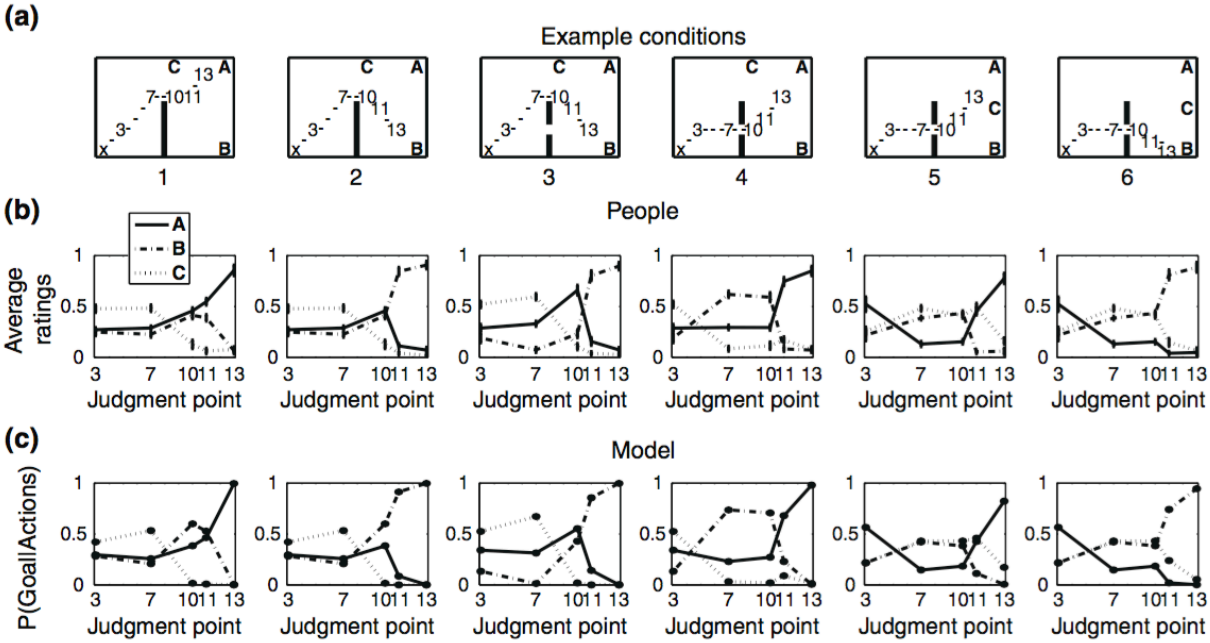


Fig. 4. Example conditions, data and model predictions from Experiment 1. (a) Stimuli illustrating a range of conditions from the experiment. Each condition differs from adjacent conditions by one feature – either the agent's path, the obstacle, or the goal locations. (b) Average subject ratings with standard error bars for the above stimuli. (c) Predictions of inverse planning model M2 with parameters $\beta = 2.0$, $\gamma = 0.25$.

- bootstrap cross-validation (BSCV) を用いて，参加者データとモデルの相関について検定を実施 (Fig. 5)
  – 適合度　　M2 > H > M3 > M1 (それぞれ最も適合度の高いパラメータを使用)

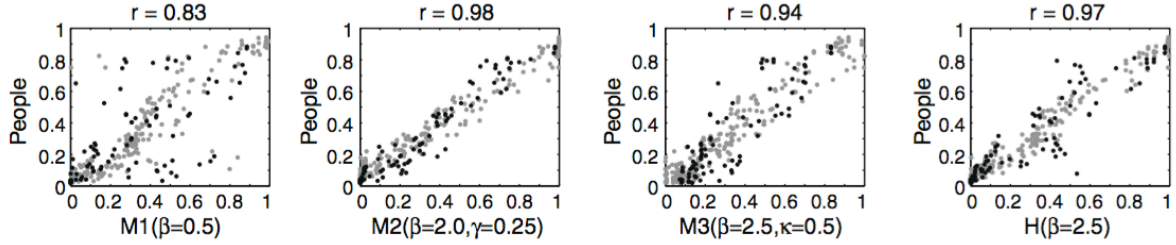


Fig. 5. Scatter plots of model predictions using best-fitting parameter settings (*X*-axes) versus people's online goal inferences (*Y*-axes) for all Experiment 1 stimuli. Points from the Gap obstacle condition at judgment points 10 and 11 are plotted in black. These trials account for most of the outliers of M1, M3 and H.

- 直近の経路のみでは推論できない条件を用いて，targeted analysis を実施 (Fig. 6)
  – ポイント 10 の時点では A・B どちらとも言えない
  > 条件 1　　ポイント 11 もまだ不確実であるため，参加者は過去に遡って推論
  M1・M2・M3 はこの参加者データをほぼ説明
  > 条件 2　　ポイント 11 で B の方へ向かうため，参加者は B と推論
  M2・H はこの参加者データをほぼ説明

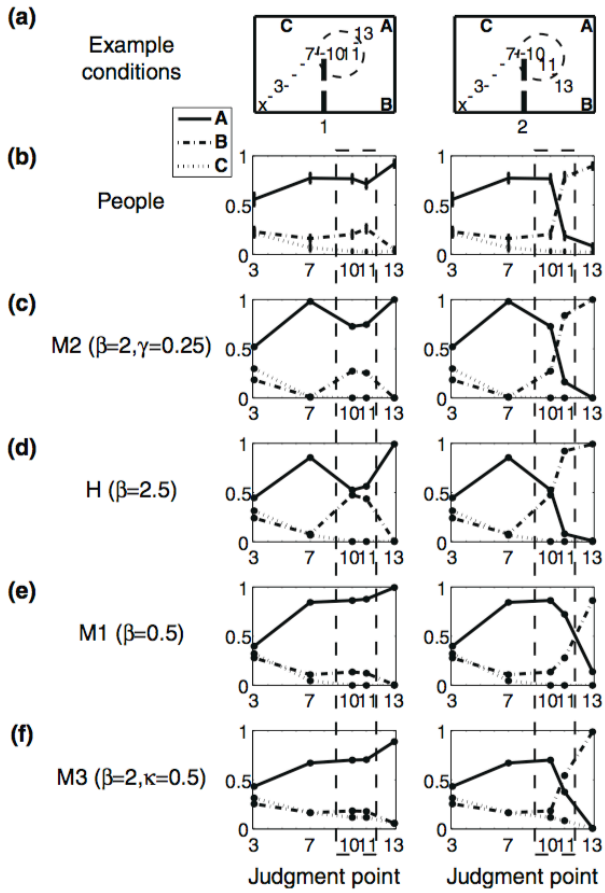


Fig. 6. Targeted analysis of Experiment 1. (a) Example conditions from our targeted analysis. (b) Participants' ratings from the above conditions. (c) Predictions of M2 ($\beta = 2.0, \gamma = 0.25$). (d) Predictions of H ($\beta = 2.5$). (e) Predictions of M1 ($\beta = 0.5$). (f) Predictions of M3 ($\beta = 2.0, \kappa = 0.5$).

3.3. Discussion

- その時点 (現在) における参加者の推論を最も説明したモデルは M2
  - 低めの目標変更パラメータ β がエージェントの行動と適合した
  - 直近の行動が不確実な場合，それ以前に遡って目標を推定する参加者の行動を説明した

4. Experiment 2

- 実験 2 では，途中で目標が変わる刺激を用いて，過去にどのように推論していたかを検証する
  - 人間の行動理解はどちらのモデルにより近いか？
    - > M2・H (目標変更可)　　過去の時点では現在と別の目標を推定できる
    - > M1・M3 (目標変更不可)　　系列全体の単一目標を推定せざるをえない

4.1. Method

4.1.1. Participants

- MIT 参加者プールから 16 名 (男性 6 名，女性 10 名)

4.1.2. *Stimuli*

- エージェントの経路が先に提示され，赤色の“+”で示されたポイント時点の目標を推定 (Fig. 2b)

4.1.3. *Design*

- 刺激は実験 1 と同様 (Fig. 3)

4.1.4. *Procedure*

- “+マークの時点では，その生物はどの目標を目指していたかを評定”

4.1.5. *Modeling*

- 方程式(1)の回想あり版を用いた M1・M2・M3・H で比較

4.2. ***Results***

- 参加者データと，最も適合度の高かった M2 ($\beta = 0.5$, $\gamma = 0.65$) の結果 (Fig. 7)
  - M2 は過去と未来の情報を統合し，各ポイントの目標をよく推定していた

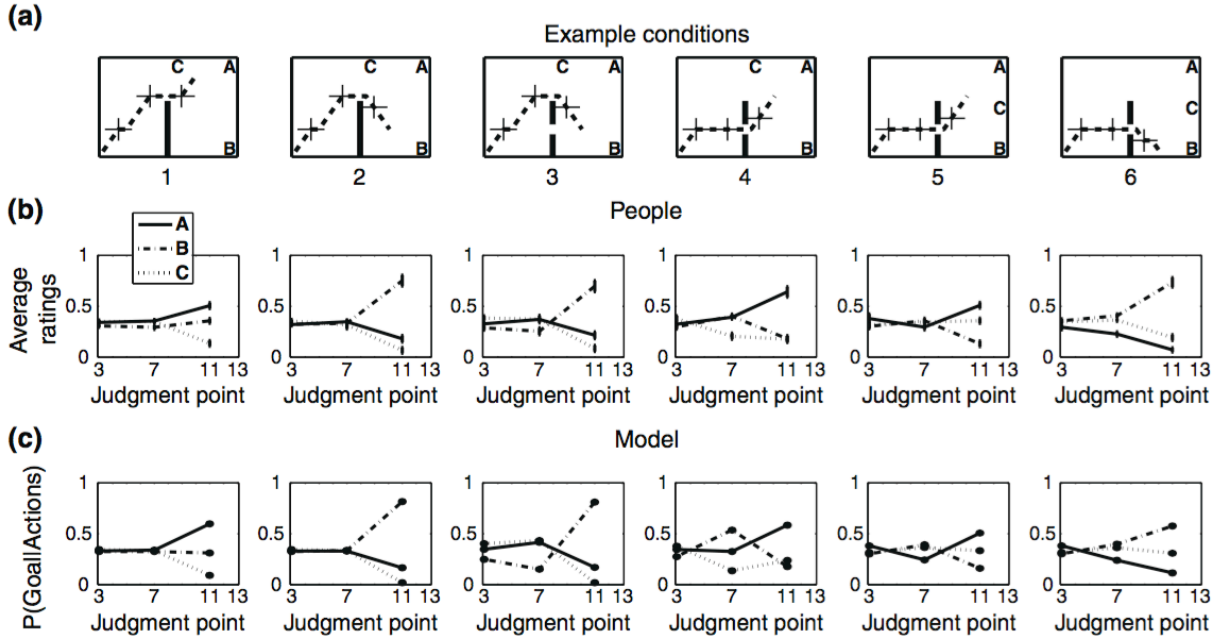


Fig. 7. Example conditions, data and model predictions from Experiment 2. (a) Stimuli illustrating a range of conditions from the experiment. These stimuli directly correspond to the Experiment 1 stimuli in Fig. 4a. Dashed lines correspond to the movement subjects saw prior to rating the likelihood of each goal at each judgment point, which are marked by black +'s. (b) Average subject ratings with standard error bars for the above stimuli. (c) Predictions of inverse planning model M2 with parameters $\beta = 0.5$, $\gamma = 0.65$.

- 参加者データとモデルの相関 (Fig. 8)
  - BSCV の適合度　　M2 > H > M1 ≒ M3 (それぞれ最も適合度の高いパラメータを使用)
  - 実験 1 に比べて，β は低く γ は高かった
    - 参加者はエージェントのノイズを多く見積もり，目標を変更したがっていた

– M1・M3 は，目標が変わっているにもかかわらず，全ポイントで同じ推論を行っていた

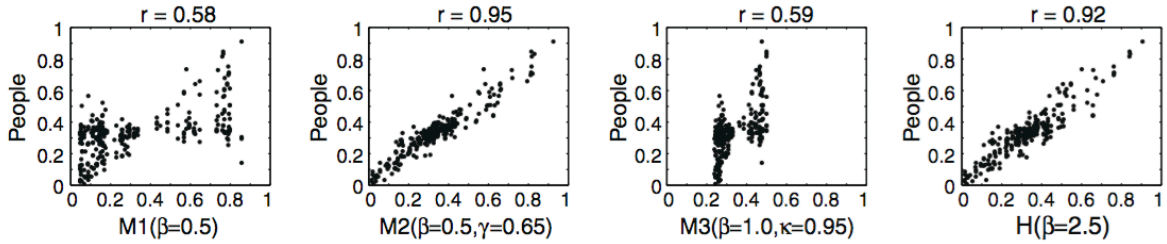


Fig. 8. Scatter plots of model predictions using best-fitting parameter values (*X*-axes) versus people's retrospective goal inferences (*Y*-axes) for all Experiment 2 stimuli.

4.3. Discussion

- 目標変更可能なモデルは，参加者の回想的な目標推論を正確に説明することができた
  - ノイズが大きく見積もられた原因は，現在より過去の推定のほうが挑戦的なため？
- 回想的な目標推論には，不要な要素を適応的に忘却する機能が不可欠

5. Experiment 3

- 実験 3 では，下位目標の推定が必要と思われる複雑な目標表象を扱う
  - 最短経路からの乖離が常に一定にした経路を通るなら，下位目標を持っている可能性が高い
    - 例: 会社帰りはいつも，最短経路から数ブロック離れたスーパーに寄る
  - 別の地点から開始しても，また大回りであってもそれが観察されるなら，さらに可能性は高い
    - 例: 別の現場から帰るときであっても，大回りしてでも，スーパーに寄る

- エージェントが迂回して同じ中間地点を通る経路を繰り返し提示
  - “エージェントが下位目標を持っている”と参加者は推論するか？

5.1. Method

5.1.1. Participants

- MIT 参加者プールから 23 名 (男性 9 名，女性 14 名)

5.1.2. Stimuli

- 実験 3 で用いた刺激 (Fig. 9)
  - 最終目標は橙色の三角形で表示

5.1.3. Design

- 2 (例示経路: 直接 / 間接)　　例示フェーズの経路が中間地点を経由するか否か
  ×2 (障害物: あり / なし)　　障害物の有無
  ×2 (下位目標の位置: 遠 / 近)　　最終目標 (最短経路) までの距離

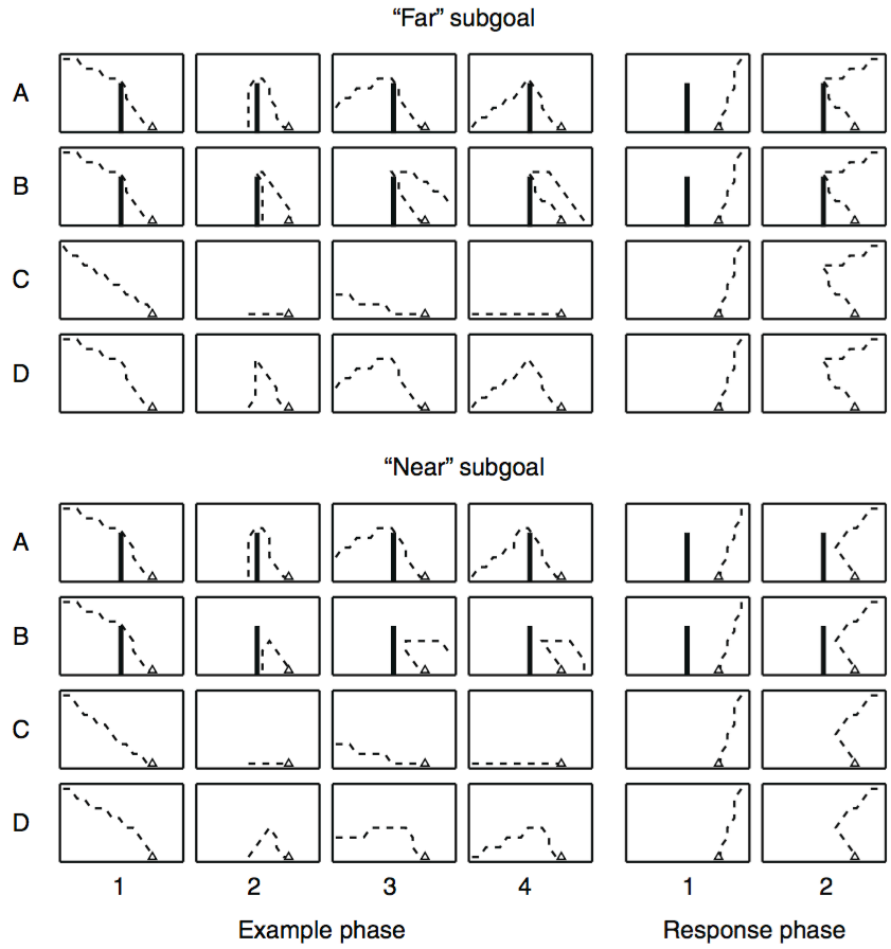


Fig. 9. All stimuli from Experiment 3. The experiment followed a 2 × 2 × 2 design, which varied the directness of the agent's path, the presence or absence of an obstacle, and the location of potential subgoals. Each condition had four trials, which each presented a different example path, and then asked subjects to predict which of two response paths was more likely given all previously displayed example paths from that condition.

5.1.4. Procedure

- 例示フェーズ 1 回・反応フェーズ 1 回で 1 試行とし，4 試行を実施
  - 例示フェーズ
    - 開始地点 4 種類 (Fig. 9 の 1 / 2 / 3 / 4) のいずれかを提示
  - 反応フェーズ
    - "次に提示する 2 つの経路のうち，この生物の動きとしてどちらがより可能性が高いか"
      - 例示フェーズと開始地点が異なる経路 (直接・間接)
      - 9 件法 ("明らかに経路 1" 〜 "明らかに経路 2")
    - これまでに提示された経路は履歴として常時表示

- 上記手順を 8 条件すべてで繰り返す
  - 条件変更時には，"色の違う個体は別の生物であるため，いちから考えること"と教示

5.1.5. Modeling

- 方程式(2)を用いた M1・M2・M3・H で比較
  - 下位目標を設定可能な M3 が有利？

- モデルを比較するため，参加者データとモデルの相関係数を算出
  - モデルから算出される事後オッズ比の対数を，シグモイド標準累積密度関数を通してスケール調整

5.2. Results and discussion

- 参加者データと，最も適合度の高かった M3 ($\beta = 5$, $\kappa = 0.6$) の結果 (Fig. 10)
  - M3 は蓄積される事例に対しても正確に説明した
  - 実験 1・2 より β が高いことから，参加者はノイズの少なく決定的なエージェントを想定

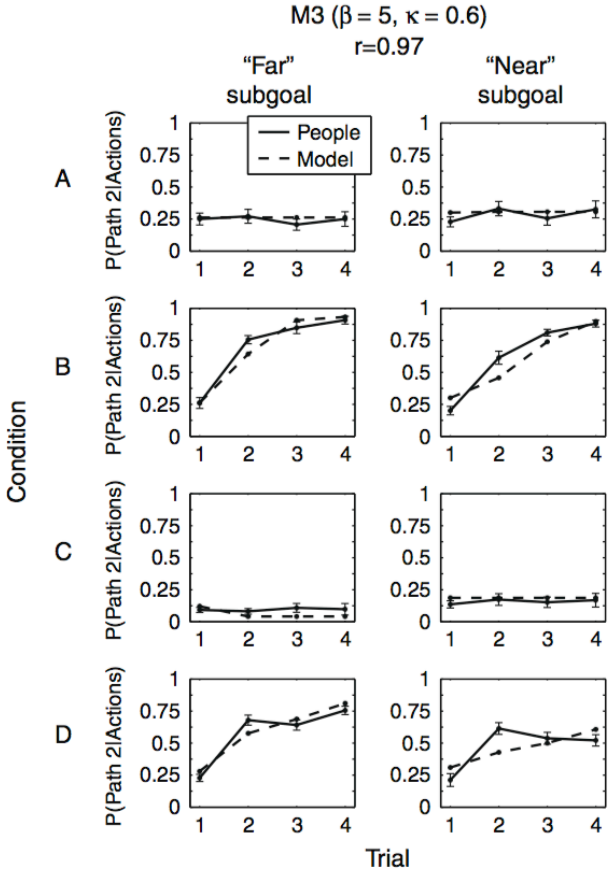


Fig. 10. Subjects versus M3 with the best-fitting parameters for all stimuli from Experiment 3. Ratings correspond directly to conditions from Fig. 9.

- 行動系列の観察によって，最短経路からの乖離をもとに下位目標の解釈を行っていたことが分かった
  - 共通して経由する中間地点が下位目標の根拠として蓄積された
  - 一方，単一の目標と想定した場合，乖離の大きい行動を“気まぐれ”と捉えなければならない

- 下位目標推論は，以下の 2 つの刺激要因によって説明される
  - 想定される下位目標に対応する経路の数，それらの経路と最短経路との乖離の大きさ

6. General discussion

- 本研究から得られた知見
  - MDPs における逆ベイズ推定を通して，人間の行動理解を定式化した
  - エージェントの目標を人間がどう捉えているかを，モデルが正確に説明することができた
    - ヒューリスティックスが苦手とする場面にも適合
  - 異なる環境では異なる種類の目標表象が使用されることを明らかにした
    - 現時点の推論 (実験 1)・回想的な推論 (実験 2)　変更可能な表象 (M2)
    - 複雑な目標の将来的な推論 (実験 3)　下位目標を持つ表象 (M3)

- 観察された刺激に対して，参加者はどの目標表象 (モデル) を使うことが適切と考えていたか？
  - 階層ベイズモデルを使用　$\log P(Stimuli|Model)$
  - 分析の結果，参加者はエージェントの説明に最適な目標表象を用いていた (Table 3)

Table 3
Log marginal likelihood of models and heuristic given all Experimental stimuli. Higher log-likelihood indicates a better fit, and the highest value for each experiment is shown in bold.

	$\log P$(Stimuli|Model)			
	M1	M2	M3	H
Experiment 1	−773.6	**−641.9**	−651.6	−778.5
Experiment 2	−1118.5	**−832.4**	−860.1	−1068.1
Experiment 3	−457.2	−294.3	**−236.5**	−391.5

- 本フレームワークの今後
  - エージェントがより複雑になった場合 (factored MDPs; Guestrin, Koller, Parr, & Venkataraman, 2003)
  - エージェントの信念を推論する場合 (partially observable MDPs; Kaelbling et al., 1998)

7. Conclusion

- 理想的な観察者・理想的な推論エージェントのモデルは，人間の心的表象がどれほど現実とかけ離れているかを明らかにしてきた
  - 人間の行動は嫌らしいほど複雑で予測がつかず，時に不合理である (Hernstein & Prelec, 1991; Kahneman, Slovic, & Tversky, 1982)
- 本実験のフレームワークを様々な課題に適用することで，目標表象の柔軟性を示すことができる
  - 合理性原則と意図表象を組み合わせたモデルを用いることで，直観心理学はさらに発展するだろう