Privacy and Recommender Systems

Adaptive Decision Support Systems

Today's reading

Privacy for recommender systems

- Architectural, algorithmic, and regulatory solutions

Recommender systems for privacy

- Help users make the right privacy decisions with User Tailored Privacy

Today's lecture will focus on the latter

"Good" privacy decisions?

People weigh the risks and benefits of disclosure.

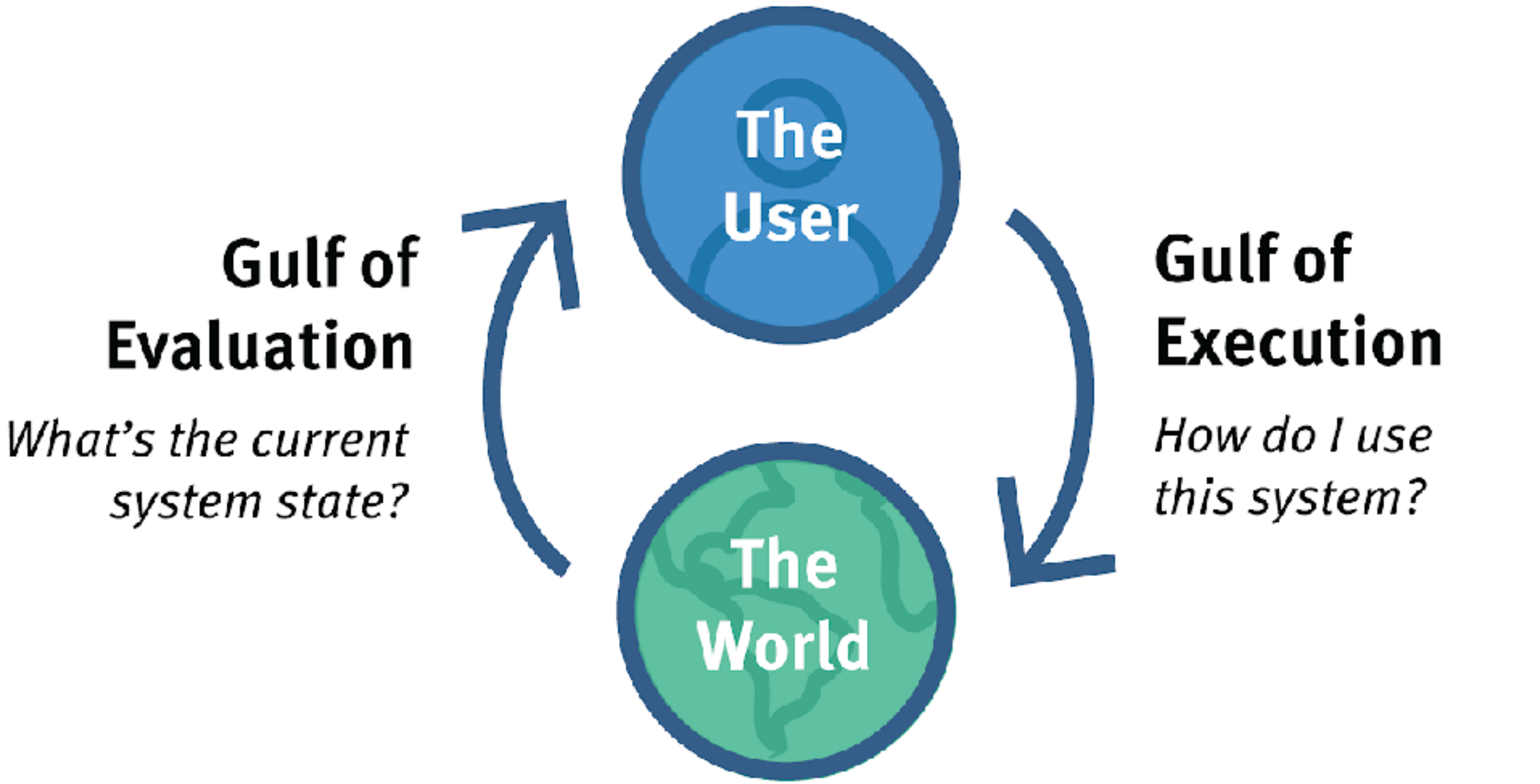
Laufer and Wolfe (1977)

Privacy Calculus

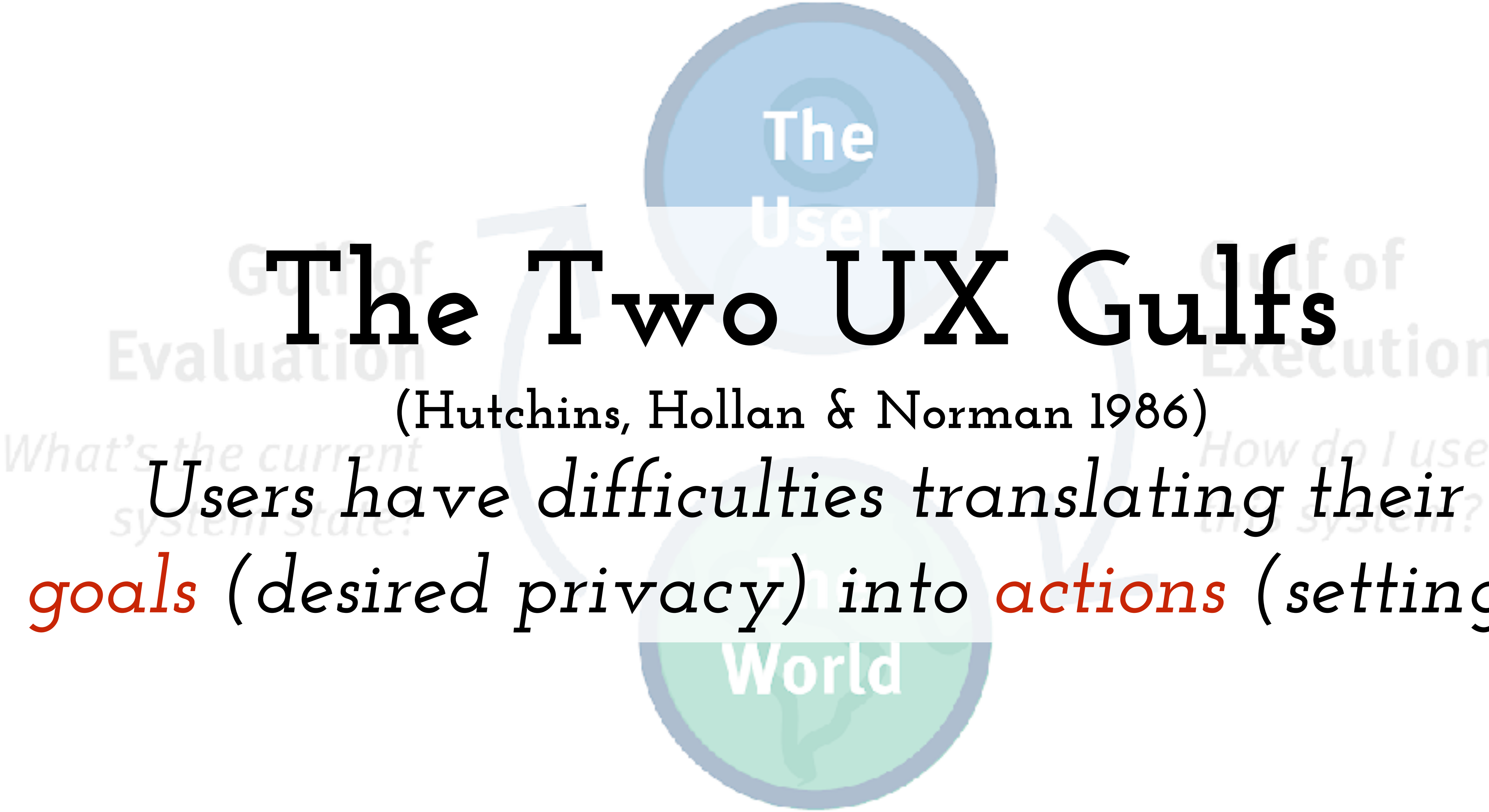
Culnan and Bie (1993)

"Good" privacy decisions optimize the balance between risks and benefits.

But how do we determine these risks and benefits? And how do we optimize them?



The User
Gulf of Evaluation
What's the current system state?
Gulf of Execution
How do I use this system?
The World
nngroup.com NN/g



The Two UX Gulfs

(Hutchins, Hollan & Norman 1986)

Users have difficulties translating their goals (desired privacy) into actions (settings).

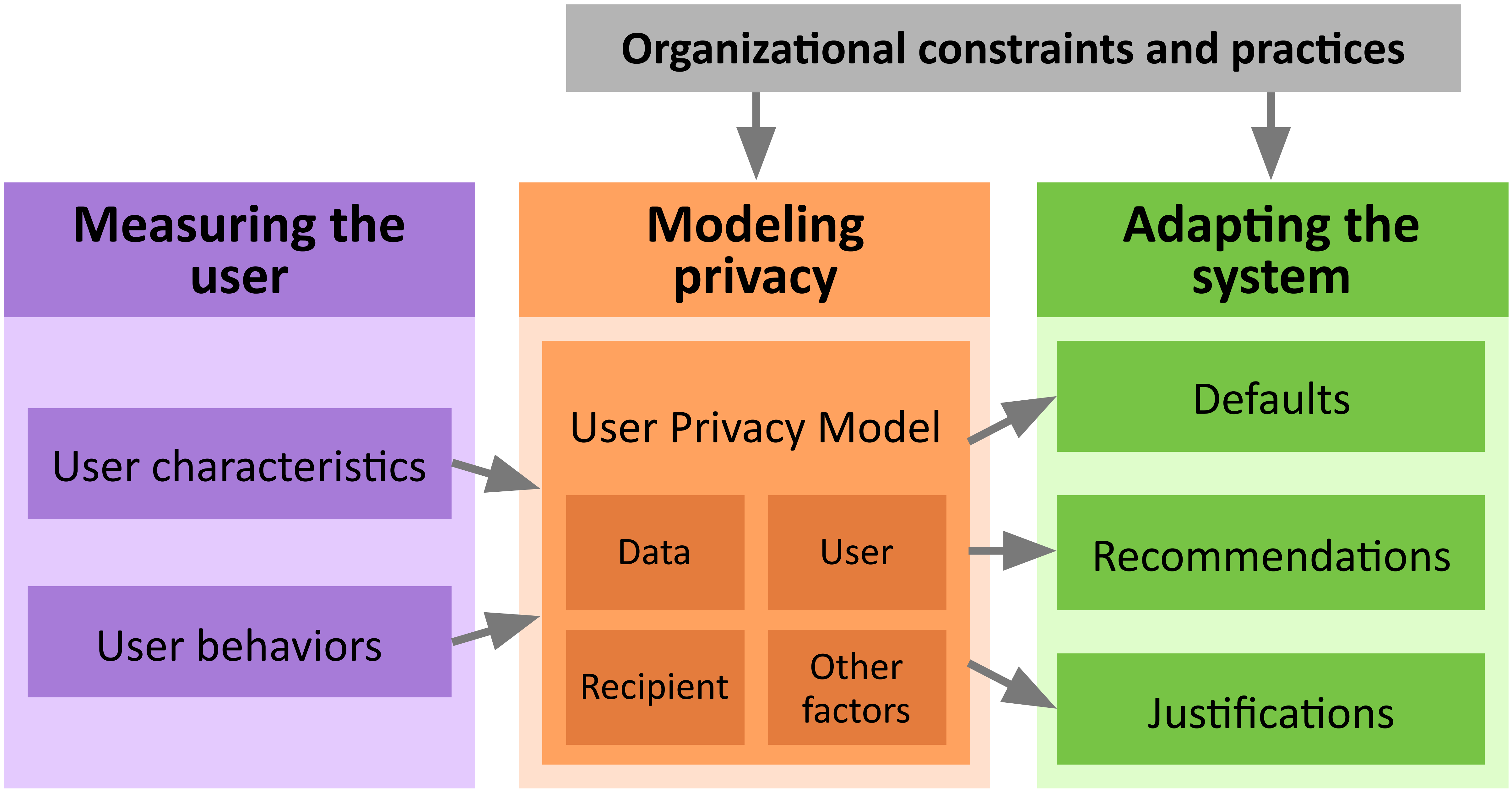
Why user-tailored privacy?

Problems with transparency and control,
and with privacy nudges.

User-Tailored Privacy

Privacy recommendations: Figure out what people want, then help them do that.

User-Tailored Privacy

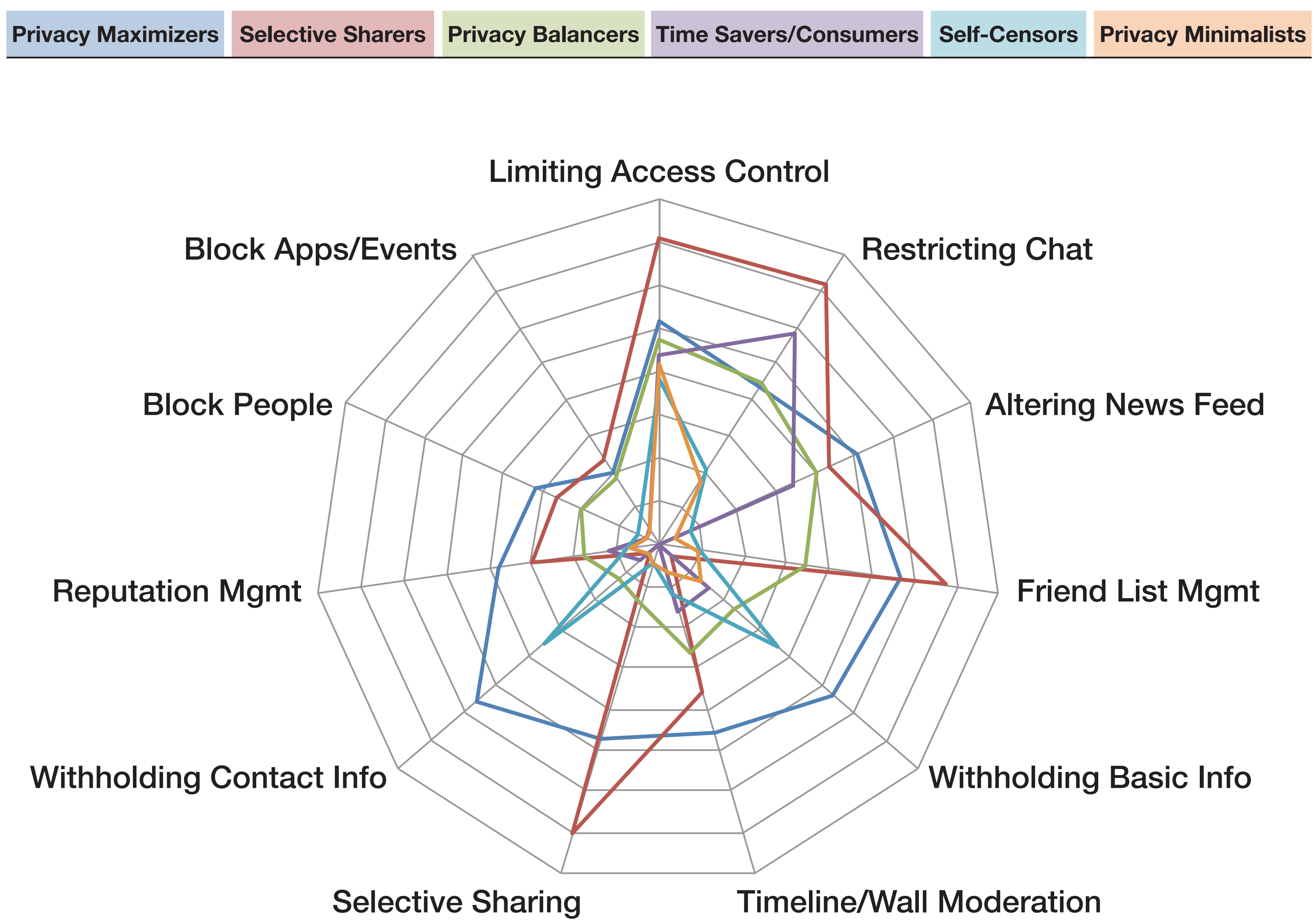


Example 1:

Facebook privacy management practices

32 individual privacy behaviors that Facebook users could perform using the native Facebook interface

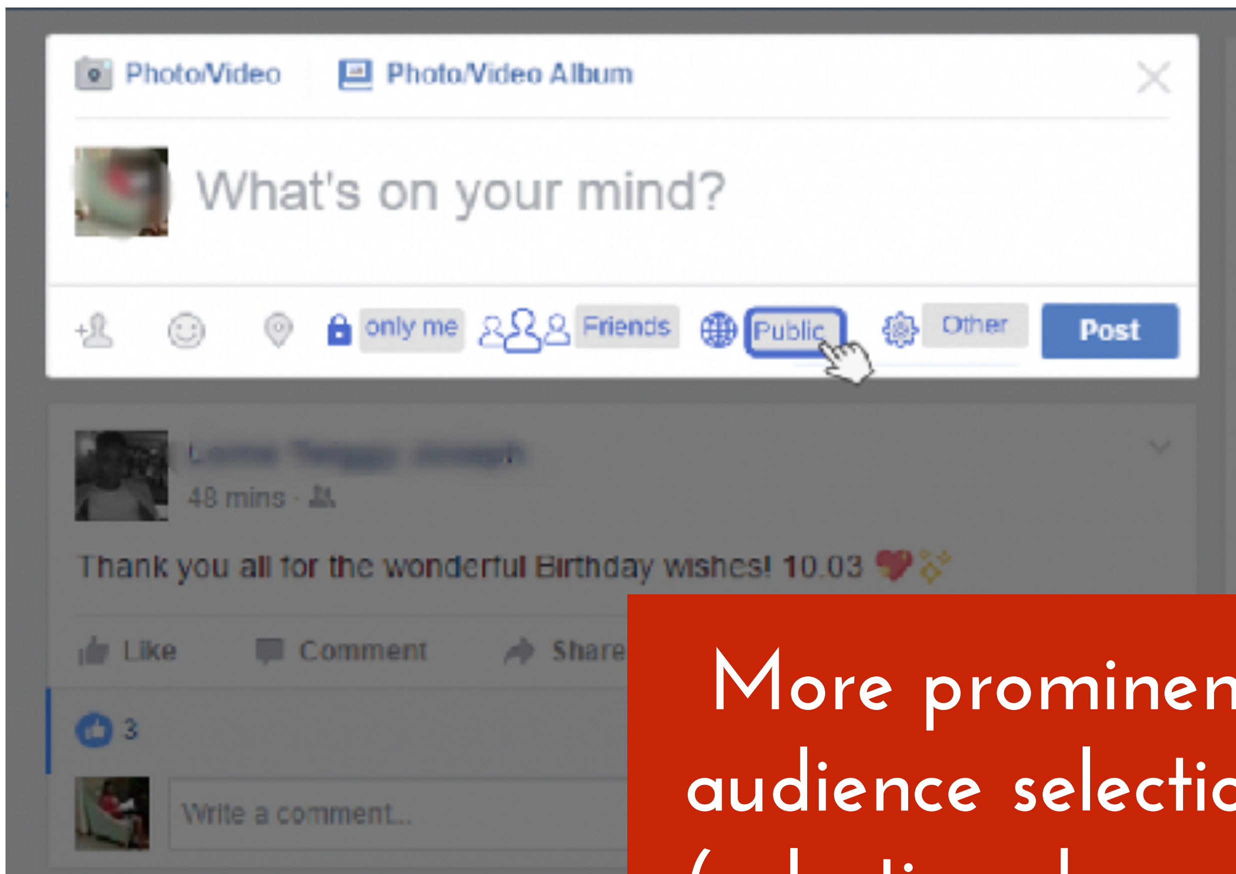
Create privacy profiles



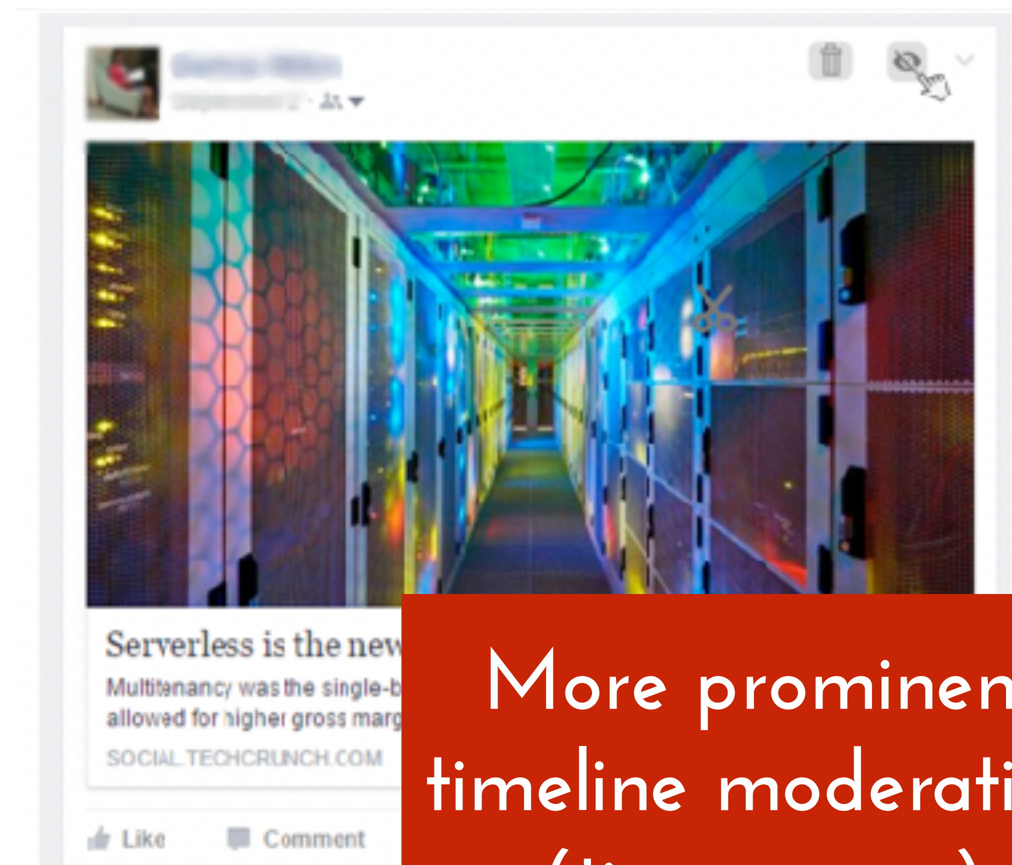
Wisniewski et al., "Making privacy personal", IJHCS 2017

see www.usabart.nl/chart

Adapt the interface



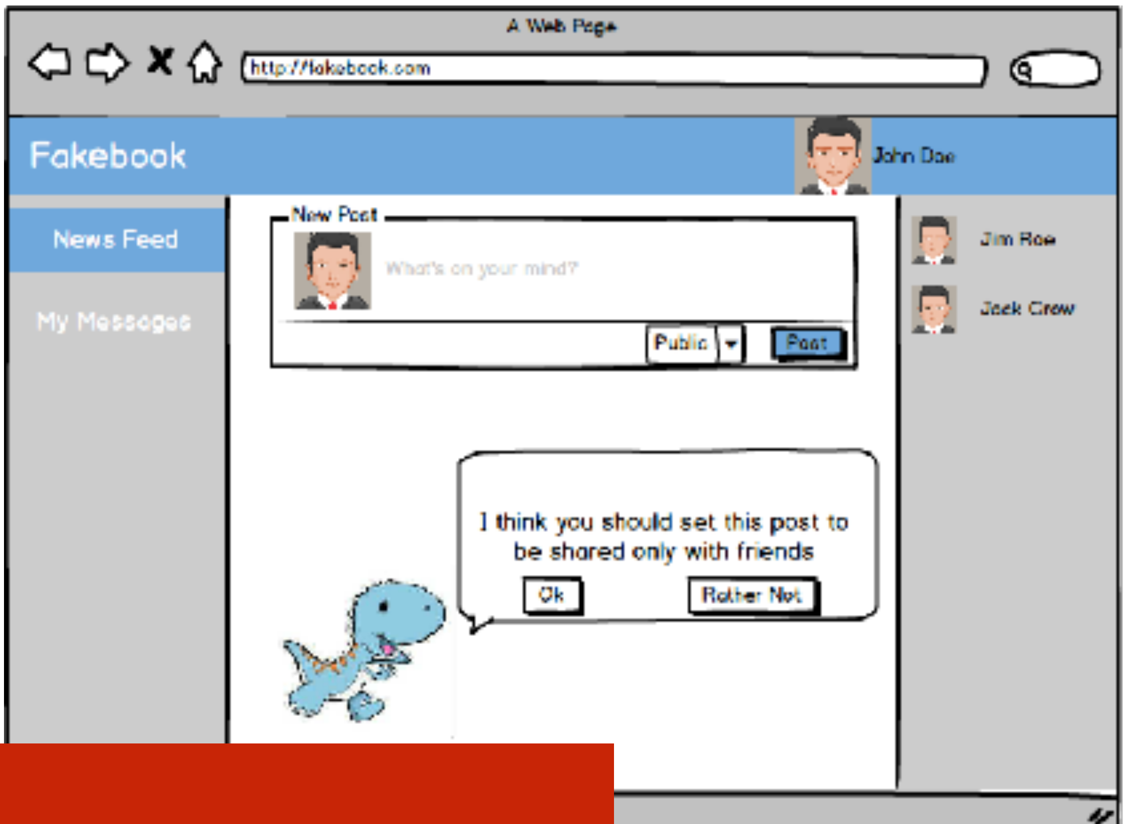
More prominent audience selection (selective sharers)



More prominent timeline moderation (time savers)

Wilkinson et al., "User-Tailored Privacy by Design", USEC 2017

Give recommendations

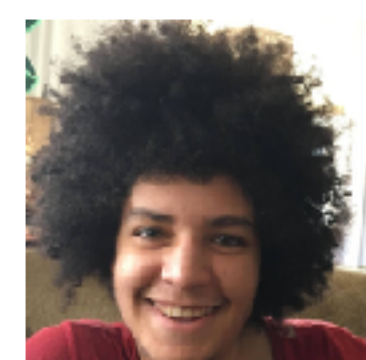


Audience suggestion (selective sharers)

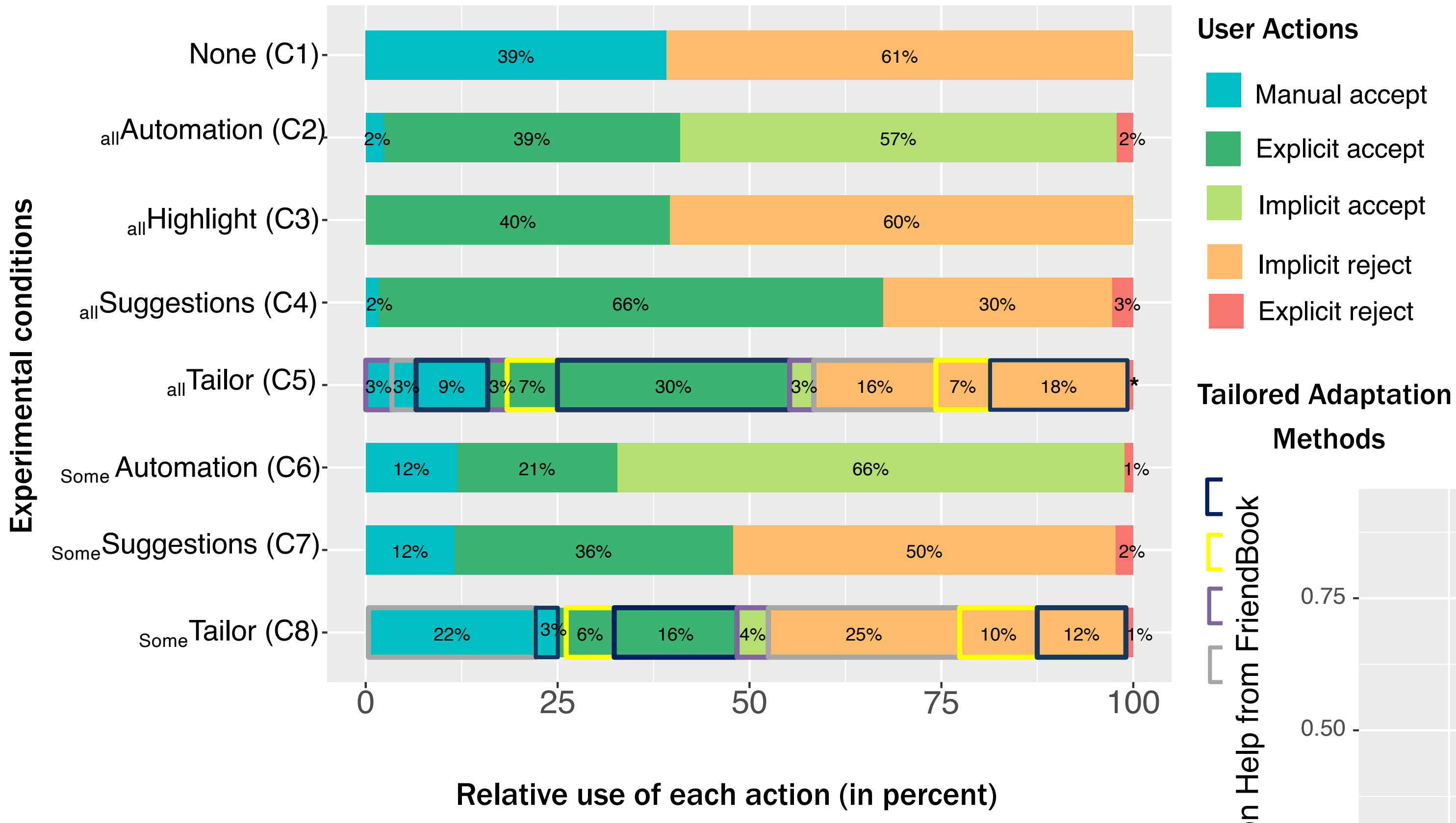
Automatically turn off chat (time savers)

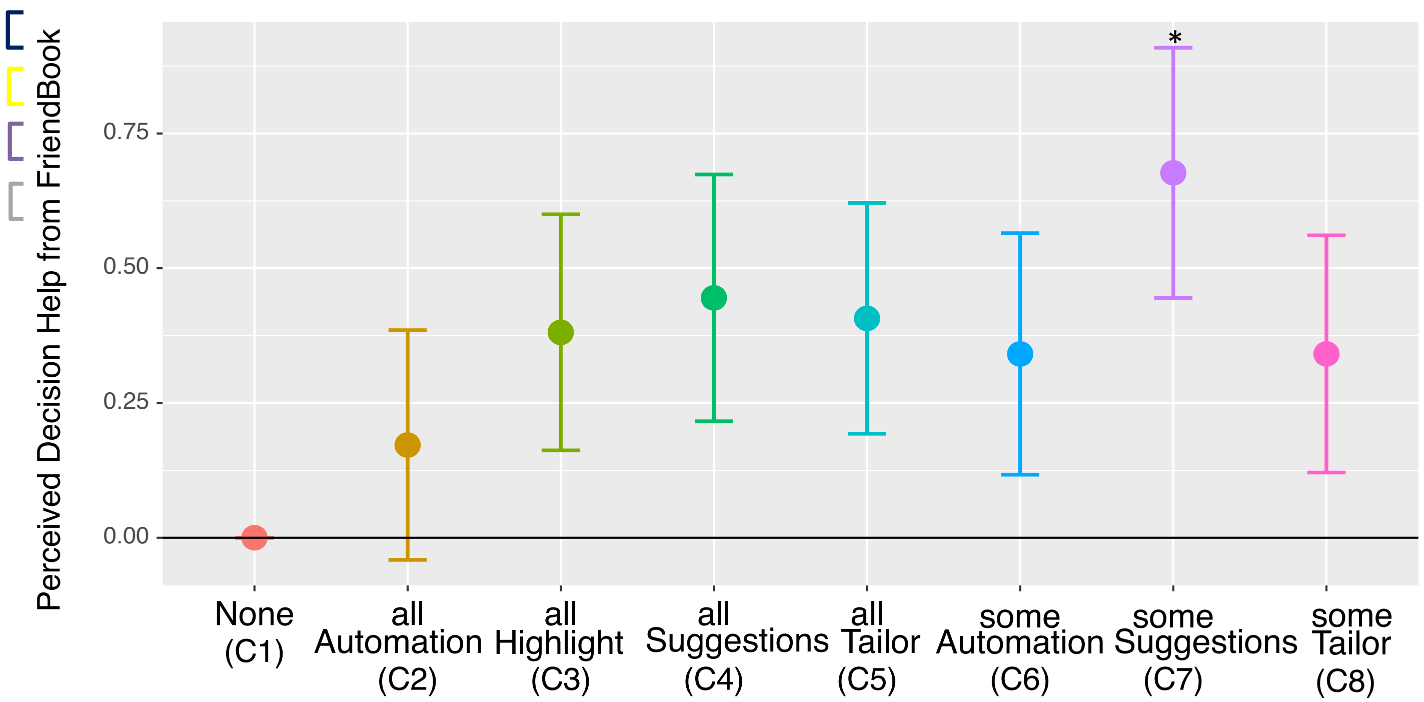
Namara et al., "The Potential for User-Tailored Privacy on Facebook", PAC 2018

Namara et al., "The Effectiveness of Adaptation Methods in Improving User Engagement and Privacy Protection on Social Network Sites", PoPETS 2022

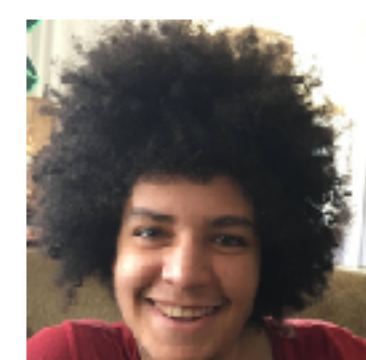


Give recommendations





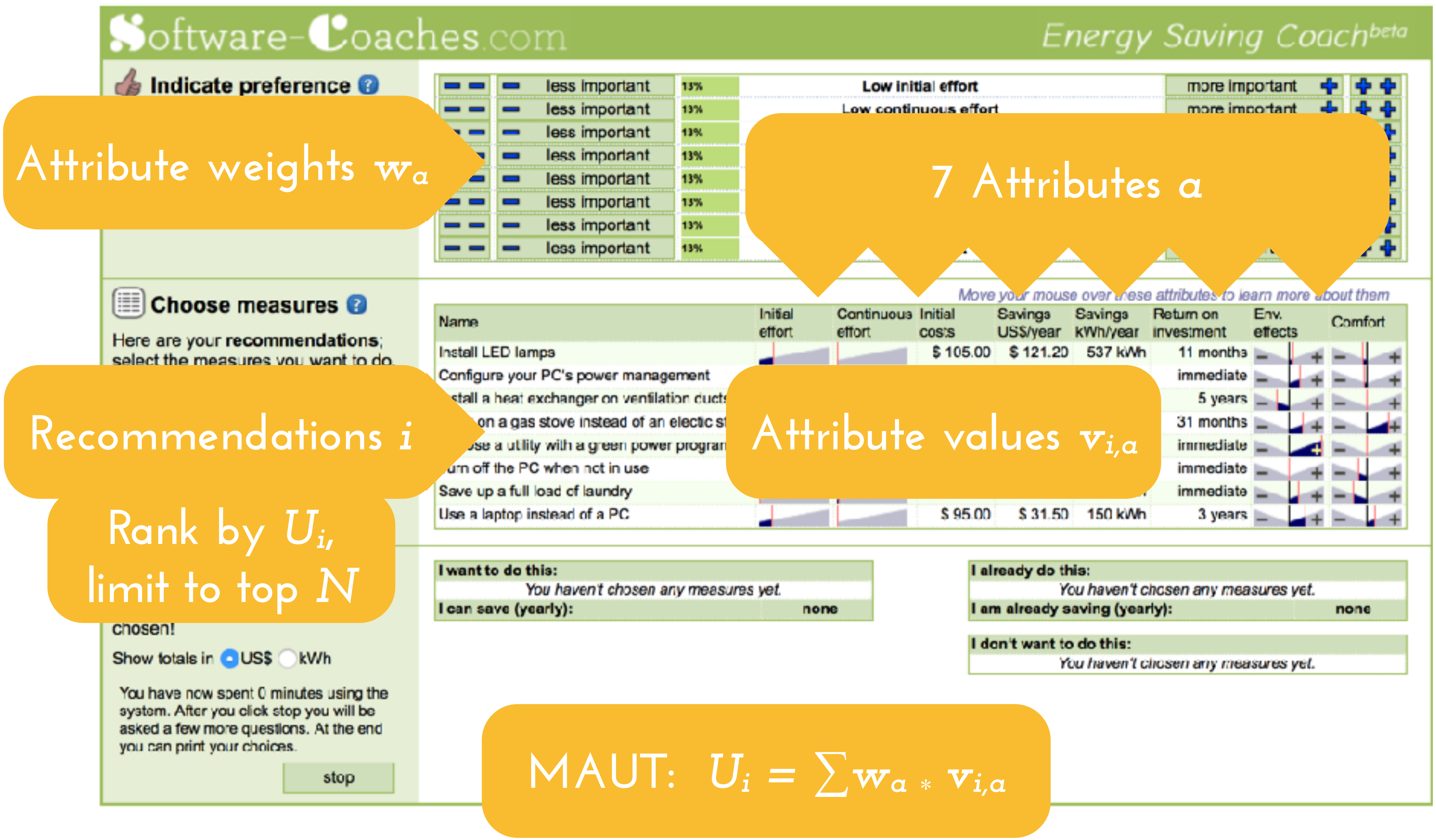
Namara et al., "The Effectiveness of Adaptation Methods in Improving User Engagement and Privacy Protection on Social Network Sites", PoPETS 2022



Example 2:

A demographics-based recommender system
preserving privacy by implementing an
adaptive request order

Host: recommender system



Preference elicitation

Attribute-based PE: users directly indicate the importance of each of the attributes with which choice options are described.

Case-based PE: discover attribute weights by analyzing users' evaluation of exemplary choice options.

Needs-based PE: users express their preferences in terms of consumer needs.

Implicit PE: infers the attribute weights as a by-product of the user's browsing behavior.

Hybrid PE: combines implicit PE with attribute-based PE.

Even simpler: **Top-N** (items ranked by popularity) and **Sort** (items ranked by one of the attributes).

Preference elicitation

The best preference elicitation method (PE-method) depends on users' **domain knowledge**.

E.g. energy-saving (Knijnenburg and Willemsen 2009, 2010; Knijnenburg et al. 2011, 2014).

Our studies show:

- Energy-saving experts prefer systems that allow direct control over attribute weights (attribute-based and hybrid PE).
- Novices prefer systems that are tailored to their needs (needs-based PE), provide limited or no control (sort, top-N).
- This does not only increase satisfaction; it also makes users save more energy.

New: demographics-based PE

New: demographics-based PE

Demographics are an important determinant of preferences in the domains of energy and health.

- Needed: an algorithm that translates answers to demographic questions into attribute weights.
- Based on these weights I can then recommend items as usual.

Demographics-based PE:

- May be most beneficial for domain novices (known and easy to report).
- May be more privacy-sensitive than other PE-methods (Ackerman et al. 1999).

"Privacy-personalization paradox"

Adaptive request order

Which item to ask first?

- Not all items are equally **useful** to the recommender.
- Not all demographic items are equally **sensitive**.
- Not everyone is equally **private** regarding their demographics.

Adaptive request order: dynamically weigh predicted privacy and benefit.

- Learn users' disclosure tendency (on the fly)
- Dynamic forecasting of benefit based on changes to the user model

Result: ask the most useful question that is not too sensitive.

Three studies

Pre-study:

- Link demographic answers to attribute weights.
- Investigate sensitivity of demographic items.

Study 1:

- Test demographics-based PE against attribute-based PE.

Study 2:

- Manipulate demographic question request order to see if we can do better.

Pre-study

Goal: link demographic answers to attributes, investigate sensitivity of demographic items.

Sample: 200 MTurk participants, 100 in energy, 100 in health condition.

Method: collect data about:

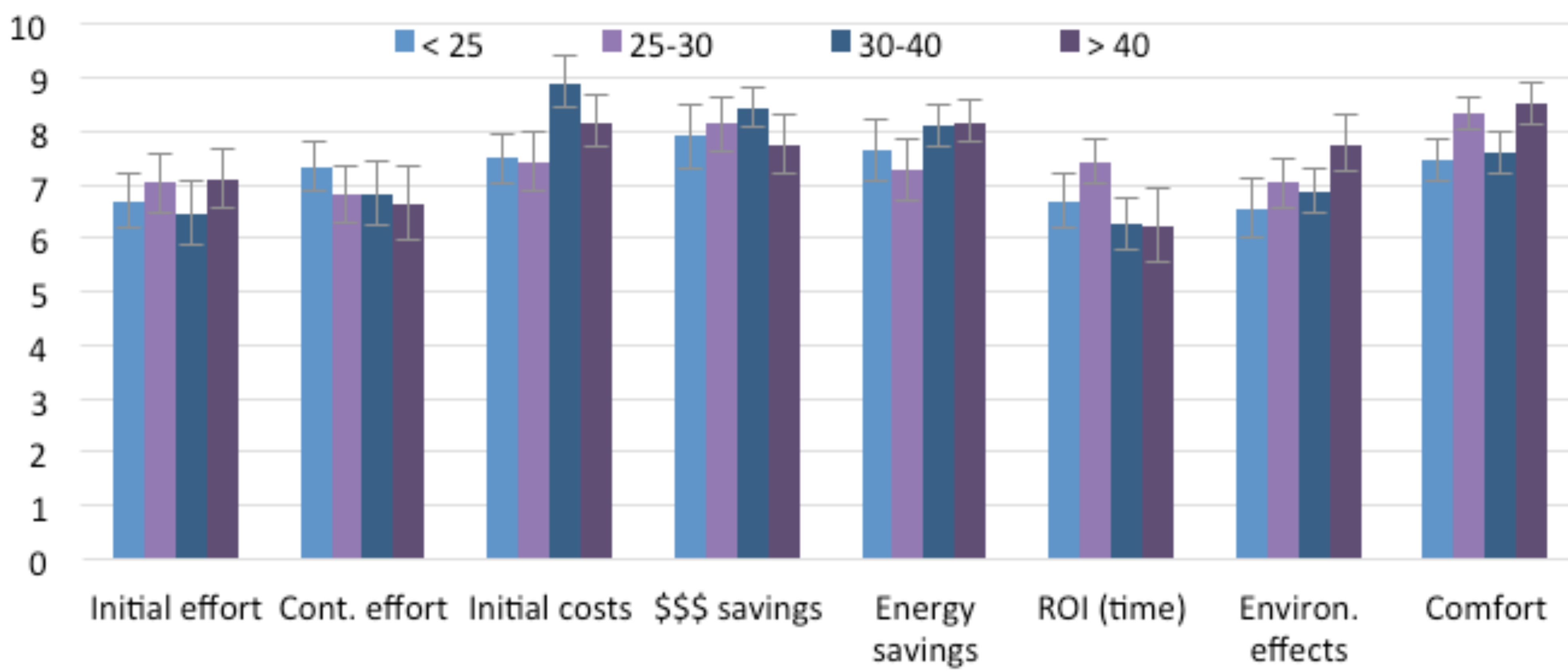
- 57 demographic items (multiple choice);
- 7-8 recommender attribute weights (scale: 0–very unimportant to 10–very important);
- perceived privacy risk of 57 items (1–very safe to 7–very risky).

Pre-study

Outcome: An algorithm that translates demographics into preferences:

- For each question, for each answer option: calculate the mean attribute weights of participants who chose that answer option.
- If two options are almost the same: combine.
- Calculate the deviance of each attribute weight from the grand mean.
- If deviance > threshold, then turn it into a "preference update rule".

Example: age


< 25
25-30
30-40
> 40
10
9
8
7
6
5
4
3
2
1
0
Initial effort
Cont. effort
Initial costs
$$$ savings
Energy savings
ROI (time)
Environ. effects
Comfort

Example: age

id	action	obj	details ▲ 1	model	modelaspect ▲ 2	modelvalue	experiment
3647	demo	26	1	utility	comfort	-2.93939394	1
2497	demo	26	1	utility	costonce	-2.80000000	1
3433	demo	26	1	utility	ecoweight	-2.79393939	1
2255	demo	26	1	utility	effortcont	2.37575758	1
2023	demo	26	1	utility	effortonce	-0.84242424	1
2746	demo	26	1	utility	realsavingsE	-1.14545455	1
3648	demo	26	2	utility	comfort	-2.93939394	1
2498	demo	26	2	utility	costonce	-2.80000000	1
3434	demo	26	2	utility	ecoweight	-2.79393939	1
2256	demo	26	2	utility	effortcont	2.37575758	1
2024	demo	26	2	utility	effortonce	-0.84242424	1
2747	demo	26	2	utility	realsavingsE	-1.14545455	1
3649	demo	26	3	utility	comfort	2.47619048	1
2499	demo	26	3	utility	costonce	-3.22857143	1
2025	demo	26	3	utility	effortonce	1.28095238	1
3196	demo	26	3	utility	returninv	4.23809524	1
2978	demo	26	3	utility	savingsKWH	-2.81904762	1
3650	demo	26	4	utility	comfort	-2.12121212	1
2500	demo	26	4	utility	costonce	5.65454545	1
3435	demo	26	4	utility	ecoweight	-0.88484848	1
2026	demo	26	4	utility	effortonce	-2.20606061	1
2748	demo	26	4	utility	realsavingsE	2.12727273	1
3197	demo	26	4	utility	returninv	-2.69696970	1
2979	demo	26	4	utility	savingsKWH	2.01212121	1

Pre-study

Outcome: Demographic question sensitivity model:

- In contrast to our previous work, perceived risk is unidimensional.
- This means we can model users' **privacy tendency** on a single dimension, rather than their multidimensional privacy profile.
- One advantage: we can use a Rasch model to dynamically track users' privacy tendency. (cf. TOEFL, GRE).

Study 1

Research questions:

- Is demographics-based PE more accurate and easier to use than attribute-based PE (possibly for novices only)?
- Does demographics-based PE incur more privacy threat?
- How do these to aspects interplay to determine system satisfaction, outcome satisfaction, and choice behavior?

Sample: 403 MTurk participants (46 removed).

Manipulations:

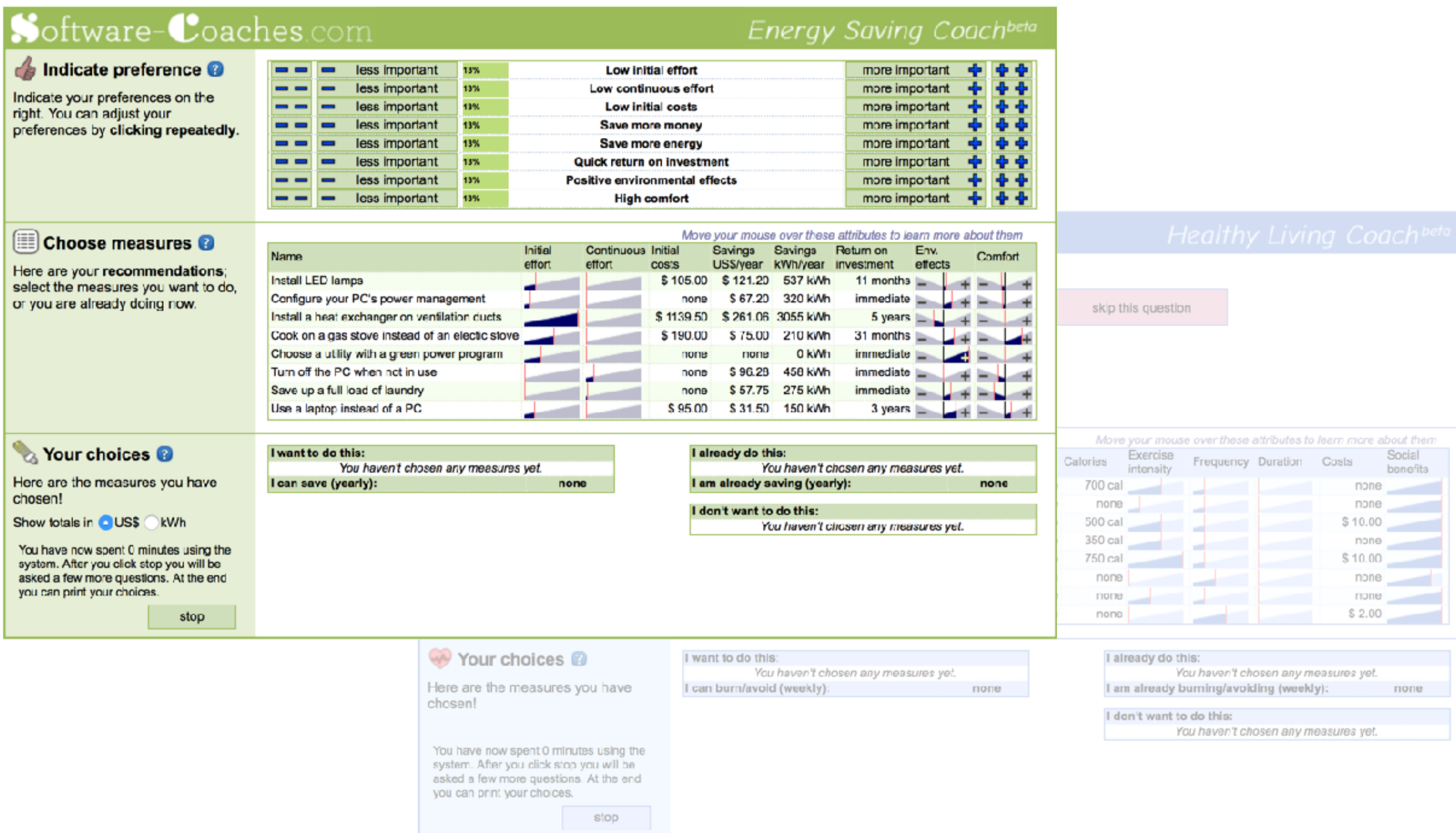
- domain (energy vs. health)
- PE-method (attribute-based vs. demographics-based)

Study 1

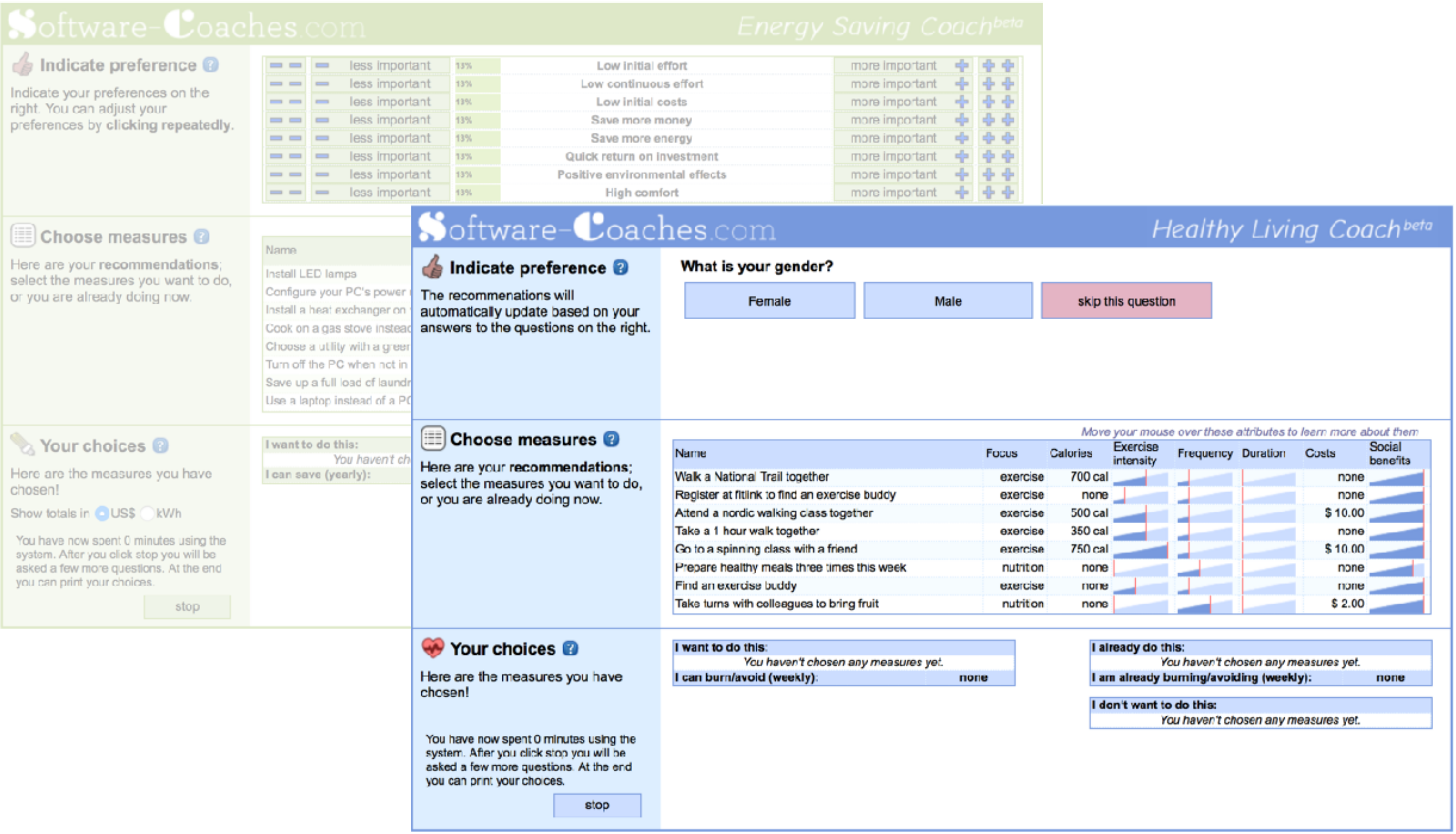
Method:

- measure domain knowledge;
- show tutorial video;
- let participant use the recommender;
- measure choice behavior, questions answered (demographics-based condition only), system and choice satisfaction, recommendation quality, understandability, perceived control, trust in provider, general online privacy concern, and familiarity with recommenders;
- use domain knowledge and privacy concerns as moderators.

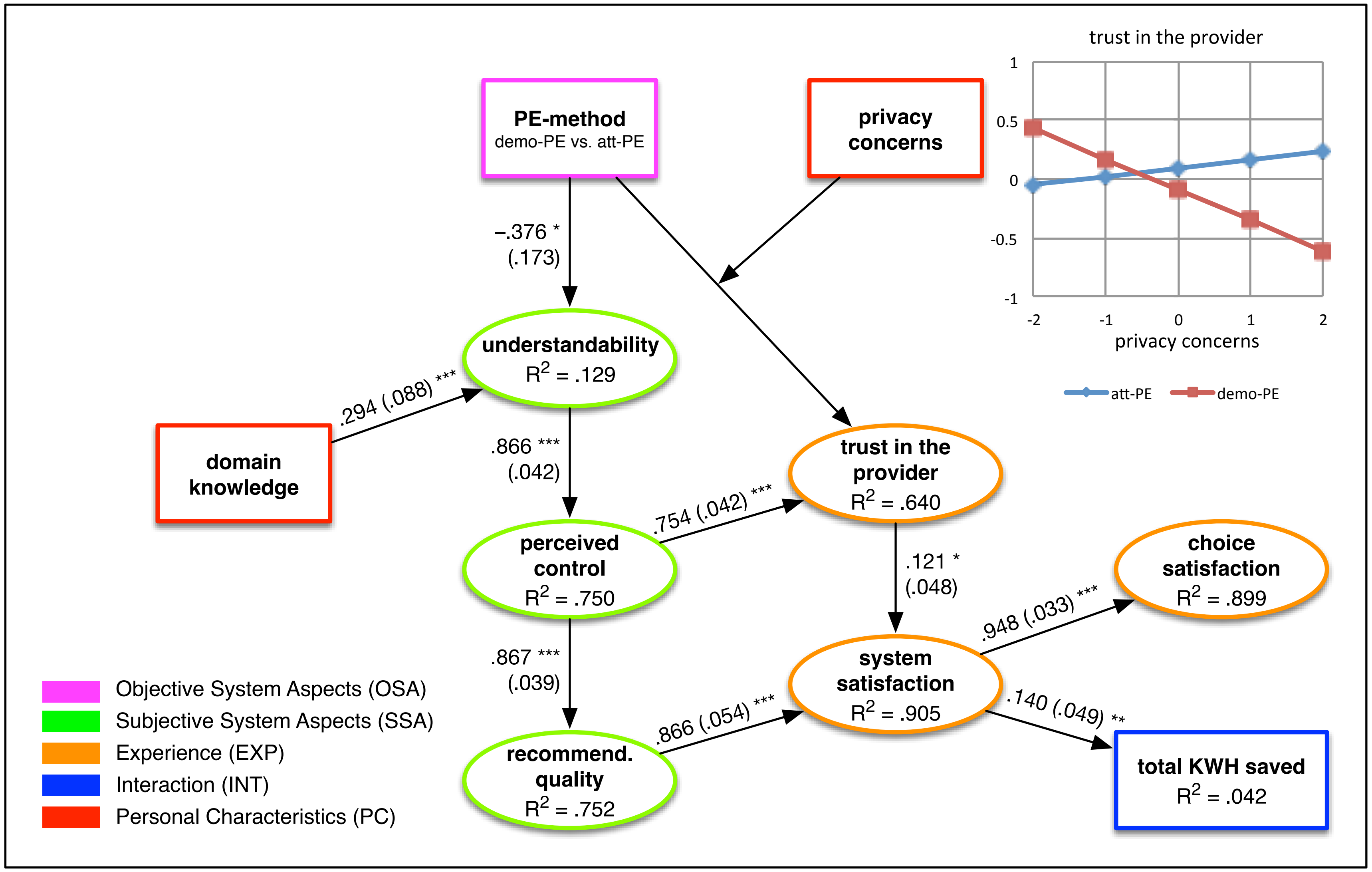
Study 1 conditions



Study 1 conditions

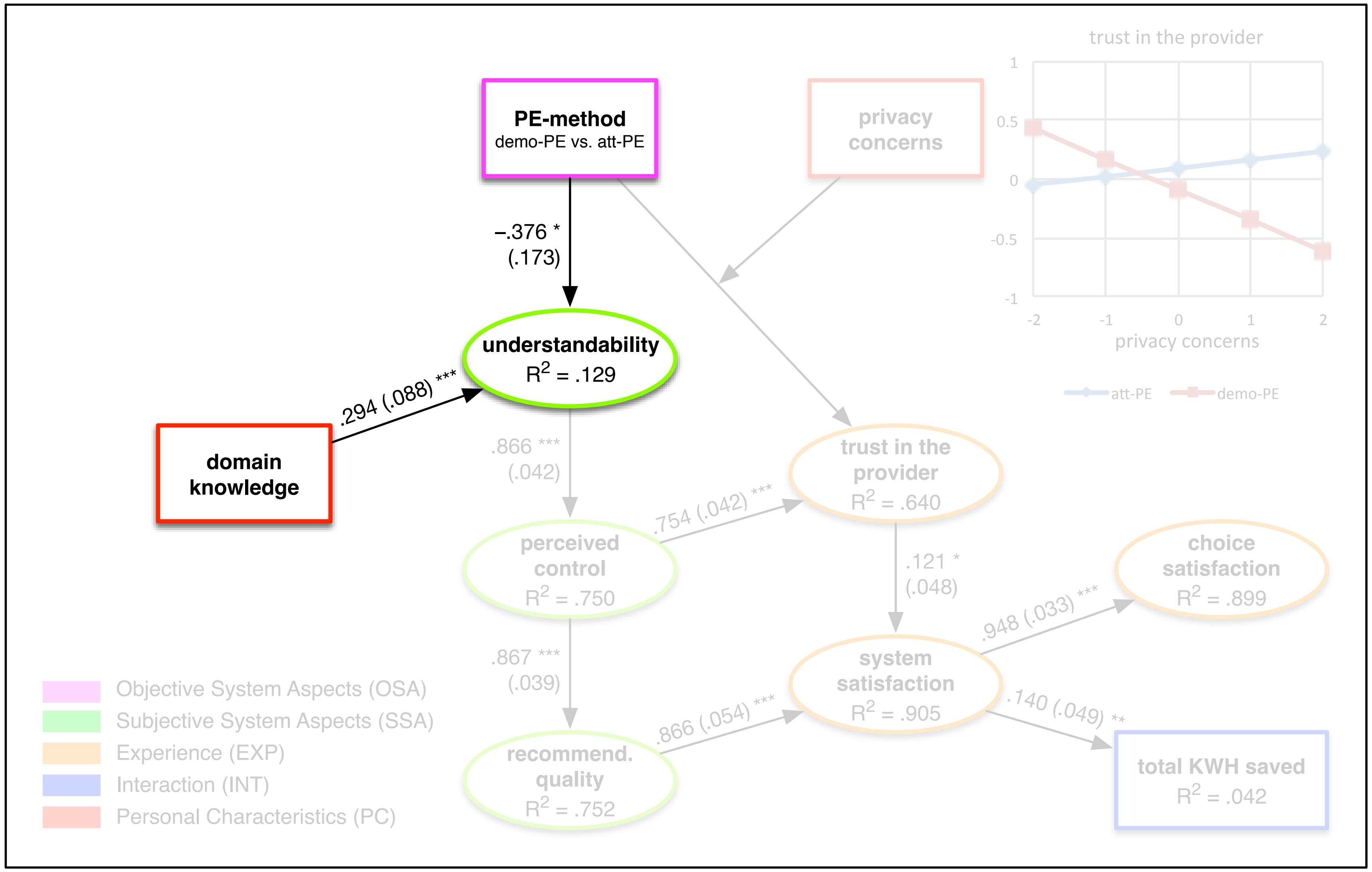


Study 1 results - Energy



Fit: *chi-sq*(805) = 1307, *p* < .001; *RMSEA* = 0.059, 90% CI: [0.053, 0.065], *CFI* = 0.974, *TLI* = 0.972

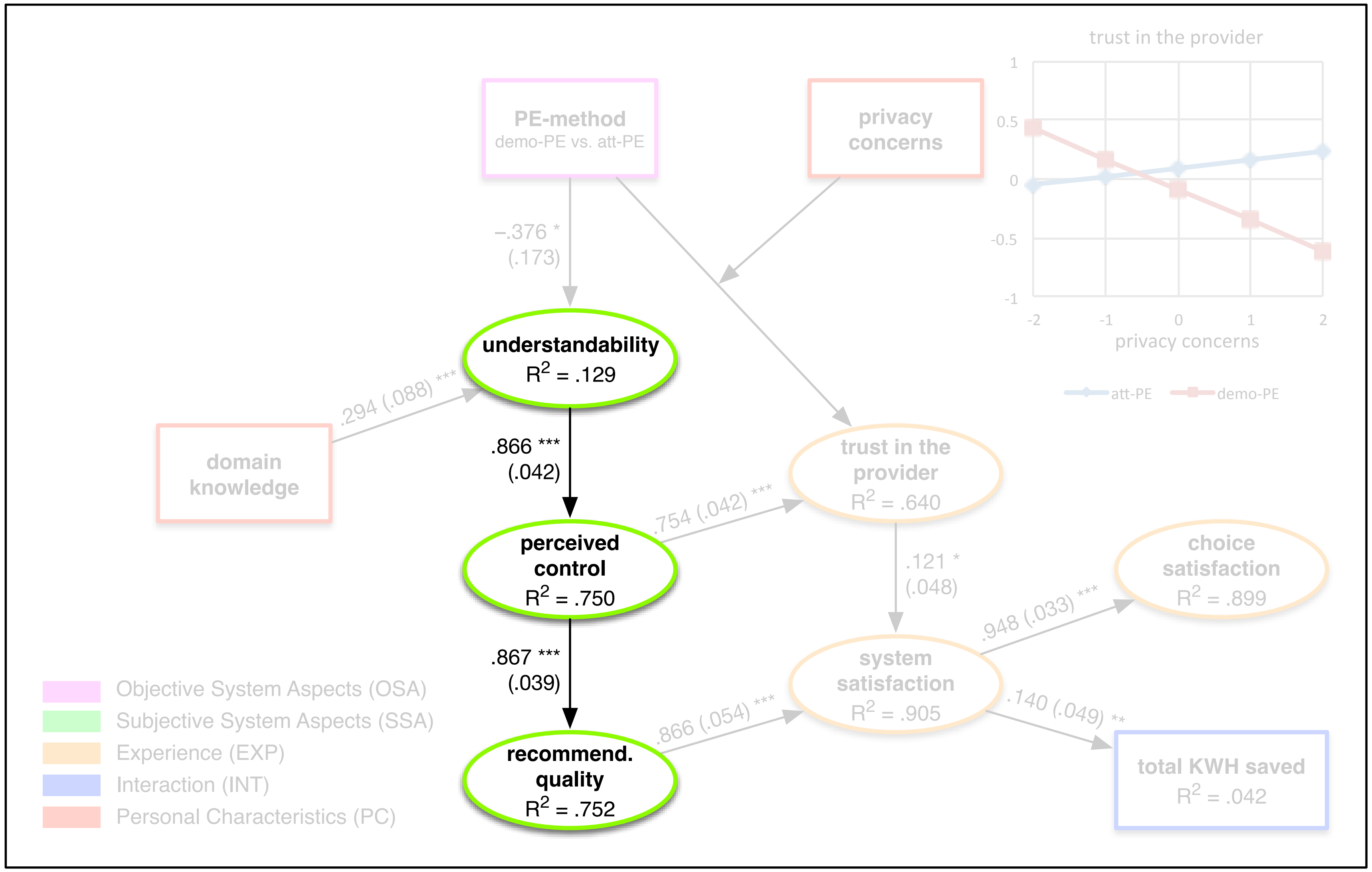
Study 1 results - Energy



Demographics-PE is less understandable than attribute-PE.

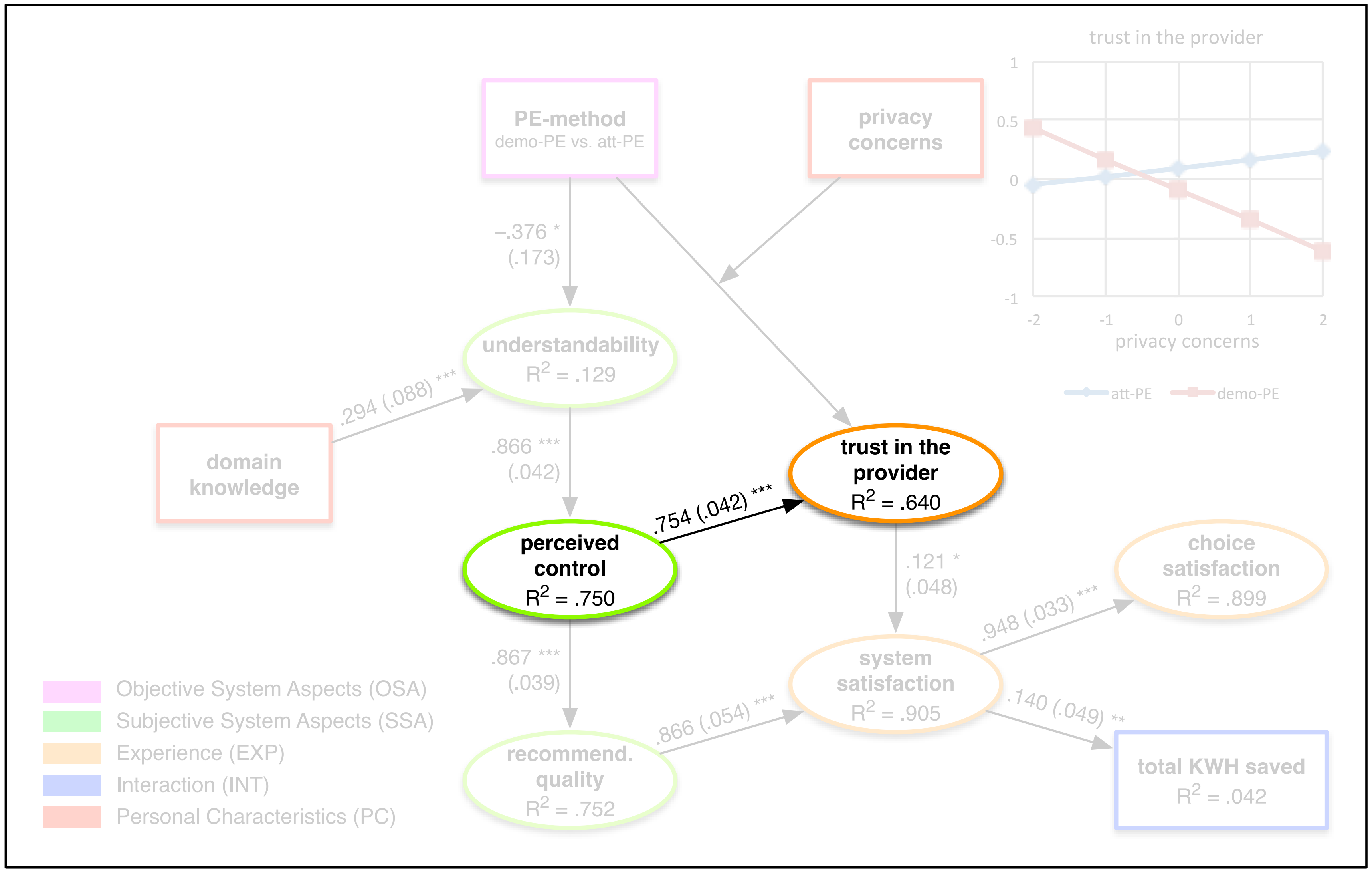
Domain experts understand the system better than domain novices.

Study 1 results - Energy



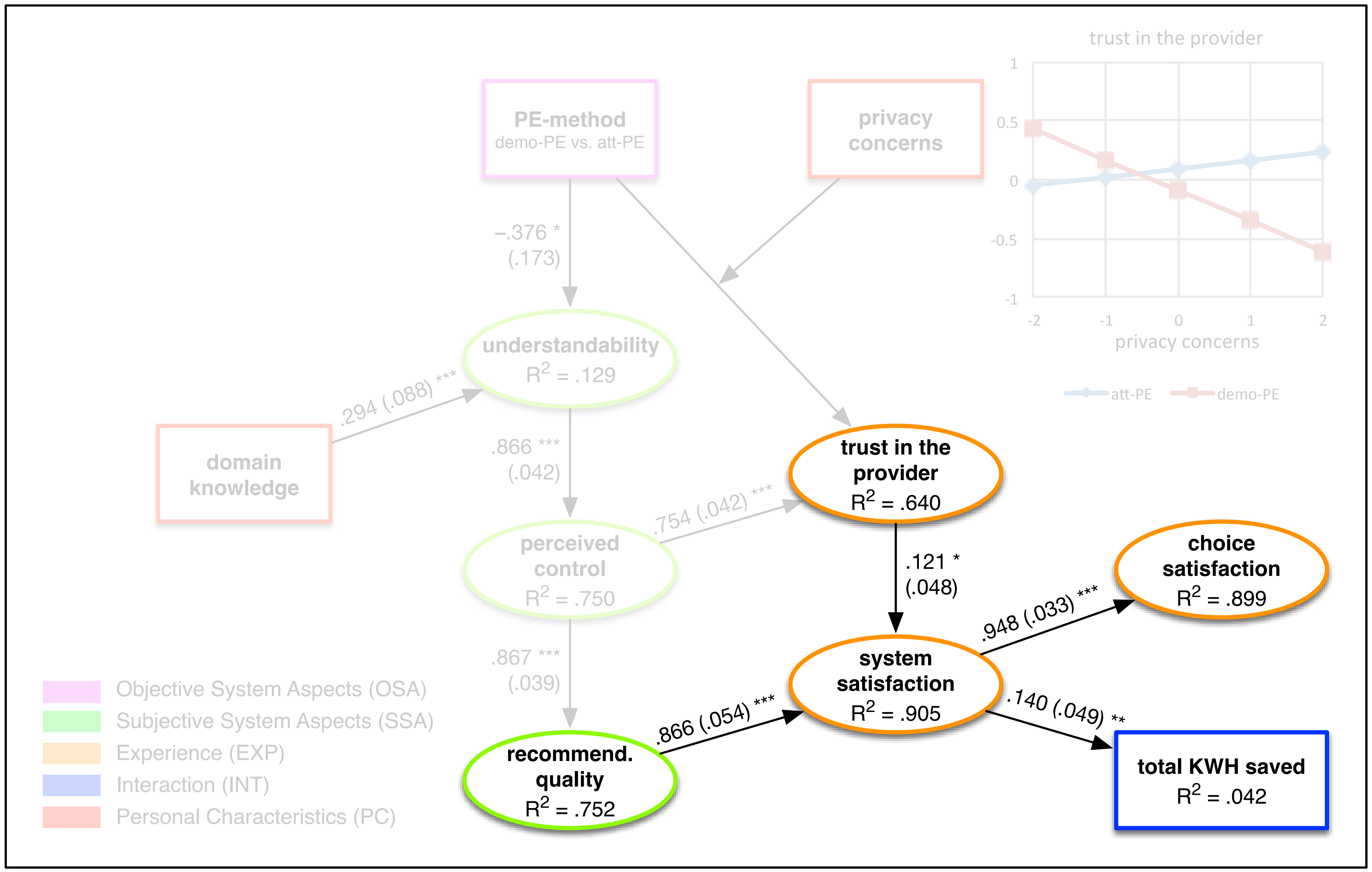
Understandability is positively related to perceived control, which is in turn positively related to recommendation quality.

Study 1 results - Energy



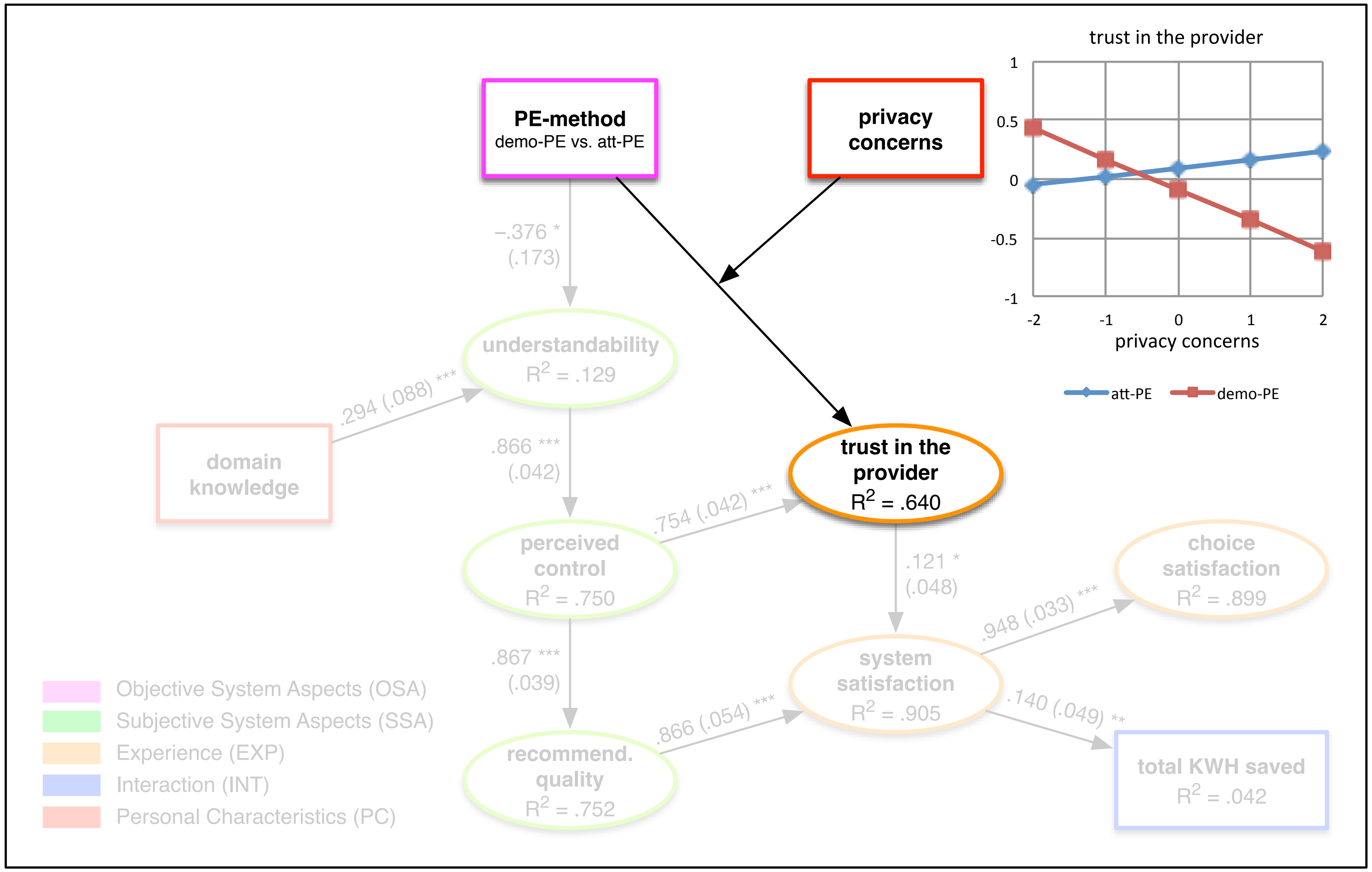
Control is positively related to trust. Interesting, because trust is privacy-related, while control is recommendation-related!

Study 1 results - Energy



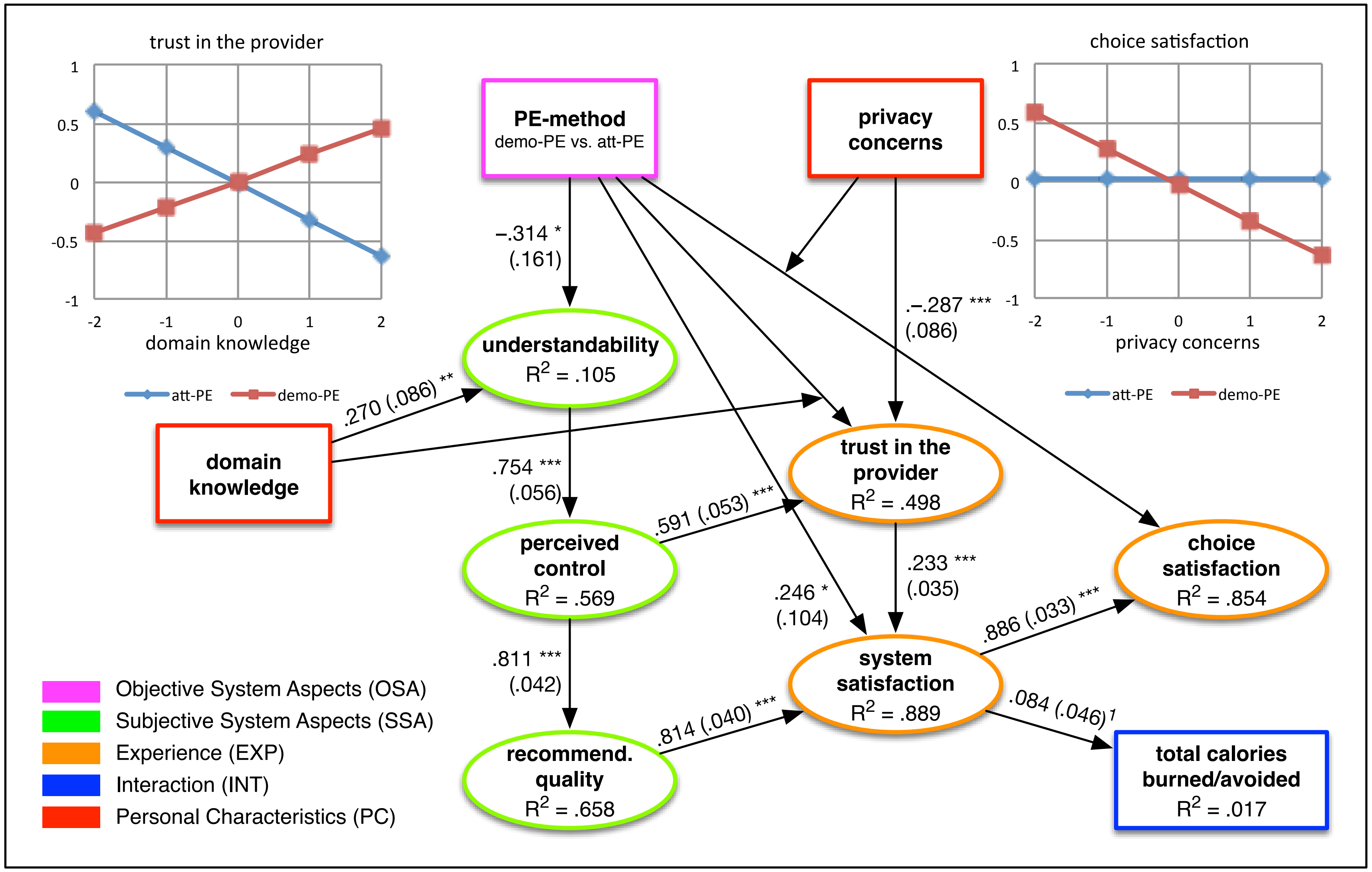
Trust and recommendation quality positively influence system satisfaction, which in turn increases choice satisfaction and the total amount of energy saved by the selected measures.

Study 1 results - Energy



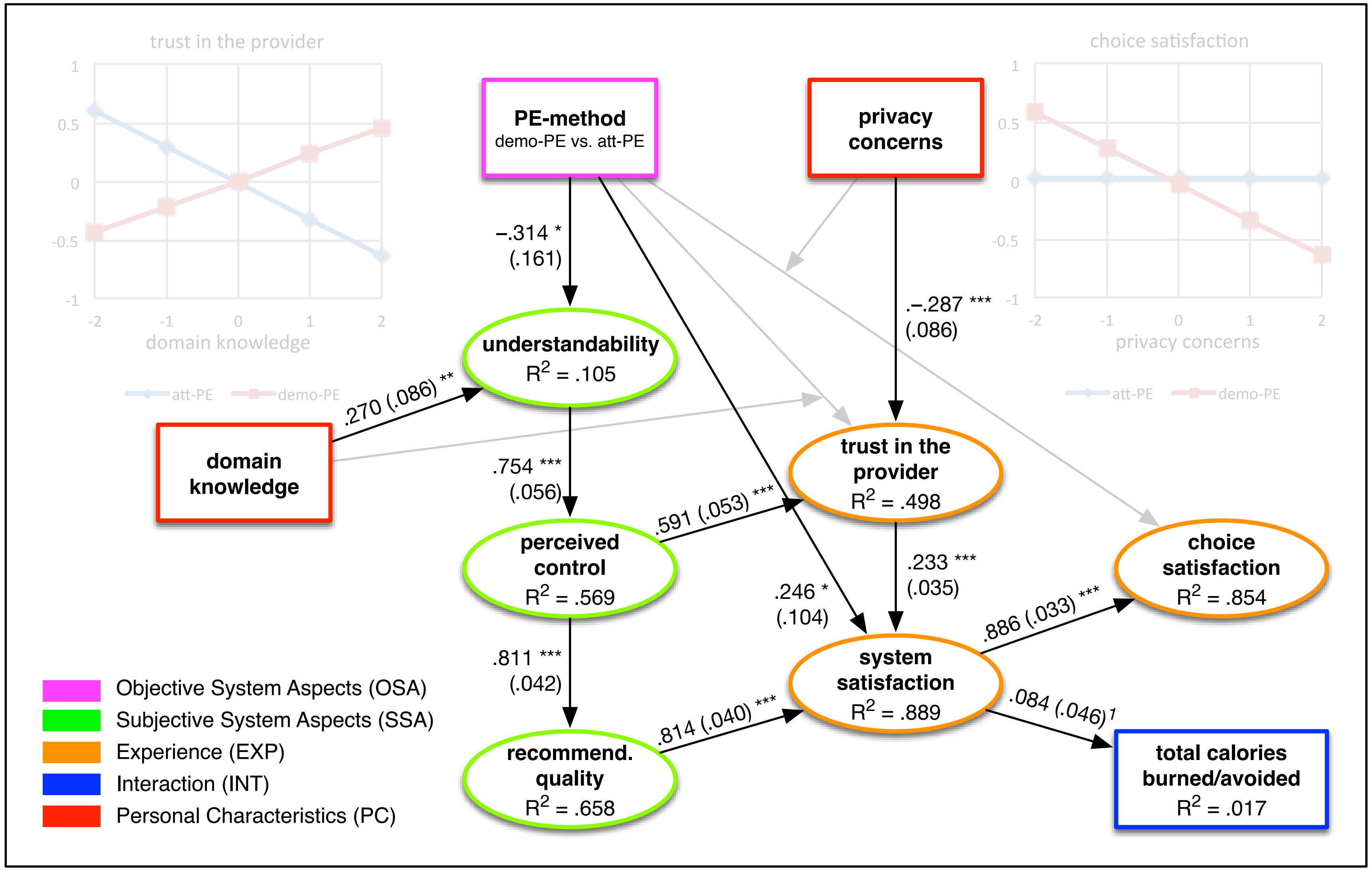
Users of demographics-PE with high concerns trust the provider less, while there is no such effect for users of attribute-PE (interaction: beta = –.330, p = .044).

Study 1 results - Health



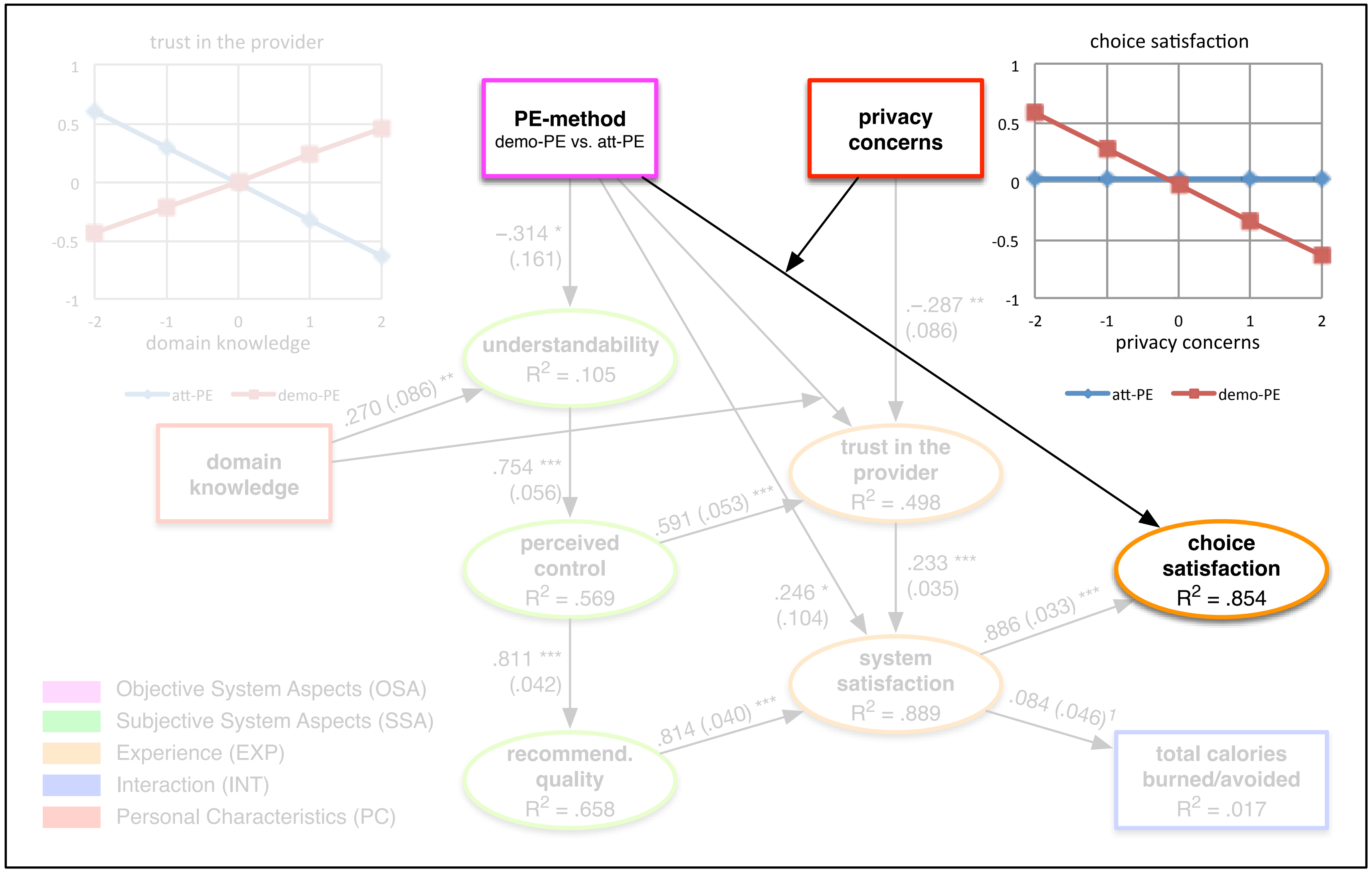
Fit: *chi-sq*(881) = 1450, *p* < .001; *RMSEA* = 0.060, 90% CI: [0.055, 0.066], *CFI* = 0.965, *TLI* = 0.963

Study 1 results - Health



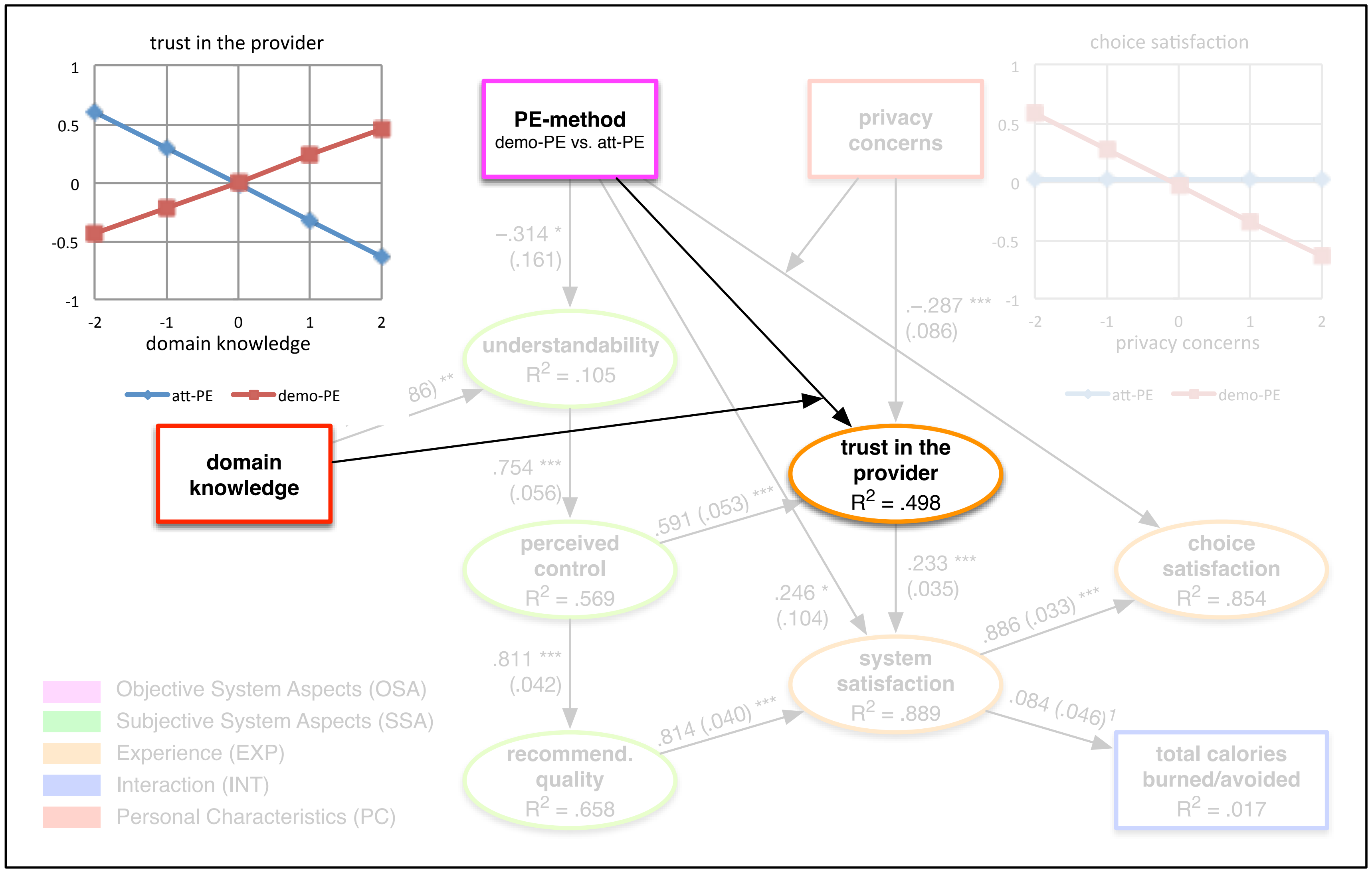
General model is almost the same as for Energy.

Study 1 results - Health



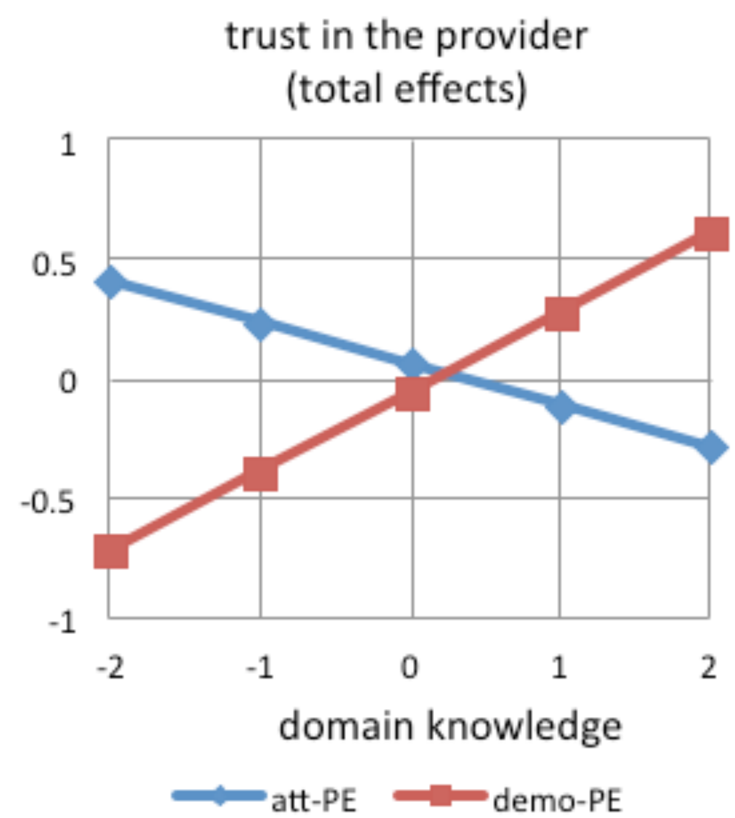
Users of demographics-PE with high concerns are less satisfied with their choices, while there is no such effect for users of attribute-PE.

Study 1 results - Health



Domain novices trust the provider more when using attribute-PE, while domain experts trust the provider more when using demographics-PE.

Study 1 results - Health



The total effect is most pronounced for the demographics-based PE method

Study 1 qualitative feedback

"Hi. I just finished this survey. I LOVED the concept!"

"This is very cool!"

"I would definitely take advantage of a program like this. This would be a good app also."

"I like the Healthy Living coach System and many of the suggestions provided. I think it would be very helpful in maintaining my motivation and tracking my progress towards my goals. Great suggestions like taking turns bringing fruit to work and finding an exercise buddy."

Study 1 qualitative feedback

“It asked questions that I didn't see how they were connected to my health.”

“I don't think this test is to check the health habits of a person, rather it has some other motive for sure. Otherwise you don't have to ask about whom do you vote in elections, questions related to sex (what is this? a dating website?), size of the beer bottle, what kind of toilet paper use, do you download movies illegally?!”

Study 1 qualitative feedback

“Additional data - I am an avid gardener, I am currently gardening more than 90 minutes a day, 6 days a week. I walk for exercise, having deliberately given up owning a car, to reduce my carbon footprint.

The form seemed too generic. I could not enter that I am coping with gout as an after effect of a bad kidney infection in 1999. High impact activities like running are not good for me, and a spinning class would bore me out of my mind! I like to walk to a location, not round and round in a mall, or peddle a stationary bike. My balance is not good enough to ride a regular bike anymore, and I prefer walking anyway.

I am overweight due to emotional issues that I am aware of, and I can live with them. I typically gain in winter when I cannot garden, and start losing again as soon as I can work the soil.”

Discussion

Demographics-PE did not live up to the expectations.

Especially for domain novices and users with high privacy concerns.

This is likely due to the random request order.

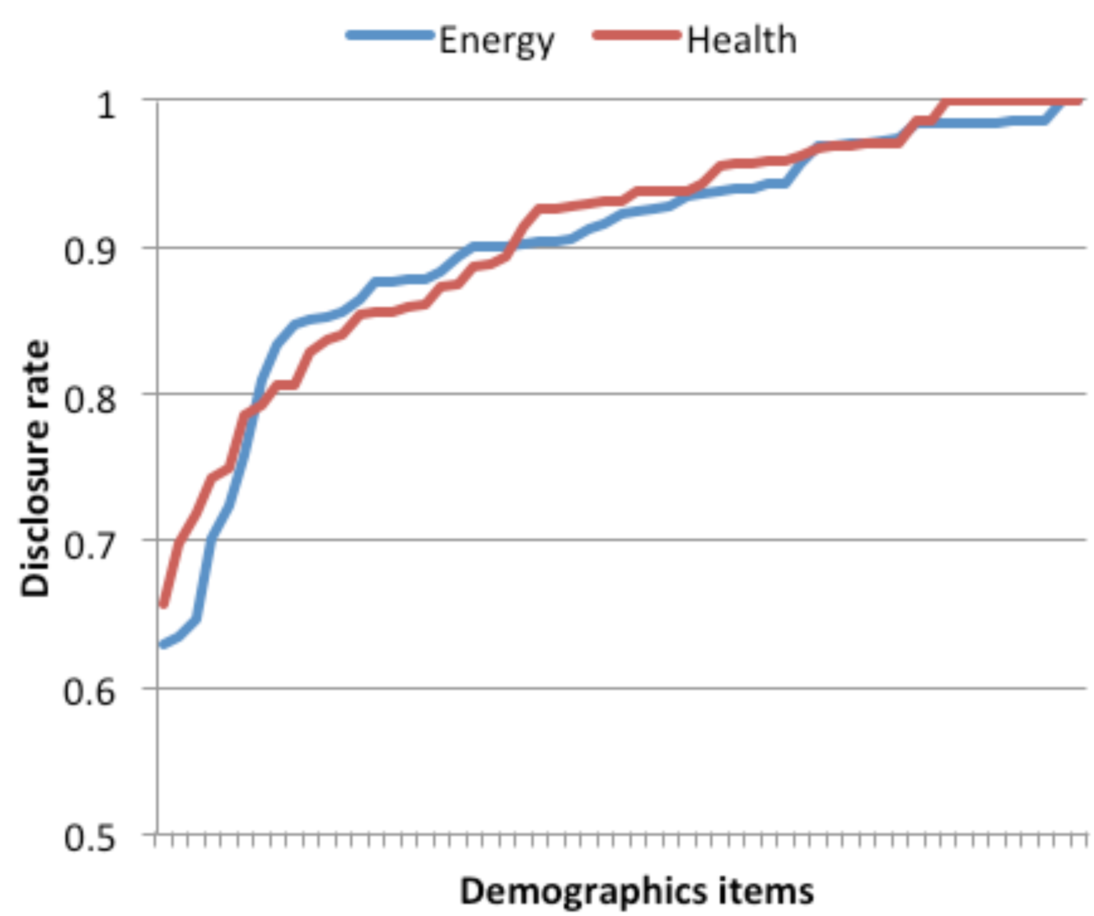
- Disregards usefulness (decreasing understandability and choice satisfaction).
- Disregards sensitivity (decreasing trust and choice satisfaction).
- Disregards disclosure tendency (creating different outcomes for people with high and low concerns).

Let's make a more intelligent request order!

But first...

We restrict study 2 to the healthy-living domain.

- More prominent trade-off between usefulness and privacy.
- More opportunities to "fix" demographics-PE for novices and people with high concerns.
- Variance in item disclosure levels in study 1 is highest in the health domain.



Better request orders

Most-useful-first:

- More efficient, could require fewer answers to demographic questions.
- Useful items may be more “obvious”; could increase understandability and trust.

Least-sensitive-first:

- Could increase trust, disclosure, and choice satisfaction.

(Static) trade-off between usefulness and sensitivity:

- Compromise; better when usefulness and sensitivity are positively correlated.

Adaptive request order:

- Prioritizes usefulness for unconcerned users; sensitivity for concerned users.

Defining a trade-off

Two strategies for trading off sensitivity **δ_i** and usefulness **u_i**:

Weighted adding: $r_i = u_i - \alpha\delta_i$

Non-compensatory: $r_i = \begin{cases} u_i & if \quad \delta_i < \alpha, \\ -\delta_i & if \quad \delta_i > \alpha. \end{cases}$

Where **r_i** is the request priority.

Defining a trade-off

Non-compensatory strategy seems less elegant, but:

- Computationally less intensive.
- Never asks questions that are very sensitive early on (even if they are very useful).
- Always shows the most sensitive item last (unless the threshold is higher than the most sensitive item).
- Defaults to most-useful-first when the threshold is very high, and to least-sensitive-first when the threshold is very low.

Therefore, we choose the non-compensatory trade-off strategy.

Item usefulness

Generally speaking:

- Questions that are likely to cause more changes in attribute weights are more useful.
- Updates to “moderate” attribute weights are more useful than changes to “extreme” attribute weights.

Item usefulness

Usefulness of an item: sum of the usefulness of its options, weighted by probability of occurrence:

$$u_i = \sum_{o_i} p_o u_o$$

Usefulness of an option: sum of the update rules, weighted by inverse deviance (**d**$_{an}$) of the attribute:

$$u_o = \sum_{r_{oa}} \frac{v_r}{d_{an}} \quad \text{where} \quad d_{an} = abs\left(w_{an} - \overline{w}_n\right) + .0001$$

Item sensitivity

A Rasch model defines the probability that user **n** discloses item **i** as follows:

$$p_{ni} = \frac{e^{\beta_n - \delta_i}}{1 + e^{\beta_n - \delta_i}}$$

where $\boldsymbol{\beta}_n$ is the user's disclosure tendency, and $\boldsymbol{\delta}_i$ is the item sensitivity.

We estimate the item sensitivities based on disclosure in study 1, and set the mean to zero.

Threshold of the trade-off

Can either be static (**α**), or dynamically estimated for each user (**α_n**).

Adaptive version can be based on disclosure tendency **β_n**.

"On the fly" estimation is supported by the PROX algorithm:

$$\beta_n = \mathrm{mean}_n(\delta) + \sqrt{1 + \mathrm{var}_n(\delta)/2.9} * \ln\left(\frac{|D_n|}{|L_n| - |D_n|}\right)$$

where **L_n** is the set of items presented to user **n**, **D_n** is the subset of disclosed items, and mean_n(**δ**) and var_n(**δ**) are the mean and variance of the sensitivity of the presented items.

Threshold of the trade-off

We don't want to set $\alpha_n = \beta_n$, because then items below the threshold may have a disclosure probability of only 50%!

After extensive simulations, we choose the following two thresholds:

$$\alpha_n^H = \beta_n - 1.5 \quad \text{and} \quad \alpha_n^L = \beta_n - 2.5$$

Assuming that β_n is accurately estimated, the user discloses items below the high threshold with a probability of at least 81.8%, and items below the low threshold with a probability of at least 92.4%.

Threshold of the trade-off

For the static threshold, we simply set β_n to the average disclosure tendency in study 1.

This means that 41/57 items fall below the high static threshold, while 18 items fall below the low static threshold.

In study 1 almost all participants disclosed more than 18 items, while only half of them disclosed more than 41 items.

Participants in the low threshold condition are thus much more likely to end up in the least-sensitive-first fallback scenario.

Overview of components

Trade-off formula ✔

Definition of Usefulness ✔

Definition of Sensitivity ✔

Dynamic threshold ✔

Static threshold ✔

Study 2

Research question: Is there an effect of request order on rate of disclosure, recommendation quality, privacy concerns, and user satisfaction?

Sample: 672 MTurk participants (58 removed).

Method: Experiment similar to study 1, but only for health domain.

Conditions: see next slide…

Study 2

Attribute-based PE: baseline condition.

Most-sensitive-first: items ordered by decreasing sensitivity.

Least-sensitive-first: items ordered by increasing sensitivity.

Most-useful-first: items ordered by decreasing usefulness.

Study 2

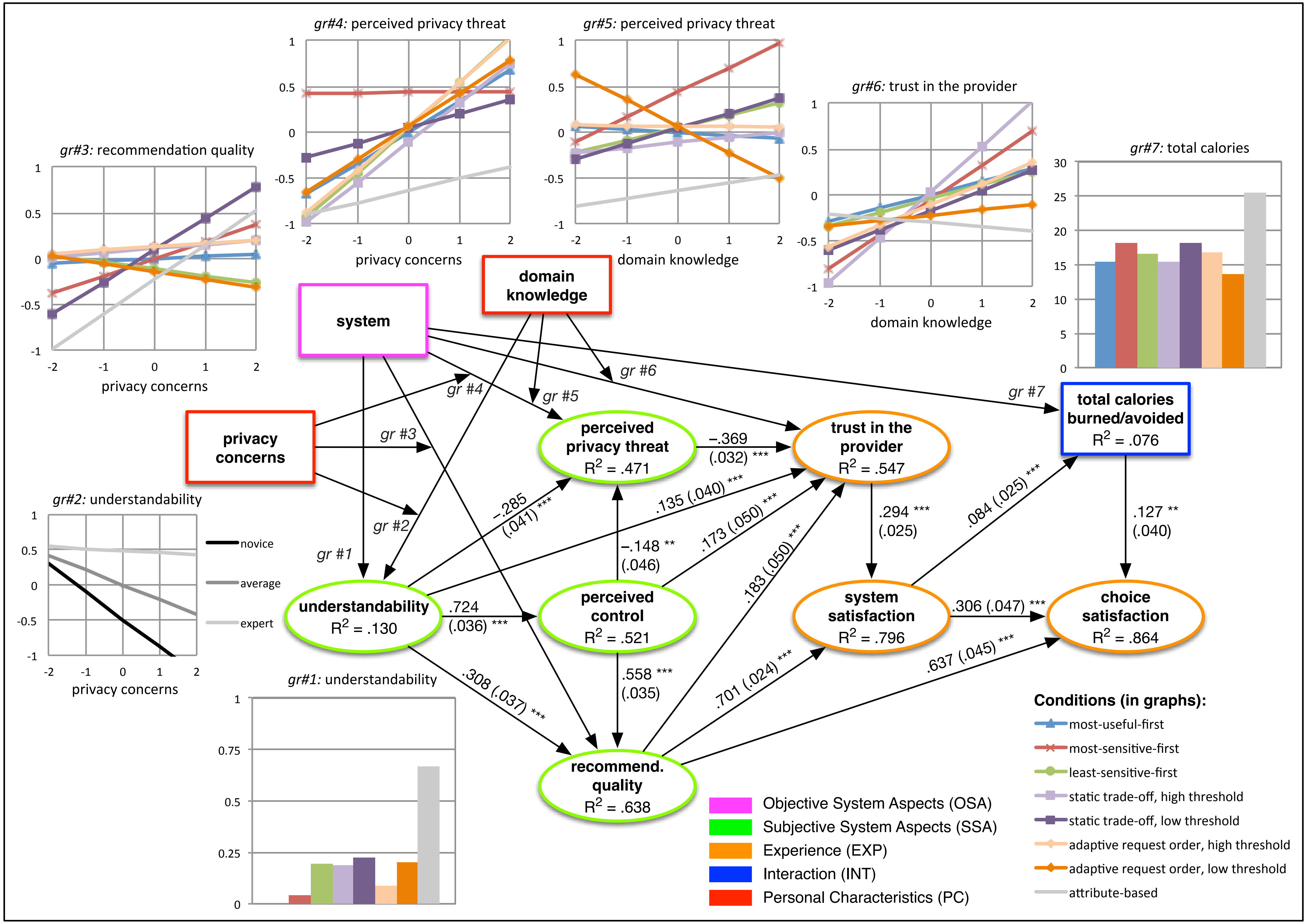
Static trade-off, low threshold: most-useful-first for items below the threshold (18 least sensitive items); least-sensitive-first for items above the threshold (remaining 39 items).

Static trade-off, high threshold: same, but with 41 items below the threshold, 16 items above the threshold.

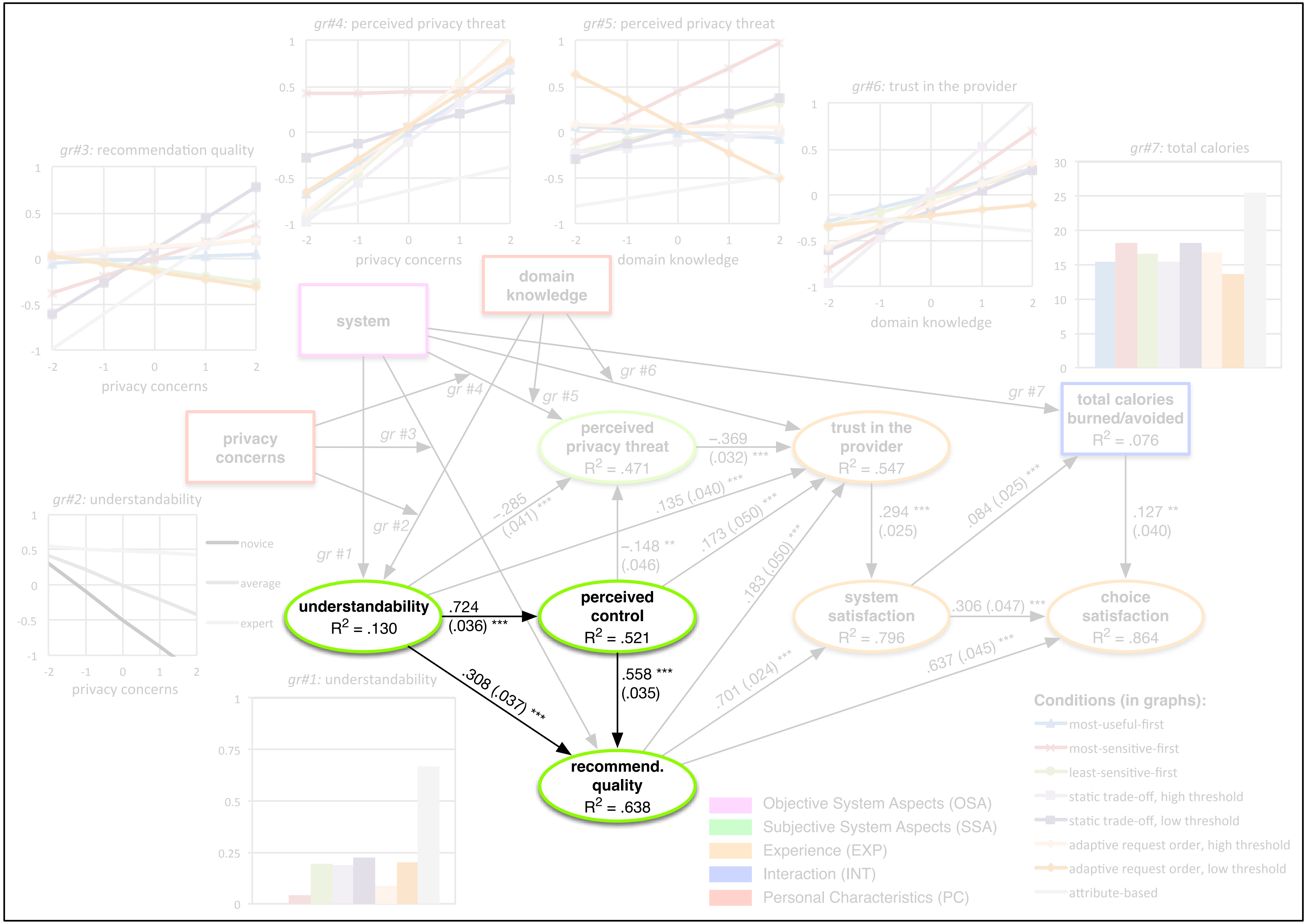
Study 2

Adaptive request order, low threshold: most-useful-first for items below the threshold; least-sensitive-first for items above the threshold. The threshold is dynamically adapted to the user's disclosure tendency – 2.5.

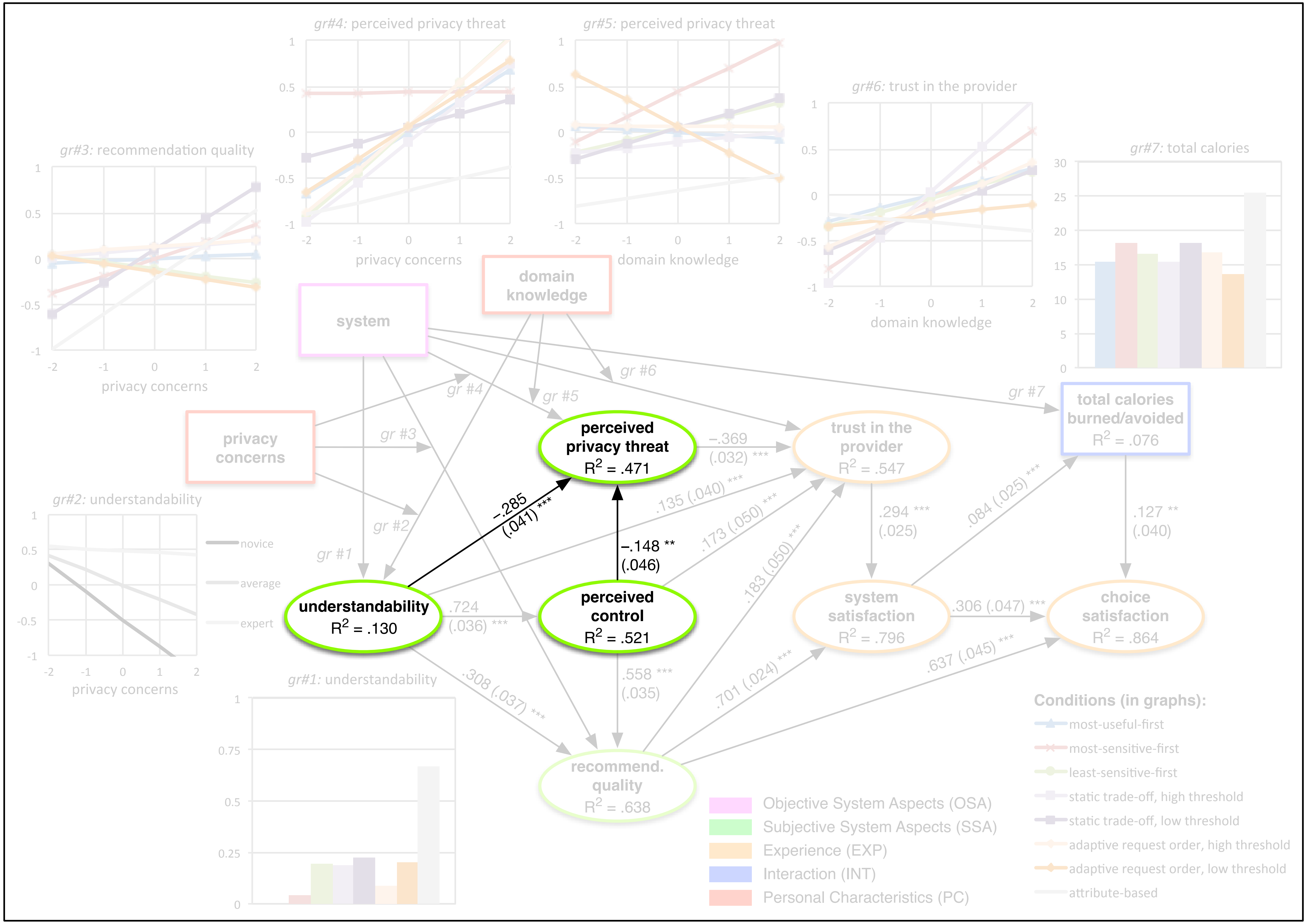
Static trade-off, high threshold: same, but with the threshold dynamically adapted to the user's disclosure tendency – 1.5.



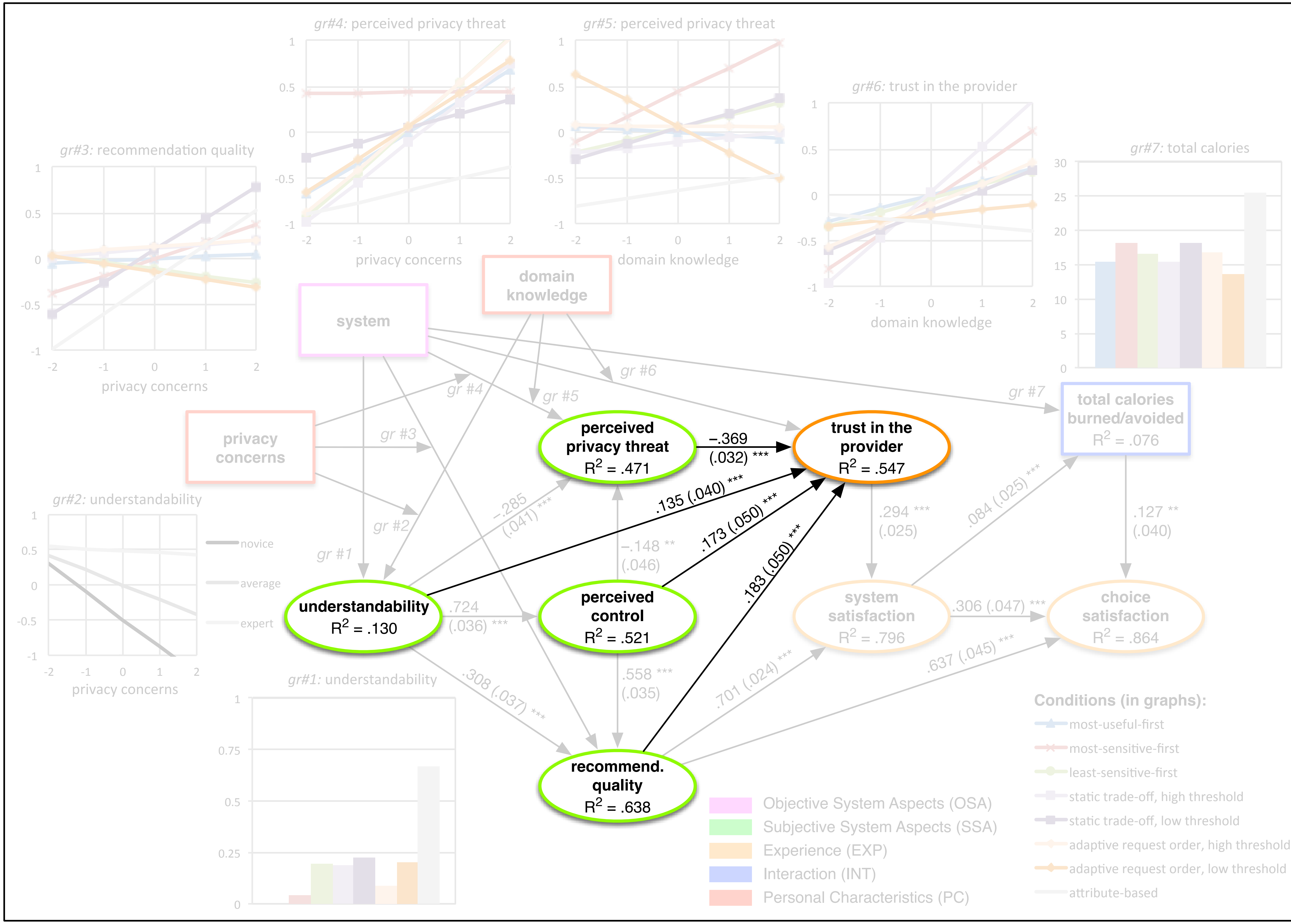
Fit: *chi-sq*(2009) = 3239, *p* < .001; *RMSEA* = 0.032, 90% CI: [0.030, 0.034], *CFI* = 0.984, *TLI* = 0.983



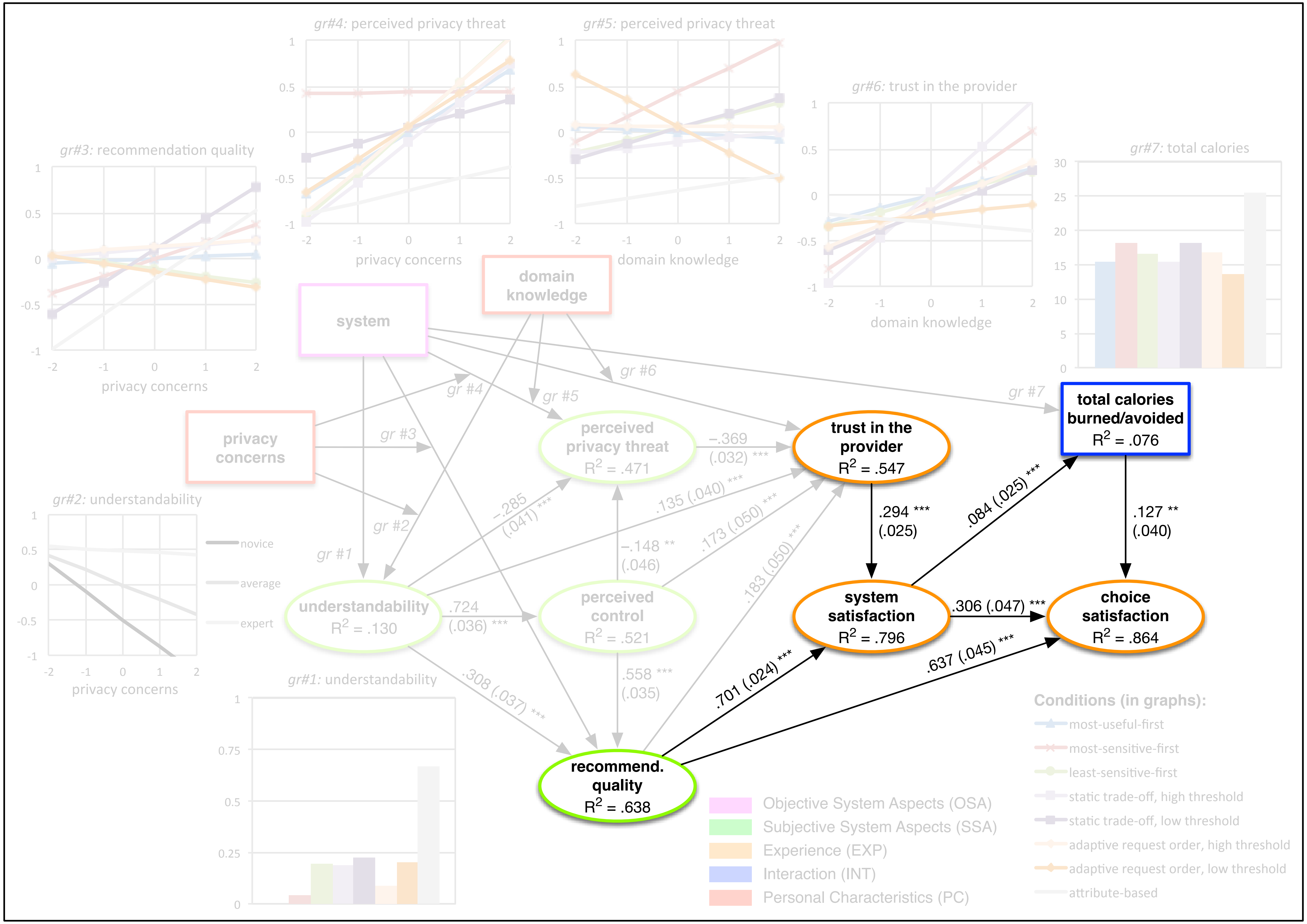
Understandability is positively related to perceived control, which is positively related to recommendation quality. Also a direct effect.



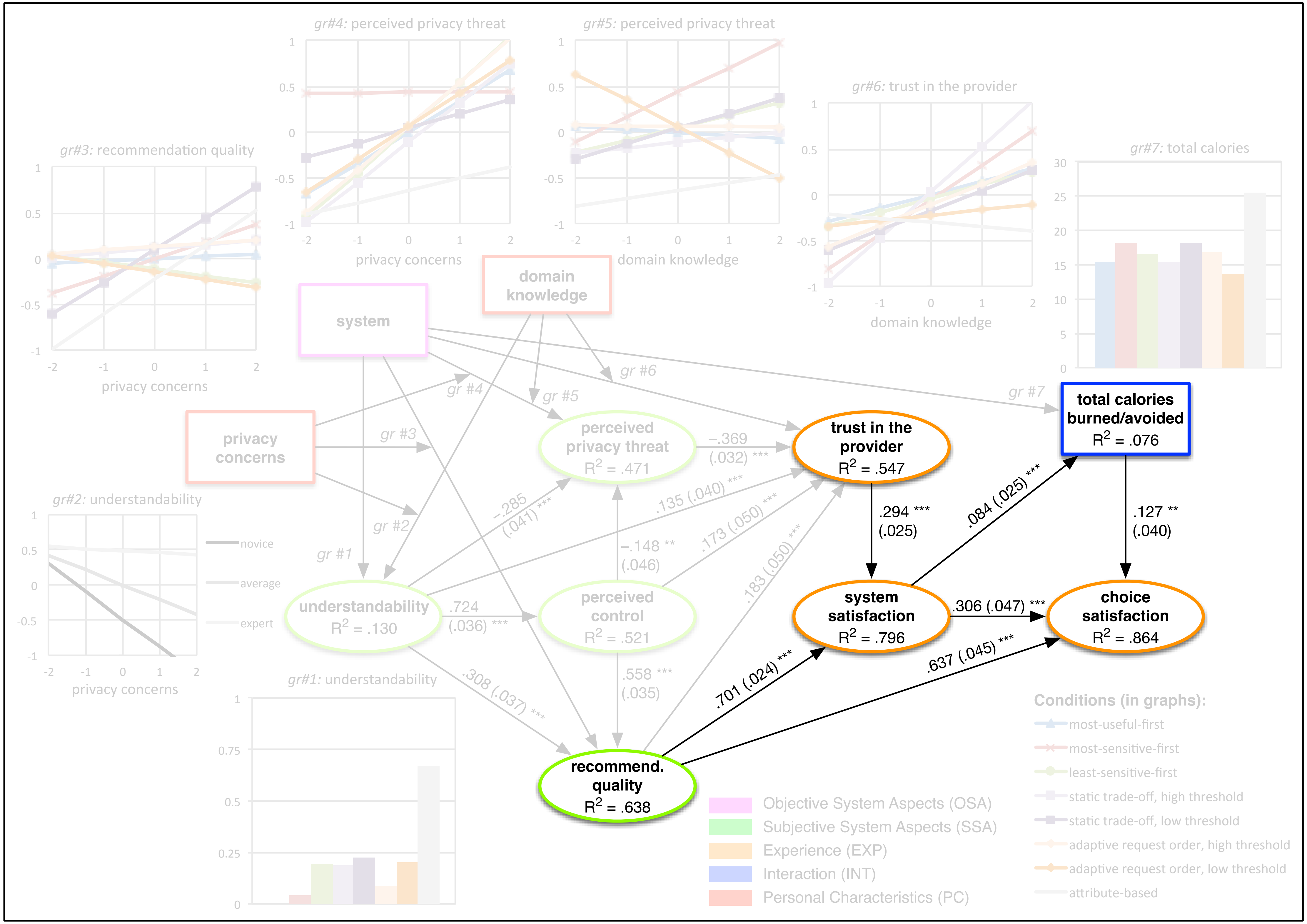
Understandability and control are negatively related to perceived threat, which which partially mediates the effect on trust in the provider. Recommendation quality also increases trust.



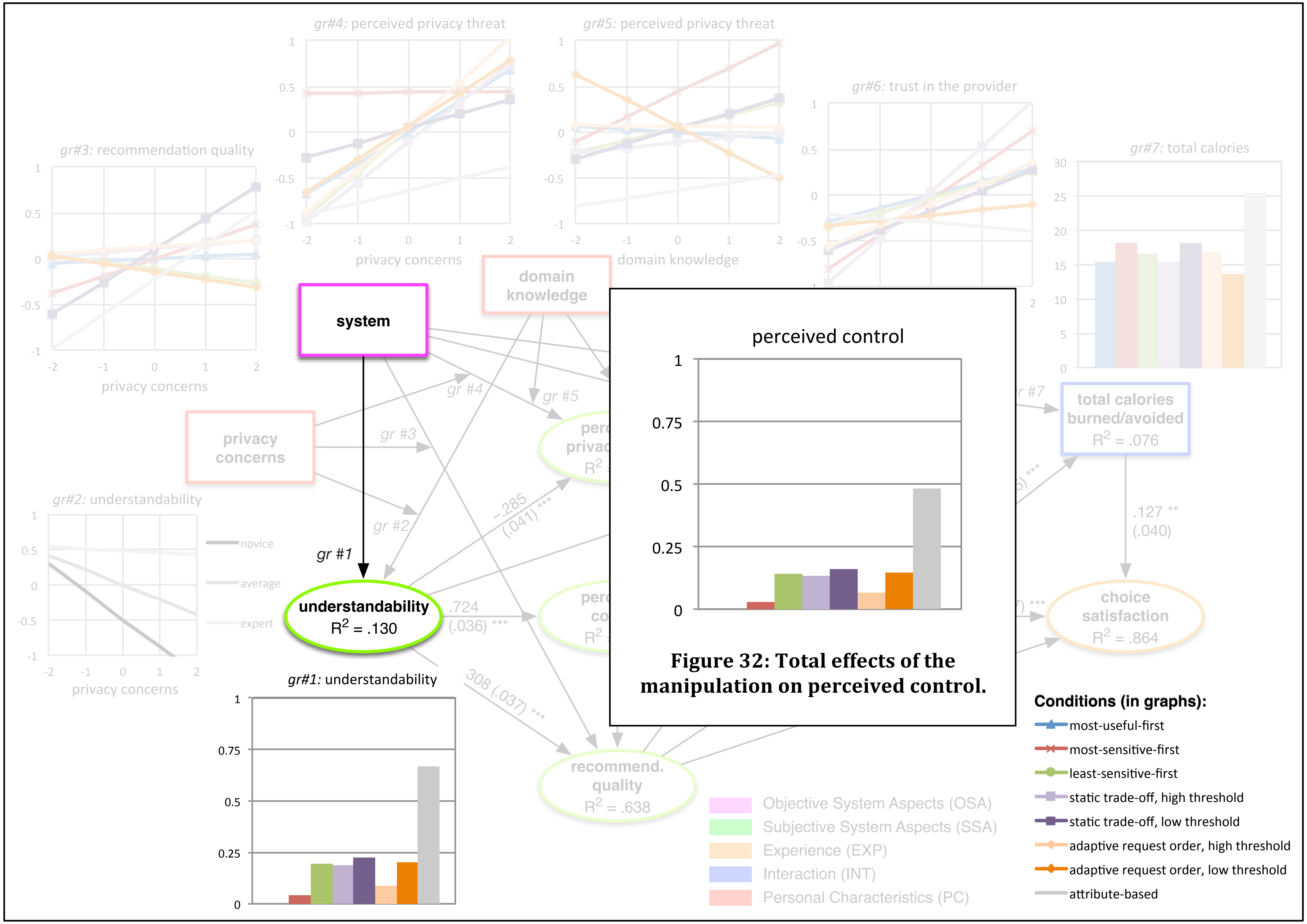
Understandability and control are negatively related to perceived threat, which which partially mediates the effect on trust in the provider. Recommendation quality also increases trust.



Both trust and recommendation quality positively influence system satisfaction, which in turn increases choice satisfaction and the total amount of calories burned/avoided.



The total amount of calories burned/avoided also has a positive effect on choice satisfaction, as does recommendation quality.



Attribute-PE (grey) is more understandable than any of the demographics-PE conditions. This also reflects on perceived control (total effect).

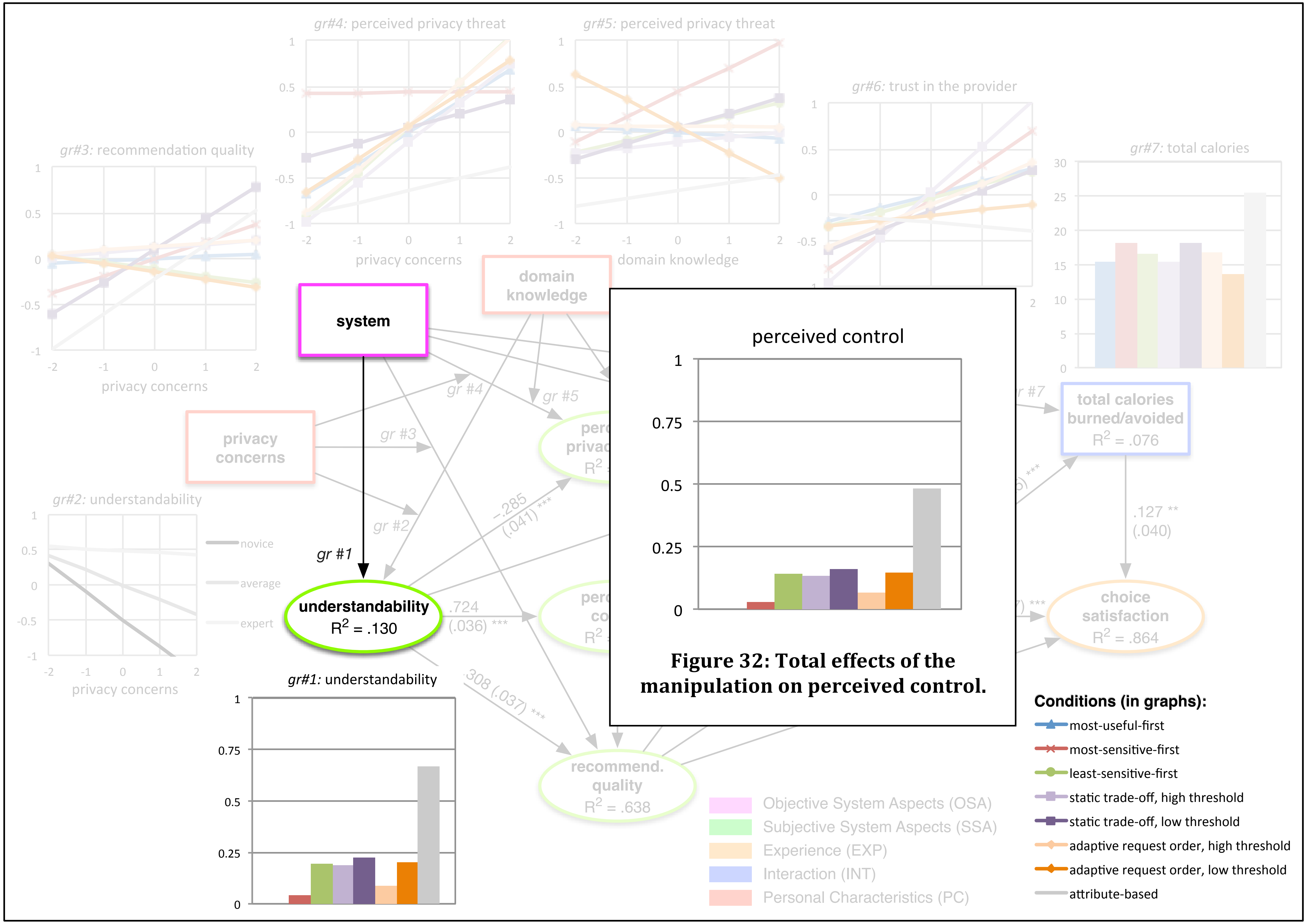
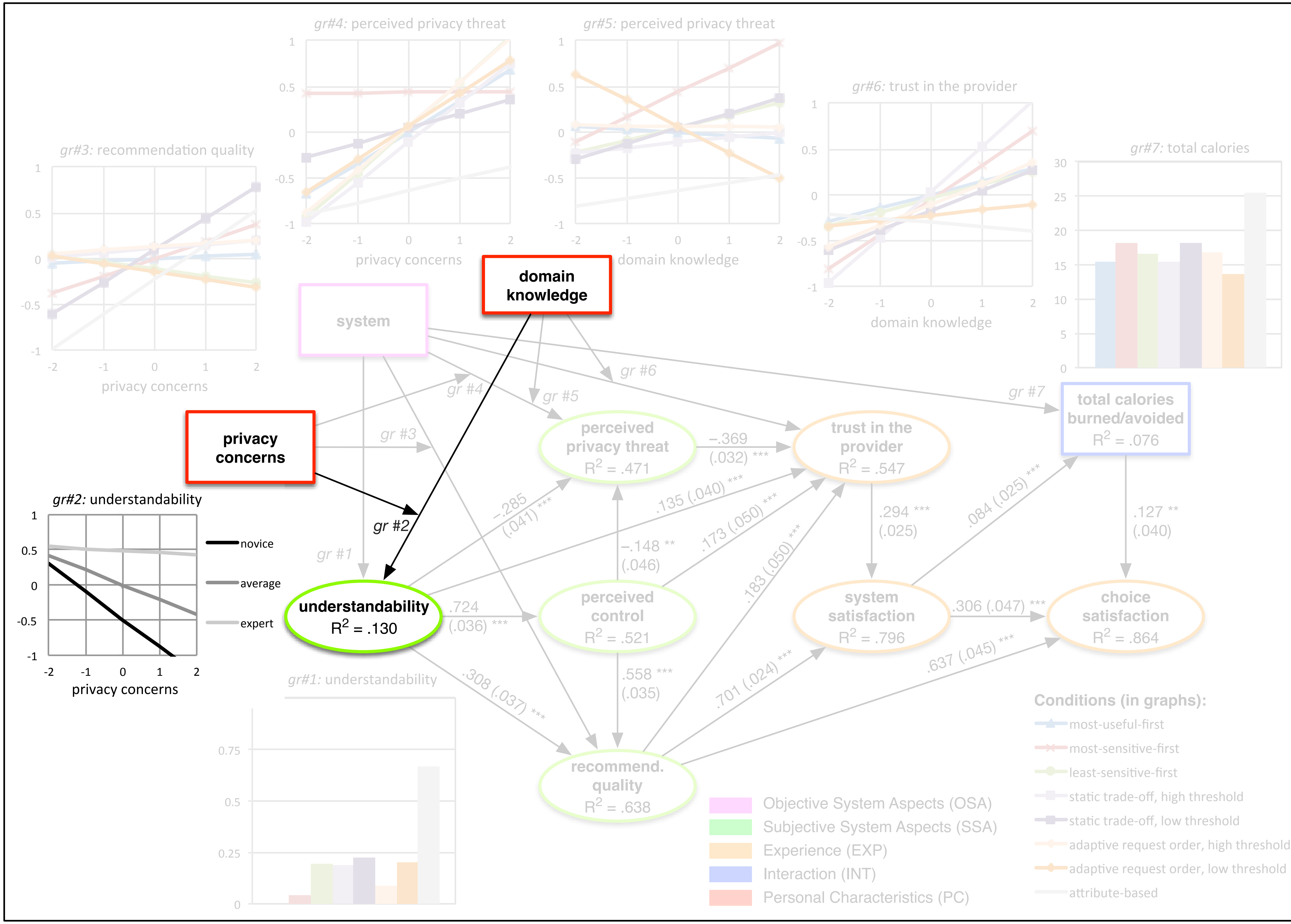


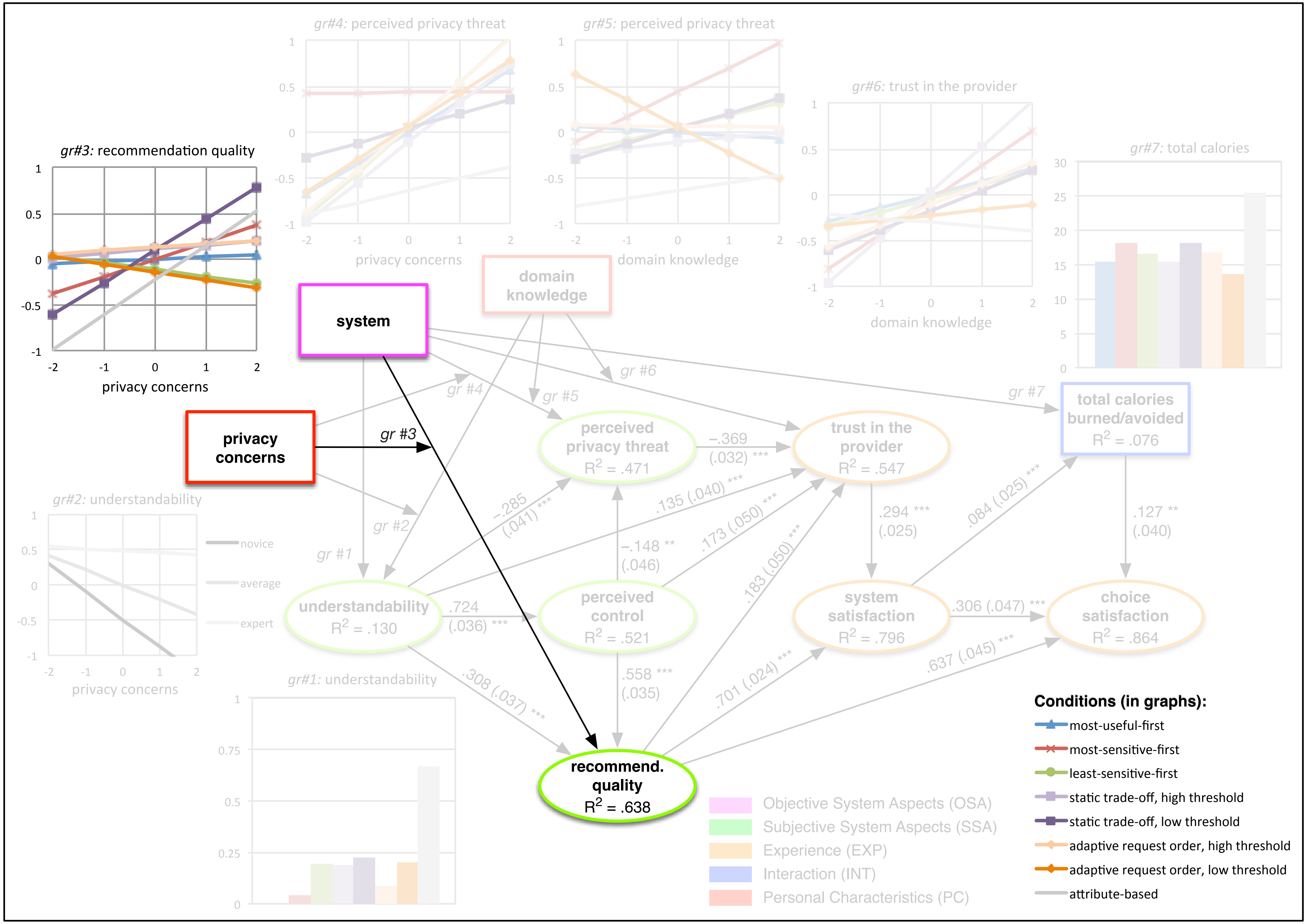
Figure 32: Total effects of the manipulation on perceived control.

Changing the request order did not increase understandability.

Possible solution: explanations/justifications, potentially adapted to the user.



Users with high concerns find the system less understandable unless they are experts, and novices find the system less understandable unless they have low concerns.



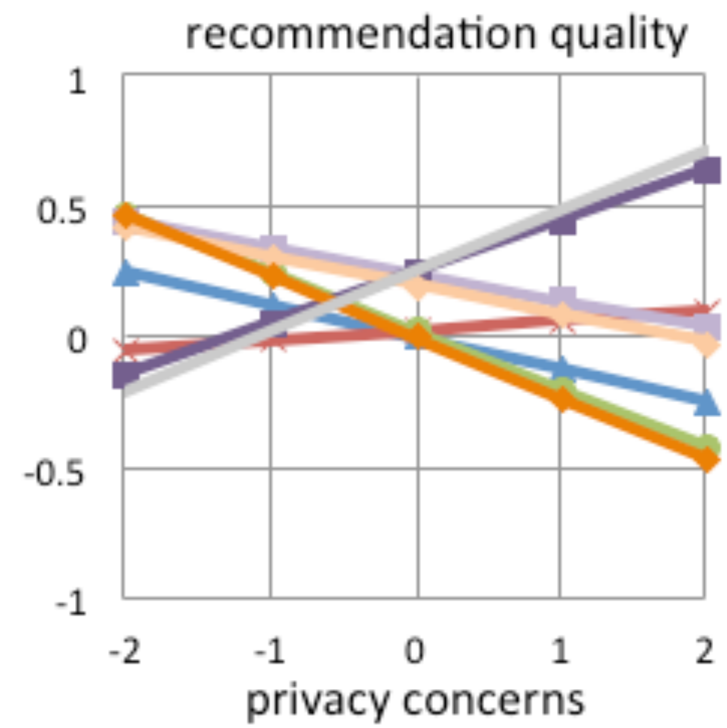
There is a interaction effect of privacy concerns and PE-method on recommendation quality.

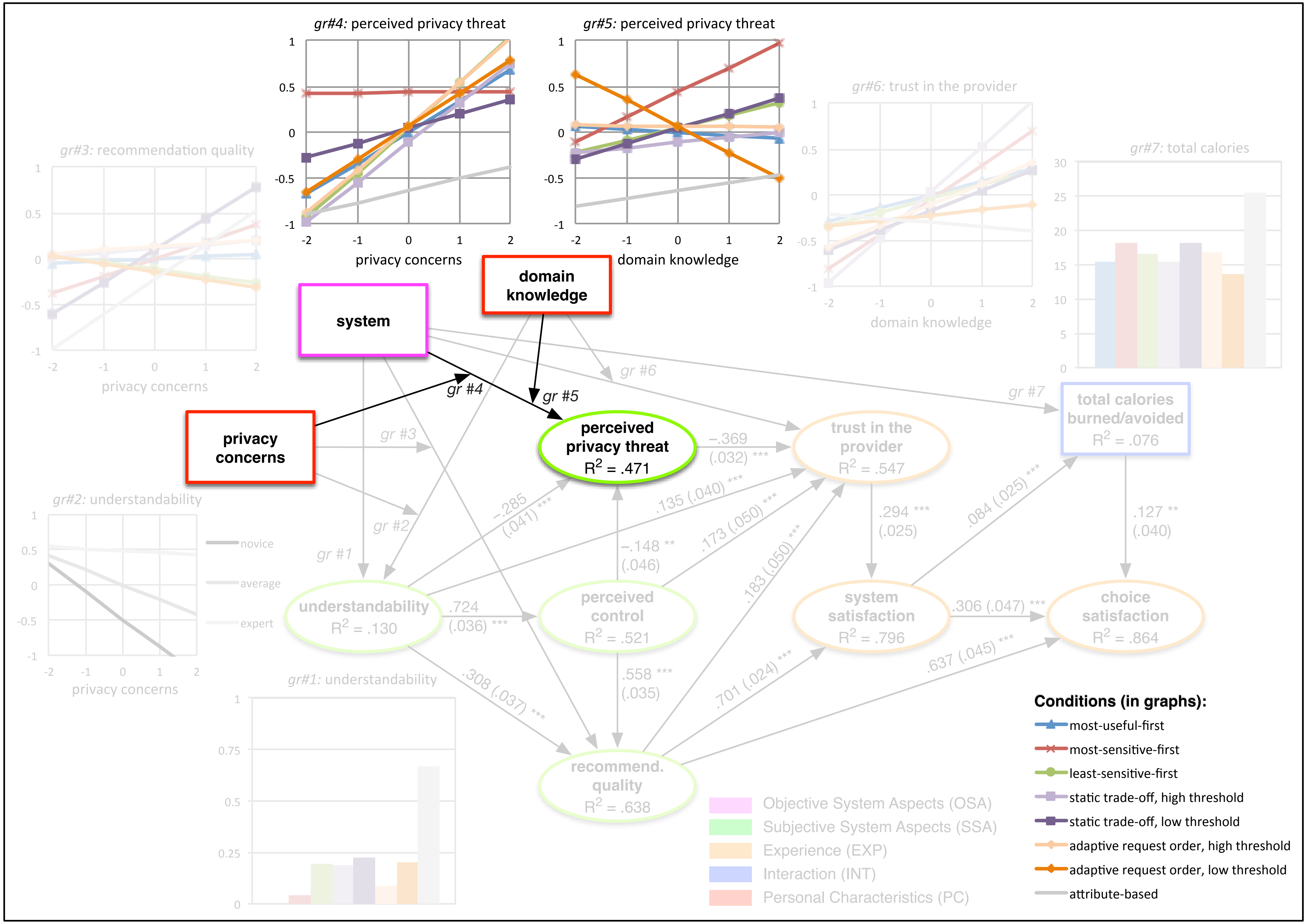
Total effect:

Attribute-PE (grey) and static trade-off, low threshold (purple) are best for people with high concerns.

Least-sensitive-first (green); static trade-off, high threshold (light purple); both adaptive request orders (orange and light orange) are best for people with low concerns.

Most-useful-first (blue) did not result in the highest overall recommendation quality! Recommendations quickly reach a "good enough" level, so users stop answering ("satisficing").





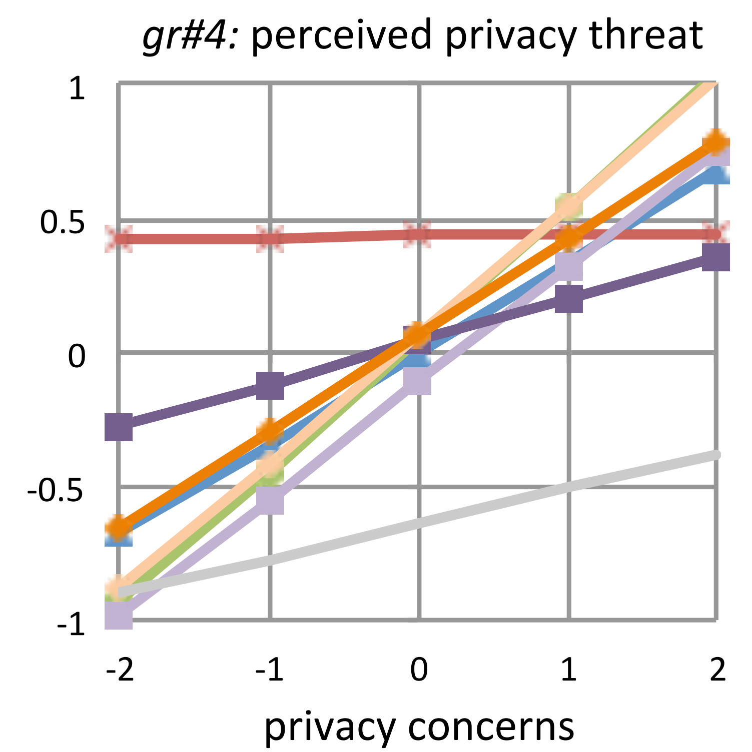
There are interactions of privacy concerns and PE-method, and domain knowledge and PE-method, on perceived privacy threat.

Interaction w/ privacy concerns:

People with higher concerns perceive more threat, except in attribute-PE (grey), most-sensitive-first (red), and static trade-off, low threshold (purple).

This makes attribute-PE by far the best for people with high concerns.

For people with low concerns, any version will do, except for most-sensitive-first (red) and static trade-off, low threshold (purple).

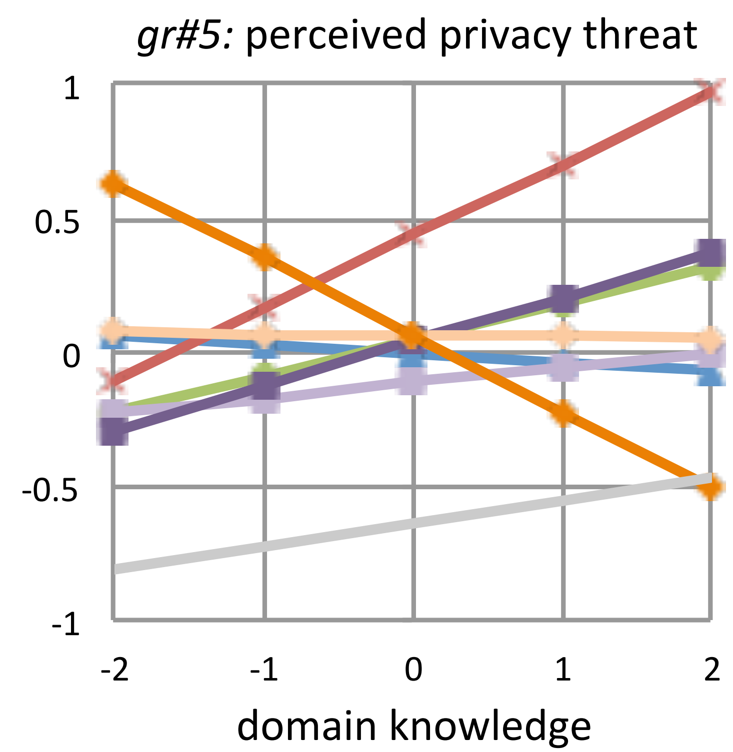


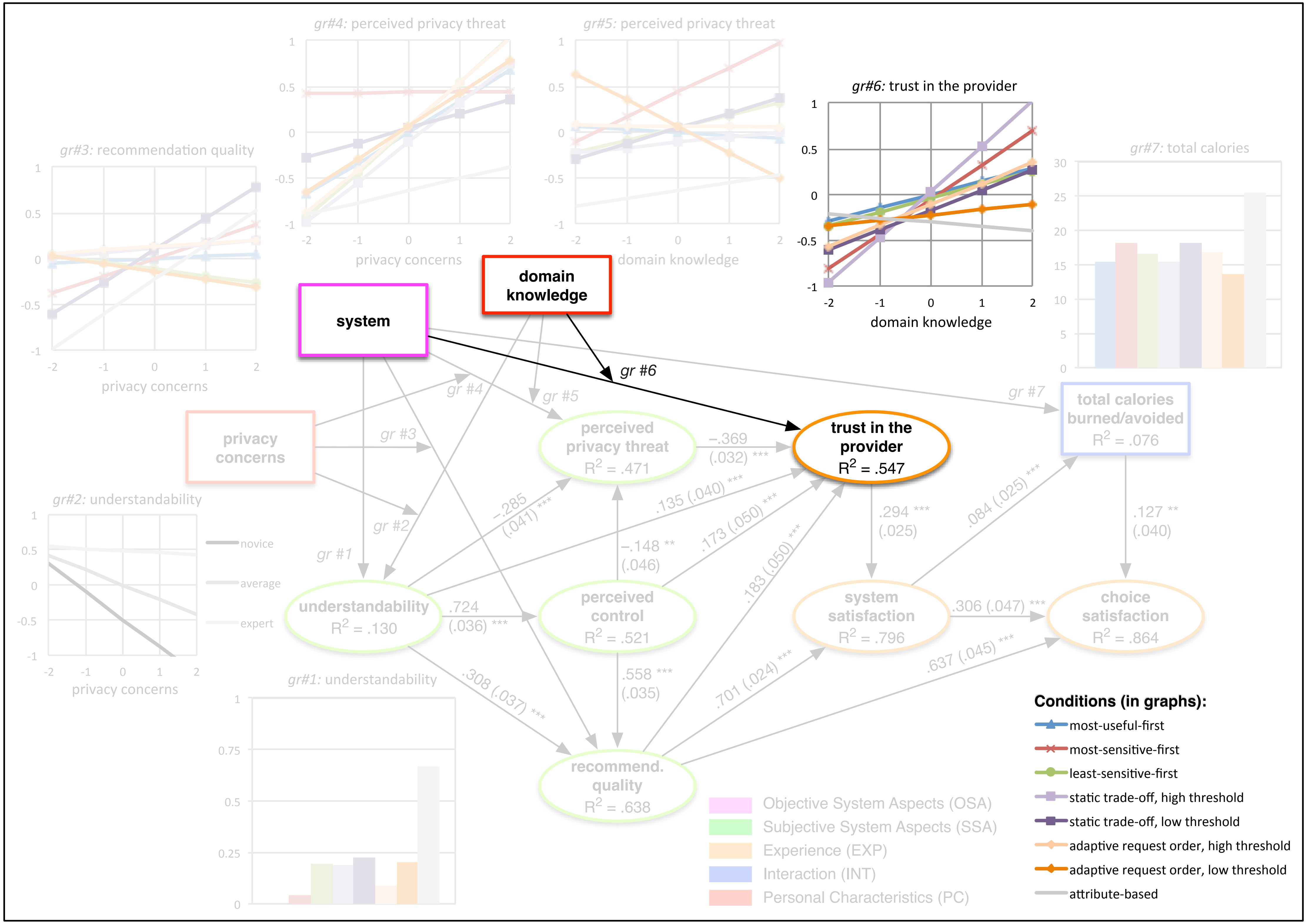
Interaction w/ domain knowledge:

Novices perceive more threat from adaptive request order, low threshold (orange), while experts perceive more threat from most-sensitive-first (red).

Most-sensitive-first: lower threat levels for novices!?

(Adaptive) justifications might fix threat assessment among novices





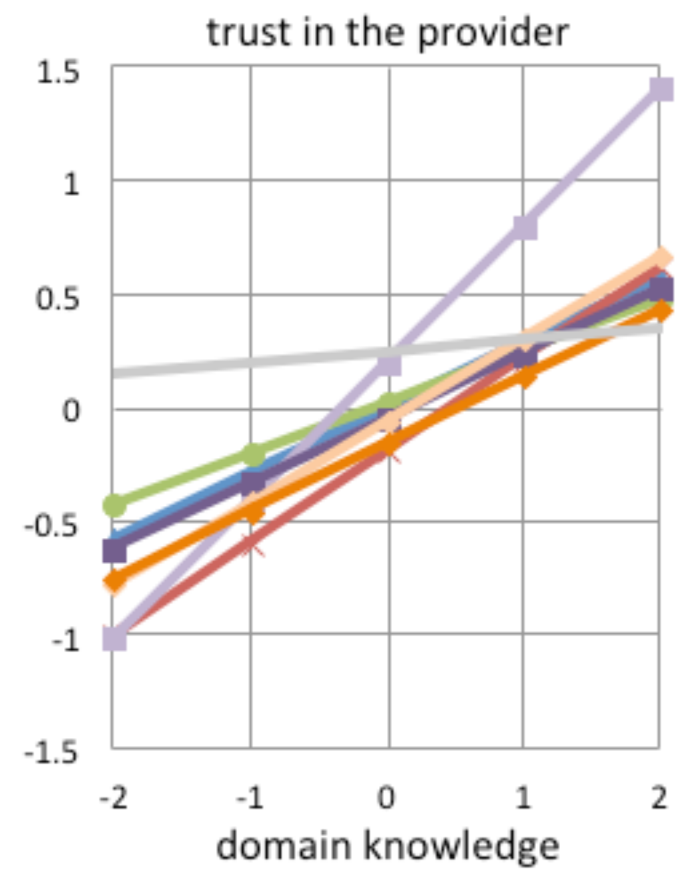
There is an interaction effect of domain knowledge and PE-method on trust in the provider.

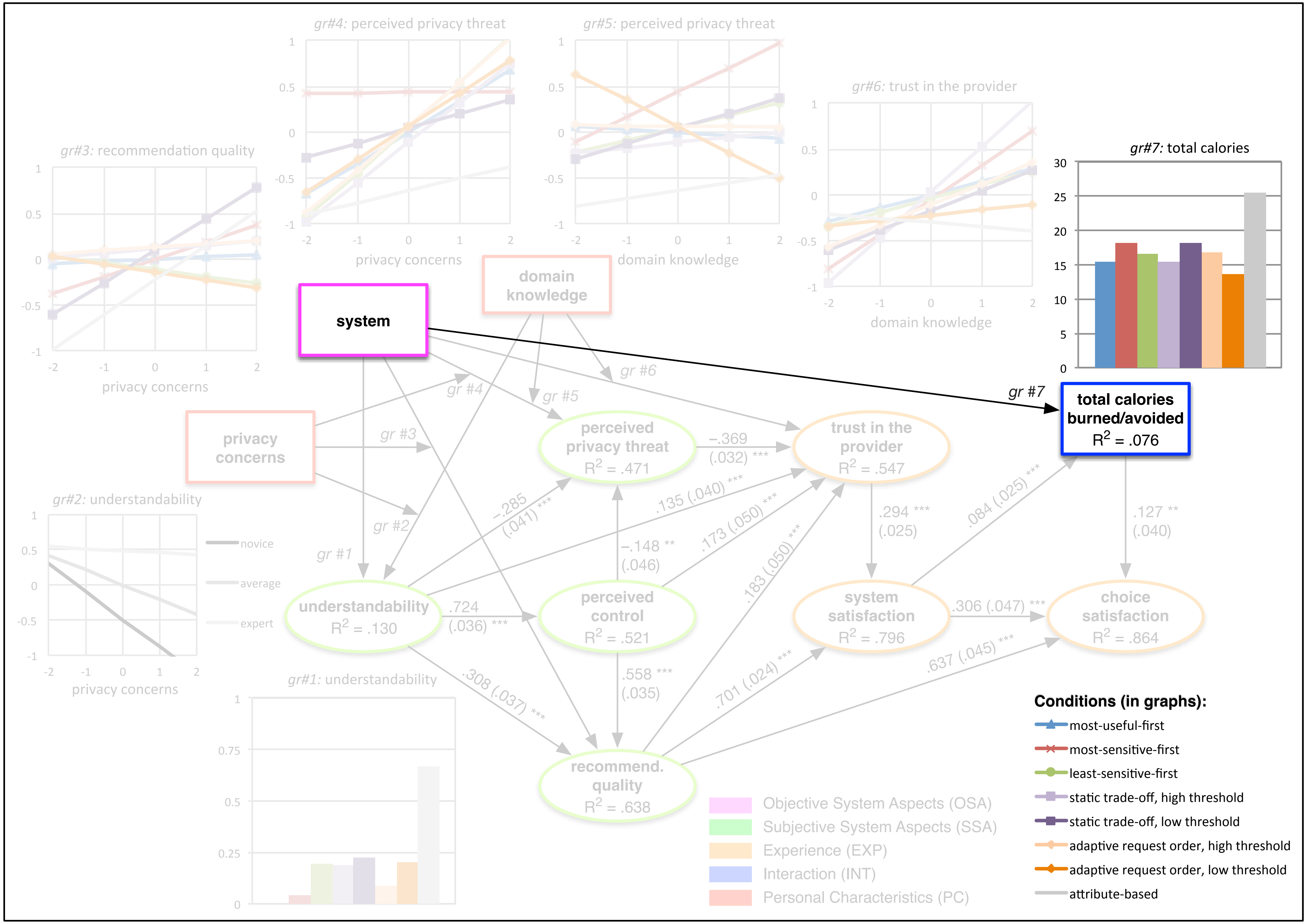
Total effect:

Experts are more trusting of static trade-off, high threshold.

Novices don't distinguish among different request orders; trust demographic-based system more.

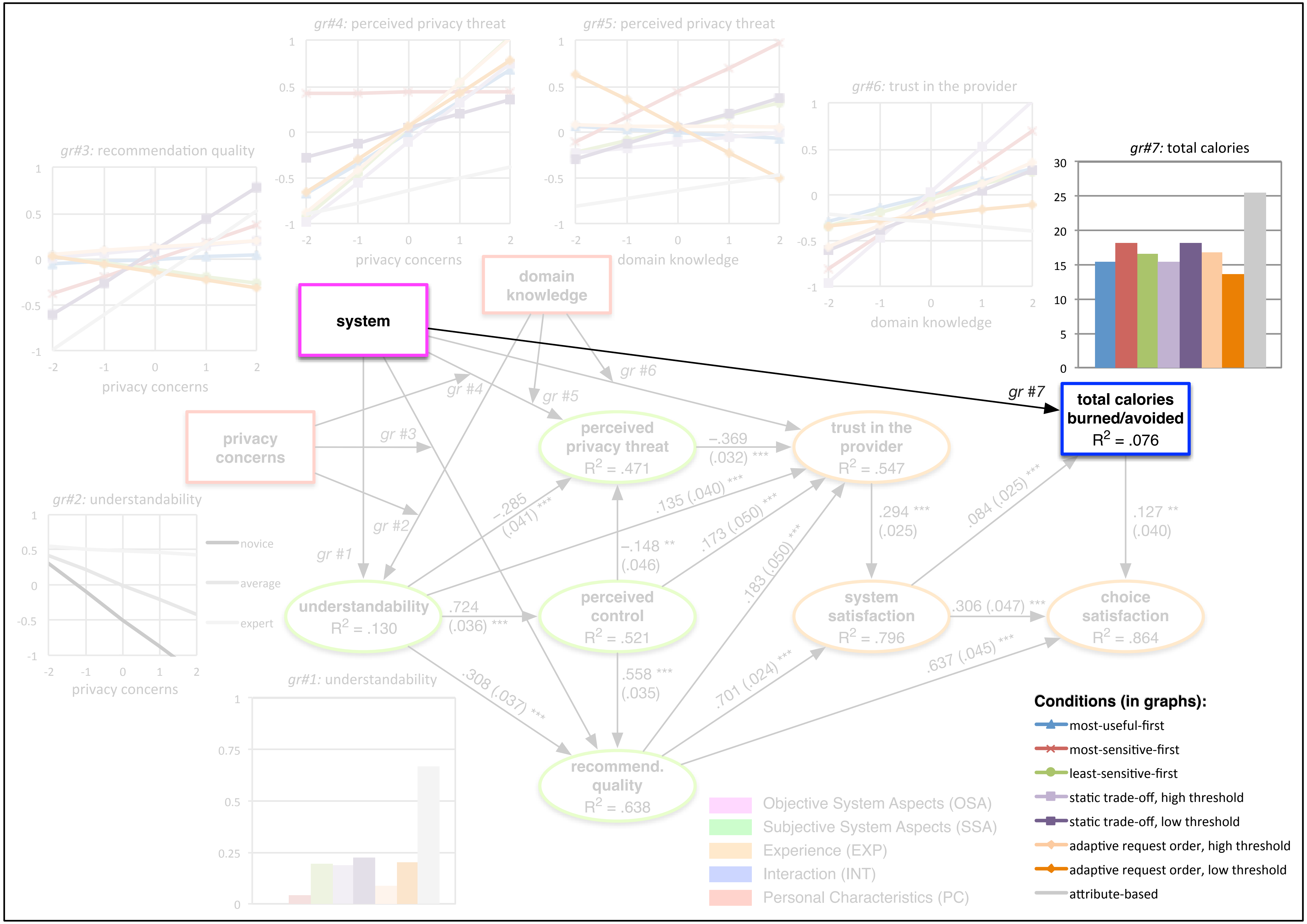
(Adaptive) justifications might fix trust assessment among novices.





Participants in attribute-PE burn/avoid substantially more calories.

Users in the demographics-PE may be distracted by the questions, and thus spend less time selecting measures.



Possible solution: a hybrid method with demographics-PE first, then attribute-PE.

Could be adaptive: ask more questions to novices and privacy-unconcerned users before switching.

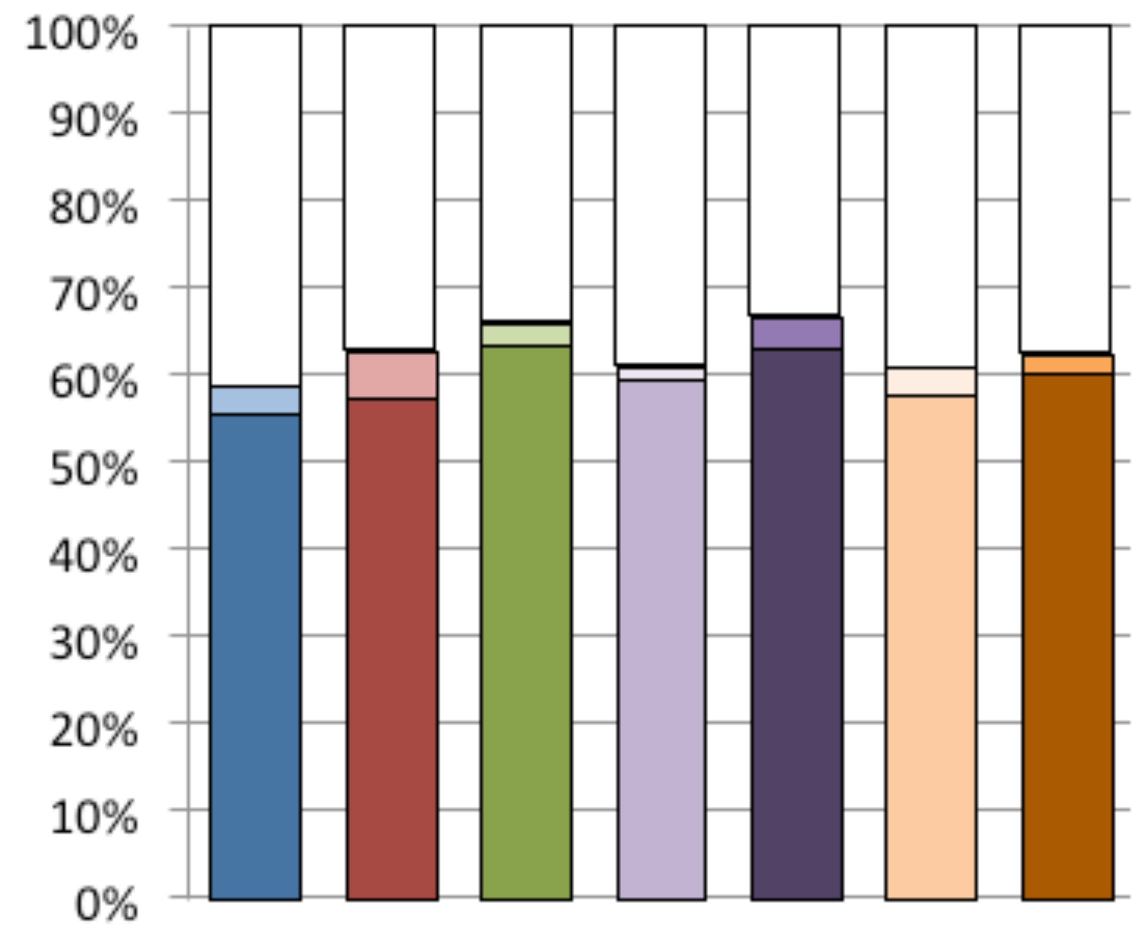
Disclosure results

Participants see fewest items in most-useful-first (blue)

Arguably due to satisficing.

Remedy: nudge users to answer (or at least review) more questions.

Or: adaptive nudge that stops when all remaining questions are above the threshold.

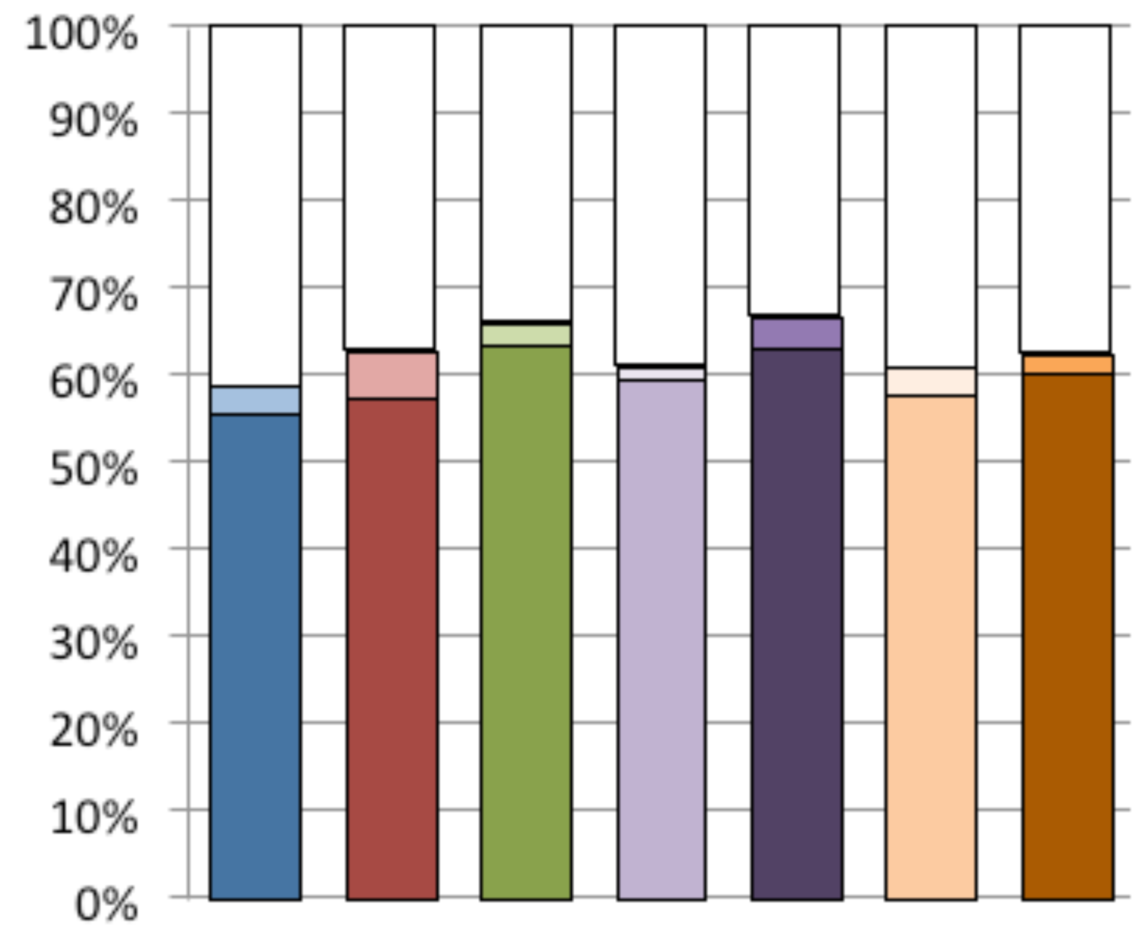


colored parts: seen
darker part: disclosed
lighter part skipped.

Disclosure results

Participants see most items in least-sensitive-first (green) and static trade-off, low threshold (purple).

Both of these conditions show very non-sensitive items upfront.



colored parts: seen
darker part: disclosed
lighter part skipped.

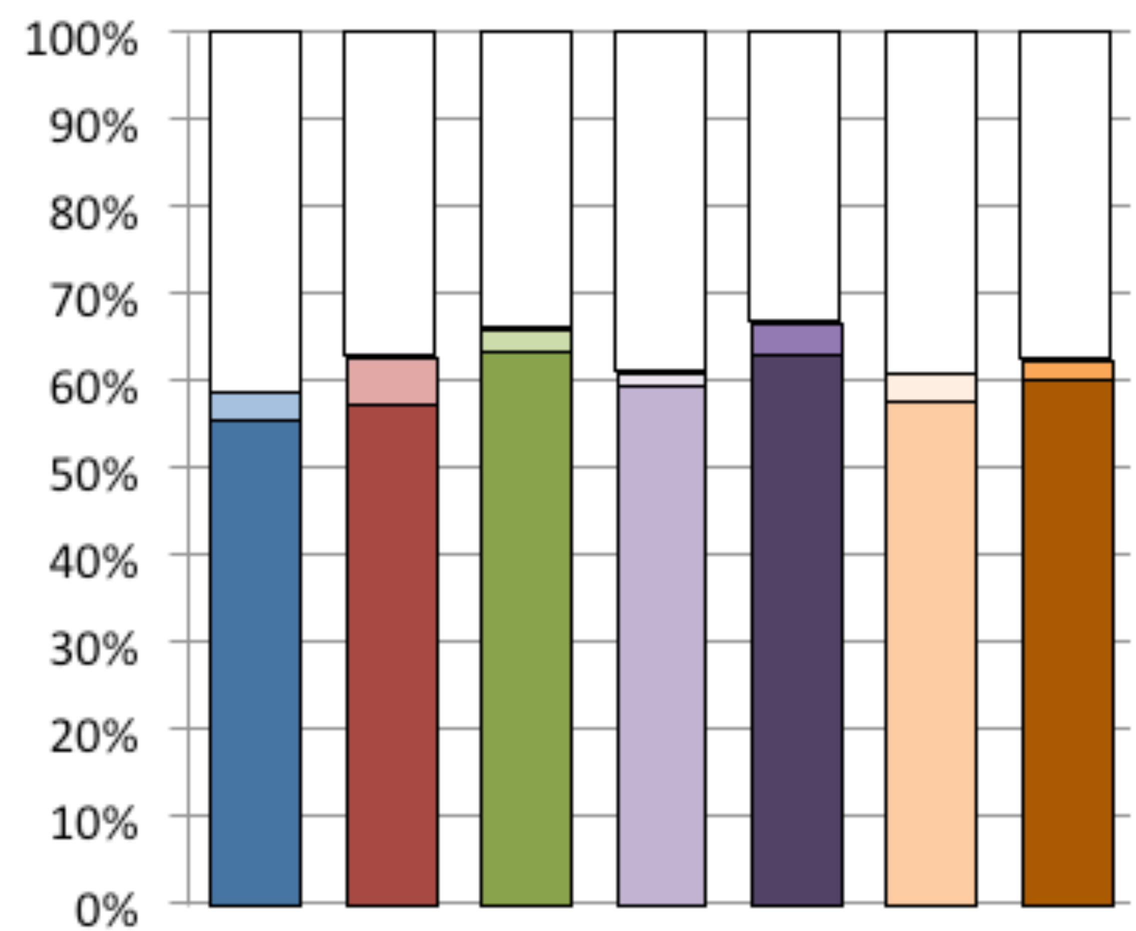
Disclosure results

Of the items that they see, participants disclose most in static trade-off, high threshold (light purple)

> This condition is thus the most "efficient", meaning that the least requests are wasted on items that users will not disclose anyway.

They disclose least in most-sensitive-first (red).

> In line with perceived threat.

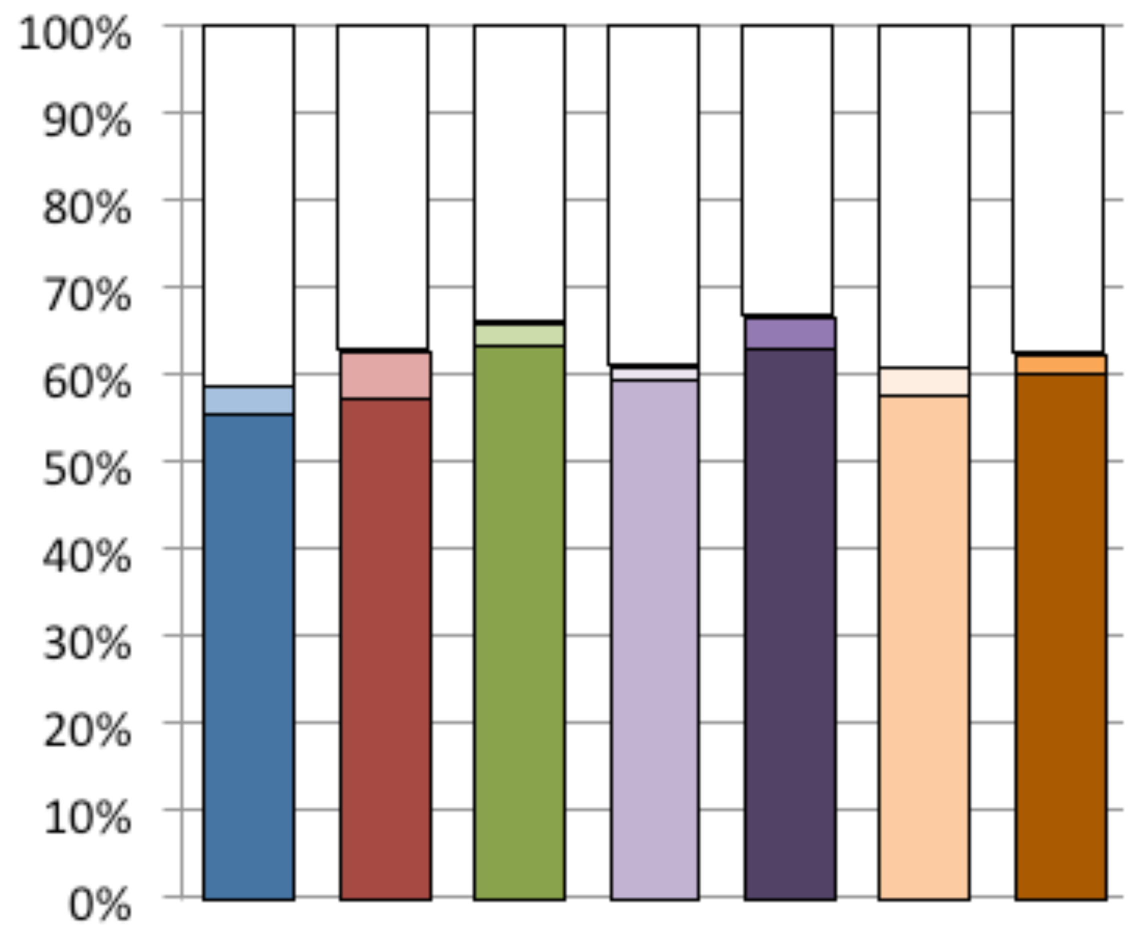


colored parts: seen
darker part: disclosed
lighter part skipped.

Disclosure results

Overall level of disclosure is highest in least-sensitive-first (green) and static trade-off, low threshold (purple).

Additional results show that the latter is true for experts but not for novices.



colored parts: seen
darker part: disclosed
lighter part skipped.

Study 2 conclusion

"Best" condition for each type of user:

- **Novices:** attribute-based PE.
- **Experts:** demographics-based PE with static trade-off, high threshold.
- **Low concerns:** any method, except most-sensitive-first; the static trade-off, low threshold; and in some cases attribute-based PE.
- **High concerns:** the attribute-based PE and demographics-based PE with static trade-off, low threshold.

Study 2 conclusion

Adaptive request orders did not end up among the "best" versions.

Static trade-off versions are better: this may be because they provide a **guaranteed upper bound** on sensitivity.

Possible remedy: put an upper bound on the adaptive threshold.

Other improvements:

- **Adaptive justifications** that increase understandability and trust
- **Adaptive nudges** to encourage users to explore more demographics questions
- **Adaptive hybrid recommender** that starts with demographics-PE and then switches to attribute-PE.

Reflection

User-Tailored Privacy:

- Relieves some of the burden of controlling privacy, while at the same time respecting each individual's preferences
- Provides **realistic empowerment**: the right amount of transparency and the right amount of control
- Refrains from making moral judgments about what the "right" level of privacy should be

The best way forward to support people's privacy decisions!