Discrete choice analysis & taboos

Caspar Chorus

Professor of choice behavior modeling

Head: Engineering Systems department

TU Delft
Delft University of Technology

Challenge the future

European Research Council

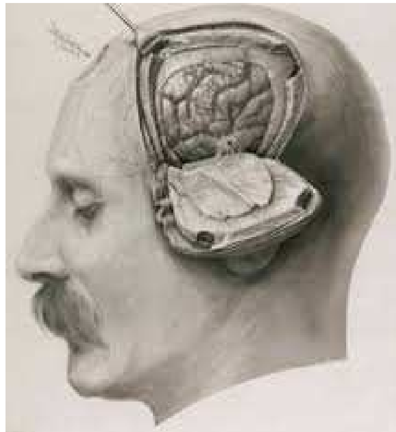
Discrete choice analysis in one slide

If you observe my choices, you may learn my preferences, desires, goals, motivations…

And once you know those, you may predict my future choices
→ **market demand, policy effects**

Discrete choice analysis in one slide

If you observe my choices, you may learn my preferences, desires, goals, motivations…



And once you know those, you may predict my future choices
→ **market demand, policy effects**

Originates from **empirical social sciences:**
Econometrics, Mathematical Psychology;
Transport, Environment, Health, Marketing…

Commonly applied in practical situations:
e.g. transport-infrastructure planning,
consumer product pricing, …

Discrete choice analysis & (computational) social choice

DCA has three aims:

1. Behavioral inference (trade-offs, weights, decision rules)
   - *e.g. travel time and cost*
2. Prediction of market shares, policy response
   - *e.g. use of new train service, highway*
3. Economic appraisal: monetary welfare effects of policies
   - *e.g. based on 1., 2.: monetary benefits of new infra.*

Clearly, there is a connection with CSC, but note:

- DCA not much concerned with individual preferences
  - *More focus on model parsimony + noise term*
- Nor with preference aggregation / ranking, ordering
  - *(cardinal) Welfare economics – no voting theory*

Discrete choice analysis & (computational) social choice

DCA has three aims:

1. Behavioral inference (trade-offs, weights, decision rules)
   - *e.g. travel time and cost*
2. Prediction of market shares, policy response
   - *e.g. use of new train service, highway*
3. Economic appraisal: monetary welfare effects of policies
   - *e.g. based on 1., 2.: monetary benefits of new infra.*

Clearly, there is a connection with CSC, but note:

- DCA not much concerned with individual preferences
  - *More focus on model parsimony + noise term*
- Nor with preference aggregation / ranking, ordering
  - *(cardinal) Welfare economics – no voting theory*

My (team's) research aims

Improving behavioral realism of choice models

Decade since PhD 2007 (NWO Veni, Vidi):

- Capturing bounded **rationality**
- E.g. random regret minimization model

Since 2017 (ERC Consolidator):

Capturing the **morality** of human choice behavior:
- Representation of heuristics, norms, obfuscation
- Use moral choice models for simulating artificial societies (study emergence of moral norms), and developing 'human-inspired moral compass for AI'.

My (team's) research aims

Improving behavioral realism of choice models

Decade since PhD (NWO Veni, Vidi):

- Capturing bounded **rationality**
- E.g. random regret minimization model

Since 2017 (ERC Consolidator):

Capturing the **morality** of human choice behavior:
- Representation of heuristics, norms, obfuscation
- Use moral choice models for simulating artificial societies (study emergence of moral norms), and developing 'human-inspired moral compass for AI'.

http://behave.tbm.tudelft.nl/

The BEHAVE-Team (a selection)

Philosophy, econometrics, symbolic AI, Socio-Physics & more

Taboo trade-off aversion: A discrete choice model and empirical analysis

Caspar Chorus, N. Mouter, B. Pudane, D. Campbell

Chorus, C. G., Pudāne, B., Mouter, N., & Campbell, D. (2018). Taboo trade-off aversion: A discrete choice model and empirical analysis. *Journal of Choice Modelling* , 27, 37-49

What is a taboo trade-off?

Willing to sacrifice an hour of travel time to meet a friend, inform how he is doing.

My Value of Time = €20 / hour

NOT willing to pay him €20 to come over to me instead...

'paying' in terms of time, attention: OK. In terms of money: taboo.

Why?

- Time, friendship belong to the same sphere (social relations)
- Money belongs to a different sphere (economic transactions)

What is a taboo trade-off? (II)

(Harel and Porat, *Cornell Law Review* 2011)

Government WtP in terms of health care or investment in dikes, equals 2M€ per human life saved;

Government allows torture to save a human life;

Conclusion? Government should allow torture to save 2M€

NO: “Human rights should never be violated for monetary gains”

Taboo trade-offs challenge transitivity axiom underlying Economics.

What is a taboo trade-off? (III)

(The Economist, 17 March 2017)

± 700,000 USD per identified and repatriated remains of a single US soldier (MIA).

"You cannot associate a dollar value with this national imperative," says General Spindler.

The mere idea of trading off the anguish of left-behind families against budget constraints, is awkward and politically dangerous.

What is a taboo trade-off? (IV)

Key concept in Moral Psychology (**Tetlock**), Economic Law (Radin)

People express 'moral outrage' when asked to trade off 'sacred' values with non-sacred ones (usually money):

- Love *versus money*
- Health of one's child *versus money*
- Loyalty to one's country *versus money*
- Wellbeing of others *versus money*

Challenge for Discrete Choice Analysis

Since Lancaster (1966), Keeney & Raiffa (1976), trade-offs at the core of decision theory, microeconomic consumer theory.

Discrete Choice Theory pendants:

- Compensatory models (linear-additive utility max.)
- Semi-compensatory models (e.g. regret, loss aversion)

This study:

1. Discrete choice model that captures taboo trade-off aversion.
2. Empirical analyses based on dataset collected for this purpose.

Empirical context

Support or oppose comprehensive national infrastructure plan.

Effects in terms of increase or decrease in:

• Vehicle ownership tax (€)	300	p. year	TAX
• Travel time (min.)	20	p. working day	TIME
• Non-fatal traffic injuries	100	p. year	INJ
• Traffic fatalities	5	p. year	FAT

Some examples of trade-offs

TAX ↓ & TIME ↑ : Secular trade-off

TAX ↓ & FAT ↑ : **Taboo** trade-off

INJ ↓ & FAT ↑ : Tragic trade-off

Data

Specifically designed Stated Choice survey (see earlier slide)

Experimental design: full factorial

- Ensures (theoretical) identification of taboo-penalties and tastes
- Drawback: seemingly illogical combinations (e.g. INJ ↓ & FAT ↑)

9 out of 16 tasks contained (1, 2, 3 or 4) taboo trade-offs

Sample of 99 representative regular car commuters, 16 choice tasks

First: pilot study (20 people), interviews with respondents.

Final data collected February 2017, random sample Dutch >18.

Example choice task

	Proposed Transport Policy
Vehicle ownership tax (per year, for each car owner including yourself)	300 euro **<u>less</u>** tax
Travel time (per working day, for each car commuter including yourself)	20 minutes **<u>less</u>** travel time
Number of seriously injured in traffic (per year)	100 seriously injured **<u>more</u>**
Number of traffic fatalities (per year):	5 traffic fatalities **<u>more</u>**
YOUR CHOICE	□ I support the proposed policy □ I oppose the proposed policy

A conventional linear RUM model

- Policy variant j constitutes change w.r.t. Status Quo (V_{SQ} = utility of doing nothing, i.e. of opposing the policy)

- Binary choice, 'referendum format'

$$V_j = \sum_m \beta_m \cdot x_{jm} = \beta_{tax} \cdot tax_j + \beta_{time} \cdot time_j + \beta_{fat} \cdot fat_j + \beta_{inj} \cdot inj_j$$

$$P(j) = \frac{\exp(V_j)}{\exp(V_j) + \exp(V_{SQ})} = \frac{\exp(\sum_m \beta_m x_{jm})}{\exp(\sum_m \beta_m x_{jm}) + \exp(V_{SQ})}$$

- m and n denote attributes, x attribute-values, β attribute weights

- Linear utility function, implies fully compensatory decision making.

Modeling taboo trade-off aversion (III)

- The following, generic specification is adopted:

$$V_j^{TTOA} = \sum_m \beta_m \cdot x_{jm} + \tau_G \cdot \max_{(m,n) \in T} I_{m \to n}$$

- T represents the set of ordered pairs (m, n) where m is a 'sacred' attribute and n is a 'secular' attribute
- I indicates taboo trade-off: a worse value is accepted for m to obtain a better value for n
- τ_G is generic taboo-penalty associated with having **one or more** taboo-trade offs embedded in the policy alternative

Results – benchmark linear RUM

Mixed Logit (Panel), 4000 draws
(all parameters $\sim N$.)

Null log-likelihood: -1098
Final log-likelihood: -598

Name	Value	Rob.SE	Rob.t	Rob. p
V_SQ	2.24	0.328	6.83	0.00
BETA_Fat	-1.81	0.246	-7.38	0.00
BETA_Inj	-2.60	0.361	-7.19	0.00
BETA_Tax	-2.09	0.301	-6.94	0.00
BETA_Time	-1.09	0.195	-5.58	0.00
SIGMA_OPPOSE	1.79	0.276	6.49	0.00
SIGMA_Fat	1.17	0.226	5.16	0.00
SIGMA_NonFat	1.68	0.281	5.98	0.00
SIGMA_Tax	1.69	0.288	5.88	0.00
SIGMA_Time	1.28	0.254	5.06	0.00

Results – Taboo trade-off aversion

Mixed Logit (Panel), 4000 draws
(all parameters $\sim N$.)

Null log-likelihood: -1098
Final log-likelihood: -589

Name	Value	Rob.SE	Rob.t	Rob. p
V_SQ	1.48	0.354	4.19	0.00
BETA_Fat	-1.52	0.234	-6.50	0.00
BETA_Inj	-2.19	0.310	-7.07	0.00
BETA_Tax	-2.27	0.330	-6.87	0.00
BETA_Time	-1.25	0.227	-5.50	0.00
SIGMA_OPPOSE	1.36	0.336	3.70	0.00
SIGMA_Fat	1.03	0.249	4.12	0.00
SIGMA_NonFat	1.75	0.384	4.57	0.00
SIGMA_Tax	1.58	0.253	6.23	0.00
SIGMA_Time	1.31	0.272	4.82	0.00
BETA_Taboo	**-1.02**	**0.473**	**-2.16**	**0.03**
SIGMA_Taboo	**2.14**	**0.499**	**4.29**	**0.00**

Effects on parameters, choice probs.

Parameters

- Relative to Taboo-model, linear RUM **overestimates** importance of traffic fatality, injury parameters (both 19% inflated)
- Correlation found between weights of injuries and fatalities, but **not** between these weights and taboo penalty!

Choice probabilities

- Relative to linear RUM, Taboo model assigns lower support for policies which contains taboo trade-off(s)
- On our data, Taboo model predictions much closer to observed support-levels

Take-away, work to be done

Take-away

- Taboo trade-offs can be relatively easily modeled in DCA-context
- Could play a non-trivial role in choice experiments, situations
- Much heterogeneity: deontologists, utilitarians, ‘I don’t-careans’

Work to be done

- Replication (or not) on other data sets
- Derive welfare-implications. Tricky:
  - Indifference curve does not apply
  - Role of agency: is a trade-off taboo when forced upon you?
- Insights from (computational) social choice?

Thank you!

Back-up slides

Latent Class choice models

To conceptualize moral uncertainty of an AI

Caspar Chorus

Delft University of Technology

MacAskill (2014) and Bogosian (2017): Moral uncertainty for AI

AI should be morally uncertain, base its decisions on expected maximum utility calculus:

Choiceworthiness of an action =

Choiceworthiness of the action under a given moral theory

Multiplied by

Credence of the theory

Summed over all theories

→

To a choice modeller, their math resembled a **Latent Class model**

E(Utility) of an action =

Utility of the action for a given class of people

Multiplied by

The size of the class

Summed over all classes

Example:

An AI that needs to decide on a transport policy, e.g. rank all $2^4=16$ policies implied by the previously discussed choice experiment

	Proposed Transport Policy
Vehicle ownership tax (per year, for each car owner including yourself)	300 euro **less** tax
Travel time (per working day, for each car commuter including yourself)	20 minutes **less** travel time
Number of seriously injured in traffic (per year)	100 seriously injured **more**
Number of traffic fatalities (per year):	5 traffic fatalities **more**
YOUR CHOICE	□ I support the proposed policy □ I oppose the proposed policy

In a taboo-trade-off-aversion context, this could include classes of utilitarians, deontologists, I-don't-careans, ...

Results in utility of each policy as perceived by an agent that is representative of a society with heterogeneous morality.

Reframes moral *uncertainty* as moral *heterogeneity*…

How does it work? Store choice

Take a theory of behavior, e.g. compensatory utility max.

Develop a choice model based on that theory:

Probability that particular store is chosen →

$$P(i) = P\left(V_i + \varepsilon_i > V_j + \varepsilon_j, \forall j \neq i\right) = \frac{\exp(\sum_m \beta_m x_{im})}{\sum_{j=1..J} \exp(\sum_m \beta_m x_{jm})}$$

- People choose the store which brings them highest utility
- Utility depends on / is a linear combination of:
  - Assortment (proxy: m^2 floorspace), travel time – KNOWN
  - **How each of these aspects is weighted – UNKNOWN**
- And don't forget the randomness

How does it work? (II)

Collect data:

Real choices
(Counts at different locations)

Or **Hypothetical choices**
(Choice experiments – see further)

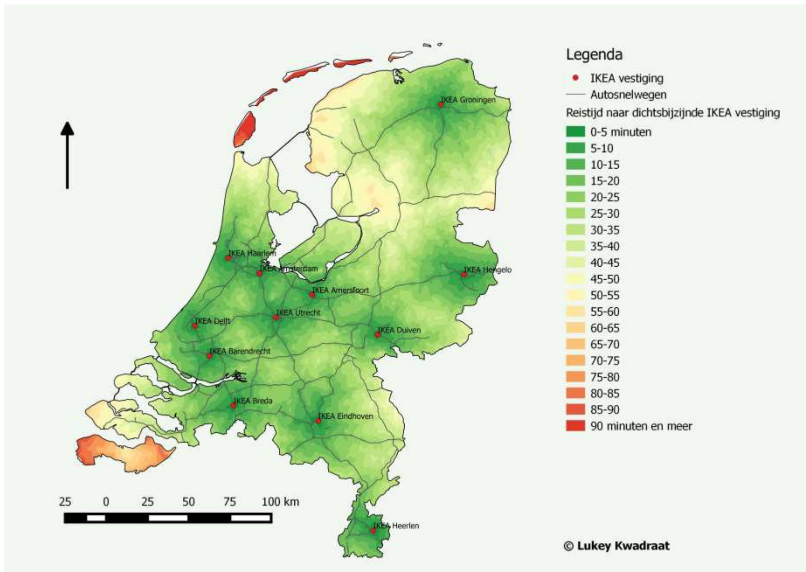
How does it work? (III)

Estimate model, interpret results

- Obtain weights for all factors (travel time, floor-space)
- Ratio gives willingness to travel <?> seconds to get 1 extra m^2

Apply them for forecasting, decision support

- Where to open/close a store?
- Of which size?



Results – other models

Hosts of other (Mixed) Logit specifications tested

1. LogNormal distribution of taste-, taboo aversion
2. Full covariance matrix for random parameters
3. Sociodemographic interactions with taboo-penalty
4. Scale effects (of taboo trade offs)

Results:

1. Taboo-aversion best modeled using $\sim N$, tastes $\sim LN$
2. No covariance between taboo aversion and tastes
3. Older, female: larger penalty for taboos (p=0.10, 0.06)
4. No effects of taboo trade off on scale

Just Counting approach (summary)

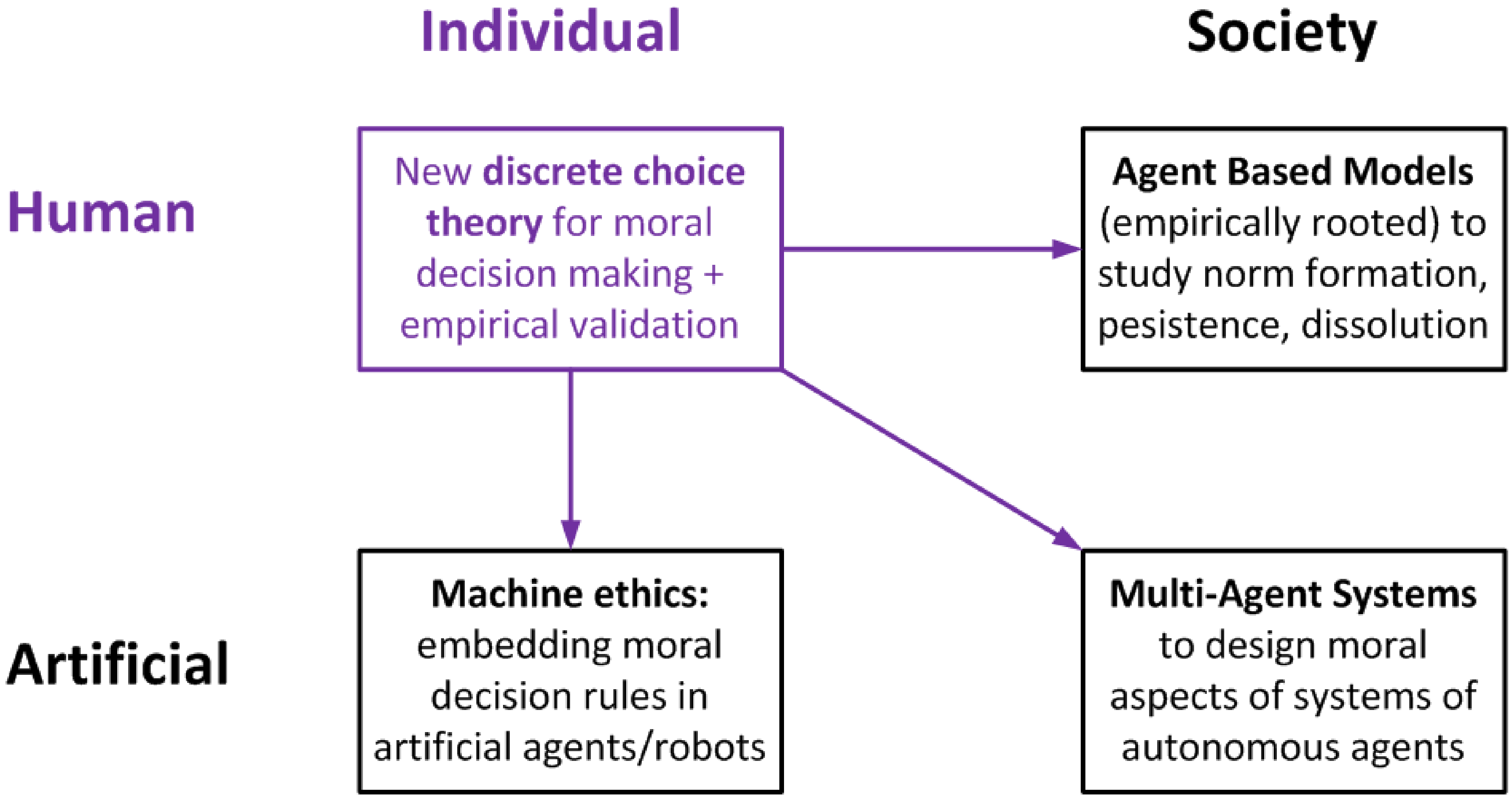
How many individuals always opposed a policy that contained TTO?

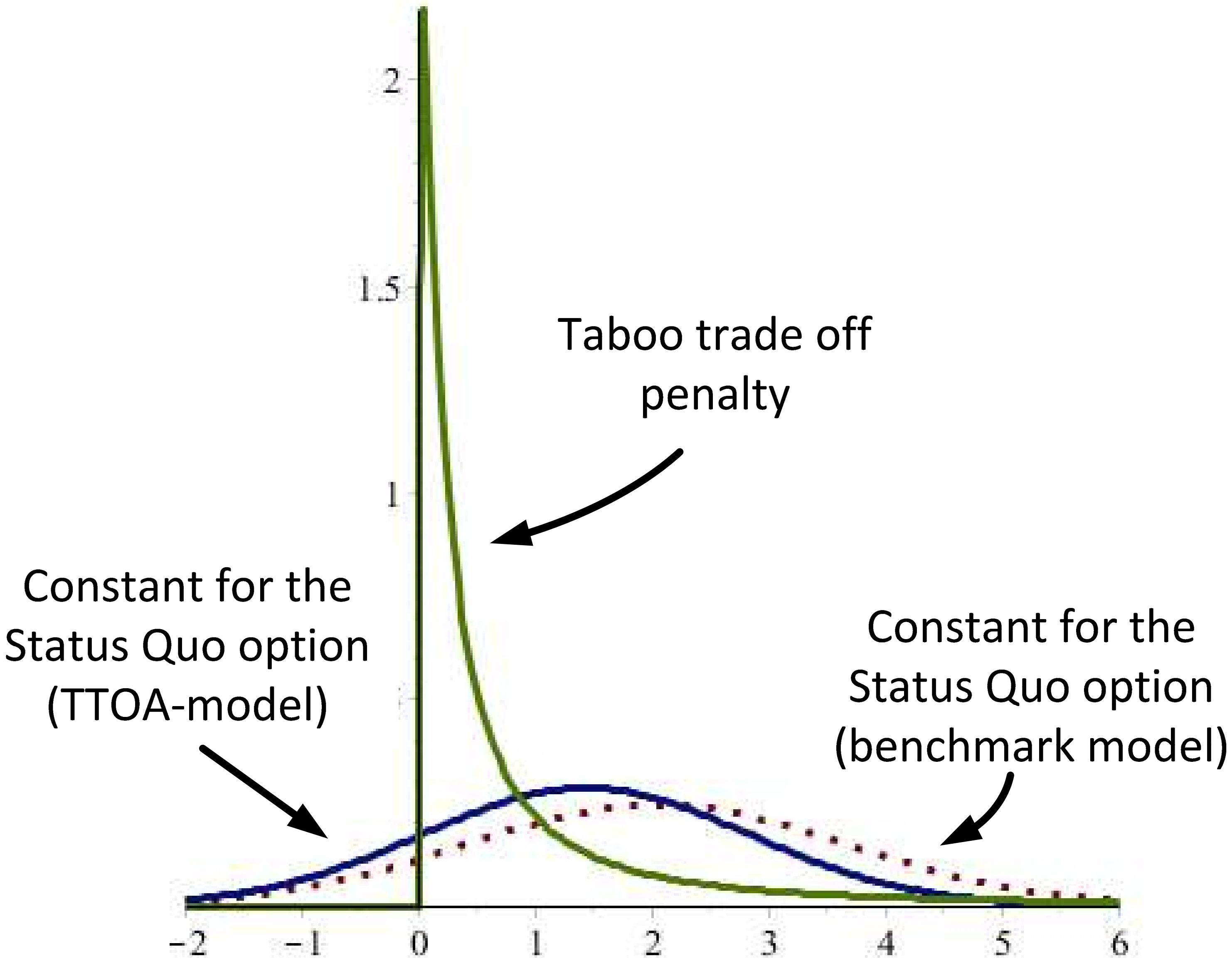
Note: includes people who opposed those policies for other reasons!

Such as dislike of policies per se, or extreme focus on safety

Results: 16 out of 99 (so, 17% upper bound of deontologists…)

Not just *human* agents, not just *individual* agents...





Taboo trade off penalty
Constant for the Status Quo option (TTOA-model)
Constant for the Status Quo option (benchmark model)
2
1.5
1
−2
−1
0
1
2
3
4
5
6

Tax	Time	Injury	Fatality	Taboo	True share of support for infrastructure investment schemes	Predicted share of support benchmark model	Predicted share of support TTOA model
-1	-1	-1	-1	**0**	98%	92%	94%
-1	1	-1	-1	**0**	81%	81%	83%
-1	-1	-1	1	**1**	69%	71%	70%
1	-1	-1	-1	**0**	63%	63%	64%
-1	-1	1	-1	**1**	46%	56%	56%
-1	1	-1	1	**1**	42%	46%	43%
1	1	-1	-1	**0**	44%	37%	37%
-1	1	1	-1	**1**	30%	31%	29%
1	-1	-1	1	**1**	22%	26%	23%
-1	-1	1	1	**1**	29%	21%	24%
1	-1	1	-1	**1**	15%	15%	13%
1	1	-1	1	**0**	7%	11%	13%
-1	1	1	1	**1**	11%	9%	9%
1	1	1	-1	**0**	4%	6%	7%
1	-1	1	1	**1**	5%	4%	4%
1	1	1	1	**0**	2%	1%	2%
Mean absolute deviation from true share of support (percentage points; all choice tasks)						**3.28**	**3.03**
Mean absolute deviation from true share of support (percentage points; choice tasks involving taboo trade-offs)						**3.65**	**2.68**

Modeling taboo trade-off aversion (IV)

If signs of β are known, $I_{m \to n}$ may be specified as a dummy-var.

Alternatively, a smooth step function may be adopted ($\mu \gg 0$):

$$I_{m \to n} = \frac{\exp(\mu \cdot -\Delta m)}{1 + \exp(\mu \cdot -\Delta m)} \cdot \frac{\exp(\mu \cdot \Delta n)}{1 + \exp(\mu \cdot \Delta n)} = \frac{\exp(\mu \cdot (\Delta n - \Delta m))}{(1 + \exp(\mu \cdot -\Delta m)) \cdot (1 + \exp(\mu \cdot \Delta n))}$$

where Δm denotes $\beta_m \cdot \left(x_{jm} - x_{im}\right)$ and Δn denotes $\beta_n \cdot \left(x_{jn} - x_{in}\right)$

(this avoids having to pre-process the data)

Synthetic data: identifiable model, equivalence dummy, smoothstep

Results – Disclaimer

This is a **confirmatory** research effort:

We try to see if there is Taboo trade-off aversion in our data.

Although the effort is explorative in covering several ways to model the aversion…

We do not claim that there no other ways to improve model fit / predictive performance

- such as interactions between attributes

However, our set-up designed to rule out some related hypotheses

- loss aversion (two-level attributes)
- regret aversion (binary choice set)

Latent Class approach (summary)

Assumes:

homogeneity within classes in terms of parameters
heterogeneity between classes in terms of parameters
Parameters & class sizes estimated – emerge from data

Depending on specification:

- Sizeable portion of population rejects TTOs ("deontologists")
- Sizeable portion of population penalizes TTOs ("utilitarians")
- Sizeable portion of population does not care about TTOs.

Latent Class approach (summary II)

Formulation of two classes, both with linear utility function.

Specification 2:

All parameters equal (between classes) except for TTO-Penalty:

- Class 1 (**77%**): TTO-Penalty fixed to zero
- Class 2 (**23%**): TTP-Penalty fixed to very large neg. (→ deontol.)

Suggests:

- Three quarters of population don't care about TTOs.
- Other (about) quarter does care a lot, behaves deontological