(De)Noise: Moderating the Inconsistency Between Human Decision-Makers

Nina Grgić-Hlača, **Junaid Ali**, Krishna P. Gummadi and Jennifer Wortman Vaughan

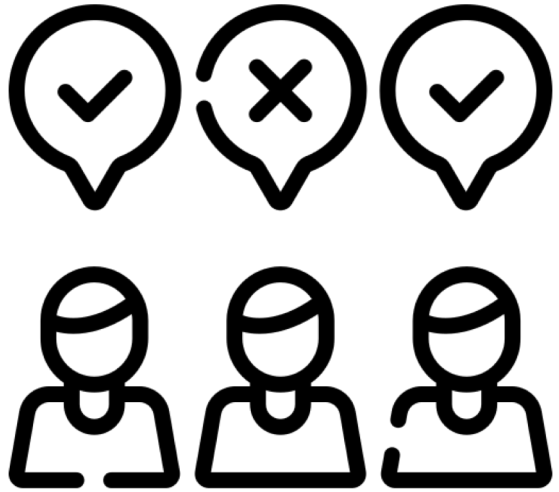
MAX PLANCK INSTITUTE
FOR SOFTWARE SYSTEMS

MAX PLANCK INSTITUTE
FOR RESEARCH ON
COLLECTIVE GOODS

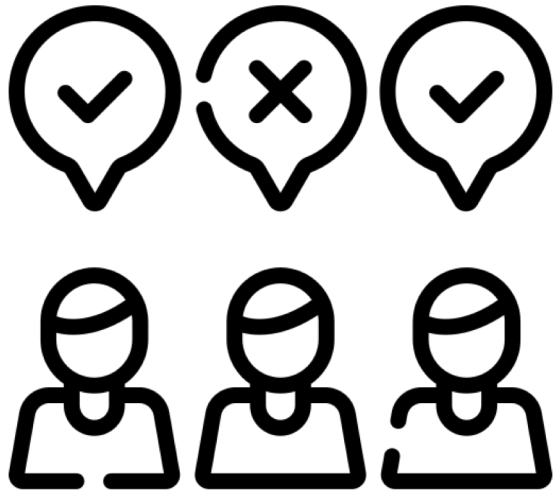
Microsoft Research

Presented with the same information, **different people** can **make different decisions**

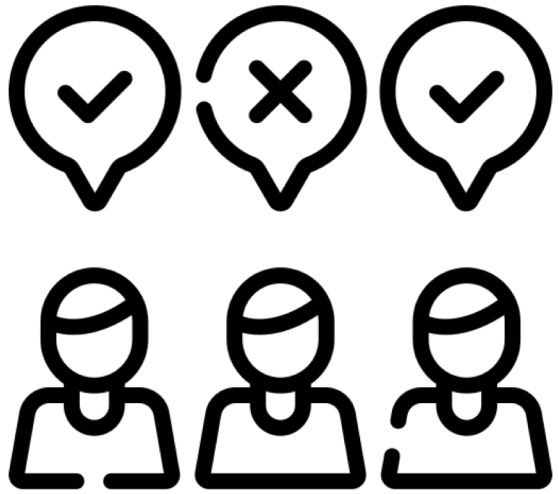
Presented with the same information, **different people** can **make different decisions**



Presented with the same information, **different people** can **make different decisions**



Presented with the same information, **different people** can **make different decisions**



How to **reduce inconsistency** between decision-makers?

- With the help of **algorithmic decision aids**!

Can algorithmic assistance moderate human (in)consistency?

Can algorithmic assistance moderate human (in)consistency?

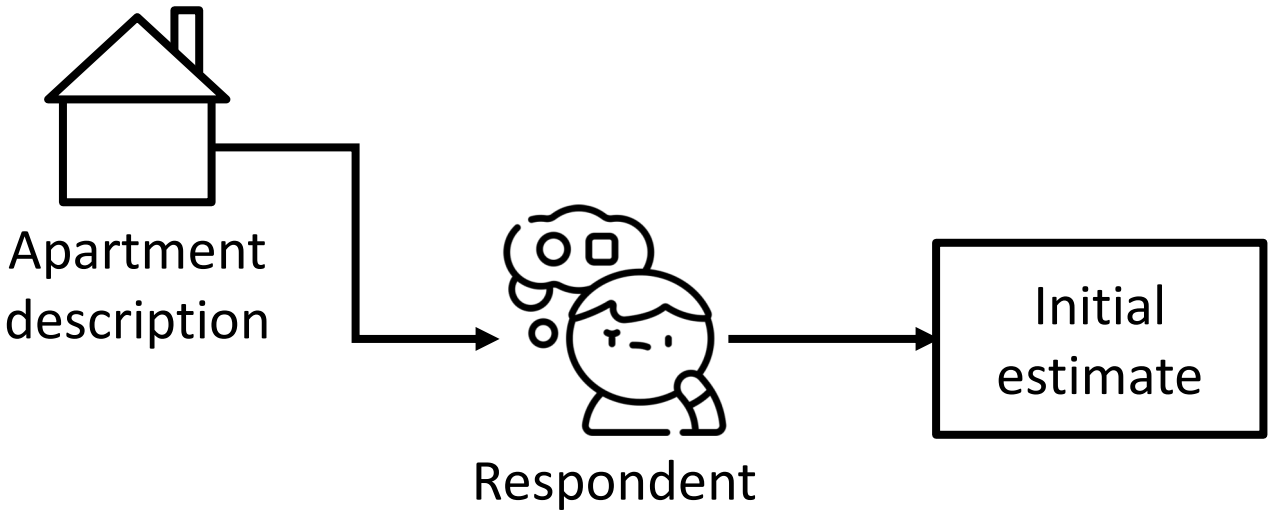
Human-subject experiment: **660** participants recruited via Prolific

Decision-making task: **real estate appraisal**

Can algorithmic assistance moderate human (in)consistency?

Human-subject experiment: **660** participants recruited via Prolific

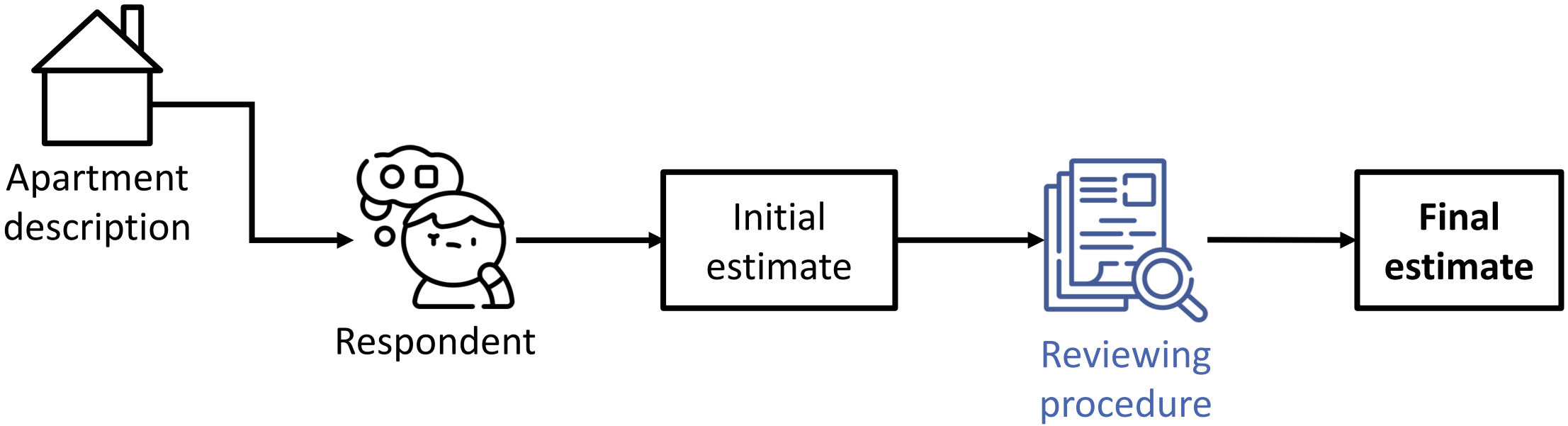
Decision-making task: **real estate appraisal**



Can algorithmic assistance moderate human (in)consistency?

Human-subject experiment: **660** participants recruited via Prolific

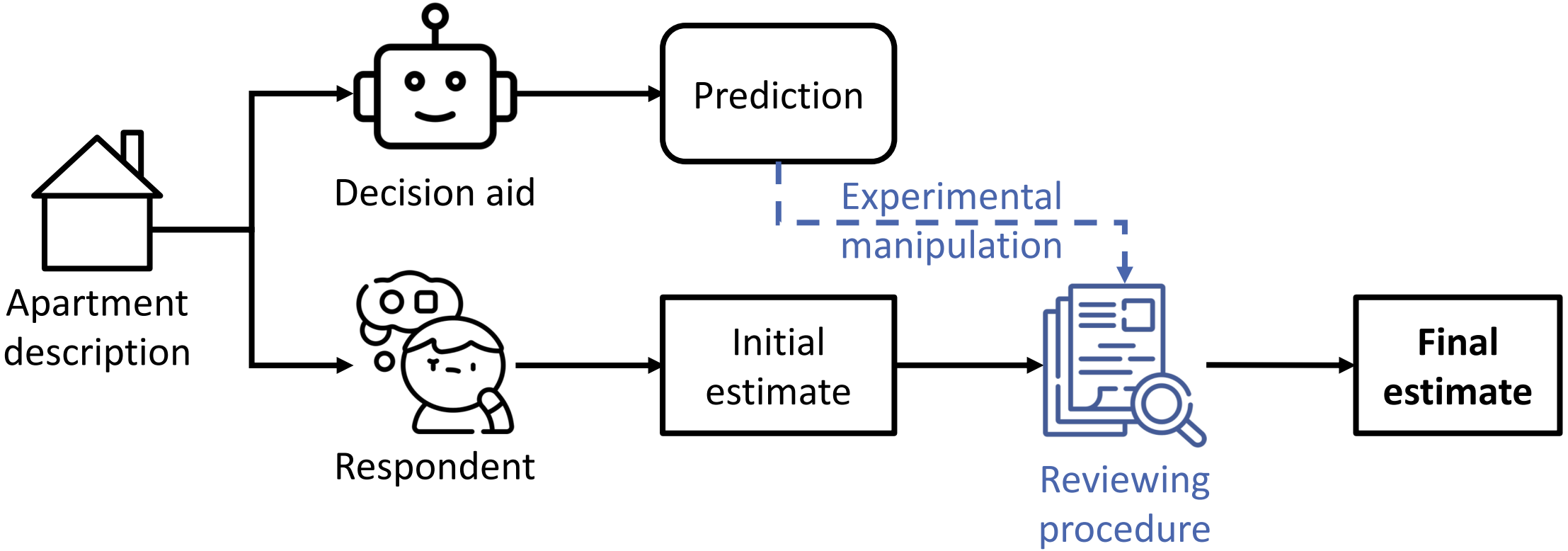
Decision-making task: **real estate appraisal**



Can algorithmic assistance moderate human (in)consistency?

Human-subject experiment: **660** participants recruited via Prolific

Decision-making task: **real estate appraisal**



Experimental Conditions

Experimental Conditions

		Reviewing Procedure	Algorithmic Assistance	Algorithm Predicts
	T1	one-by-one	/	/
	T2	one-by-one	**advice**	**true price**

Experimental Conditions

	Reviewing Procedure	Algorithmic Assistance	Algorithm Predicts
T1	one-by-one	/	/
T2	one-by-one	**advice**	**true price**

For the apartment below, the computer program estimated its price to be: **$1,800,000**
You initially estimated its price to be: **$1,300,000**

Please provide your responses again below. If you wish to change your initial response, please feel free to do so.

Bedrooms	2.0
Bathrooms	2
Square footage	1320
Total rooms	4.5
Monthly maintenance fee	$1,330
Days on the market	80
Subway distance (miles)	0.168
School distance (miles)	0.225
Your initial estimate	$1,300,000

Please provide your response below:

What do you think the apartment was sold for?

Experimental Conditions

		Reviewing Procedure	Algorithmic Assistance	Algorithm Predicts
	T1	one-by-one	/	/
	T2	one-by-one	**advice**	**true price**

For the apartment below, the computer program estimated its price to be: **$1,800,000**

You initially estimated its price to be: **$1,300,000**

Please provide your responses again below. If you wish to change your initial response, please feel free to do so.

Bedrooms	2.0
Bathrooms	2
Square footage	1320
Total rooms	4.5
Monthly maintenance fee	$1,330
Days on the market	80
Subway distance (miles)	0.168
School distance (miles)	0.225
Your initial estimate	$1,300,000

Please provide your response below:

What do you think the apartment was sold for?

Experimental Conditions

Prior work in social psychology: pairwise decisions are easier than absolute [Miller et al.; Stewart et al.]

Absolute identification by relative judgment

Neil Stewart [1], Gordon D A Brown, Nick Chater

Affiliations + expand

PMID: 16262472 DOI: 10.1037/0033-295X.112.4.881

Abstract

In unidimensional absolute identification tasks, participants identify stimuli that vary along a single dimension. Performance is surprisingly poor compared with discrimination of the same stimuli. Existing models assume that identification is achieved using long-term representations of absolute magnitudes. The authors propose an alternative relative judgment model (RJM) in which the elemental perceptual units are representations of the differences between current and previous stimuli. These differences are used, together with the previous feedback, to respond. Without using long-term representations of absolute magnitudes, the RJM accounts for (a) information transmission limits, (b) bowed serial position effects, and (c) sequential effects, where responses are biased toward immediately preceding stimuli but away from more distant stimuli (assimilation and contrast).

The Magical Number Seven, Plus or Minus Two
Some Limits on Our Capacity for Processing Information

George A. Miller
Harvard University

First, the span of absolute judgment and the span of immediate memory impose severe limitations on the amount of information that we are able to receive, process, and remember. By organizing the stimulus input simultaneously into several dimensions and successively into a sequence of chunks, we manage to break (or at least stretch) this informational bottleneck.

Second, the process of recoding is a very important one in human psychology and deserves much more explicit attention than it has received. In particular, the kind of linguistic recoding that people do seems

Experimental Conditions

	Reviewing Procedure	Algorithmic Assistance	Algorithm Predicts
 T1	one-by-one	/	/
 T2	one-by-one	**advice**	**true price**
 T3	in pairs	/	/

Please provide your responses again below. If you wish to change your initial response, please feel free to do so.

	Apartment A	Apartment B
Bedrooms	1.0	2.0
Bathrooms	1	2
Square footage	1240	1160
Total rooms	3	4
Monthly maintenance fee	$1,170	$1,330
Days on the market	119	71
Subway distance (miles)	0.149	0.026
School distance (miles)	0.323	0.323
Your initial estimate	**$1,200,000**	**$1,200,000**

Please provide your response below:

What do you think apartment A was sold for?

Please provide your response below:

What do you think apartment B was sold for?

Experimental Conditions

	Reviewing Procedure	Algorithmic Assistance	Algorithm Predicts
 T1	one-by-one	/	/
 T2	one-by-one	**advice**	**true price**
 T3	in pairs	/	/
 T4	in pairs	**pair selection**	**people's pairwise comparisons**

Please provide your responses again below. If you wish to change your initial response, please feel free to do so.

	Apartment A	Apartment B
Bedrooms	1.0	2.0
Bathrooms	1	2
Square footage	1240	1160
Total rooms	3	4
Monthly maintenance fee	$1,170	$1,330
Days on the market	119	71
Subway distance (miles)	0.149	0.026
School distance (miles)	0.323	0.323
Your initial estimate	**$1,200,000**	**$1,200,000**

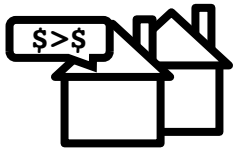
Please provide your response below:

What do you think apartment A was sold for?

Please provide your response below:

What do you think apartment B was sold for?

Experimental Conditions

		Reviewing Procedure	Algorithmic Assistance	Algorithm Predicts
	T1	one-by-one	/	/
	T2	one-by-one	**advice**	**true price**
	T3	in pairs	/	/
	T4	in pairs	**pair selection**	**people's pairwise comparisons**
	T5	in pairs	**pair selection + advice**	**people's pairwise comparisons**

For the two apartments shown below, our computer program estimated that **Apartment A is less expensive than Apartment B**.

However, you estimated that **Apartment A is equally expensive as Apartment B** .

Please provide your responses again below. If you wish to change your initial response, please feel free to do so.

	Apartment A	Apartment B
Bedrooms	1.0	2.0
Bathrooms	1	2
Square footage	1240	1160
Total rooms	3	4
Monthly maintenance fee	$1,170	$1,330
Days on the market	119	71
Subway distance (miles)	0.149	0.026
School distance (miles)	0.323	0.323
Your initial estimate	**$1,200,000**	**$1,200,000**

Please provide your response below:

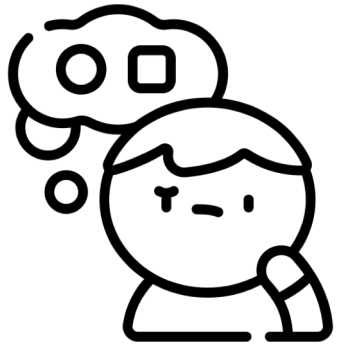
What do you think apartment A was sold for?

Please provide your response below:

What do you think apartment B was sold for?

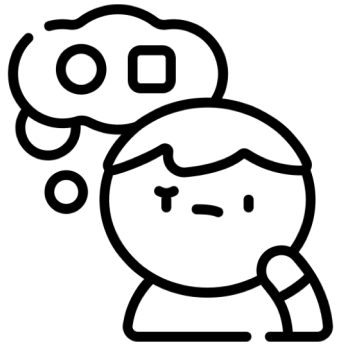
Manual reviewing

(T1, T3)



Manual reviewing

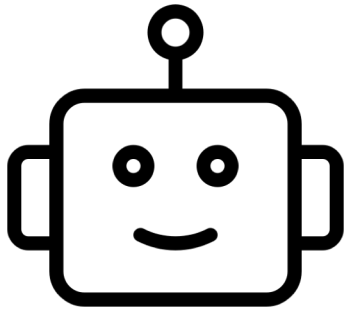
(T1, T3)



VS

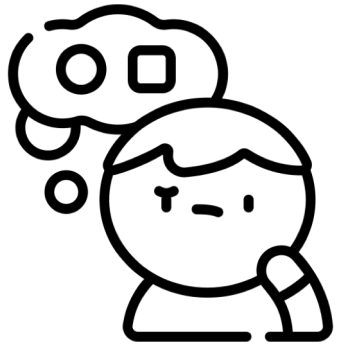
Reviewing with
algorithmic assistance

(T2, T4, T5)



Manual reviewing

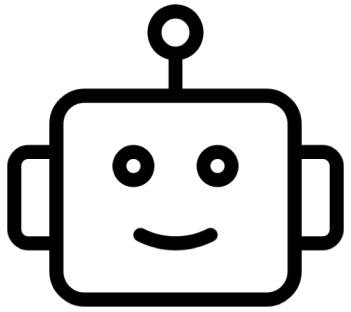
(T1, T3)



VS

Reviewing with **algorithmic assistance**

(T2, T4, T5)

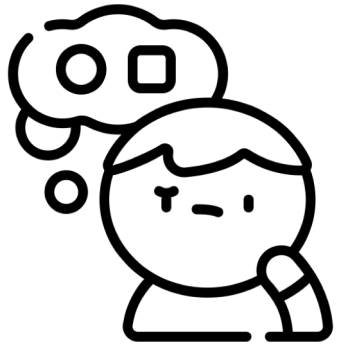


Which reviewing procedure will lead to

- H1: More **updates** in initial estimates?

Manual reviewing

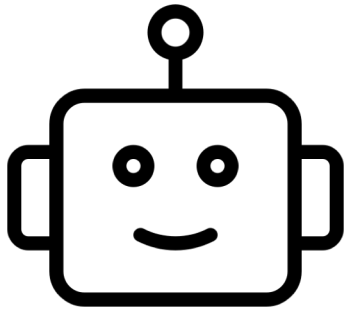
(T1, T3)



VS

Reviewing with **algorithmic assistance**

(T2, T4, T5)

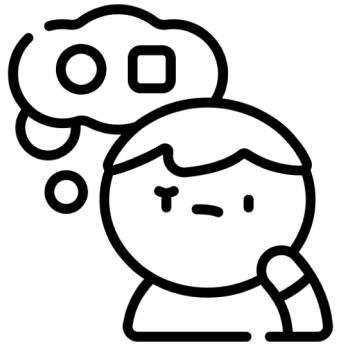


Which reviewing procedure will lead to

- H1: More **updates** in initial estimates?
- H2: More **accurate** final estimates?

Manual reviewing

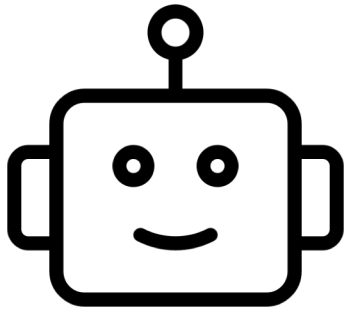
(T1, T3)



VS

Reviewing with **algorithmic assistance**

(T2, T4, T5)



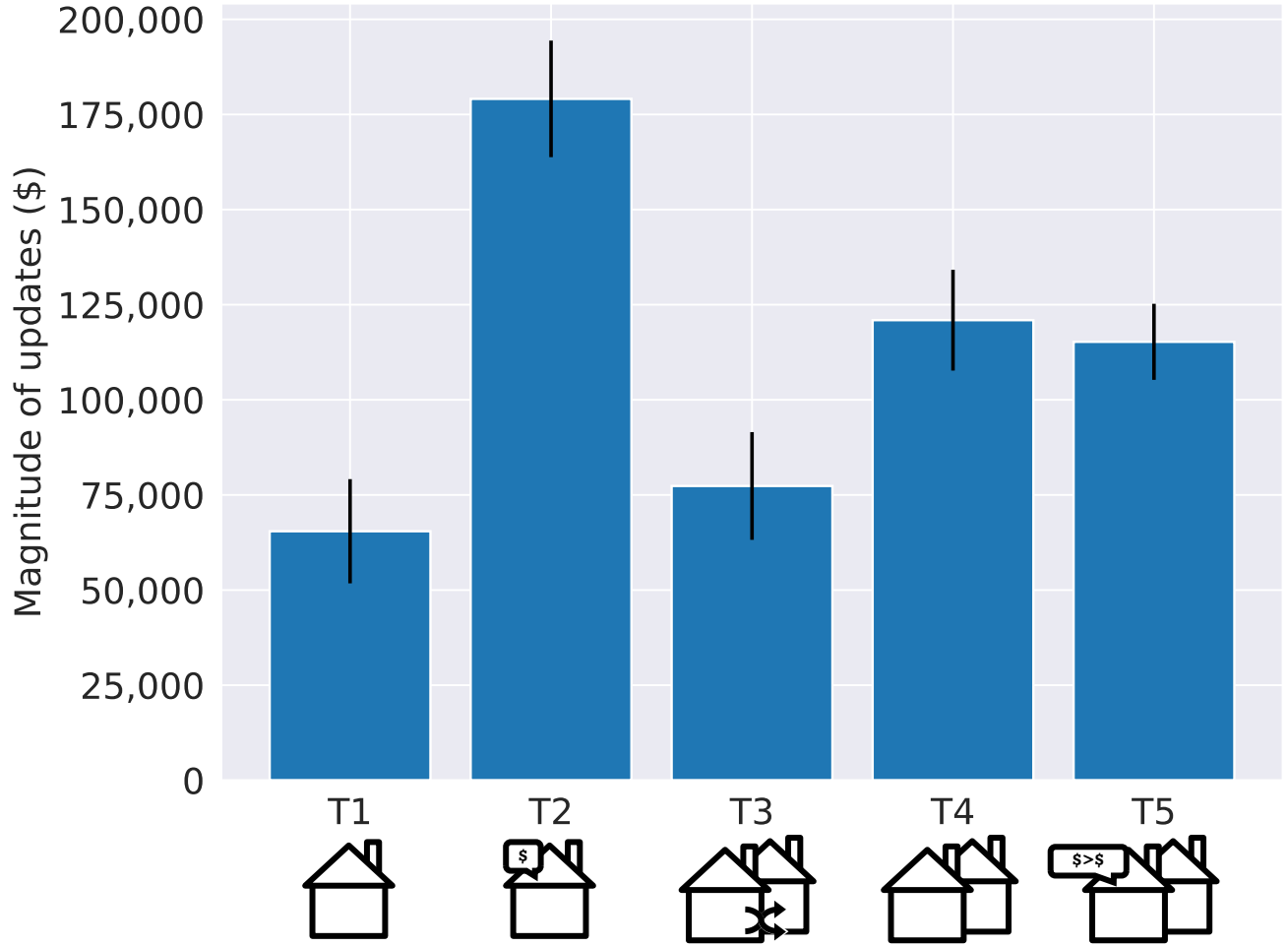
Which reviewing procedure will lead to

- H1: More **updates** in initial estimates?
- H2: More **accurate** final estimates?
- H3: More **consistency** between participants' estimates?

Compared to **manual reviewing** (T1, T3), **reviewing with algorithmic assistance** (T2, T4, T5) leads to:

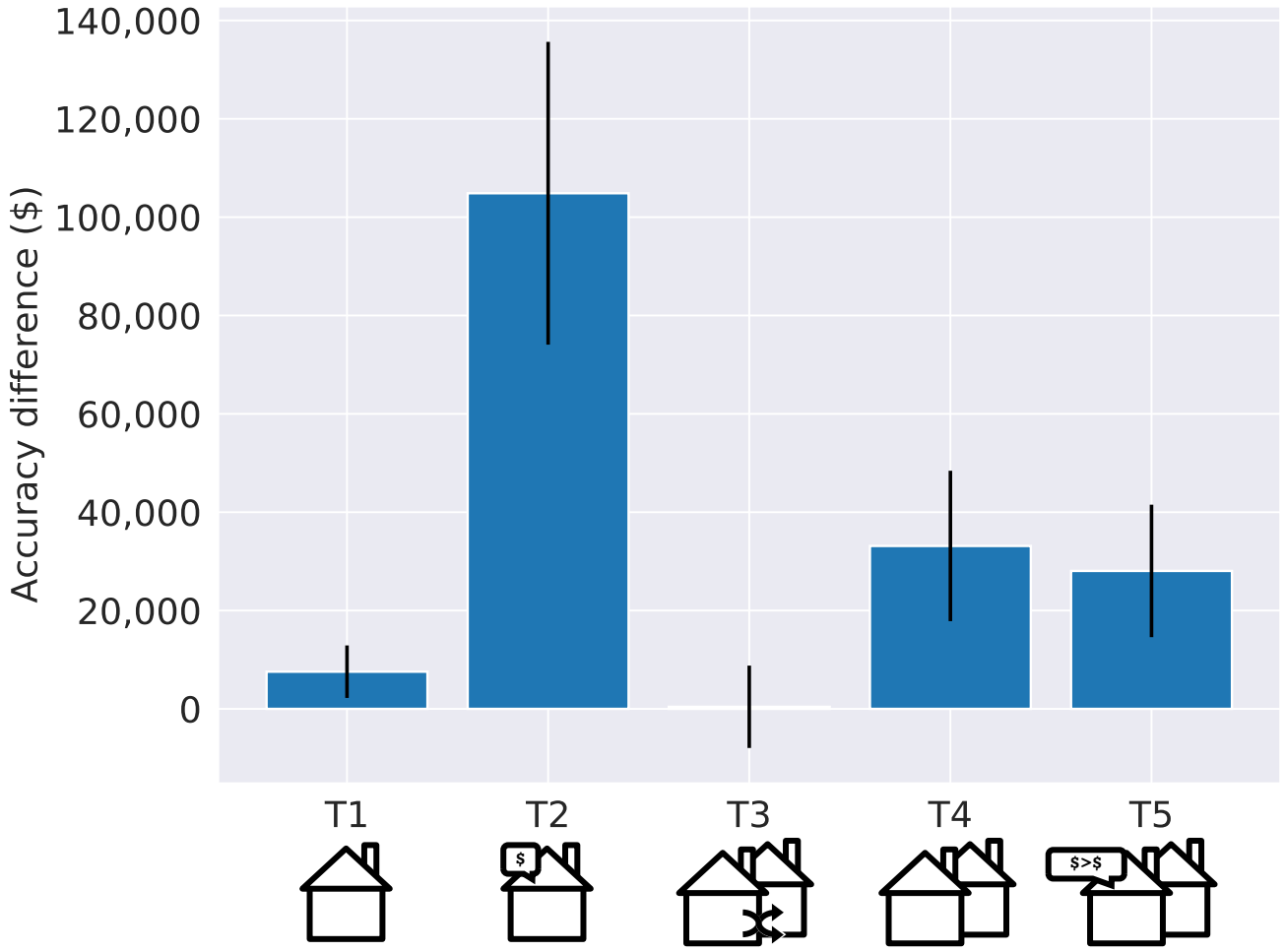
Compared to **manual reviewing** (T1, T3), **reviewing with algorithmic assistance** (T2, T4, T5) leads to:

- H1: A **larger magnitude** of **updates** of estimates



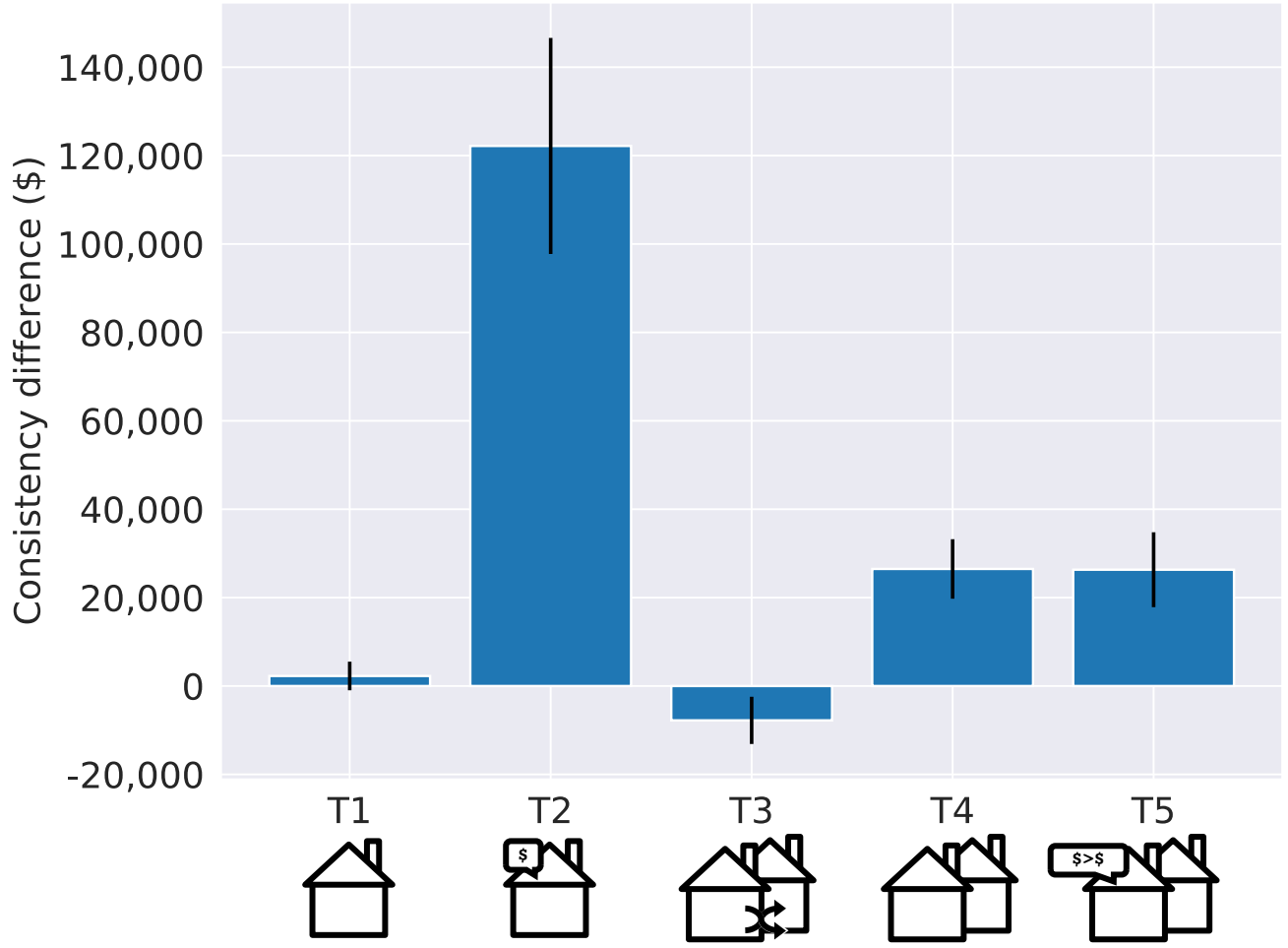
Compared to **manual reviewing** (T1, T3), **reviewing with algorithmic assistance** (T2, T4, T5) leads to:

- H1: A **larger magnitude** of **updates** of estimates
- H2: a **larger increase** in **accuracy** of estimates



Compared to **manual reviewing** (T1, T3), **reviewing with algorithmic assistance** (T2, T4, T5) leads to:

- H1: A **larger magnitude** of **updates** of estimates
- H2: a **larger increase** in **accuracy** of estimates
- H3: a **larger increase** in **consistency** with others' estimates



Compared to **manual reviewing** (T1, T3), **reviewing with algorithmic assistance** (T2, T4, T5) leads to:

- H1: A **larger magnitude** of **updates** of estimates
- H2: a **larger increase** in **accuracy** of estimates
- H3: a **larger increase** in **consistency** with others' estimates

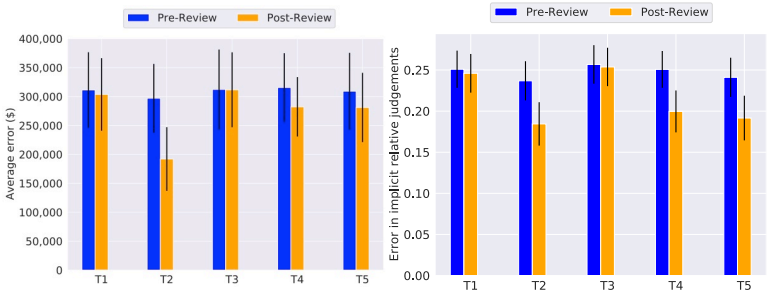
Compared to **manual reviewing** (T1, T3), **reviewing with algorithmic assistance** (T2, T4, T5) leads to:

- H1: A **larger magnitude** of **updates** of estimates
- H2: a **larger increase** in **accuracy** of estimates
- H3: a **larger increase** in **consistency** with others' estimates

} Are these findings **robust**?

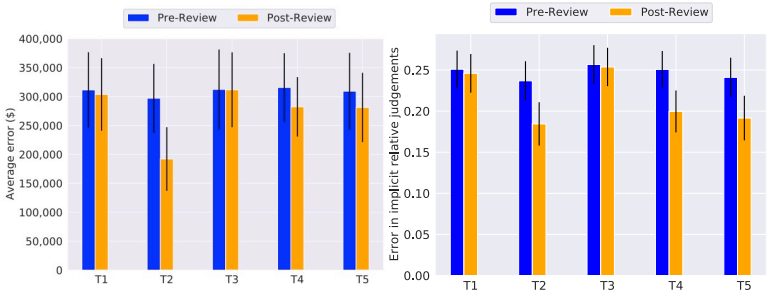
Our findings are **robust** wrt **various different measures** of

Accuracy

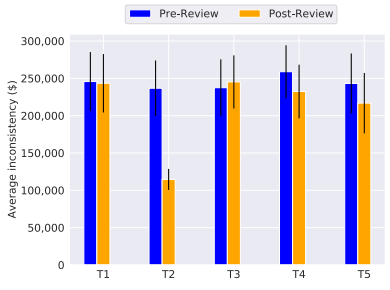


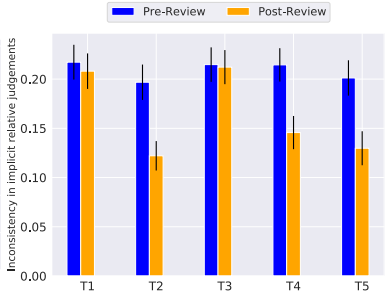
Our findings are **robust** wrt **various different measures** of

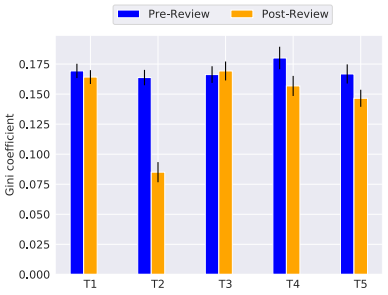
Accuracy



Between-respondent **consistency**

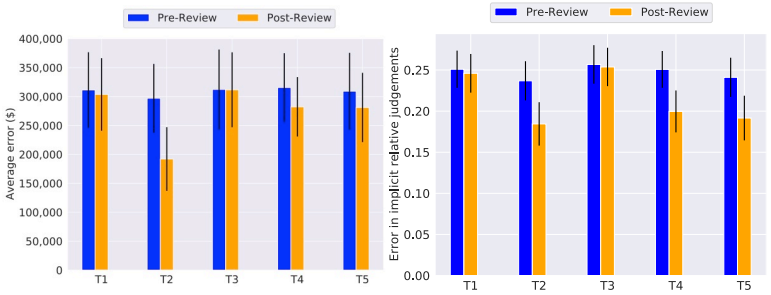




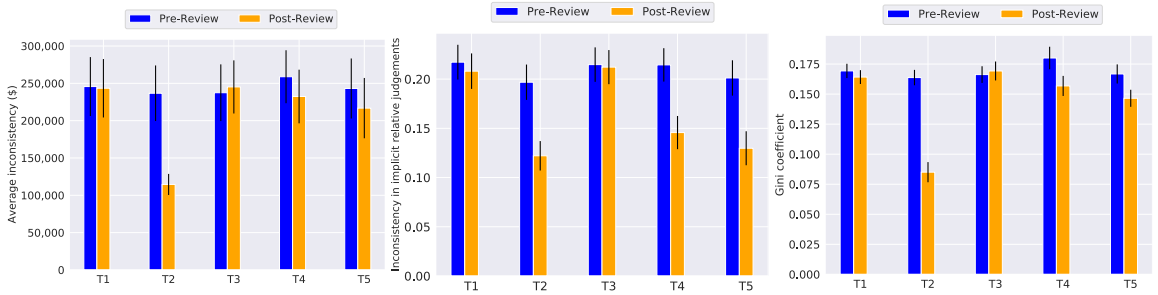


Our findings are **robust** wrt **various different measures** of

Accuracy



Between-respondent consistency

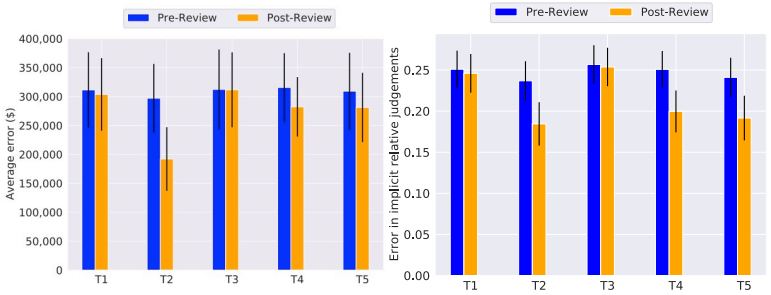


Future work:

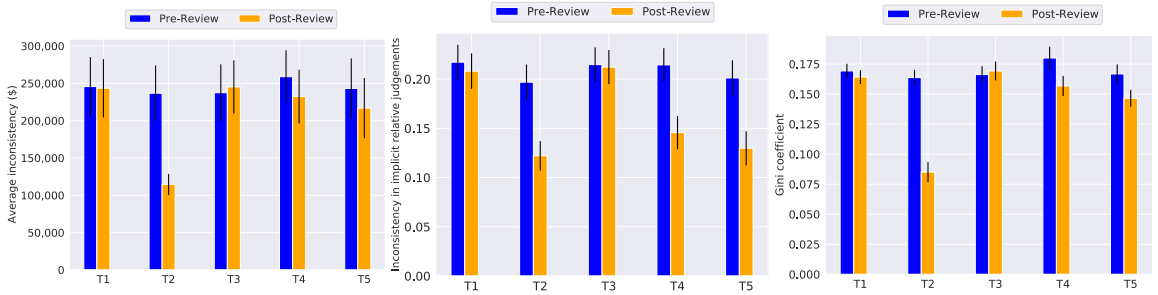
- Robustness
  - Decision-making tasks
  - Respondent samples
  - Types of algorithmic assistance

Our findings are **robust** wrt **various different measures** of

Accuracy



Between-respondent consistency



Future work:

- Robustness
  - Decision-making tasks
  - Respondent samples
  - Types of algorithmic assistance
- Different notions of consistency
  - Consistency of an **individual decision-maker** across **time** & **inputs**

Takeaways

Algorithmic decision aids **increase** respondents' **accuracy** and **consistency**

Takeaways

Algorithmic decision aids **increase** respondents' **accuracy** and **consistency**

With access to **ground truth** data

- **T2** is most effective

Takeaways

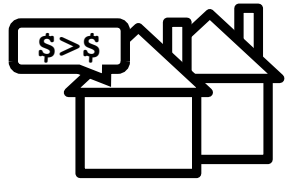
Algorithmic decision aids **increase** respondents' **accuracy** and **consistency**

With access to **ground truth** data

- **T2** is most effective

Without access to **ground truth** data

- **T4** and **T5** are still applicable and effective!



Takeaways

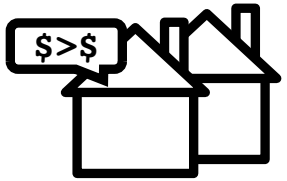
Algorithmic decision aids **increase** respondents' **accuracy** and **consistency**

With access to **ground truth** data

- **T2** is most effective

Without access to **ground truth** data

- **T4** and **T5** are still applicable and effective!



Without providing **explicit advice**

- **T4** is still applicable and effective!

Takeaways

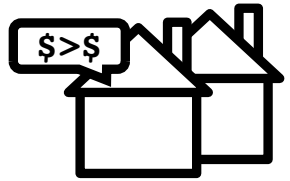
Algorithmic decision aids **increase** respondents' **accuracy** and **consistency**

With access to **ground truth** data

- **T2** is most effective

Without access to **ground truth** data

- **T4** and **T5** are still applicable and effective!



Without providing **explicit advice**

- **T4** is still applicable and effective!

Read the full paper here!

illustrations: Flaticon.com