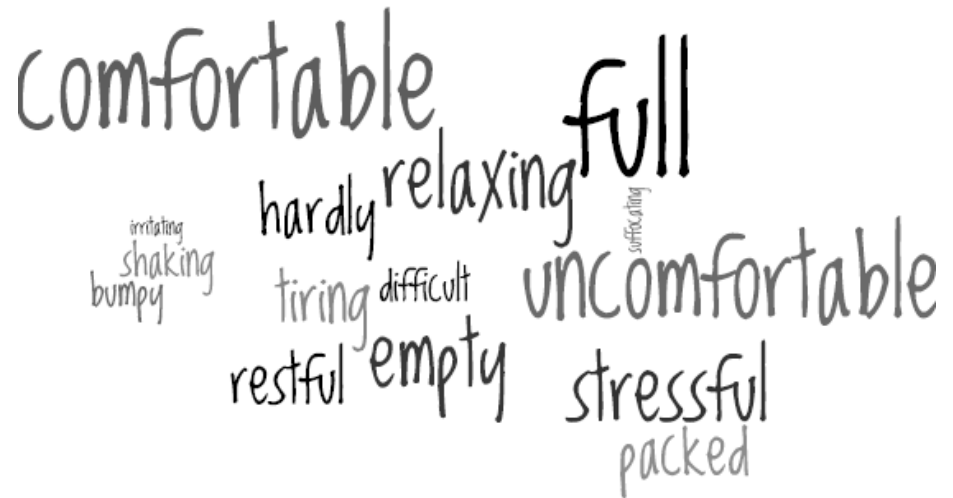


# THE USE OF WORD DATA TO MEASURE PERCEPTION IN HYBRID CHOICE MODELS

---

Aurélie Glerum

Michel Bierlaire



13<sup>th</sup> International Conference on Travel Behavior Research  
17th July 2012

## Introduction & motivation

### The data

- RP survey
- Adjective quantification survey

### The integrated model framework

- Discrete choice model
- Latent variable model
- Quantification model

### Application example

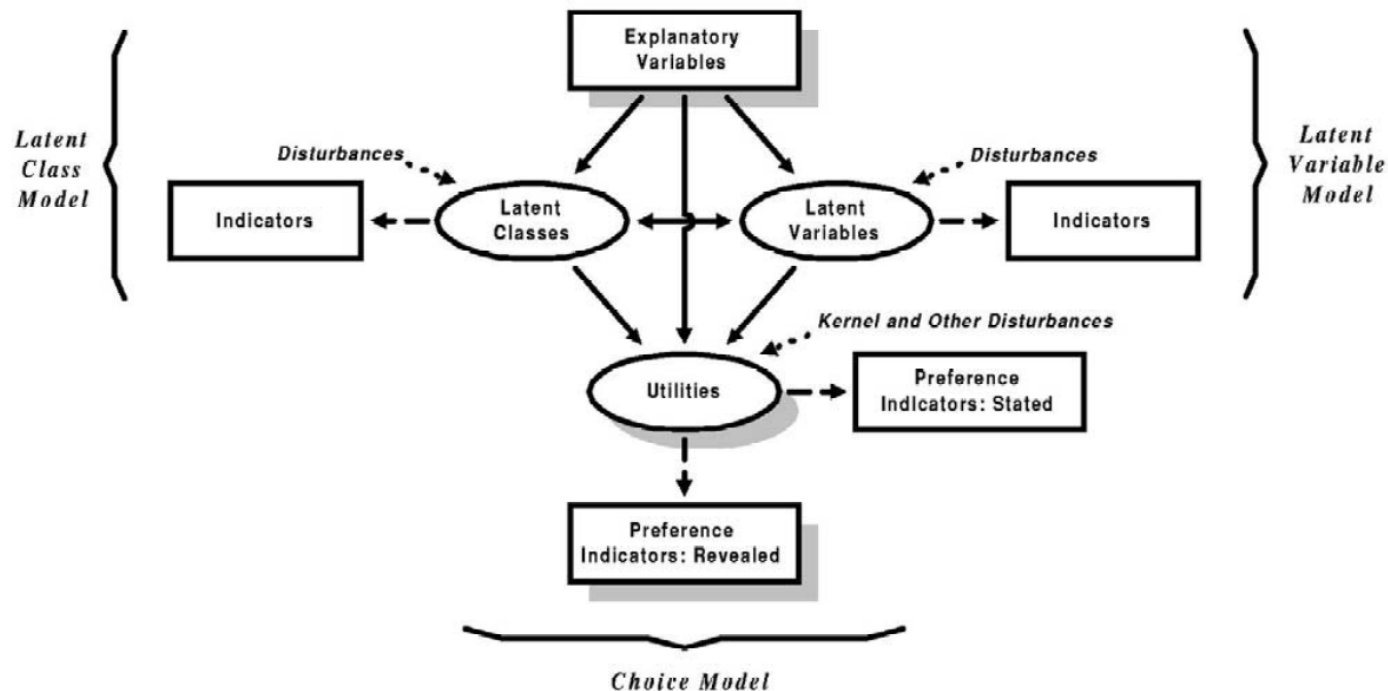
- Quantification model

## Conclusion

## Recent developments in discrete choice modeling (DCM)

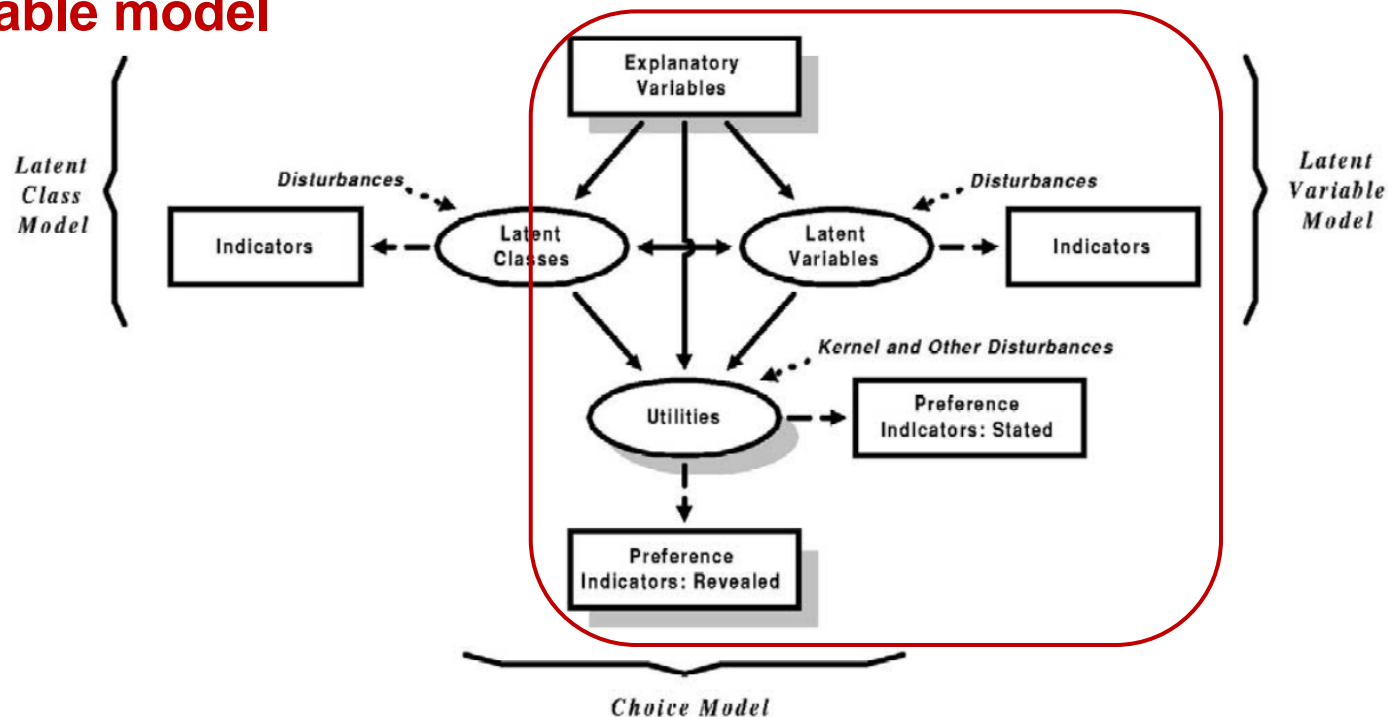
- Choice cannot only be explained by economic indicators (travel duration, price of a trip, etc.)
- **Psychological constructs (attitudes, perceptions, etc.) play important role in choice behavior:** need to be integrated in an appropriate way into DCMs.
- Framework handling this issue:  
**hybrid choice model (HCM)** framework  
(Walker, 2001; Ben-Akiva et al., 2002)

Hybrid choice model (HCM): DCM with latent constructs.



**Hybrid choice model (HCM):** DCM with latent constructs.

In this research: focus on the **integration of choice model and latent variable model**



Issues related to the integration of latent variables into choice models:

## 1. **Measurement of latent variable**

⇒ How to obtain the most realistic and accurate measure of a perception?

## 2. **Integration of the measurement into the choice model**

⇒ How to incorporate this information in the choice modeling framework?

## 1. Measurement of latent variable:

- Use of **opinion statements**  
Five-point Likert scale } Usual way in literature  
(Likert, 1932; Bearden and Netemeyer, 1999)

- **Recent technique** developed in **social sciences:**

Respondents report **adjectives** characterizing a variable of interest (Kaufmann et al., 2001; Kaufmann et al., 2010)

Reflects **spontaneous** perceptions of individuals  
(≠ survey designer's conception of the perception)

## 2. Integration of the measurement into the choice model:

- **Structural equation model (SEM)** framework used to characterize latent variable and relate it to its measurement indicators (e.g. Bollen, 1989).
- Latent variable model embedded into DCM  $\Rightarrow$  **HCM framework**  
Many studies (Espino et al., 2006; Abou-Zeid et al., 2010; Van Acker et al., 2011; Daziano and Bolduc, 2011; Atasoy et al., forthcoming)
- **Integration of measurements into HCM framework:**
  - Easy for models with opinion statements
  - Needs an **additional modeling step** for model with **adjectives**



## Purpose of the research:

Develop an HCM that uses adjectives as measurements of latent construct

## Steps:

1. Collection of **choice data** & psychometrics in the form of **adjectives**
2. **Quantification of adjectives**:
  1. **Survey** to obtain ratings of adjectives
  2. **Quantification model**
  3. **Integration** of the quantification model into the **HCM** framework

## Purpose of the research:

Develop an HCM that uses adjectives as measurements of latent construct

## Steps:

1. Collection of **choice data** & psychometrics in the form of **adjectives**

2. **Quantification of adjectives:**

1. **Survey** to obtain ratings of adjectives

2. **Quantification model**

3. **Integration** of the quantification model into the **HCM** framework

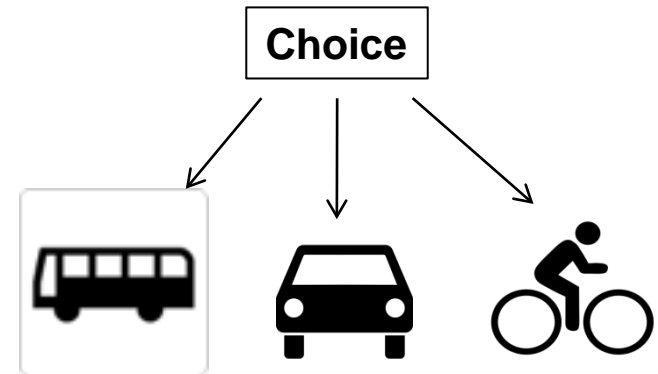
**FOCUS OF THIS  
PRESENTATION**

## Two surveys:

- Revealed preferences (RP) survey
- Survey with evaluators (adjective quantification survey)

### RP survey

- **Mode choice study**
- Conducted between 2009-2010 in low-density areas of Switzerland
- Conducted with PostBus (major bus company in Switzerland, operates in low-density areas)
- Info on **all trips performed by inhabitants in one day**:
  - Transport mode
  - Trip duration
  - Cost of trip
  - Activity at destination
  - Etc.
- **1763 valid questionnaires** collected



## Adjective data for perception of transport modes:

*For each of the following transport modes, give three adjectives that describe them best according to you.*

		Adjective 1	Adjective 2	Adjective 3
1	The car is:			
2	The train is:			
3	The bus, the metro and the tram are:			
4	The post bus is:			
5	The bicycle is:			
6	The walk is:			

### Adjective data for perception of transport modes:

*For each of the following transport modes, give three adjectives that describe them best according to you.*

		Adjective 1	Adjective 2	Adjective 3
1	The car is:	<b>convenient</b>	<b>comfortable</b>	<b>expensive</b>
2	The train is:	<b>relaxing</b>	<b>punctual</b>	<b>restful</b>
3	The bus, the metro and the tram are:	<b>fast</b>	<b>frequent</b>	<b>cheap</b>
4	The post bus is:	<b>punctual</b>	<b>comfortable</b>	<b>cheap</b>
5	The bicycle is:	<b>stimulating</b>	<b>convenient</b>	<b>cheap</b>
6	The walk is:	<b>healthy</b>	<b>relaxing</b>	<b>independent</b>

### Extraction of information on perceptions

#### 1. Classification into themes:

- Perception of cost
- Perception of time
- Difficulty of access
- Flexibility
- Comfort, etc.

#### 2. Focused on adjectives related to one theme only and one mode only:

### Comfort in public transportation (PT)

---

#### Comfort

---

hardly full

packed

bumpy

comfortable

hard

irritating

tiring

unsuitable with bags

uncomfortable

bad air

...

---

## Extraction of information on perceptions

### 1. Classification into themes:

- Perception of cost
- Perception of time
- Difficulty of access
- Flexibility
- Comfort, etc.

### 2. Focused on adjectives related to one theme only and one mode only:

**Comfort in public transportation (PT)**



**LATENT VARIABLE  
WE STUDY**

### Comfort

hardly full

packed

bumpy

comfortable

hard

irritating

tiring

unsuitable with bags

uncomfortable

bad air

...

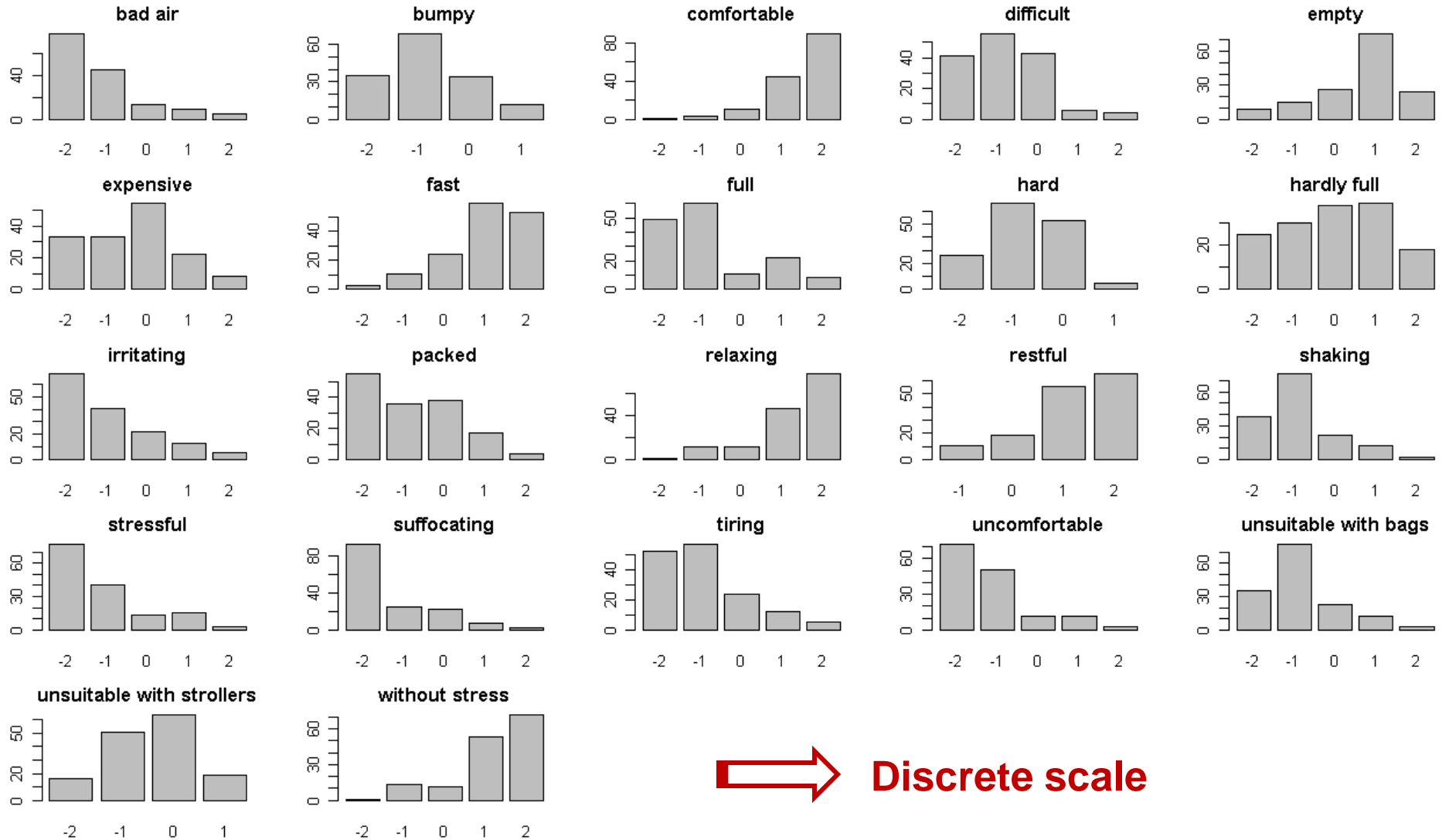


## ADJECTIVE QUANTIFICATION SURVEY

### Adjective quantification survey

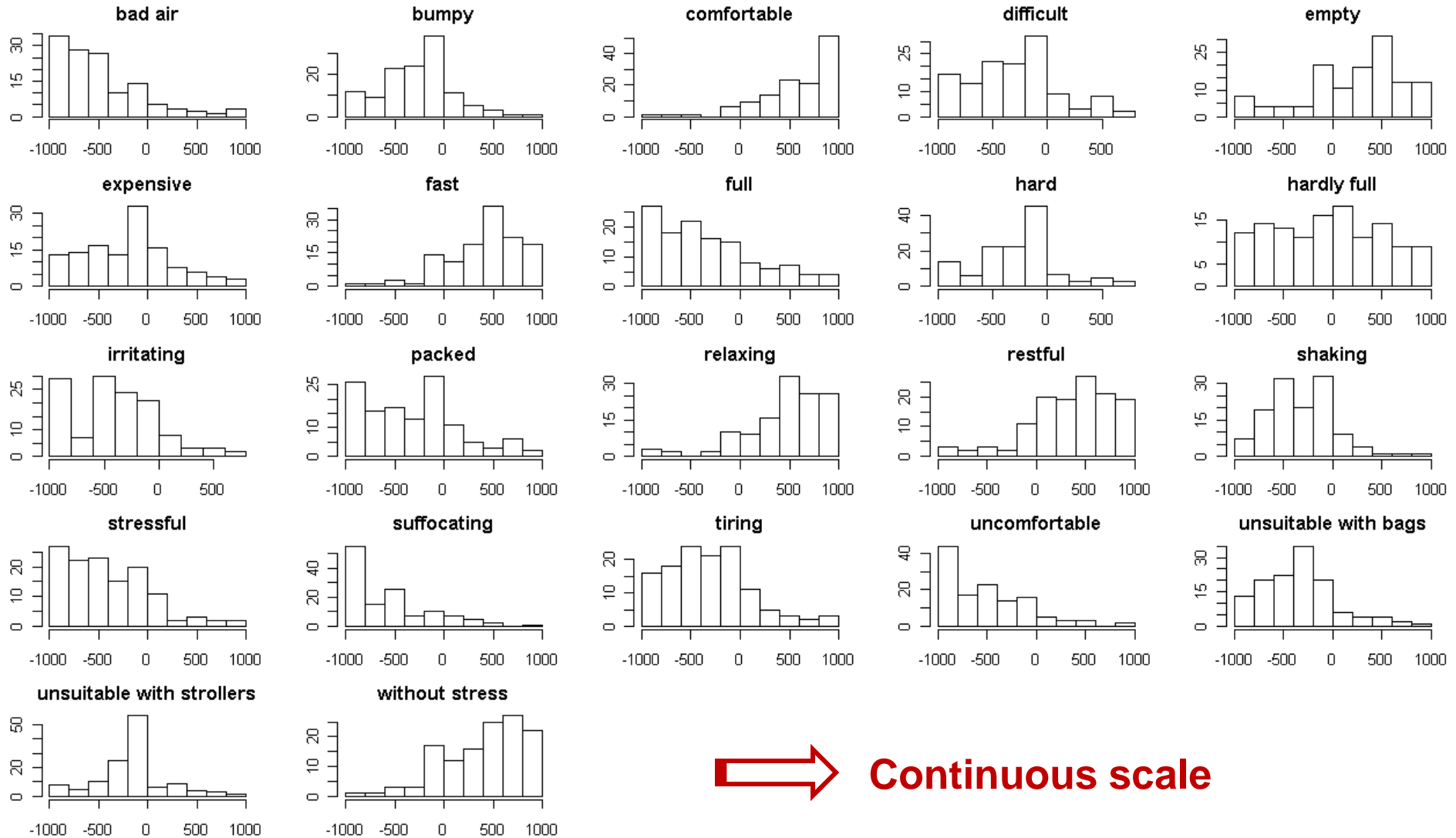
- Asked external evaluators to rate the adjectives on scale of comfort.
- Two scales:
  - Discrete scale: ratings from -2 to 2.
  - Continuous scale: ratings from -1000 to 1000.
- Number of evaluators: 277

## ADJECTIVE QUANTIFICATION SURVEY



**Discrete scale**

## ADJECTIVE QUANTIFICATION SURVEY



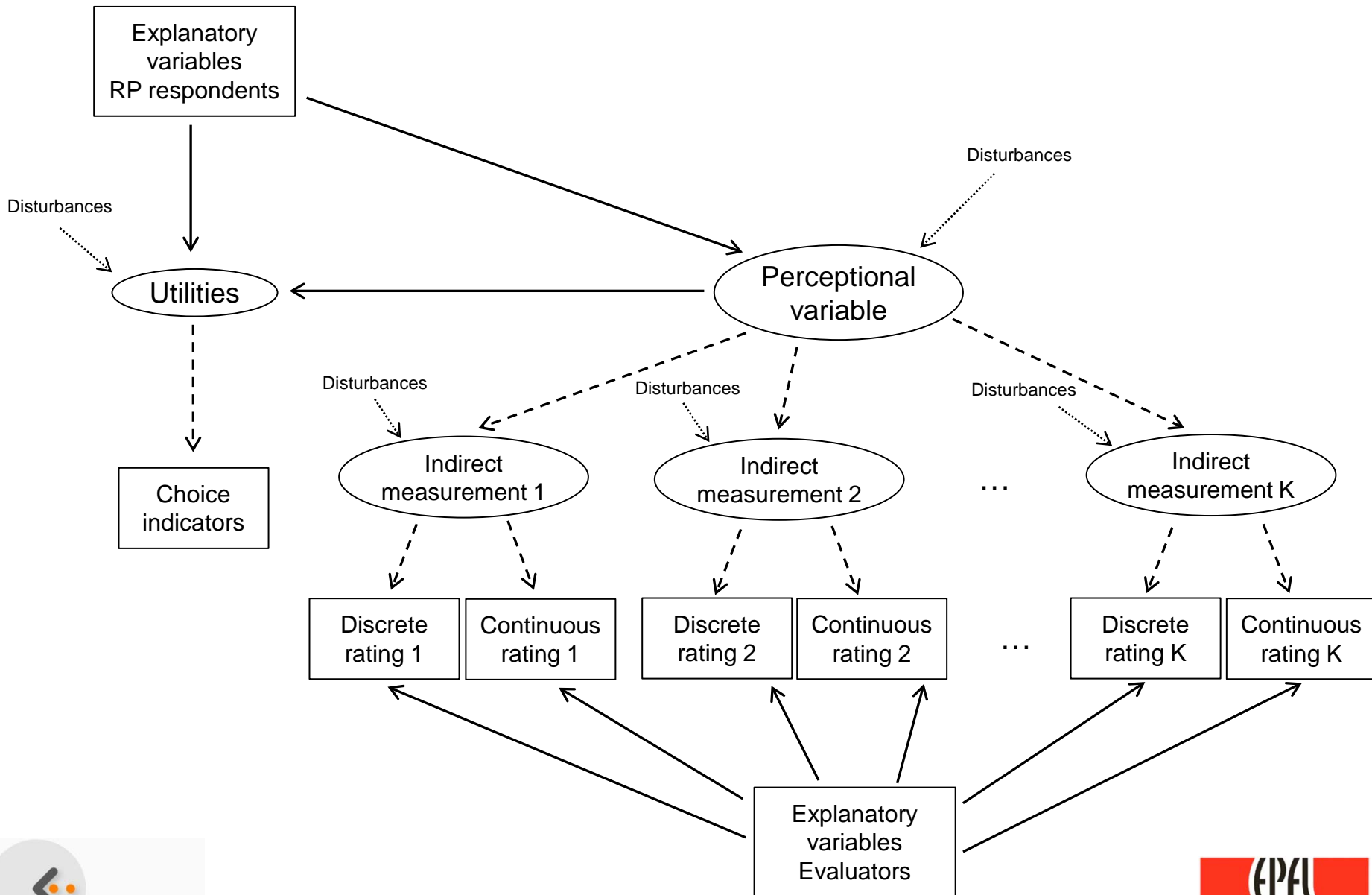
## Goal of HCM framework:

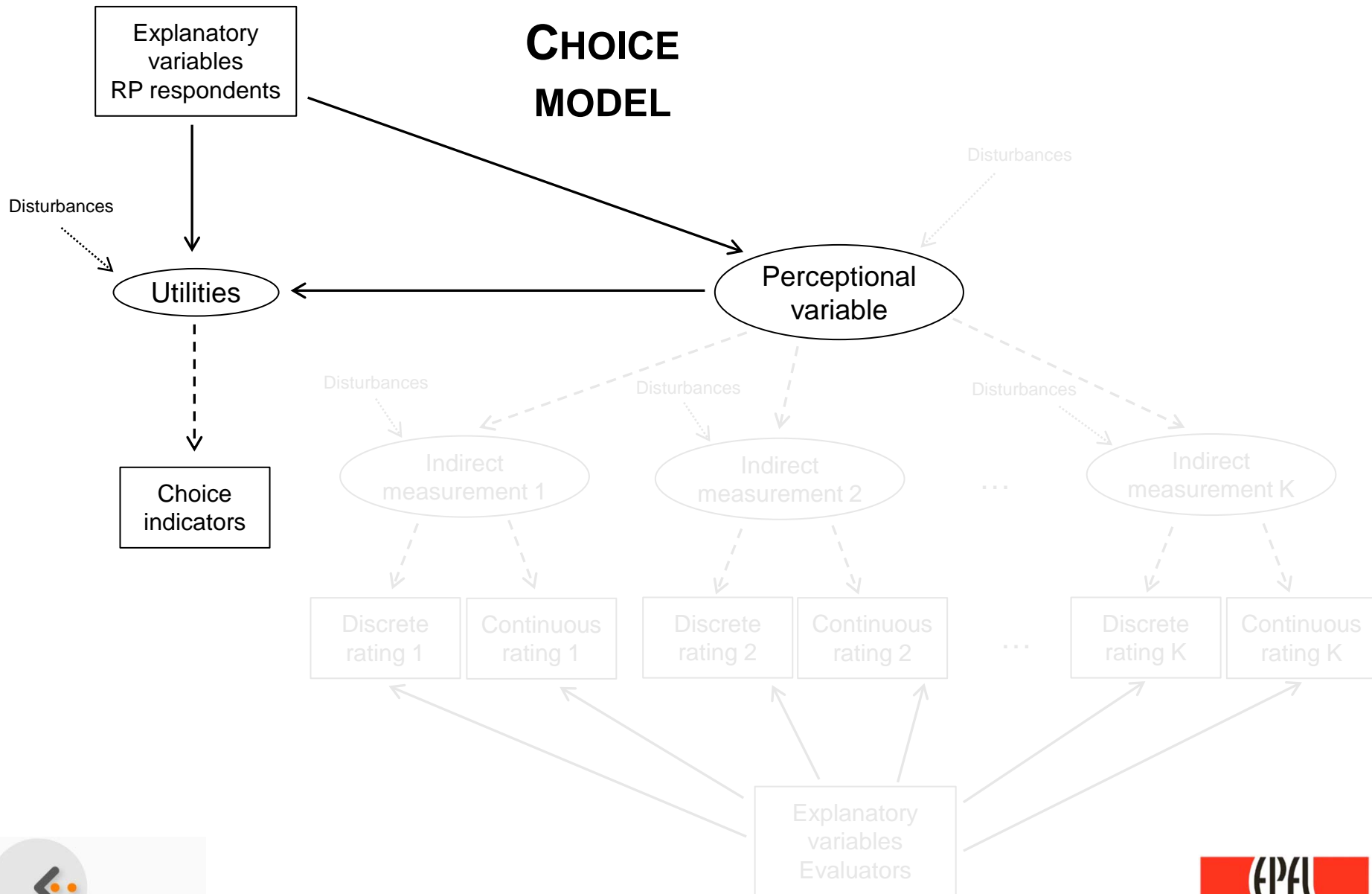
Assess impact of perception on choice.

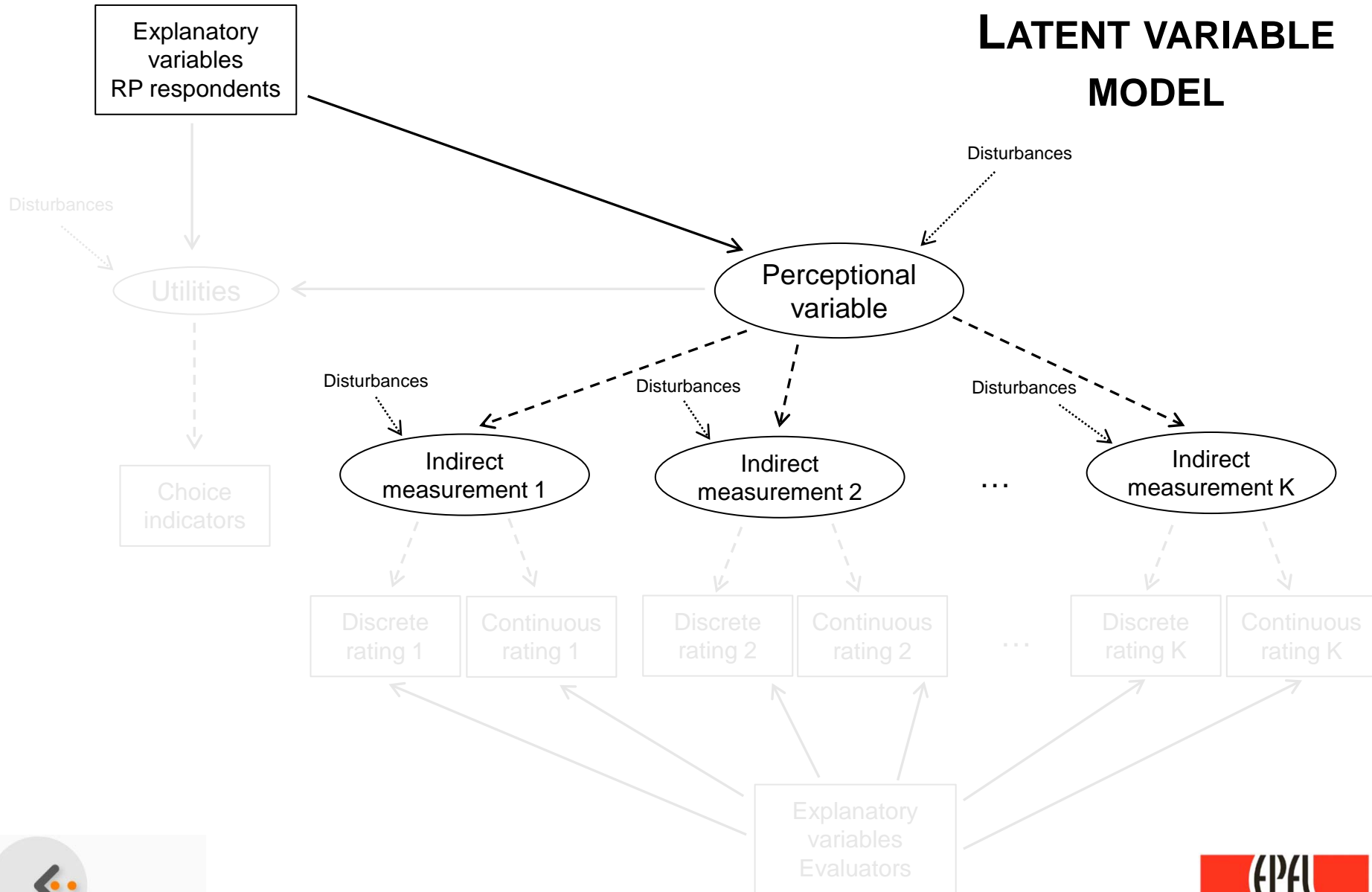
Using adjective data  $\Rightarrow$  need following integrated framework.

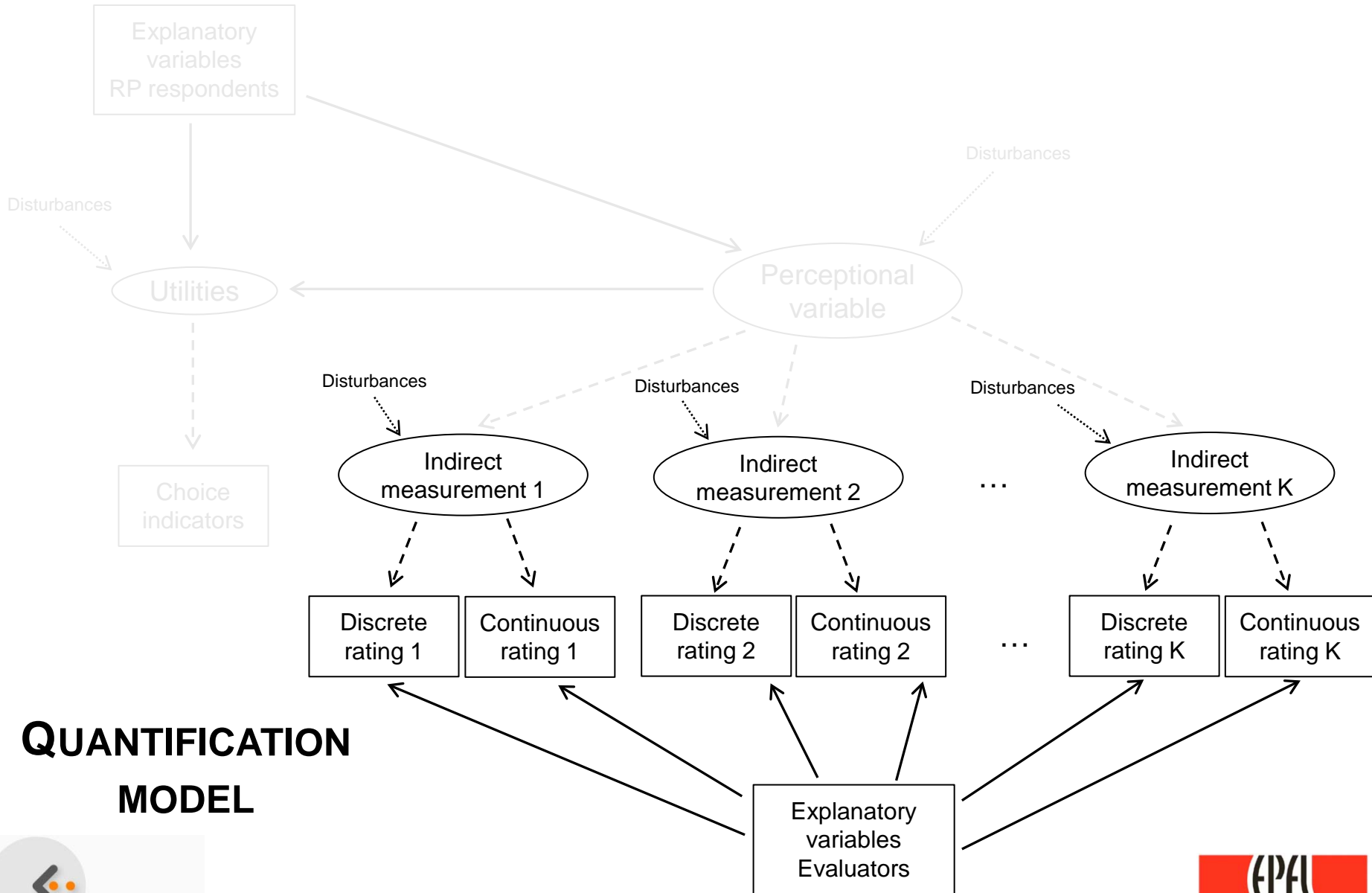
## Framework involves three components:

- Discrete choice model
- Latent variable model for the perception
- Quantification model for the indicators of the latent variable

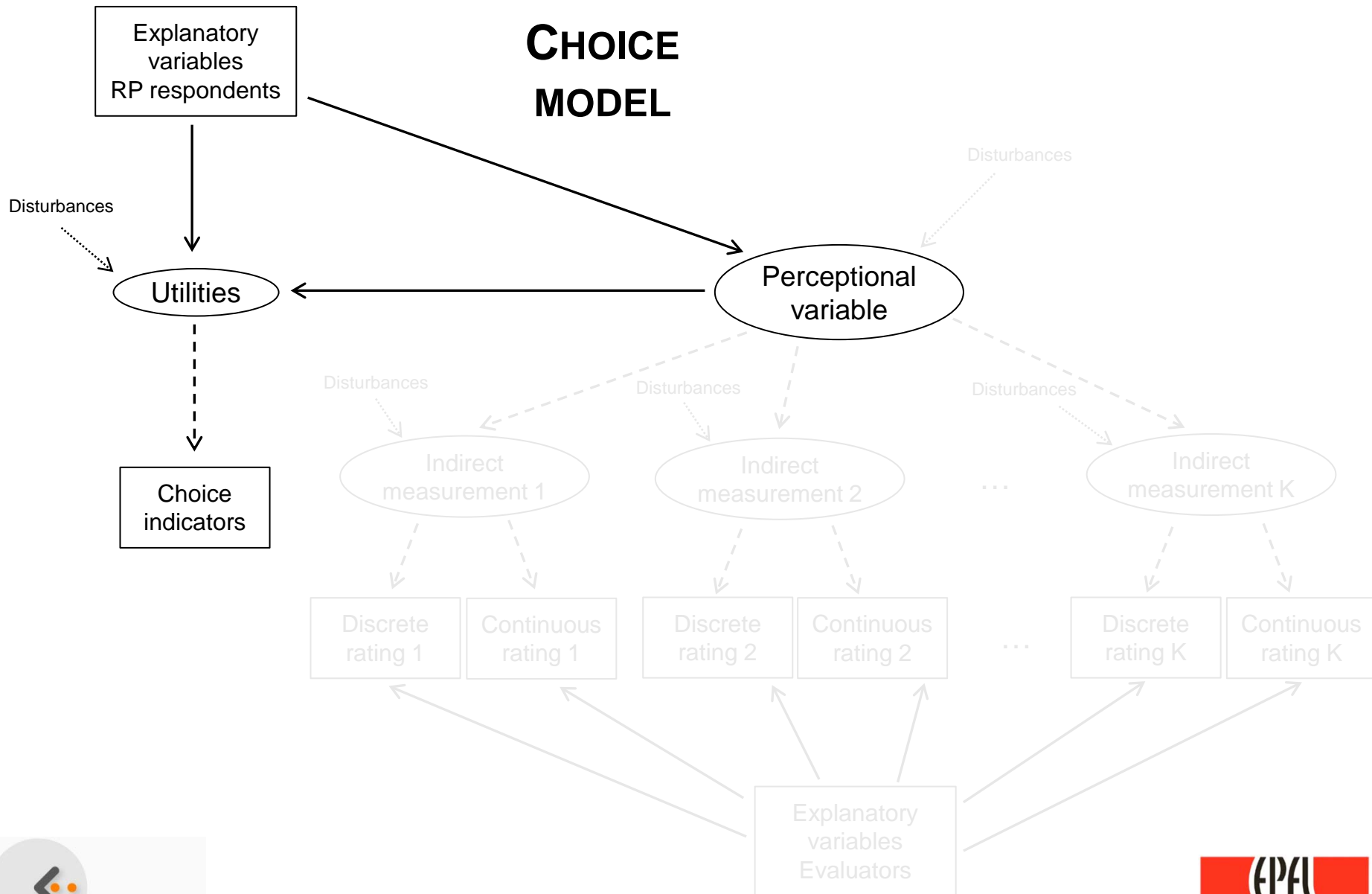






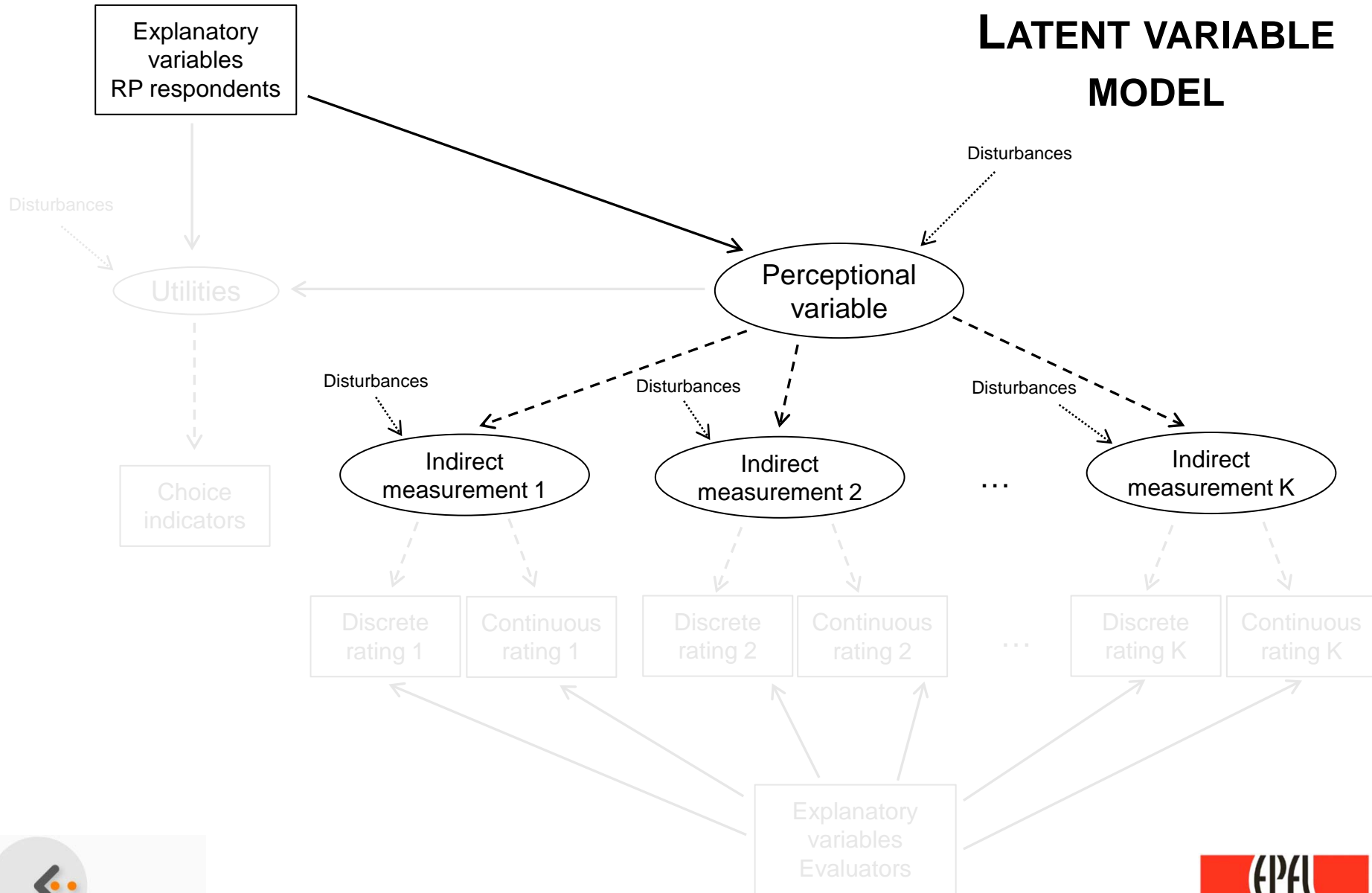






Discrete choice model is standard:

$$U_{in} = V(X_{in}, X_n^*; \beta) + \varepsilon_{in} \quad \text{with } \varepsilon_{in} \sim EV(0, 1)$$



Latent variable model of perception (SEM):

**Structural equation:**

$$X_n^* = h(X_n; \mu) + \omega_n, \quad \text{with } \omega_n \sim \mathcal{N}(0, \sigma_\omega)$$

**Measurement equation:**

$$I_{kn}^* = r_k(X_n^*; \eta_k) + v_{kn}, \quad \text{with } v_{kn} \sim \mathcal{N}(0, \sigma_k)$$

Latent variable model of perception (SEM):

**Structural equation:**

$$X_n^* = h(X_n; \mu) + \omega_n, \quad \text{with } \omega_n \sim \mathcal{N}(0, \sigma_\omega)$$

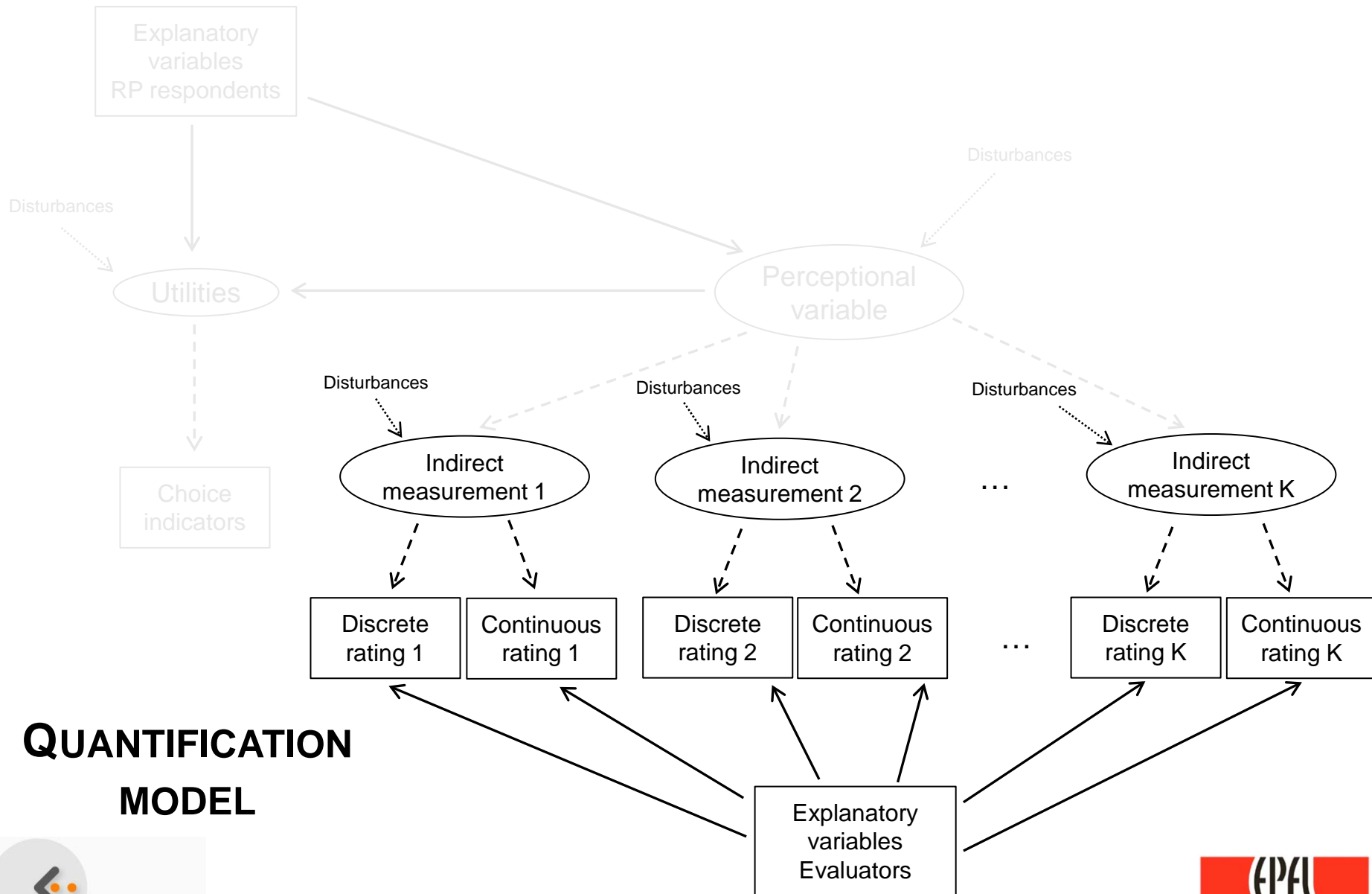
**Measurement equation:**

$$\boxed{I_{kn}^*} = r_k(X_n^*; \eta_k) + v_{kn}, \quad \text{with } v_{kn} \sim \mathcal{N}(0, \sigma_k)$$

Unobservable score of  
indicator  $k$  for individual  $n$



**Indirect measurement** of perception  $X_n^*$ ,  
which is treated as a latent variable



## QUANTIFICATION MODEL

Quantification model (SEM):

**Structural equation:** Score of adjective  $l$  by individual  $m$

$$\boxed{J_{lm}^*} = c_l + \delta_\gamma, \quad \text{with } \delta_\gamma \sim \mathcal{N}(0, \sigma_\gamma)$$

**Measurement equation:**

Discrete:  $\tilde{J}_{lm}^D = \lambda_D \cdot J_{lm}^* + \beta_{Xl}^D \cdot X_m + \delta_D$ , with  $\delta_D \sim \text{Logistic}(0, 1)$

$$J_{lm}^D = \begin{cases} -2 & \text{if } -\infty < \tilde{J}_{lm}^* \leq \tau_{1l} \\ -1 & \text{if } \tau_{1l} < \tilde{J}_{lm}^* \leq \tau_{2l} \\ 0 & \text{if } \tau_{2l} < \tilde{J}_{lm}^* \leq \tau_{3l} \\ 1 & \text{if } \tau_{3l} < \tilde{J}_{lm}^* \leq \tau_{4l} \\ 2 & \text{if } \tau_{4l} < \tilde{J}_{lm}^* \leq +\infty \end{cases}$$

Continuous:  $J_{lm}^C = \alpha_C + \lambda_C \cdot J_{lm}^* + \beta_{Xl}^C \cdot X_m + \delta_C$ , with  $\delta_C \sim \mathcal{N}(0, \sigma_C)$

Quantification model (SEM):

**Structural equation:** Adjective-specific  
constant to be estimated

$$J_{lm}^* = \boxed{c_l} + \delta_\gamma, \quad \text{with } \delta_\gamma \sim \mathcal{N}(0, \sigma_\gamma)$$

**Measurement equation:**

Discrete:  $\tilde{J}_{lm}^D = \lambda_D \cdot J_{lm}^* + \beta_{Xl}^D \cdot X_m + \delta_D$ , with  $\delta_D \sim \text{Logistic}(0, 1)$

$$J_{lm}^D = \begin{cases} -2 & \text{if } -\infty < \tilde{J}_{lm}^* \leq \tau_{1l} \\ -1 & \text{if } \tau_{1l} < \tilde{J}_{lm}^* \leq \tau_{2l} \\ 0 & \text{if } \tau_{2l} < \tilde{J}_{lm}^* \leq \tau_{3l} \\ 1 & \text{if } \tau_{3l} < \tilde{J}_{lm}^* \leq \tau_{4l} \\ 2 & \text{if } \tau_{4l} < \tilde{J}_{lm}^* \leq +\infty \end{cases}$$

Continuous:  $J_{lm}^C = \alpha_C + \lambda_C \cdot J_{lm}^* + \beta_{Xl}^C \cdot X_m + \delta_C$ , with  $\delta_C \sim \mathcal{N}(0, \sigma_C)$



Quantification model (SEM):

**Structural equation:**

$$J_{lm}^* = c_l + \delta_\gamma, \quad \text{with } \delta_\gamma \sim \mathcal{N}(0, \sigma_\gamma)$$

**Measurement equation:**

Discrete:  $\tilde{J}_{lm}^D = \lambda_D \cdot J_{lm}^* + \beta_{Xl}^D \cdot X_m + \delta_D$ , with  $\delta_D \sim \text{Logistic}(0, 1)$

$$J_{lm}^D = \begin{cases} -2 & \text{if } -\infty < \tilde{J}_{lm}^* \leq \tau_{1l} \\ -1 & \text{if } \tau_{1l} < \tilde{J}_{lm}^* \leq \tau_{2l} \\ 0 & \text{if } \tau_{2l} < \tilde{J}_{lm}^* \leq \tau_{3l} \\ 1 & \text{if } \tau_{3l} < \tilde{J}_{lm}^* \leq \tau_{4l} \\ 2 & \text{if } \tau_{4l} < \tilde{J}_{lm}^* \leq +\infty \end{cases}$$

Continuous:  $J_{lm}^C = \alpha_C + \lambda_C \cdot J_{lm}^* + \beta_{Xl}^C \cdot X_m + \delta_C$ , with  $\delta_C \sim \mathcal{N}(0, \sigma_C)$

- Socio-economic information of the evaluator is introduced into measurement equation.
- Heterogeneity in response behavior is handled.

### Estimation of the quantification model alone:

- Likelihood for an adjective  $l$ :

$$\mathcal{L}_l = \prod_{m=1}^M \int_{J_{lm}^*} f(J_{lm}^C | J_{lm}^*, X_m; \alpha_C, \lambda_C, \beta_X^C, \sigma_C) f(J_{lm}^D | J_{lm}^*, X_m; \lambda_D, \beta_X^D, \tau_1, \tau_2, \tau_3, \tau_4) f(J_{lm}^* | c_l, \sigma_\gamma) dJ_{lm}^*$$

- **Score of adjective  $l$  by individual  $m$**  is inferred.

$$\boxed{\hat{J}_{lm}^*} = c_l, \forall m$$

- The obtained scores are then introduced as measurements of the perceptual variable.

### Integration of the 3 model components:

- Simultaneous estimation of the DCM and LVM of perception
- Likelihood

$$\mathcal{L} = \prod_{n=1}^N \int_{X_n^*} \prod_{i=1}^I P(y_{in} | X_{in}, X_n^*; \beta)^{y_{in}} \cdot f(X_n^* | X_n; \mu, \sigma_\omega) \cdot \prod_{k=1}^K f(\hat{I}_{kn}^* | X_n^*; \eta_k; \sigma_k) dX_n^*$$

# APPLICATION EXAMPLE

## QUANTIFICATION MODEL

### Specification

Structural equation:

$$J_{lm}^* = c_l + \delta_\gamma, \quad \text{with } \delta_\gamma \sim \mathcal{N}(0, \sigma_\gamma)$$

Measurement equations:

Discrete  $\tilde{J}_{lm}^D = \lambda_D \cdot J_{lm}^* + \beta_{\text{Educ},l}^D \cdot \text{Educ}_m + \delta_D$ , with  $\delta_D \sim \text{Logistic}(0, 1)$

$$J_{lm}^D = \begin{cases} -2 & \text{if } -\infty < \tilde{J}_{lm}^* \leq \tau_{1l} \\ -1 & \text{if } \tau_{1l} < \tilde{J}_{lm}^* \leq \tau_{2l} \\ 0 & \text{if } \tau_{2l} < \tilde{J}_{lm}^* \leq \tau_{3l} \\ 1 & \text{if } \tau_{3l} < \tilde{J}_{lm}^* \leq \tau_{4l} \\ 2 & \text{if } \tau_{4l} < \tilde{J}_{lm}^* \leq +\infty \end{cases}$$

Observation from exploratory analysis:

Evaluators with higher education level give higher scores.

Continuous  $J_{lm}^C = \alpha_C + \lambda_C \cdot J_{lm}^* + \beta_{\text{Educ},l}^C \cdot \text{Educ}_m + \delta_C$ , with  $\delta_C \sim \mathcal{N}(0, 1)$

# APPLICATION EXAMPLE

Estimation results for 5 adjectives:  
consistent with expectations

## QUANTIFICATION MODEL

Name	Value	t-test	Name	Value	t-test
$C_{\text{bad air}}$	-1.11	-44.56	$C_{\text{comfortable}}$	-2.83	-36.16
$C_{\text{comfortable}}$	1.21	113.35	$\tau_{1, \text{comfortable}}$	-4.2	-4.13
$C_{\text{difficult}}$	-0.563	-31.08	$\delta_{1, \text{comfortable}}$	1.45	1.64
$C_{\text{empty}}$	0.348	29.52	$\delta_{2, \text{comfortable}}$	1.51	3.28
$C_{\text{full}}$	-0.661	-42.91	$\delta_{3, \text{comfortable}}$	2.11	7.29
$\beta^C_{\text{Educ, bad air}}$	-0.07	-3.53	$C_{\text{difficult}}$	-2.66	-26.21
$\beta^D_{\text{Educ, bad air}}$	-0.622	-3.32	$\tau_{1, \text{difficult}}$	-1.88	-7.99
$\beta^C_{\text{Educ, comfortable}}$	0.0151	1.25	$\delta_{1, \text{difficult}}$	1.92	8.63
$\beta^D_{\text{Educ, comfortable}}$	0.371	1.90	$\delta_{2, \text{difficult}}$	2.35	6.88
$\beta^C_{\text{Educ, difficult}}$	0.00496	0.44	$\delta_{3, \text{difficult}}$	1.02	2.45
$\beta^D_{\text{Educ, difficult}}$	-0.292	-1.67	$C_{\text{empty}}$	-2.74	-29.32
$\beta^C_{\text{Educ, empty}}$	0.245	24.29	$\tau_{1, \text{empty}}$	-2.72	-7.30
$\beta^D_{\text{Educ, empty}}$	0.372	2.08	$\delta_{1, \text{empty}}$	1.23	3.99
$\beta^C_{\text{Educ, full}}$	-0.0709	-6.03	$\delta_{2, \text{empty}}$	1.16	5.49
$\beta^D_{\text{Educ, full}}$	-0.555	-3.09	$\delta_{3, \text{empty}}$	2.85	10.21
$C_{\text{bad air}}$	-2.67	-25.92	$C_{\text{full}}$	-2.7	-26.82
$\tau_{1, \text{bad air}}$	-1.28	-5.90	$\tau_{1, \text{full}}$	-1.82	-7.83
$\delta_{1, \text{bad air}}$	1.79	7.49	$\delta_{1, \text{full}}$	2.15	8.94
$\delta_{2, \text{bad air}}$	0.888	3.76	$\delta_{2, \text{full}}$	0.504	3.45
$\delta_{3, \text{bad air}}$	1.22	3.04	$\delta_{3, \text{full}}$	1.7	4.83

Loglikelihood: - 1731

# APPLICATION EXAMPLE

Estimation results for 5 adjectives:  
consistent with expectations

Constants have expected signs:

- Adjectives with positive meaning have + sign
- Adjectives with negative meaning have - sign

## QUANTIFICATION MODEL

Name	Value	t-test	Name	Value	t-test
$C_{\text{bad air}}$	-1.11	-44.56	$C_{\text{comfortable}}$	-2.83	-36.16
$C_{\text{comfortable}}$	1.21	113.35	$\tau_{1, \text{comfortable}}$	-4.2	-4.13
$C_{\text{difficult}}$	-0.563	-31.08	$\delta_{1, \text{comfortable}}$	1.45	1.64
$C_{\text{empty}}$	0.348	29.52	$\delta_{2, \text{comfortable}}$	1.51	3.28
$C_{\text{full}}$	-0.661	-42.91	$\delta_{3, \text{comfortable}}$	2.11	7.29
$\beta^C_{\text{Educ, bad air}}$	-0.07	-3.53	$C_{\text{difficult}}$	-2.66	-26.21
$\beta^D_{\text{Educ, bad air}}$	-0.622	-3.32	$\tau_{1, \text{difficult}}$	-1.88	-7.99
$\beta^C_{\text{Educ, comfortable}}$	0.0151	1.25	$\delta_{1, \text{difficult}}$	1.92	8.63
$\beta^D_{\text{Educ, comfortable}}$	0.371	1.90	$\delta_{2, \text{difficult}}$	2.35	6.88
$\beta^C_{\text{Educ, difficult}}$	0.00496	0.44	$\delta_{3, \text{difficult}}$	1.02	2.45
$\beta^D_{\text{Educ, difficult}}$	-0.292	-1.67	$C_{\text{empty}}$	-2.74	-29.32
$\beta^C_{\text{Educ, empty}}$	0.245	24.29	$\tau_{1, \text{empty}}$	-2.72	-7.30
$\beta^D_{\text{Educ, empty}}$	0.372	2.08	$\delta_{1, \text{empty}}$	1.23	3.99
$\beta^C_{\text{Educ, full}}$	-0.0709	-6.03	$\delta_{2, \text{empty}}$	1.16	5.49
$\beta^D_{\text{Educ, full}}$	-0.555	-3.09	$\delta_{3, \text{empty}}$	2.85	10.21
$C_{\text{bad air}}$	-2.67	-25.92	$C_{\text{full}}$	-2.7	-26.82
$\tau_{1, \text{bad air}}$	-1.28	-5.90	$\tau_{1, \text{full}}$	-1.82	-7.83
$\delta_{1, \text{bad air}}$	1.79	7.49	$\delta_{1, \text{full}}$	2.15	8.94
$\delta_{2, \text{bad air}}$	0.888	3.76	$\delta_{2, \text{full}}$	0.504	3.45
$\delta_{3, \text{bad air}}$	1.22	3.04	$\delta_{3, \text{full}}$	1.7	4.83

Loglikelihood: - 1731

# APPLICATION EXAMPLE

Estimation results for 5 adjectives:  
consistent with expectations

Results from exploratory analysis confirmed:

The higher the level of education, the higher the scores in absolute value.

## QUANTIFICATION MODEL

Name	Value	t-test	Name	Value	t-test
$C_{\text{bad air}}$	-1.11	-44.56	$C_{\text{comfortable}}$	-2.83	-36.16
$C_{\text{comfortable}}$	1.21	113.35	$\tau_{1, \text{comfortable}}$	-4.2	-4.13
$C_{\text{difficult}}$	-0.563	-31.08	$\delta_{1, \text{comfortable}}$	1.45	1.64
$C_{\text{empty}}$	0.348	29.52	$\delta_{2, \text{comfortable}}$	1.51	3.28
$C_{\text{full}}$	-0.661	-42.91	$\delta_{3, \text{comfortable}}$	2.11	7.29
$\beta^C_{\text{Educ, bad air}}$	-0.07	-3.53	$C_{\text{difficult}}$	-2.66	-26.21
$\beta^D_{\text{Educ, bad air}}$	-0.622	-3.32	$\tau_{1, \text{difficult}}$	-1.88	-7.99
$\beta^C_{\text{Educ, comfortable}}$	0.0151	1.25	$\delta_{1, \text{difficult}}$	1.92	8.63
$\beta^D_{\text{Educ, comfortable}}$	0.371	1.90	$\delta_{2, \text{difficult}}$	2.35	6.88
$\beta^C_{\text{Educ, difficult}}$	0.00496	0.44	$\delta_{3, \text{difficult}}$	1.02	2.45
$\beta^D_{\text{Educ, difficult}}$	-0.292	-1.67	$C_{\text{empty}}$	-2.74	-29.32
$\beta^C_{\text{Educ, empty}}$	0.245	24.29	$\tau_{1, \text{empty}}$	-2.72	-7.30
$\beta^D_{\text{Educ, empty}}$	0.372	2.08	$\delta_{1, \text{empty}}$	1.23	3.99
$\beta^C_{\text{Educ, full}}$	-0.0709	-6.03	$\delta_{2, \text{empty}}$	1.16	5.49
$\beta^D_{\text{Educ, full}}$	-0.555	-3.09	$\delta_{3, \text{empty}}$	2.85	10.21
$C_{\text{bad air}}$	-2.67	-25.92	$C_{\text{full}}$	-2.7	-26.82
$\tau_{1, \text{bad air}}$	-1.28	-5.90	$\tau_{1, \text{full}}$	-1.82	-7.83
$\delta_{1, \text{bad air}}$	1.79	7.49	$\delta_{1, \text{full}}$	2.15	8.94
$\delta_{2, \text{bad air}}$	0.888	3.76	$\delta_{2, \text{full}}$	0.504	3.45
$\delta_{3, \text{bad air}}$	1.22	3.04	$\delta_{3, \text{full}}$	1.7	4.83

Loglikelihood: - 1731

## Main findings:

- **Alternative approach** to measure perceptions
- Main **advantage** over classical opinion statements: **spontaneity** of respondents captured.
- Difficulty: code and integrate these measurements in choice model.  
The proposed **model**:
  - **Quantifies adjectives**
  - **Accounts for subjectivity** inherent to quantification method:
    - Use a fairly large sample of evaluators
    - Account for bias linked to different education levels

## Further research:

- Estimate the three parts of the integrated model together.



# Thanks!