"What if?" Predicting Individual Users' Smart Home Privacy Preferences and Their Changes

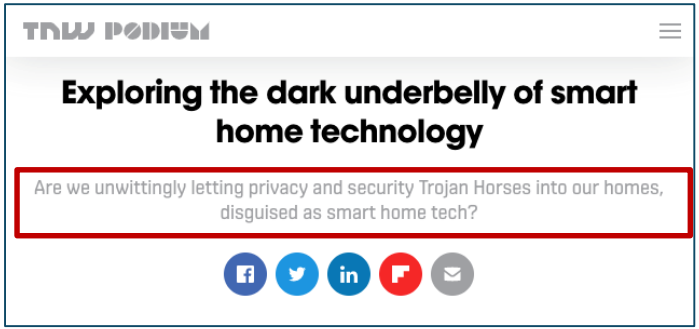
Natã M. Barbosa, Joon S. Park, **Yaxing Yao**, Yang Wang

[download: http://bit.do/what-if-smart-home]

PRIVACYCON

Motivation

- While smart home adoption grows, privacy is a concern:
  - Secondary use
  - Appropriation



- Violation of the home's privacy norms
- Spanning of long-settled boundaries

Our Goal

Enable developers to derive actionable steps toward respecting the privacy of smart home users in a personalized way.

How?

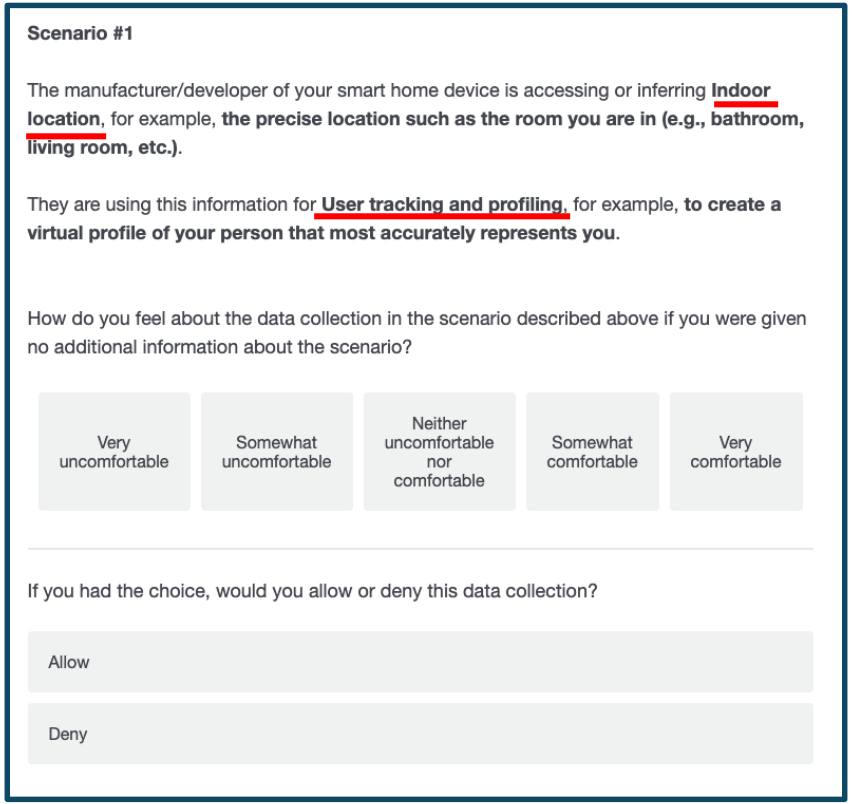
1. Predict Allow/Deny preferences

2. Identify preference-changing circumstances

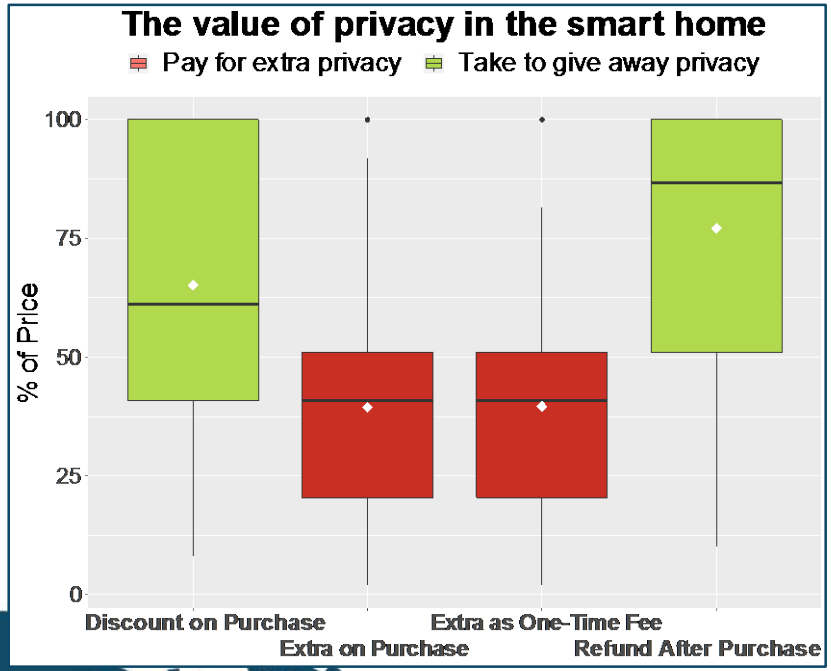
3. Predict dollar value of smart home privacy

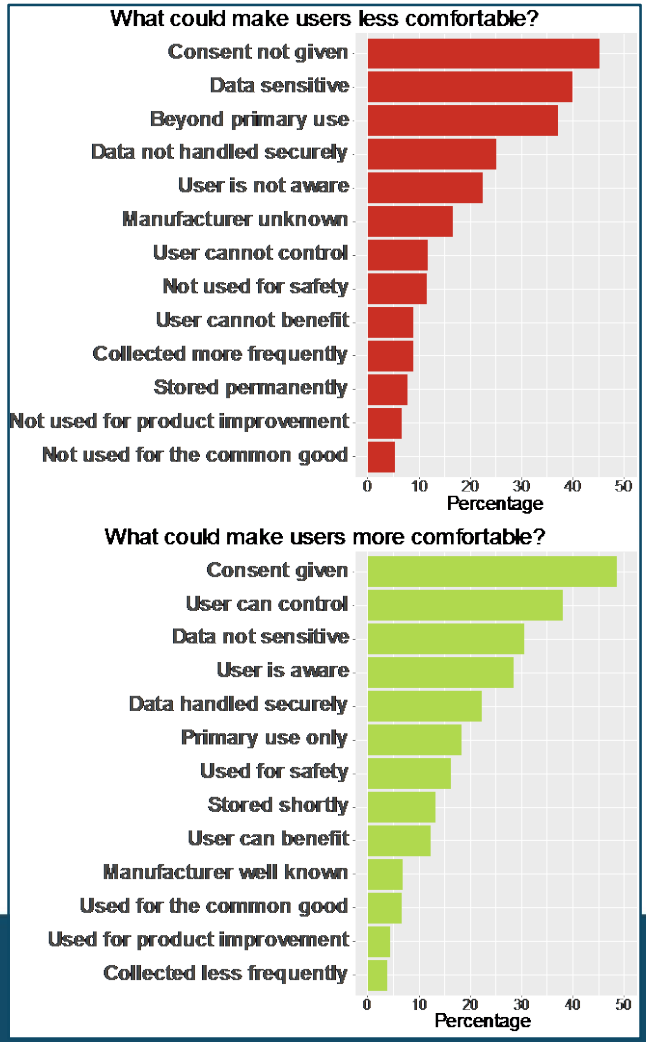
Scenario-Based Survey

on Amazon Mechanical Turk (N=698)



Survey Data → Actionable Steps





PRIVACYCON

Machine Learning Models

Built with PySpark and scikit-learn

PRIVACYCON

Actionable steps from model predictions

#	Attribute	Purpose	What if... (situational factor selected)	% more comfortable	% less comfortable
1	Any	Any	user can control or not?	**84.6%**	15.4%
2	Any	Any	data handled securely or not?	43.9%	**56.1%**
3	Any	Any	used only for primary purposes or not?	32.9%	**67.1%**
4	Any	Any	user is aware or not?	**63.3%**	36.7%
5	Any	Any	used for safety or not?	**69.8%**	30.2%
6	Any	Targeted ads	user can control or not?	**95.5%**	4.5%
7	Any	Targeted ads	user is aware or not?	**79.9%**	20.1%
8	Indoor location	Any	used only for primary purposes or not?	36.2%	**63.8%**
9	Indoor location	Any	user can control or not?	**92.1%**	7.9%
10	Indoor location	Any	user has consented or not?	**57.1%**	42.9%
11	Indoor location	Any	manufacturer well known or not?	**51.8%**	48.2%
12	Age of people at home	Any	manufacturer well known or not?	**73.6%**	26.4%
13	Any	Home safety	manufacturer well known or not?	13.8%	**86.2%**
14	Any	Home safety	used only for primary purposes or not?	5.5%	**94.5%**
15	Energy use	Targeted ads	user is aware or not?	**68%**	32%
16	Energy use	Targeted ads	user can benefit or not?	**86.8%**	13.2%

Table 2. Model predictions for the "average" user in the scenarios data set. Percentages indicate the number of scenarios for which "more" or "less" comfortable was predicted when the situational factor is present i.e. "checked." Comfort changes can be identified using combinations within information flows considering specific attributes, purposes, or devices, as well as for any levels of such factors.

Actionable steps from model predictions

#	Attribute	Purpose	What if... (situational factor selected)	% more comfortable	% less comfortable
1	Any	Any	user can control or not?	**84.6%**	15.4%
2	Any	Any	data handled securely or not?	43.9%	**56.1%**
3	Any	Any	used only for primary purposes or not?	32.9%	**67.1%**
4	Any	Any	user is aware or not?	**63.3%**	36.7%
5	Any	Any	used for safety or not?	**69.8%**	30.2%
6	Any	Targeted ads	user can control or not?	**95.5%**	4.5%
7	Any	Targeted ads	user is aware or not?	**79.9%**	20.1%
8	Indoor location	Any	used only for primary purposes or not?	36.2%	**63.8%**
9	Indoor location	Any	user can control or not?	**92.1%**	7.9%
10	Indoor location	Any	user has consented or not?	**57.1%**	42.9%
11	Indoor location	Any	manufacturer well known or not?	**51.8%**	48.2%
12	Age of people at home	Any	manufacturer well known or not?	**73.6%**	26.4%
13	Any	Home safety	manufacturer well known or not?	13.8%	**86.2%**
14	Any	Home safety	used only for primary purposes or not?	5.5%	**94.5%**
15	Energy use	Targeted ads	user is aware or not?	**68%**	32%
16	Energy use	Targeted ads	user can benefit or not?	**86.8%**	13.2%

Table 2. Model predictions for the "average" user in the scenarios data set. Percentages indicate the number of scenarios for which "more" or "less" comfortable was predicted when the situational factor is present i.e. "checked." Comfort changes can be identified using combinations within information flows considering specific attributes, purposes, or devices, as well as for any levels of such factors.

Actionable steps from model predictions

#	Attribute	Purpose	What if... (situational factor selected)	% more comfortable	% less comfortable
1	Any	Any	user can control or not?	**84.6%**	15.4%
2	Any	Any	data handled securely or not?	43.9%	**56.1%**
3	Any	Any	used only for primary purposes or not?	32.9%	**67.1%**
4	Any	Any	user is aware or not?	**63.3%**	36.7%
5	Any	Any	used for safety or not?	**69.8%**	30.2%
6	Any	Targeted ads	user can control or not?	**95.5%**	4.5%
7	Any	Targeted ads	user is aware or not?	**79.9%**	20.1%
8	Indoor location	Any	used only for primary purposes or not?	36.2%	**63.8%**
9	Indoor location	Any	user can control or not?	**92.1%**	7.9%
10	Indoor location	Any	user has consented or not?	**57.1%**	42.9%
11	Indoor location	Any	manufacturer well known or not?	**51.8%**	48.2%
12	Age of people at home	Any	manufacturer well known or not?	**73.6%**	26.4%
13	Any	Home safety	manufacturer well known or not?	13.8%	**86.2%**
14	Any	Home safety	used only for primary purposes or not?	5.5%	**94.5%**
15	Energy use	Targeted ads	user is aware or not?	**68%**	32%
16	Energy use	Targeted ads	user can benefit or not?	**86.8%**	13.2%

Table 2. Model predictions for the "average" user in the scenarios data set. Percentages indicate the number of scenarios for which "more" or "less" comfortable was predicted when the situational factor is present i.e. "checked." Comfort changes can be identified using combinations within information flows considering specific attributes, purposes, or devices, as well as for any levels of such factors.

Actionable steps from model predictions

#	Attribute	Purpose	What if... (situational factor selected)	% more comfortable	% less comfortable
1	Any	Any	user can control or not?	**84.6%**	15.4%
2	Any	Any	data handled securely or not?	43.9%	**56.1%**
3	Any	Any	used only for primary purposes or not?	32.9%	**67.1%**
4	Any	Any	user is aware or not?	**63.3%**	36.7%
5	Any	Any	used for safety or not?	**69.8%**	30.2%
6	Any	Targeted ads	user can control or not?	**95.5%**	4.5%
7	Any	Targeted ads	user is aware or not?	**79.9%**	20.1%
8	Indoor location	Any	used only for primary purposes or not?	36.2%	**63.8%**
9	Indoor location	Any	user can control or not?	**92.1%**	7.9%
10	Indoor location	Any	user has consented or not?	**57.1%**	42.9%
11	Indoor location	Any	manufacturer well known or not?	**51.8%**	48.2%
12	Age of people at home	Any	manufacturer well known or not?	**73.6%**	26.4%
13	Any	Home safety	manufacturer well known or not?	13.8%	**86.2%**
14	Any	Home safety	used only for primary purposes or not?	5.5%	**94.5%**
15	Energy use	Targeted ads	user is aware or not?	**68%**	32%
16	Energy use	Targeted ads	user can benefit or not?	**86.8%**	13.2%

Table 2. Model predictions for the "average" user in the scenarios data set. Percentages indicate the number of scenarios for which "more" or "less" comfortable was predicted when the situational factor is present i.e. "checked." Comfort changes can be identified using combinations within information flows considering specific attributes, purposes, or devices, as well as for any levels of such factors.

Actionable steps from model predictions

#	Attribute	Purpose	What if... (situational factor selected)	% more comfortable	% less comfortable
1	Any	Any	user can control or not?	**84.6%**	15.4%
2	Any	Any	data handled securely or not?	43.9%	**56.1%**
3	Any	Any	used only for primary purposes or not?	32.9%	**67.1%**
4	Any	Any	user is aware or not?	**63.3%**	36.7%
5	Any	Any	used for safety or not?	**69.8%**	30.2%
6	Any	Targeted ads	user can control or not?	**95.5%**	4.5%
7	Any	Targeted ads	user is aware or not?	**79.9%**	20.1%
8	Indoor location	Any	used only for primary purposes or not?	36.2%	**63.8%**
9	Indoor location	Any	user can control or not?	**92.1%**	7.9%
10	Indoor location	Any	user has consented or not?	**57.1%**	42.9%
11	Indoor location	Any	manufacturer well known or not?	**51.8%**	48.2%
12	Age of people at home	Any	manufacturer well known or not?	**73.6%**	26.4%
13	Any	Home safety	manufacturer well known or not?	13.8%	**86.2%**
14	Any	Home safety	used only for primary purposes or not?	5.5%	**94.5%**
15	Energy use	Targeted ads	user is aware or not?	**68%**	32%
16	Energy use	Targeted ads	user can benefit or not?	**86.8%**	13.2%

Table 2. Model predictions for the "average" user in the scenarios data set. Percentages indicate the number of scenarios for which "more" or "less" comfortable was predicted when the situational factor is present i.e. "checked." Comfort changes can be identified using combinations within information flows considering specific attributes, purposes, or devices, as well as for any levels of such factors.

Value Prediction

- Average user in data set, $49 voice assistant

	Willing to Protect	Willing to Accept
After Purchase	$28.01	$44.48
Before Purchase	$31.24	$38.03

Value Prediction

- Average user in data set, $49 voice assistant

	Willing to Protect	Willing to Accept
After Purchase	$28.01	$44.48
Before Purchase	$31.24	$38.03

Value Prediction

- Average user in data set, $49 voice assistant

	Willing to Protect	Willing to Accept
After Purchase	$28.01	$44.48
Before Purchase	$31.24	$38.03

Value Prediction

- Average user in data set, $49 voice assistant

	Willing to Protect	Willing to Accept
After Purchase	$28.01	$44.48
Before Purchase	$31.24	$38.03

Discussion

- Secondary uses not OK, circumstances do apply
- Consumers are loss averse, but many expect privacy by default
- *“So what? Adoption will grow regardless”*
- **Prevent the home from becoming a place where privacy is no longer included by default**

Takeaway

A smart home developer can reproduce our work to identify [in]appropriate data practices and take actionable steps towards respecting the privacy of their user base at scale.

Thank you!

- PETS paper: http://bit.do/what-if-smart-home
- Thanks to:
  - NSF
  - SALT Lab
  - Daniel Acuña
  - Participants, reviewers, and shepherd

Natã Barbosa
nmbarbos@syr.edu

Joon S. Park
jspark@syr.edu

Yaxing Yao
yyao08@syr.edu

Yang Wang
ywang@syr.edu