

質問項目に対する回答困難感を反映する パラデータ指標の探索的研究

分寺 杏介

神戸大学

馬場 絢子

奈良女子大学

福島 綾

一宮メンタルクリニック



bunji@bear.kobe-u.ac.jp

日本行動計量学会第53回大会

2025年8月30日

知らない人と話すのを躊躇することがある

☐

まったく
当てはまらない

☐

当てはまらない

☐

あまり
当てはまらない

☐

やや
当てはまる

☒

当てはまる

☐

とても
当てはまる

「あまり」ってどれくらいだ.....?



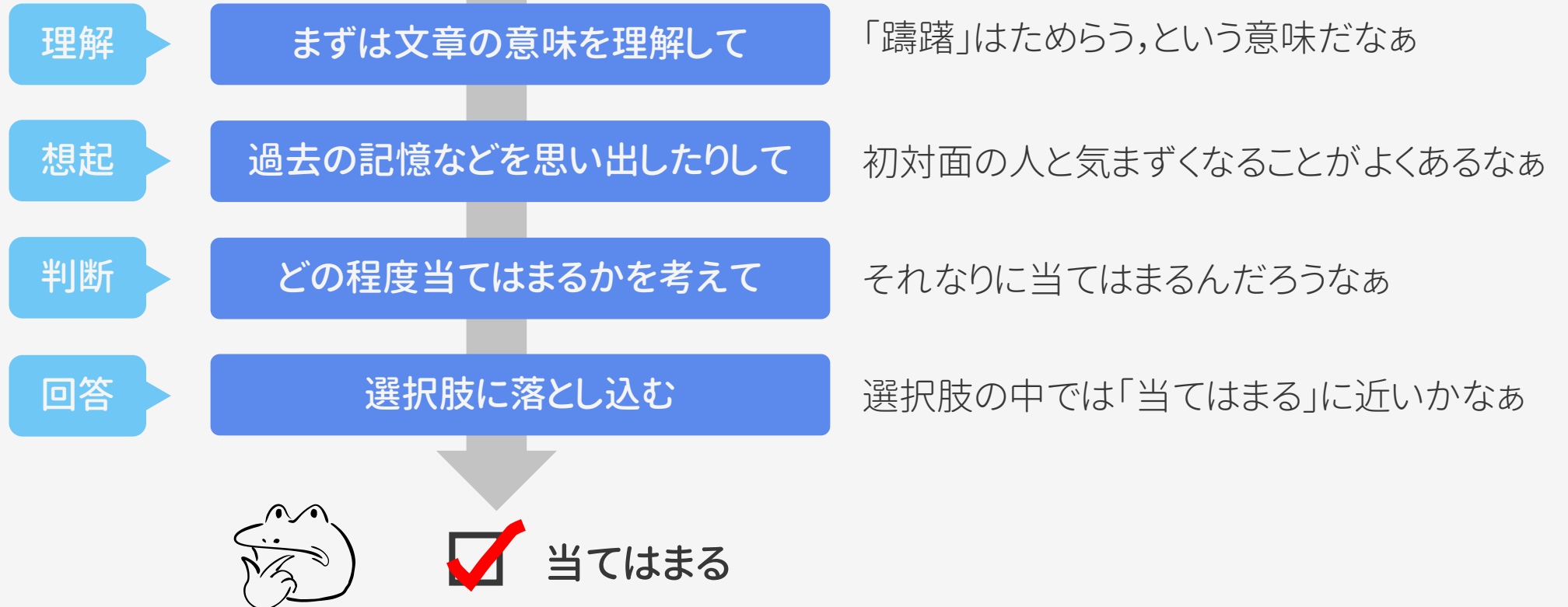
「躊躇」ってなんて読むんだ.....?

どんな場面を想定してるんだ.....?

■ 様々な認知過程を経て回答を決定している (e.g., Tourangeau, 2018)

知らない人と話すのを躊躇することがある

Tourangeau et al.(2000)
のフレームワーク



- 「回答者は、心理尺度を**正しく理解して**(まじめに)回答している」

あるいは、正しく理解して回答したものとして扱っている

- 仮定の逸脱は様々な問題に繋がらうる

【研究】データの質が低下するかもしれない

【臨床】回答をもとに判断してしまうと誤った診断を下してしまうかもしれない



- 様々なレベルでの回答困難感が報告されている (馬場他, 2025)

(臨床場面での心理尺度の利用に関して)

「項目の意味について質問する」「項目の指す状況について質問する」「状況によって異なる場合の回答方法について質問する」「程度表現について質問する」「未経験事項の回答方法について質問する」「特定の項目の回答時間が長すぎる」**etc.**

- できるならば,異なる回答困難感を弁別したい
 - (臨床場面ならば) つまずいているポイントに応じて個別最適なサポートを提供する
 - ▶ その人の現状をより正確に表す回答を得られるようになる
- 回答者個人が自覚的に弁別することは難しそう (分寺他, 2024; 2025, 日本心理学会大会)

「答えにくさ」評価項目	
総合	A1 この項目は回答しやすいと感じた
	A2 この項目はたいていの人が簡単に回答できだろうと感じた
	A3 この項目に正しく回答できたと確信している
個別	B1 この項目にはわかりにくい言葉が入っていると感じた
	B2 この項目への回答は,その時々で変わりやすいと感じた
	B3 この項目の意味がわかりにくかった
	B4 この項目は選択肢の違いがわかりにくかった
	B5 この項目が指す内容は想像しにくかった
	B6 この項目に関連することを思い返すのが大変だった
	B7 この項目に正直に回答するのをためらった

【マルチレベル相関】項目をレベル2として:対角成分は級内相関係数

	A1	A2	A3	B1	B2	B3	B4	B5	B6	B7
A1	0.453	0.751	0.696	-0.332	-0.201	-0.504	-0.371	-0.577	-0.488	-0.394
A2	0.998	0.425	0.614	-0.322	-0.177	-0.452	-0.325	-0.508	-0.432	-0.331
A3	0.995	0.991	0.395	-0.299	-0.174	-0.432	-0.356	-0.508	-0.455	-0.381
B1	-0.879	-0.864	-0.917	0.570	0.167	0.507	0.399	0.397	0.355	0.368
B2	-0.532	-0.514	-0.520	0.376	0.185	0.169	0.231	0.219	0.269	0.202
B3	-0.928	-0.913	-0.949	0.959	0.438	0.594	0.494	0.578	0.414	0.392
B4	-0.926	-0.911	-0.943	0.914	0.481	0.987	0.324	0.419	0.381	0.376
B5	-0.988	-0.992	-0.986	0.878	0.449	0.936	0.932	0.471	0.588	0.427
B6	-0.920	-0.938	-0.906	0.701	0.578	0.745	0.763	0.911	0.416	0.414
B7	-0.970	-0.955	-0.965	0.882	0.510	0.931	0.924	0.949	0.834	0.238

Within

Between

聞き方を変えてもかなり高い相関が得られた

■ まずは、回答困難感の有無だけでも検出したい

「答えにくかったか」を逐一尋ねるわけにもいかないこともあるので、
観測された情報からどこまで迫れるかを検証してみる

■ パラデータの利用の可能性を検証する

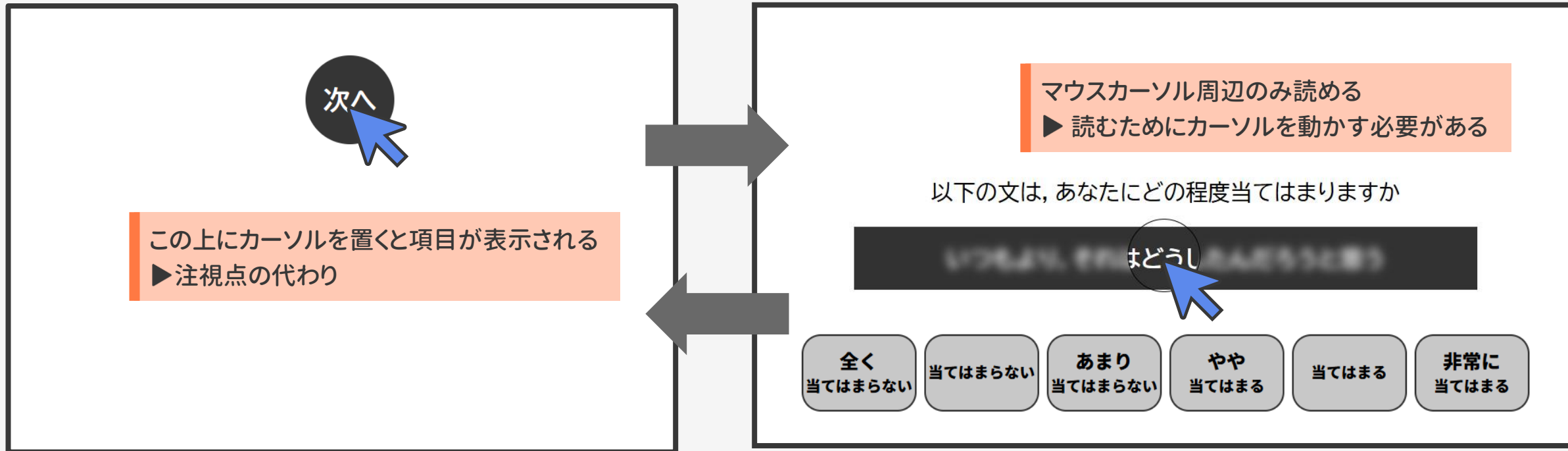
回答内容だけではもちろん難しいので、他に得られる情報を使ってより深く迫りたい
今回は**回答時間とマウスカーソルの軌跡**を使用します

※探索的検討なので、意味付けは明確にせずにとりあえず特徴量を作成してみます

クラウドソーシング (Crowdworks) にて実施 ($n = 122$)

■ 手順

1. 「回答が困難になりうる要素」を含んだ項目に回答



■ 「回答が困難になりうる要素」を含んだ項目

回答困難項目	
1	生活が 光風霽月 でないと感じることはない
2	出かけるときは 身体が重い
3	いつもより,それはどうしたんだろうと思う
4	野鳥観察を すると心が安らぐ
5	飛行機を操縦するとき,緊張より楽しみのほうが大きい
6	先週の水曜日は 良い一日を過ごせたと思う
7	オペラを見るなど,趣味や好きなことをしたい気持ちがある
8	友人知人に嫌味を言いたくなる ことがある
9	読書をしたり,体を動かしたりする のが好きだ
10	いつも朝食を しっかり食べている
S	私はパソコンを持っている

attention checkかつ
「回答しやすい項目」として使用

■ 手順

2. 回答困難感に関する質問に回答

読書をしたり、体を動かしたりするのが好きだ

全く
当てはまらない

当てはまらない

あまり
当てはまらない

やや
当てはまる

当てはまる

非常に
当てはまる

上記の項目について、以下の質問に感じたことを
1(まったくそう思わない)から5(非常にそう思う)までの
5段階評価の数字で教えてください。

すべての質問に、考え込まないで直感的に評価してください。
また、答えられない質問には3をつけてください。

教示および質問項目の一部は
システムユーザビリティスケール日本語版 (山内, 2015)
を参考に作成

	まったくそう思わない			非常にそう思う	
	1	2	3	4	5
この項目に正しく回答できたと確信している	☐	☐	☐	☐	☐
この項目は回答しやすいと感じた	☐	☐	☐	☐	☐
この項目はたいていの人が簡単に回答できると感じた	☐	☐	☐	☐	☐

■ 回答困難感に関する質問項目

「まったくそう思わない」から「非常にそう思う」の5件法

	評価項目
総合	A1 この項目は回答しやすいと感じた A2 この項目はたいていの人が簡単に回答できるだろうと感じた A3 この項目に正しく回答できたと確信している
個別	B1 この項目にはわかりにくい言葉が入っていると感じた B2 この項目への回答は、その時々で変わりやすいと感じた B3 この項目の意味がわかりにくかった B4 この項目は選択肢の違いがわかりにくかった B5 この項目が指す内容は想像しにくかった B6 この項目に関連することを思い返すのが大変だった B7 この項目に正直に回答するのをためらった

3項目に対する
1因子の因子分析の得点を
outcomeとして使用

弁別できなかったので不使用

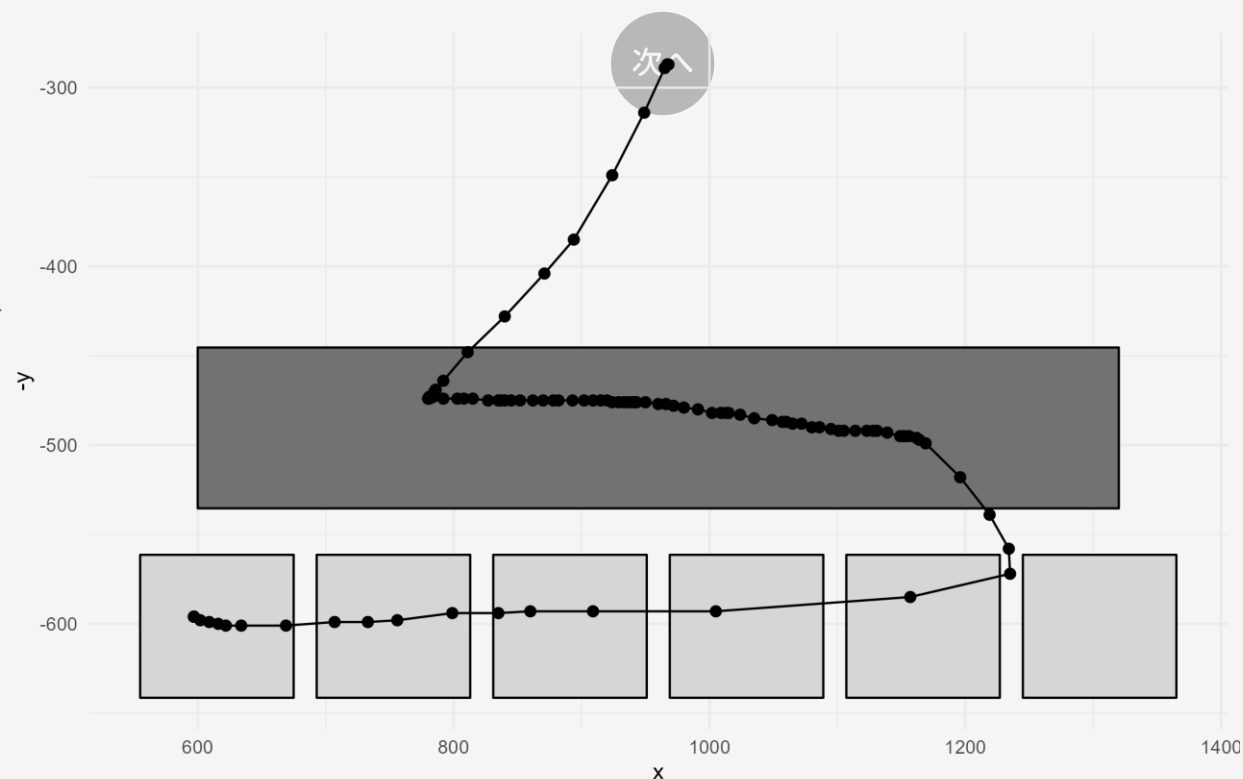
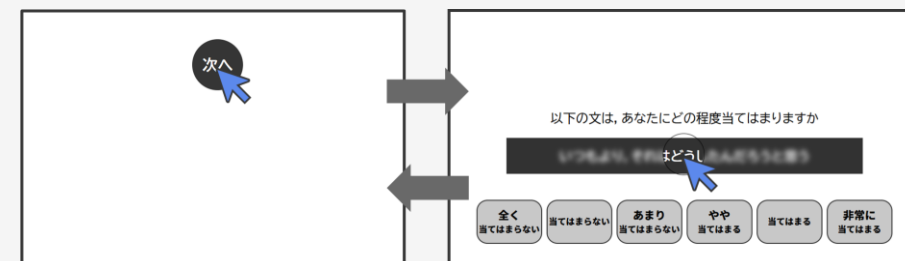
■ およそ20-30msごとに座標を記録

```
> mt_i
```

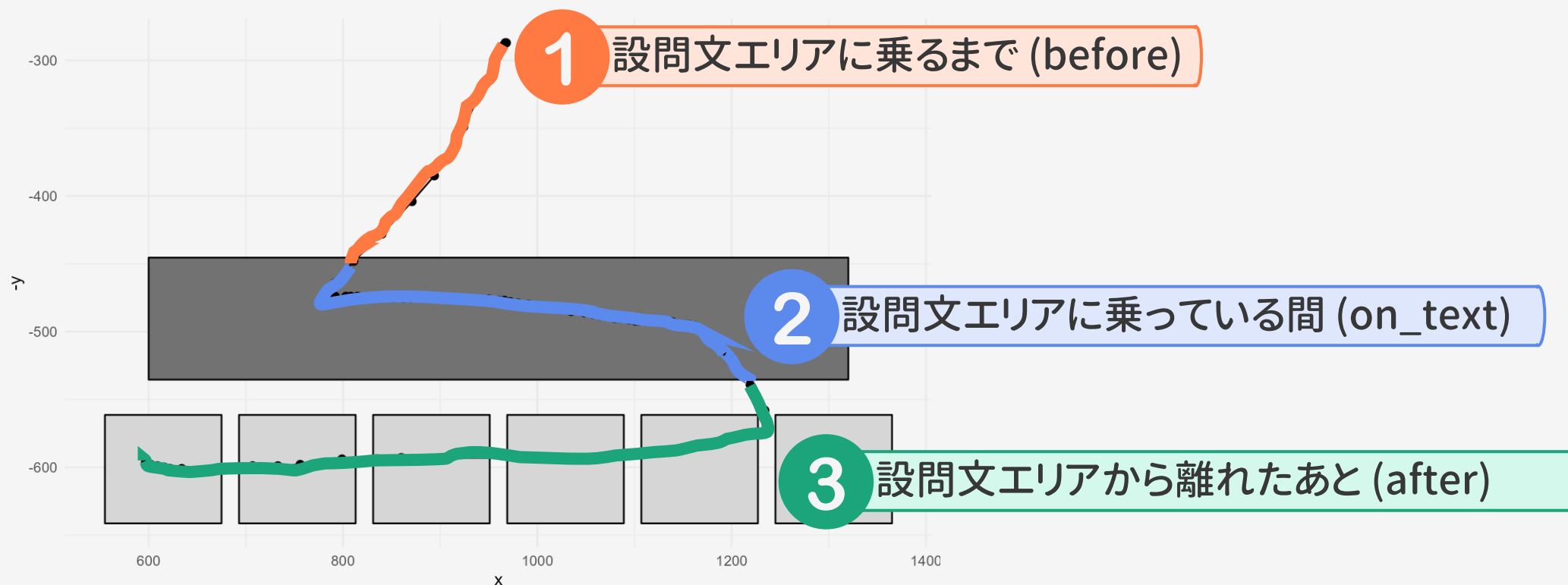
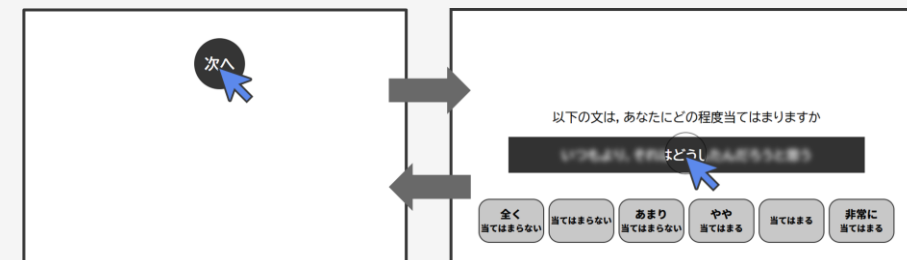
	x	y	t	event
1	968	287	44	pointerup
2	967	287	366	pointermove
3	965	289	386	pointermove
4	949	314	420	pointermove
5	924	349	453	pointermove
6	894	385	487	pointermove
7	871	404	520	pointermove
8	840	428	554	pointermove
9	811	448	587	pointermove

調査システムはjsPsychで作成

▶ mouse-tracking extensionをもとに記録



カーソルの軌跡をもとに 3つの状態に分割して考える



区分	名称	説明
基本	itemid	項目ID (ダミー変数化)
	response	回答 (1-6, ダミー変数化)
回答時間	rt_total	合計のRT
	rt_on_text	設問文エリアにカーソルがある時間
	rt_after	設問文エリアからカーソルが離れたあとの時間
	rt_on_choice	実際に選んだ選択肢の上にカーソルがある時間
	rt_text_again	設問文エリアからカーソルが離れた後に, 再び設問文エリアに入った時間
	idle_time_on_text	設問文エリアでカーソルが停止していた時間

※ rt_total 以外の rt は, すべてマウスカーソル軌跡があるために区別可能な特徴量

カーソルに関する特徴量の一部はKieslich, Henninger, et al. (2019); Pino et al. (2025)を参考に作成

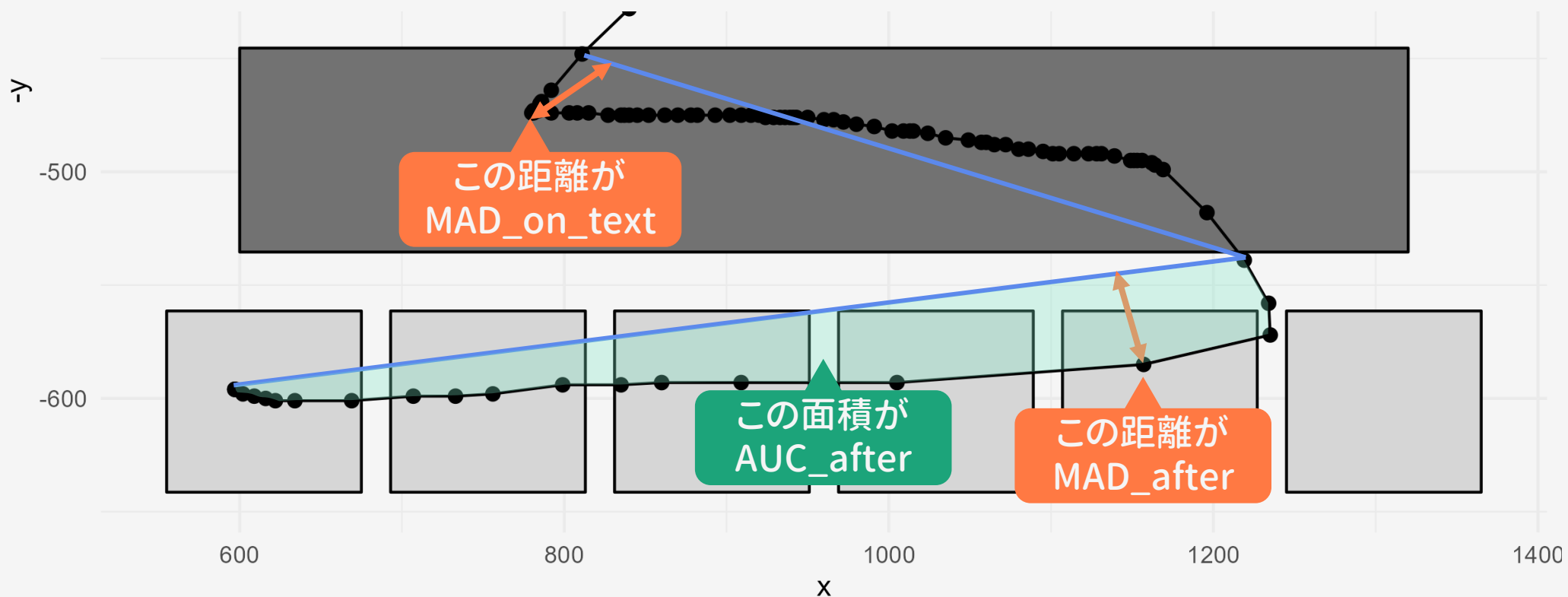
区分	名称	説明
カーソル	MAD_xx	xxでのMAD
	MAD_above_xx	xxでの, 直線よりも上側でのMAD
	MAD_below_xx	xxでの, 直線よりも下側でのMAD
	AUC_xx	xxでの, AUC
	total_dist_xx	xxでのカーソルの総移動距離
	vel_max_xx	xxでのカーソルの最高速度
	acc_max_xx	xxでのカーソルの加速度の最大値
	xpos_flips_after	afterで, カーソルが横方向に反転した回数
	n_option_change	カーソルが異なる選択肢の上に移動した回数
	target_reentry	最終的に選んだ選択肢にカーソルを乗せ直した回数

※ xx には, on_textまたはafterがそれぞれ入ります

(補足)MADとAUC

- 始点から終点までを結んだ直線からの距離の最大値がMAD

(例) MAD_on_text と MAD_after と AUC_after



■ LASSO(線形モデル)とXGBoost(非線形モデル)を使用

■ 使用する特徴量の違いにより,3つのモデルを比較

1. パラデータ不使用(itemid, responseのみ)
2. 1. に rt のみを追加(マウスカーソルなしで追加できるパラデータ)
3. すべての特徴量を使用

■ データ分割

8:2 に分割するため,92人分を訓練データに,30人分をテストデータに

■ パラメータチューニング

LASSO: penalty

XGBoost: eta, max_depth, colsample_bytree, subsample, min_child_weight

5-fold CVを用いたgrid searchによって最適な組み合わせを探索(不十分かも)

■ マウスカーソルに関する特徴量をいれると予測精度が低下した

		LASSO		XGBoost	
model	特徴量	RMSE	相関	RMSE	相関
1	itemid + response	0.671	0.738	0.684	0.723
2	1+RT	0.658	0.751	0.687	0.727
3	(全特徴量)	0.679	0.731	0.738	0.669

itemidの効果が強すぎて
他の特徴量が軒並みノイズに
なっているのかも？

ただ、ややover-fittingの気配もあり

■ (参考) itemidを入れない場合

		LASSO		XGBoost	
model	特徴量	RMSE	相関	RMSE	相関
1	itemid + response	0.877	0.473	0.876	0.475
2	1+RT	0.793	0.610	0.773	0.630
3	(全特徴量)	0.773	0.629	0.775	0.627

※ itemidあり, モデル3での結果

特徴量	標準化回帰係数
(itemidのダミーが上位を独占, [0.676, 1.555])	
response_1	0.450
response_2	0.259
rt_after	-0.176
response_3	-0.154
MD_above_on_text	-0.128
vel_max_on_text	0.119
rt_on_text_std	-0.119
response_4	-0.095
vel_max_on_text_std	-0.087

【response】(6が基準カテゴリ)

端の選択肢を選ぶほうが答えやすそう

【rt】

特に読み終わったあとの時間が長いほど
困難感を覚えていそう

【vel_max】(最高速度)

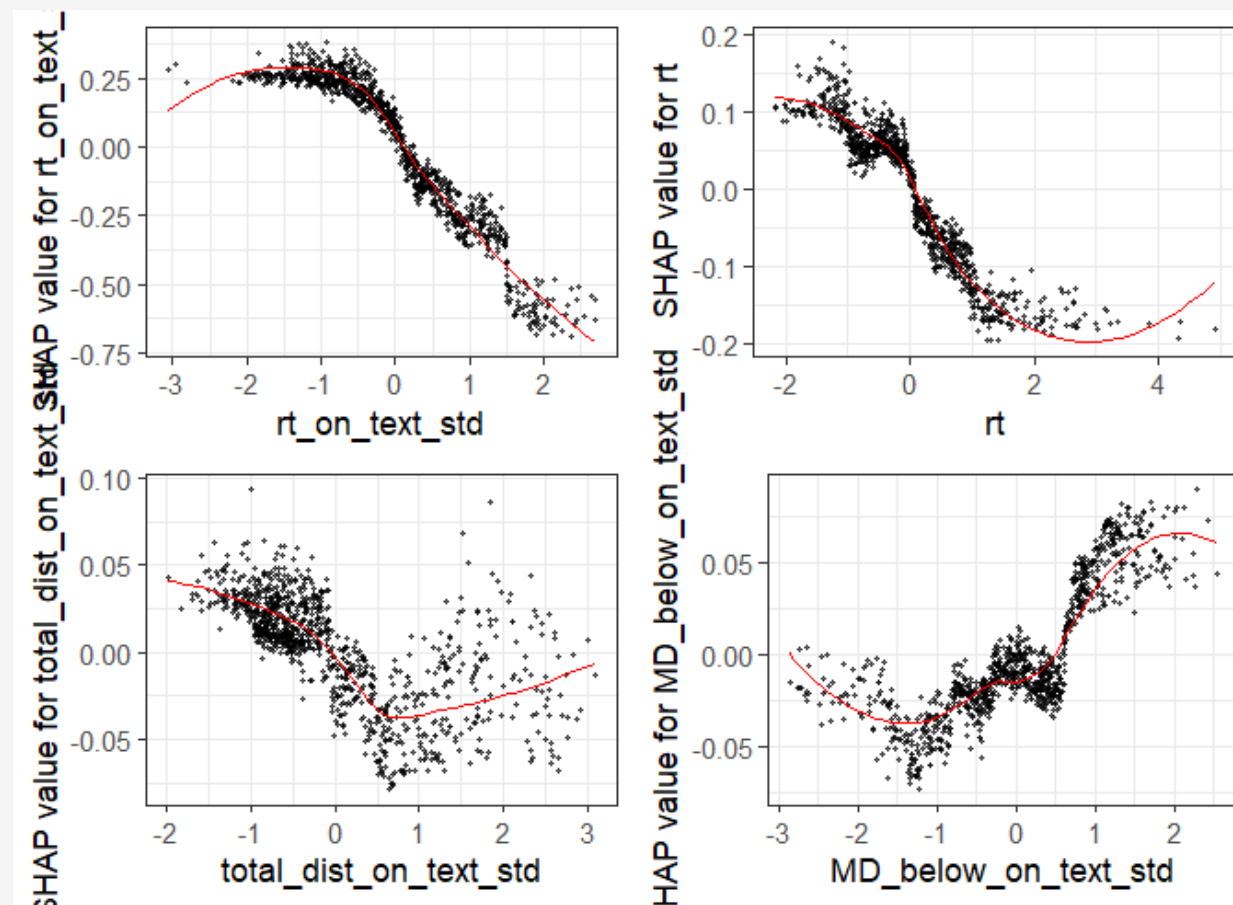
スラスラ読めているほど困難感は低そう

※ itemidあり, モデル3での結果

特徴量	Gain	SHAP
rt_on_text_std	0.163	0.233
response_1	0.120	0.202
rt	0.065	0.081
itemid_3	0.044	0.080
rt_on_text	0.043	0.039
total_dist_on_text_std	0.032	0.025
total_dist_on_text	0.028	0.021
MD_below_on_text_std	0.020	0.029
rt_after_std	0.018	0.027
rt_after	0.018	0.022

itemidがあまり効いていない?

SHAP dependence plot



■ 質問項目に対する回答困難感の予測を試みたが……

マウスカーソルの軌跡に基づく特徴量は精度の改善には寄与しなかった

RTがあれば説明がついてしまいそうな特徴量しか効果がなかった

基本的には項目IDによって予測できてしまう部分がかなり大きかった

▶ 個人ごとの差異を検出することは難しい？

■ 今後の課題

質問項目による結果の違いの検討 ▶ 広く用いられている尺度項目ではどうなるか

サンプルサイズ ▶ over-fittingであるかどうかの確認のため

より高次元特徴量の抽出 ▶ リッチな情報を引き出すため