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

Chorus, Caspar G.; Pudane, Baiba; Mouter, Niek; Campbell, Danny

DOI
10.1016/j.jocm.2017.09.002

Publication date
2017

Document Version
Final published version

Published in
Journal of Choice Modelling

Journal of Choice Modelling xxx (2017) 1–13

Contents lists available at ScienceDirect

Journal of Choice Modelling

journal homepage: www.elsevier.com/locate/jocm

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

Caspar G. Chorus [a,*], Baiba Pudāne [a], Niek Mouter [a], Danny Campbell [b]

[a] Transport and Logistics Group, Department of Engineering Systems and Services, Delft University of Technology, The Netherlands
[b] Department of Economics, Stirling Management School, University of Stirling, UK

ABSTRACT

An influential body of literature in moral psychology suggests that decision makers consider trade-offs morally problematic, or taboo, when the attributes traded off against each other belong to different 'spheres', such as friendship versus market transactions. This study is the first to model and empirically explore taboo trade-off aversion in a discrete choice context. To capture possible taboo trade-off aversion, we propose to extend the conventional linear in parameters logit model by including penalties for taboo trade-offs. Using this model, we then explore the presence (and size) of taboo trade-off aversion in a data set specifically collected for this purpose. Results, based on estimation of a variety of (Mixed) Logit models with and without taboo trade-off penalties, suggest that there is indeed a significant and sizeable taboo trade-off aversion underlying choice behaviour of respondents.

1. Introduction

Ever since the notion of taboo trade-off aversion was systematically explored by moral psychologists, most notably Philip Tetlock and co-workers (e.g., Fiske and Tetlock, 1997), it has received ample attention throughout the behavioural sciences. The idea of taboo trade-off aversion (from here on TTOA) suggests that decision makers consider some types of trade-offs morally problematic, or taboo. Indeed, much empirical evidence (see below for a brief review of key contributions to this literature) has surfaced to suggest that there are many situations where decision makers dislike making a trade-off between different attributes of choice alternatives, and may even become distressed (express moral outrage) when asked to consider such trade-offs in the first place. Generally, trade-offs are found to be considered taboo by decision makers when the two attributes being traded off against one another belong to different 'spheres'; usually one attribute belongs to the sphere of market transactions (e.g. a price attribute), while another attribute belongs to, for example, the sphere of social relations (e.g. friendship) or another sphere in which market transactions are, by many, frowned upon (e.g. healthcare, or matters of war and peace). To illustrate the concept of taboo trade-off aversion, and before we discuss notable contributions to the taboo trade off literature, we here present two brief examples.

Imagine that you plan to visit your friend, who lives 30 min away and that you have a value of time of 20 euro per hour. This implies a) that you are willing to sacrifice an hour of traveling (to your friend's home and back) to visit your friend, and b) that you would be willing to pay 20 euro to eliminate the hour of traveling to your friend's home and back. Now, consider how would you react if your friend offered you the following proposition: "don't come, but pay me 20 euro instead and I will come to your place"? It is safe to expect that most people will not like this proposition; many would even feel offended. However, transitivity laws would suggest that a rational individual would at least consider this trade-off between money and travel time. So why might this proposition feel so awkward to many of us? Taboo trade-off theory (Fiske and Tetlock, 1997) suggests an answer: the awkwardness stems from the interference of two spheres

* Corresponding author.
E-mail address: c.g.chorus@tudelft.nl (C.G. Chorus).

http://dx.doi.org/10.1016/j.jocm.2017.09.002
Received 10 May 2017; Received in revised form 13 September 2017; Accepted 13 September 2017
Available online xxxx

Please cite this article in press as: Chorus, C.G., et al., Taboo trade-off aversion: A discrete choice model and empirical analysis, Journal of Choice Modelling (2017), http://dx.doi.org/10.1016/j.jocm.2017.09.002

– monetary transactions and friendship – which most people believe should be firmly separated. We are fine with "paying attention" to a friend, but not with paying her money in exchange for an update on how she is doing.

The second example is inspired by a thought experiment of Harel and Porat (2011). Consider a government that is willing to pay two million euro per human live saved, for example, when considering how much money to spend on dike improvements which aim to protect people from floods. Now assume that the same government agrees that it is acceptable to torture a suspected terrorist if this will save a human life. Does this imply that the same government should be willing to torture a suspected terrorist in order to save two million euro worth of property damage? Again, although laws of transitivity suggest "yes", most people, and governments, would agree that this would be morally highly problematic. The reason behind this seeming inconsistency again is that elements from different spheres are compared (dignity of a human life versus money), which many people believe to be 'incomparable' or incommensurable. In contrast, the government's acceptance of torturing a terrorist to save life involves a trade-off between two elements within the same sphere (dignity of a human life versus saving a human life), which leads to what psychologists have termed a 'tragic' – but not a 'taboo' – trade-off. If both attributes belong to the secular sphere, then the trade-off is called a 'secular' trade-off or a 'routine' trade-off.

The literature on taboo trade-offs is sprawling; given word constraints and in light of the fact that there are recent review papers available on moral decision rules, including taboo trade-offs (Chorus, 2015), we here limit ourselves to a brief discussion of some seminal and some very recent contributions. Tetlock et al. (2000) presented participants with a variety of secular, tragic, and taboo trade-offs, and found that even the mere act of reading about taboo trade-offs (e.g., the buying of U.S. citizenship) led participants to respond with *moral outrage* (as measured by means of a series of answers to propositions) and with so-called *moral cleansing* (as measured by a participant's stated inclination to volunteer for a good cause). Despite this seemingly strong aversion to taboo trade-offs, Tetlock (2003) found that a simple rhetorical reframing of the trade-off could strongly reduce levels of moral outrage and moral cleansing. To cite one example from that paper, the author reports a survey held in the fictionalized context of the government deciding to allocate less tax money than originally planned to the cleaning up of a hazardous waste site. This proposition was estimated to lead to a reduction in 'saved lives' from 400 to 200, and the freed up money (100 million USD) would be used to reduce the deficit and to decrease taxes. Depending on how the question was framed, between 35% and 72% of respondents agreed with the policy. Cheap talk statements such as that the government "concludes that morally this is the right thing to do" had great positive effect on support levels. These insights are confirmed in a consumer choice context by McGraw and Tetlock (2005), who confronted participants to a fictional situation where someone ('John') was selling a ballpoint pen which he had received in the context of a particular relationship. Depending on how this relationship was framed, participants reported lower or higher levels of 'distress' when confronted with John's intention of selling of the pen; highest distress levels were reported for the frame that emphasized that John had received the pen as a token of gratitude for having helped a fellow graduate student. This frame also led to a relatively high percentage of respondents who refused to state what would be an appropriate buying price for the pen. This experiment clearly indicates that market transactions involving relational aspects are considered taboo by many. In the rather different context of ecosystem protection, Daw et al. (2015) find that taboo trade-offs – in this case, between the well-being of marginalized women in a remote Kenyan village and profits from fishery and trade – can be resolved by either making the taboo-aspect explicit, or by reframing them into routine or tragic trade-offs. The importance of rhetorical re-framing of taboo trade-offs is also confirmed by Zaal et al. (2014), in the context of public opposition against the siting of hazardous facilities. The authors find that participants react very negatively to an offered monetary compensation for accepting the safety risks associated with a hazardous facility being sited in their neighbourhood (implying a taboo trade-off between money and safety). However, when the offer is rhetorically redefined into a tragic trade-off between two forms of safety (in this case, by highlighting that the compensation could be used by the community to improve local traffic safety), this led to much higher levels of acceptance. Stikvoort et al. (2016) report that participants to an experiment which involves making taboo trade-offs regarding eco-system preservation (i.e., making a donation to help preserve a forest, or spending the money on designer furniture) made higher real donations to an environmental cause *ex post*, than respondents to the control condition which did not involve taboo trade-offs. The authors interpret this as a confirmation of the 'moral cleansing' effect predicted by earlier studies (see above).

This brief review of the empirical literature on taboo trade-offs suggests a consistent picture: trade-offs between 'sacred' and 'secular' values are considered problematic by many. When confronted with such trade-offs, people express moral outrage or distress and they react by means of acts of moral cleansing. However, simple rhetorical reframing can substantially reduce or even eliminate these negative reactions.

Given the substantial interest in the concept of taboo trade-offs in moral psychology and in different fields of application, it is somewhat surprising to see that – to the best of our knowledge – no attempt has been made to formalize the notion of taboo trade-off aversion (TTOA) in a discrete choice context. Such a formalization would open the door towards a rigorous econometric analysis of TTOA using data obtained from real life or stated choice (SC) experiments. Such formalization and empirical investigation (using SC-data) of TTOA is the aim of this paper, and constitutes the paper's main contribution to the moral psychology and choice modelling literature. Another contribution of the paper lies in the introduction of the TTOA-concept to the choice modelling community.[1]

In sum, this paper contributes to the choice modelling literature by introducing and operationalizing the notion of taboo trade-off

[1] Indeed, the notion of TTOA has not caught on in the economic sciences, including the choice modelling community. Inspired by seminal work in consumer theory and decision making (e.g., Lancaster, 1966; Keeney and Raiffa, 1976), discrete choice theory has traditionally been based on models of choice behaviour that pre-suppose that any attribute can in principle be traded off against any other attribute. This holds for conventional, linear in parameters utility maximization models (see for an introduction Ben-Akiva and Lerman, 1985; Train, 2009; Hensher et al., 2015), as well as for more recent attempts to incorporate semi-compensatory decision making in choice models like regret aversion, loss aversion, or contextual concavity models (see for an overview Leong and Hensher, 2012; Chorus, 2014). Lexicographic models (Sælensminde, 2006), in contrast, assume a general rejection of trade-offs. However, those models do not differentiate between taboo trade-offs and other trade-offs, and are therefore not well suited to study taboo trade-off aversion.

aversion in a discrete choice context. More specifically, we derive so-called TTOA-models which accommodate taboo trade-off aversion, and we empirically test the resulting models using a discrete choice experiment designed for this purpose.

Before we move on to the main part of our paper, we wish to clarify a few things. First, we do not present our TTOA-models as potentially superior to existing models that do not accommodate taboo trade-off aversion. We merely aim to explore the potential role of TTOA in a discrete choice context. Second, we present our models as descriptive, rather than normative accounts of choice behaviour. That is, we aim to model 'actual' decision making behaviour, rather than 'ideal' behaviour of rational and moral agents. Clearly, the notion of TTOA is at odds with some of the core underpinnings of rational choice theory. We consider exploration of the connection – if there is any – of our TTOA-models with neoclassical economic axioms to be an important direction for further research. However, in this paper we explore TTOA-models as behavioural models. Third, our model specifications by design stay fairly close to the conventional linear in parameters utility maximization logit models that have been the workhorse of the discrete choice community for decades. This is partly based on pragmatic considerations (e.g. we want our models to be compatible with existing discrete choice software packages), and it partly reflects a desire for tractability: we will show that the most conventional choice model specification can be rather easily extended to accommodate a notion (TTOA) which at first sight seems to be largely incompatible with classical discrete choice theory. Fourthly and finally, unlike some of the seminal papers in moral psychology cited above, we do not presume that taboo trade-offs are by definition rejected by taboo trade-off averse decision makers. Rather, we hypothesize that taboo trade-offs lead to a disutility – which we call a taboo trade-off penalty – to the decision maker. Fourth, because of word limits and since we do not want to repeat past work, our paper focuses only on aspects related to the *modelling* and *empirical analysis* of taboo trade offs in a discrete choice context; for an excellent and exhaustive *conceptual* exposition and definition of taboo trade offs and related notions, we refer interested readers to Fiske and Tetlock (1997). That paper draws from theoretical work in Law, Economics, Political Science, and Psychology to conceptualize and define taboo trade offs. In our work, we use their definition of taboo trade offs as a conceptual foundation: "By a taboo trade off, we mean any explicit mental comparison or social transaction that violates deeply held normative intuitions about the integrity, even sanctity, of certain forms of relationships and of the moral-political values that derive from those relations." (Fiske and Tetlock, 1997 – page 256).

The remainder of this paper is structured as follows: section 2 presents TTOA-models; section 3 presents the data collection effort; section 4 presents empirical analyses; section 5 presents conclusions, discussion, and avenues for further research.

2. Discrete choice models accommodating taboo trade-off aversion (TTOA)

Taboo trade-offs have been mostly studied, and are most easily understood and modelled, in a binary choice context where a decision maker has to choose between a status quo option and an alternative which embodies changes with respect to this status quo. We work from the assumption that there is no possibility to opt out of the choice situation altogether (see the end of the section for a brief discussion on this topic).

Assume an individual who faces a choice between a status quo option *i* and an alternative *j*. The alternative may perform better than the status quo in terms of some attributes, and worse than the status quo in other attributes. Assume that the alternative to the status quo is defined in terms of differences – in terms of each attribute – with respect to the status quo. After normalizing the systematic utility of the status quo option at zero (alternatively, a constant may be estimated to represent the average inclination to choose the status quo), a classical linear-in-parameter utility specification would lead to the following formulation for the systematic utility[2] of alternative *j*:

$$V_j = \sum_m \beta_m \cdot x_{jm} \tag{1}$$

Here, V denotes systematic utility, m denotes an attribute, x denotes a difference in attribute values with respect to the status quo, and β denotes the taste-parameter associated with the attribute, which is estimated and may be positive or negative.

To accommodate that the decision maker may be averse to taboo trade-offs, our first taboo trade-off aversion (TTOA) model extends the systematic utility of alternative j as follows:

$$V_j^{TTOA} = \sum_m \beta_m \cdot x_{jm} + \sum_{(m,n)\in T} \tau_{m\to n} \cdot I_{m\to n} \tag{2}$$

Here, T represents the set of ordered pairs (m, n) where m is a 'sacred' attribute (such as human life) and n is a 'secular' attribute (such as travel costs or time). This combination is suggested in the moral psychology literature to be potentially taboo for decision makers.

Importantly, the notion of a taboo trade-off is directional: if m represents a sacred value and n a secular value, then trading off m against n (i.e., accepting an inferior or deteriorated value for m in exchange for a superior or improved value of n) is considered taboo, whereas trading off n against m is not considered taboo. For example, *sacrificing* the health of one's children with the aim of *earning* money is a taboo, while *sacrificing* money to *improve* one's children's' health is clearly not a taboo. This directionality of taboo trade-offs is captured in the indicator function $I_{m\to n}$ which is specified as follows:

$$I_{m\to n} = \left\{ \begin{matrix} 1 & if \ \left[\beta_m \cdot (x_{jm} - x_{im}) < 0 \text{ and } \beta_n \cdot (x_{jn} - x_{in}) > 0\right] \\ 0 & \text{otherwise} \end{matrix} \right\} \tag{3}$$

[2] We here define our models in terms of systematic utilities only. Depending on the specification of the error term, different discrete choice models can be derived. In this paper, we will focus on the most used ones (i.e., binary logit, and mixed binary logit with additive error terms).

The indicator function equals 1, if we 'take' from the sacred attribute m and 'give' to the secular attribute n. The arrow in $m \rightarrow n$ represents this. In truth-table form, the indicator function can be visualized as follows (Table 1):

Estimable parameter $\tau_{m \rightarrow n}$ is the associated taboo trade-off aversion parameter, which penalizes an alternative if choosing that alternative embodies trading off sacred attribute m against secular attribute n, compared to the status quo values of those attributes.

The behavioural interpretation of the model presented in equation (2) is as follows: a choice away from the status quo may involve making a taboo trade-off. Such a taboo trade-off, embodied in a choice for the alternative j, comes at the cost of a taboo trade-off penalty $\tau_{m \rightarrow n}$ which represents the disutility associated with trading off attribute m against n – that is, accepting an inferior (compared to the status quo) attribute value for m in exchange for a superior (compared to the status quo) attribute value for n. Note that moral psychology literature suggests that τ will be negative, given that m is 'sacred' and n is 'secular' (see references in the introduction).

Note that, from an econometric viewpoint, the step-function which is implied by the if-then formulation of the indicator function presented above may be inconvenient, as it precludes differentiation around the points where $\beta_m \cdot (x_{jm} - x_{im})$ or $\beta_n \cdot (x_{jn} - x_{in})$ are zero. We propose the following smooth-step function to circumvent this issue; note that, for purposes of readability, $\beta_m \cdot (x_{jm} - x_{im})$ is shorthanded as $\Delta\beta m$, and $\beta_n \cdot (x_{jn} - x_{in})$ as $\Delta\beta n$:

$$I_{m \rightarrow n} = \frac{\exp(\mu \cdot - \Delta\beta m)}{1 + \exp(\mu \cdot - \Delta\beta m)} \cdot \frac{\exp(\mu \cdot \Delta\beta n)}{1 + \exp(\mu \cdot \Delta\beta n)} = \frac{\exp(\mu \cdot (\Delta\beta n - \Delta\beta m))}{(1 + \exp(\mu \cdot - \Delta\beta m))(1 + \exp(\mu \cdot \Delta\beta n))} \quad (4)$$

Here, $\mu > 0$ is a smoothness parameter which is pre-set and not estimated, and which determines the smoothness (for values close to 0) or hardness (for larger values) of the step-function. If set to an appropriately high value (e.g., 10), the smooth-step function of equation (4) will very closely approximate the if-then function presented in equation (3). Fig. 1 illustrates this for the situation where $\mu = 10$, $-1 < \beta_m \cdot (x_{jm} - x_{im}) = \Delta\beta m < 1$ and $-1 < \beta_m \cdot (x_{jn} - x_{in}) = \Delta\beta n < 1$. Clearly, the smooth-step function approximates the truth table values presented above: it approaches 1 if and only if $\beta_n \cdot (x_{jn} - x_{in}) > 0$ and $\beta_m \cdot (x_{jm} - x_{im}) < 0$, and it approaches 0 otherwise.

Note that if there are strong expectations regarding the signs of the taste parameters (i.e., higher attribute values are strongly expected to be preferred over lower ones, or vice versa), then the indicator function need not be specified in the choice model itself (as above), but instead can be accommodated by means of an appropriate specification of – a series of – dummy variables in the data set. This is the case in our empirical example, see further below. Note again that parameters β and τ are to be estimated from choice data. Whereas the former reflects the 'trade-off free' weight of the attribute, the latter reflects the penalty associated with trading off two particular attributes against one another, allowing such a trade-off to be perceived as taboo by the decision-maker.

We present two other intuitive formulations which capture the notion of TTOA. The first of these is a so-called Count-TTOA model, denoted TTOA-C, which is written as follows:

$$V_j^{TTOA\_C} = \sum_m \beta_m \cdot x_{jm} + \tau_C \cdot \sum_{(m,n) \in T} I_{m \rightarrow n} \quad (5)$$

Here, τ_C is a count-taboo aversion parameter. The only difference between this utility function and that of equation (1) is that equation (5) estimates just one taboo aversion parameter which is multiplied by the total number of taboo trade-offs embodied in a choice for the alternative, rather than a single taboo aversion parameter for each of these taboo trade-offs. The more taboo trade-offs are embodied in a choice for j, the higher the penalty – and no distinction is made between these different taboo trade-offs.

The second formulation is called a generic-TTOA model, or TTOA-G:

$$V_j^{TTOA\_G} = \sum_m \beta_m \cdot x_{jm} + \tau_G \cdot \max_{(m,n) \in T} I_{m \rightarrow n} \quad (6)$$

Here, τ_G is a generic taboo aversion parameter, and the maximum operator indicates if there is at least one taboo trade-off. In other words, one generic taboo trade-off penalty is assigned to the alternative whenever there is any taboo trade-off.

As mentioned further above, we are so far assuming that the decision maker's taboo trade-off aversion is accommodated in a disutility associated with the non-status quo alternative. That is, moving away from the status quo may imply making a taboo trade-off and this comes with a penalty for the alternative. Although this specification makes sense behaviourally, and is compatible with the finding that choosing the default (or: status quo) option is a well-known strategy employed by decision makers in morally difficult choice tasks (e.g. Gigerenzer, 2010), it is not the only way in which taboo trade-off aversion can be specified.

More specifically, it might make sense to allow for an effect of taboo trade-off aversion on the decision-makers response to the choice set in general, rather than – as in the above models – on his or her (dis-)like of the non-status quo alternative in particular. There are two ways of creating such a choice set-based taboo aversion effect. A first approach would specify a heteroscedastic model where the taboo-parameter enters the scale of utility; for example, the presence of a taboo trade-off in the choice task may be hypothesized to lead to more or less random choice behaviour. We tested this specification on our data but found no significant effect of the presence of taboo trade-offs on scale. A second approach which acknowledges the choice set wide effect of a taboo trade aversion, would be to create an opt out

Table 1
Truth table for the taboo trade-off indicator function.

$I_{m \rightarrow n}$	$\beta_n \cdot (x_{jn} - x_{in}) < 0$	$\beta_n \cdot (x_{jn} - x_{in}) > 0$
$\beta_m \cdot (x_{jm} - x_{im}) < 0$	0	1
$\beta_m \cdot (x_{jm} - x_{im}) > 0$	0	0

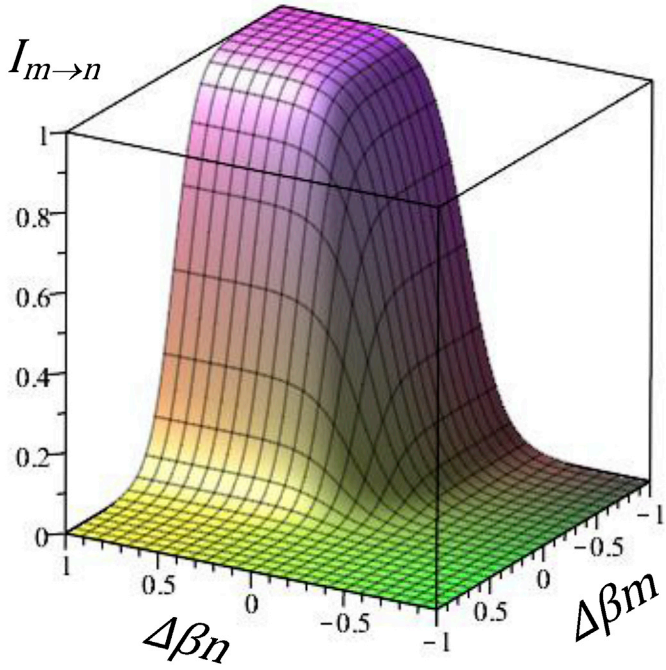

Fig. 1. Smooth-step approximation of.$I_{m\to n}$

option (in addition to the status quo and the alternative), and to assign the taboo trade-off disutility (i.e., $\sum_{(m,n)\in T} \tau_{m\to n}\cdot I_{m\to n}$, or $\tau_C\cdot \sum_{(m,n)\in T} I_{m\to n}$, or $\tau_G\cdot \max_{(m,n)\in T} I_{m\to n}$) to that opt out option. In our experiment – see next section – we have chosen to work with binary choice sets involving a status quo option and an alternative, but without the possibility to opt out of the choice task altogether. As such, in the remainder of this paper, we model taboo trade-off aversion by means of a disutility associated with a choice away from the status quo.

3. Discrete choice experiment and data collection effort

To empirically study the potential presence of taboo trade-off aversion, we designed a discrete choice experiment considering the support for or opposition against a massive national transport infrastructure investment scheme among car commuters in the Netherlands. In line with our model specifications (see previous section), we use a binary choice situation with a clear status quo (the current situation), an alternative (the proposed infrastructure investment scheme, or transport policy) and no opt out option. Also in line with the way we model taboo trade-off aversion, we specified the infrastructure investment scheme in terms of its effects (positive or negative) on a number of dimensions, relative to the status quo. Specifically, we created four attributes with two levels each. We included two 'sacred' attributes ('number of traffic fatalities' and 'number of seriously injured in traffic') and two 'secular' attributes ('vehicle ownership tax' and 'travel time') in the experiment.[3] In light of the reviewed literature (see the introduction), we *a priori* expect that respondents will experience 'travel time' as a secular attribute, although we cannot undergird this expectation with existing literature in moral psychology. The other three attributes are identified as archetypes for 'sacred' and 'secular' attributes in the moral psychology literature (see references in the introduction). The attributes and there levels are as follows:

(secular) Vehicle ownership tax (per year, for each car owner including yourself):

- 300 euro more/less than currently.

(secular) Travel time (per working day, for each car commuter including yourself):

- 20 min more/less than currently.

(sacred) Number of seriously injured in traffic (per year, for the Netherlands as a whole):

[3] As a reviewer noted, it would have been very insightful had we asked respondents to what extent they viewed these attributes as being either sacred or secular. This is related to the broader question, posed in Fiske and Tetlock (1997) "When do people treat trade-offs as taboo?" (page 256); we leave this empirical question for further research.

- 100 more/less than currently.

(sacred) Number of traffic fatalities (per year, for the Netherlands as a whole):

- 5 more/less than currently.

Given these attribute specifications, we hypothesize that a choice for the infrastructure investment scheme alternative (i.e., away from the status quo) embodied a taboo trade-off, in case one or more of the following combinations occur:

- tax goes down and the number of injured goes up
- tax goes down and the number of fatalities goes up
- travel time goes down and the number of injured goes up
- travel time goes down and the number of fatalities goes up

Note that the specific levels of each attribute were chosen such that they would be realistic given current (base) levels in the Dutch transport situation, and such that each attribute would have an approximately equal impact on utility. These values were identified, based on a pilot study with 30 participants. This pilot study involved participants completing the choice experiment, and a debriefing interview afterwards during which they were asked to reflect on the choice tasks and their responses.

As reviewers noted, we did not explicitly rule out, but neither did we emphasize, the possibility that the respondent him- or herself could be one of the fewer or more individuals that are injured or killed in traffic. As such, we cannot strictly claim that there is no element of personal (selfish) risk aversion in a participant's decision to support or oppose the transport policy. Although it seems fair to expect that such risk aversion only plays a limited role in our experiment given the very small probabilities involved, future research may explore ways to more strictly control for this aspect. Furthermore, it may be the case that participants to our experiment associate the implementation of the transport policy with other effects that are not explicitly mentioned in the survey (e.g. environmental aspects). Such effects are captured in the alternative specific constant associated with the status quo.

In order to maximize the probability of empirically identifying taboo trade-off parameters and taste parameters jointly and simultaneously, we chose to use a so-called full factorial design, consisting of $2^4 = 16$ choice tasks.[4] A disadvantage of this approach was that some combinations of attribute values were less logical, for example, an investment plan which causes less seriously injured but more traffic fatalities. Also, two choice tasks with dominant alternatives emerged in this full factorial design (e.g. an investment plan that implied a tax reduction, a travel time reduction, and safer roads).[5] We communicated to respondents that such seemingly illogical combinations might occur, and we asked them to take them seriously anyway. In our pilot study and actual experiment, this did not lead to problems. Choice tasks were presented to participants in random order, to avoid structural ordering effects. It was made clear to participants, that there are no 'right' or 'wrong' answers, and that we are interested merely in hearing the opinion of the sampled individuals. It was also communicated, that answers to the choice tasks would be processed fully anonymously, and that only aggregate results would be communicated in our publication(s). We emphasized the importance of taking the proposed policies seriously, and not considering the experiment as a game. For each attribute considered, we provided a brief explanation including metrics of the current situation (e.g., current number of traffic fatalities on the Dutch road network). It was made clear that a choice to oppose the policy would constitute a choice in favour of the status quo. Fig. 2 presents a (translated) example of a choice task.

As mentioned above, according to theory, a decrease in tax and/or travel time which comes at the cost of an increase in traffic injuries and/or fatalities is a potential taboo trade-off. Hence, our experimental set up involved seven choice tasks without any taboo trade-off (e.g. an investment plan with lower numbers of injuries and fatalities, a higher tax and a lower travel time). Four tasks contained one taboo trade-off, four contained two taboo trade-offs, and one contained four taboo trade-offs (i.e., lower taxes and travel times at the cost of more injuries and fatalities); this latter choice task is presented in Fig. 2 above.

Using a panel hosted by a large marketing research company, 99 respondents participated in the final experiment resulting in $16 * 99 = 1584$ choice observations. It is important to note, that this is a panel of individuals who have agreed to be contacted on a regular basis with the request to fill out a web-based survey. In exchange for completing a survey, they receive a small monetary payment. Importantly, subscribers to the panel receive survey requests on a variety of topics; as such, they do not have a specific interest in transport policy. The survey was administered online and took about 15 min to complete. Data were collected during the second half of February 2017. We attempted to arrive at a sample which is roughly representative for the Dutch regular car commuter (note that only respondents that commute by car at least three times per week could participate in the survey). More males (63) than females (36) joined our survey which was to be expected given the characteristics of the Dutch labour force composition; the average respondent's age was

[4] Note that we also asked respondents, after each choice task and on a five point scale, how difficult it was for them to make a choice. We did this to explore potential differences between choice tasks that did, and those that did not, contain taboo trade-offs. Statistical analyses revealed that perceived difficulty was highest for choice tasks involving more trade-offs; as expected, choice tasks containing a dominant or near-dominant alternative were considered by respondents to be relatively easy to complete. However, the presence of taboo trade-offs did not lead to (statistically significant) higher difficulty-ratings, beyond this general trade off-effect. This is consistent with our empirical finding that no scale differences exist between taboo- and non-taboo-choice tasks. It also adds credibility to our modelling approach in general, which models taboo trade off aversion at the level of the alternative, rather than at the level of the choice set (e.g. using opt out options).

[5] As advocated by Bliemer et al. (2017), we tested for scale differences between the two choice tasks containing a dominant alternative, and the other 14 tasks, but we did not find significant differences. Also note that, in line with argumentations put forward by Lancsar and Louviere (2006), we did not exclude the few observations where choices were made for a dominated alternative.

	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

Fig. 2. Example of a choice task.

49 years old; average education level was 4.8 on an eight point scale ranging from no education (0) to completed masters level education (7); average gross household income class was 40,000–52,900 euro per year. Since this study has a methodological focus, and does not aim to generate insights into transport policy preferences of the Dutch population, we do not explore the socio-demographic composition of our sample in more detail. Note that it is impossible for us to prove that (the sample drawn from) the panel is not biased in any way with respect to the level of taboo trade-off aversion (which is the core subject of study here); nonetheless, there is no indication that any such specific bias might exist.

4. Empirical analyses

4.1. Synthetic data

To show that in theory a TTOA-model based on systematic utilities as described above can be identified from choice data, we first estimated the model on synthetic data. Specifically, we generated choices for the full factorial design presented in section 3, based on a binary logit TTOA-model equipped with systematic utilities as in equation (2). Note that attribute levels were effect-coded, that is, '–1' represents a decrease in the attribute level (compared to the status quo) and '1' represents an increase. The utility of the status quo option was normalized to zero, and values of all taste and taboo parameters were set at -1 (the negative value for the taste parameters reflects that for each attribute, lower values are preferred over higher ones). For illustration, the utility of the alternative (j) in the choice task in Fig. 1 is $V_j = -4$: the taboo-free utility of j is $0 = +1+1–1–1$, and four taboo penalties of -1 are added. Note again that this is the maximum number of taboo trade-offs that can be obtained in the experiment: tax and time are reduced, and the numbers of injuries and fatalities both increase. Adding random errors from the EV Type I distribution assumed by the binary logit model, we generated 9600 choices (i.e., 600 for each of the 16 choice tasks). Subsequently, we estimated the TTOA-model presented in equation (2) with taboo penalties for trading off injuries against tax, fatalities against tax, injuries against travel time, and fatalities against travel time. Estimation results are given in Table 2.

Results show that all parameters are within two standard errors from their true values (equal to -1), suggesting that in theory, the TTOA-model in its most detailed form (which estimates a separate taboo penalty parameter for every single taboo trade-off) is identified.

4.2. Real data, binary logit models

Before presenting our model estimation results, we wish to emphasize that this is a confirmatory empirical research effort. That is, based on the moral psychology literature we hypothesize that there may be penalties associated with trading off increases in traffic injuries and fatalities against improvements (reductions) in tax and travel times. As such, we estimate purposefully designed taboo trade-off aversion models (see section 2) which were built to capture such penalties. We did not explore if there were other ways (such as including two- and three-way interactions between attributes) to improve empirical performance – our goal was not to maximize the model fit. By design, the experiment (binary choice set, two-level attributes) precluded testing regret aversion and loss aversion models.

We start by reporting results of binary logit models. Note that all models in this paper were estimated using BisonBiogeme and PythonBiogeme (Bierlaire, 2003, 2016). As somewhat expected (given that we obtained highly correlated estimates for this model even on our synthetic data set), we did not succeed in obtaining stable convergence and reliable results for the most elaborate TTOA-model presented in equation (2). Apparently, the data did not allow us to identify all four taboo trade-off parameters separately, together with all the taste parameters. The same holds for the so-called Count-TTOA model presented in equation (5), which counts how many taboo trade-offs are present in a given choice task, and multiplies this with the penalty that is estimated. However, we did obtain good results for the Generic-TTOA model presented in equation (6), which assigns one generic penalty (that needs to be estimated) to the utility of the non-status quo alternative, when that alternative embodies one or more taboo trade-offs. Estimation results are reported in Table 3. Note again that we adopted a coding scheme where the 'increase' in a level of an attribute was coded as '1', and a 'decrease' as a '-1'. As such, negative signs are expected for each of the four taste-parameters, as reductions are expected to be preferred over increases for all the attributes.

Based on these estimation results, a number of observations can be made. To start with the benchmark linear-in-parameters utility maximization model, we see that all attribute related parameters are of the expected negative sign, and are highly significant. For example, an increase (decrease) in tax compared to the stated quo would imply a 0.978 utils decrease (increase) in the utility of the transport investment scheme.

All estimates are of about the same order of magnitude, which suggests that the attribute levels which were adopted in the experiment succeeded in making all attributes approximately equally important to the average respondent. The adjusted rho-squared suggests that about one third of the initial uncertainty (entropy) faced by the analyst is explained away by the estimated model (Hauser, 1978). Moving to the Generic-TTOA model, we observe a negative (as expected) taboo trade-off penalty. Its size is non-trivial in comparison with the effects of the regular attributes; for example, having one or more taboo trade-offs embodied in the transport infrastructure investment scheme leads to a disutility which is almost as big as the disutility associated with a daily increase in travel time of 20 min. Given that the penalty's sign is as expected, a one-tailed *t*-test can be used to establish that the parameter is significant at a conventional 5% level. Note that under a two-tailed test regime, the t-value of 1.943 results in an associated p-value just above 0.05, making the parameter just insignificant at a 5% level.

Interestingly, we see non-trivial differences in the estimates for the other parameters when comparing the TTOA-model with its benchmark, linear-in-parameters model. First, we observe that the preference for the status quo – as implied by the associated constant – is lower in the TTOA-model. This is intuitive, since part of this preference away from the infrastructure investment scheme is now captured in a (negative) penalty for that scheme in case it embodies one or more taboo trade-offs, which is the case in 9 out of 16 choice tasks. More interesting are the differences in parameters associated with the attributes. Compared to the TTOA-model, the parameters for Injury and Fatality are about 10% larger (i.e., more negative) in the benchmark model, whereas the parameters for Tax and Time are about 10% smaller (i.e., less negative) in the benchmark model. These differences can be directly related to the behavioural premises underlying the taboo trade-off aversion model. That is, the TTOA model separates the attribute weights from penalties associated with particular trade-offs. For example, it allocates a certain (negative) weight to Fatalities, and assigns a further (negative) penalty to the situations where increases in Fatalities coincide with decreases in, for example, Tax. The benchmark model does not make this distinction, and 'adds together' both behavioural effects into one attribute weight, which as a consequence becomes inflated compared to its value in the TTOA model. A similar reasoning holds for the Injury attribute, and exactly opposite argumentations underlie the

Table 2
Estimation results (Synthetic data: binary Logit models).

Attributes: True parameter values between brackets.	TTOA-model	
	Est.	*Rob. SE*
ASC Status Quo (0)	0.11	0.10
Beta Tax (−1)	−0.92	0.06
Beta Time (−1)	−0.98	0.06
Beta Injury (−1)	−0.99	0.06
Beta Fatality (−1)	−1.07	0.06
Taboo Penalty Fatality-Tax (−1)	−0.74	0.15
Taboo Penalty Injury-Tax (−1)	−0.91	0.14
Taboo Penalty Fatality-Time (−1)	−0.95	0.15
Taboo Penalty Injury-Time (−1)	−0.94	0.15
Number of observations	9600	
Null-LogLikelihood	−6654	
Final LogLikelihood	−3819	

Table 3
Estimation results (binary Logit models).

	Linear-in-parameters model		Generic-TTOA model	
	Est.	*Rob. SE*	*Est.*	*Rob. SE*
ASC Status Quo	0.93	0.07	0.63	0.16
Beta Tax	−0.98	0.07	−1.05	0.08
Beta Time	−0.52	0.07	−0.57	0.07
Beta Injury	−1.11	0.07	−1.01	0.09
Beta Fatality	−0.79	0.07	−0.69	0.09
Taboo Trade-off Penalty	–	–	−0.44	0.23
Sample size	1584		1584	
Null-LogLikelihood	−1097.9		−1097.9	
Final-LogLikelihood	−721.2		−719.5	
rho-squared adjusted	0.339		0.339	

Table 4
Predicted in-sample probabilities for supporting a given policy.

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**

Bold indicates whether or not one or more taboo trade-offs are embedded in the choice task.

deflation of the 'secular' attributes (Tax and Time) in the benchmark model, compared to the TTOA model.

To explore if, and to what extent, the TTOA-model actually leads to different predictions than the benchmark model, we use the estimated models to predict, for each of the 16 choice tasks, the probability that a randomly sampled individual supports the infrastructure investment scheme. Predictions of the benchmark and TTOA models are compared with each other and with the true observed share of individuals that expressed support for the given investment schemes. Results are reported in Table 4, where the choice tasks are ordered in decreasing popularity of the investment scheme. Coded attributes are presented as well, where '1' stands for an increase in the attribute compared to the status quo, and '-1' stands for a decrease. Column Taboo indicates '1' if the infrastructure investment scheme embodies one or more taboos.

Note that with few exceptions, the TTOA-model, as expected, assigns a lower level of support for those infrastructure investment schemes that embody taboo trade-offs, compared to the benchmark model. For the schemes that do not involve taboo trade-offs, it predicts higher levels of support than the benchmark model. The mean absolute deviation of the benchmark model is higher than that of the TTOA-model, and this difference is especially pronounced for the choice tasks involving taboo trade-offs. These results are in line with the slightly better model fit[6] of the TTOA-model as reported in Table 3, and are also in line with behavioural intuition and the moral psychology literature.

To explore the potential role of sociodemographic variables in explaining support levels and taboo aversion, we allowed age, education level, income and gender to interact with the ASC for the status quo constant and with the taboo penalty. For brevity, we will not present the full table with estimation results, and provide a summary instead: we only found significant (at a 5% level) interactions between age and the status quo constant (older people were less likely to oppose a policy), and income and the status quo constant (higher income implies a greater inclination to oppose a policy). If one is willing to use a 10% significance level, we find an effect of age on the taboo penalty (older people have a larger, i.e., more negative, taboo penalty) and of gender on the taboo penalty (females have a

[6] An out of sample validation which we do not report in detail here due to space limitations, showed that the slight difference in model fit in favour of the TTOA-model was not due to overfitting.

Table 5
Estimation results Part I (Mixed Logit models; 1000 Halton draws).

	Linear-in-parameters model		Generic-TTOA model	
	Est.	*Rob. SE*	*Est.*	*Rob. SE*
ASC Status Quo (~N)	1.02	0.12	0.68	0.18
Sigma ASC Status Quo	0.80	0.15	0.17	0.29
Beta Tax	−1.08	0.11	−1.18	0.13
Beta Time	−0.58	0.09	−0.62	0.10
Beta Injury	−1.23	0.12	−1.15	0.13
Beta Fatality	−0.88	0.09	−0.74	0.10
Taboo Penalty (~N)	–	–	−0.65	0.26
Sigma Taboo Penalty	–	–	1.53	0.28
Sample size	1584		1584	
Null-LogLikelihood	−1097.9		−1097.9	
Final-LogLikelihood	−700.2		−678.1	
rho-squared adjusted	0.357		0.375	

larger, i.e., more negative, taboo penalty).

4.3. Real data, mixed binary logit models

Given that observable socio-demographic factors (such as age and income) typically explain only a small part of the heterogeneity, we estimated a series of Panel Mixed Logit models, which take into account that parameters may differ across individuals and that each individual made 16 choices. We experimented with a variety of specifications, such as different distributional assumptions for the random parameters (Normal, LogNormal), and different correlation structures between random parameters (full correlation, no correlation). Results of these estimation efforts can be summarized as follows:

- irrespective of the specification used, a large level of unobserved heterogeneity was retrieved and a substantial improvement in model fit was obtained compared to the corresponding binary logit models;
- irrespective of the specification used, the ML-TTOA model achieved a statistically significant improvement in model fit compared to the benchmark ML-model;
- irrespective of the specification used, a statistically significant taboo penalty was obtained as well as a significant sigma (representing heterogeneity in terms of the taboo penalty);
- the best model fit was found for a model that assumed a lognormal distribution for the taste parameters, and a normal distribution for the taboo penalty (implying a non-zero probability mass for a positive taboo parameter, suggesting taboo-loving), accompanied by a full correlation structure between random parameters;
- regarding the model with full covariance structure for the random parameters, we obtain significant and positive correlations between taste parameters – for example, between the parameters for Injury and Fatality (reflecting traffic safety minded individuals). However, we did not find a significant correlation between the taboo penalty and any of the taste parameters. This suggests that the taboo penalty is picking up a distinct behavioural effect which goes beyond a mere deviation in weights attached to particular attributes.

For illustration purposes, we present estimation results of two pairs of Mixed Logit models (one with and one without a taboo trade-off penalty); both presented models have no estimated correlation structure between random parameters, to facilitate representation. The first pair (Table 5) considers models in which only the constant for the status quo and – in the context of the TTOA-model – the taboo penalty were allowed to vary randomly (both with a Normal distribution); the second pair (Table 6) considers models in which all parameters are allowed to vary randomly (here, all parameters are Lognormally distributed[7] except for the Normally distributed constant for the status quo).

To start with the first pair: we observe that adding the generic taboo-penalty and its sigma leads to a substantial improvement of fit of more than 20 LL-points. More importantly, we find that the addition of the taboo penalty and its sigma reduces the size of the constant for the status quo option and makes its sigma insignificant. This means that what was initially assumed to be (heterogeneity in) a preference for the status quo option, turns out to be – to a considerable extent – a dislike for taboo trade-offs embedded in some of the non-status quo options. Like was the case in the binary logit models presented in the previous section, we again see that addition of the taboo penalty leads to more pronounced effects for secular attributes (tax and time), and less pronounced effects for sacred attributes (injuries and fatalities) in the TTOA-model. This is in line with theory, as explained in sub-section 4.2.

Table 6 shows a similar picture.

Note again that the parameters of the LogNormal distribution are in fact those of the underlying Normal distribution, and that minus-signs were placed in front of all LogNormally distributed parameters in the Biogeme-syntax, to ensure a negative effect of the attributes

[7] To ensure negative effects of increases in attribute levels – and of taboo trade-offs – on utility, a minus sign is added to the LogNormal distribution (which is by definition only defined on the positive domain) in the Biogeme-syntax.

Table 6
Estimation results Part II (Mixed Logit models; 4000 Halton draws).

Note: for LogNormally distributed parameters (∼LN), reported estimates refer to the underlying Normal distribution	Linear-in-parameters model		Generic-TTOA model	
	Est.	*Rob. SE*	*Est.*	*Rob. SE*
ASC Status Quo (∼N)	2.04	0.27	1.43	0.33
Sigma ASC Status Quo	1.63	0.23	1.40	0.37
Beta Tax (∼LN)	0.38	0.17	0.51	0.16
Sigma Tax	1.00	0.27	0.86	0.13
Beta Time (∼LN)	−0.57	0.30	−0.42	0.26
Sigma Time	1.27	0.27	1.18	0.18
Beta Injury (∼LN)	0.68	0.14	0.57	0.16
Sigma Injury	0.66	0.09	0.73	0.10
Beta Fatality (∼LN)	0.42	0.14	0.25	0.16
Sigma Fatality	0.52	0.09	0.56	0.12
Taboo Penalty (∼LN)	–	–	−0.95	0.43
Sigma Taboo Penalty	–	–	1.53	0.300
Sample size	1584		1584	
Null-LogLikelihood	−1097.9		−1097.9	
Final-LogLikelihood	−585.6		−579.0	
rho-squared adjusted	0.458		0.461	

and taboo trade-offs on utility. Also here, we see a significant (heterogeneity) in terms of taboo trade-off aversion, which ‘eats away’ from the (heterogeneity) in preference for the status quo option. To facilitate interpretation of the LogNormally distributed Taboo penalty, and to illustrate its effect on the position and spread of the Normally distributed constant for the status quo option, Fig. 3 presents the probability density functions (pdfs) for the constant before (dotted line) and after (solid line) introduction of the taboo penalty, as well as the pdf of that penalty (note that in this plot, the penalty is visualized in the positive domain; as explained earlier, a minus sign is added to the penalty in the utility function). These pdfs are the visual representation of the parameters reported in Table 6.

Taken together, our results suggest that on our data, there is a degree of taboo trade-off aversion, and that this may be captured by a taboo trade-off aversion-model which consists of a conventional linear in parameters logit model with the addition of a generic taboo penalty. Inclusion of this penalty-parameter results in improvements in model fit, as well as in changes in parameters and model predictions that are in line with the behavioural expectations.

5. Conclusions and discussion

Inspired by a large and growing body of work in moral psychology, this paper presents a series of models that accommodate so-called taboo trade-off aversion (TTOA). TTOA implies that decision makers consider trade-offs morally problematic – or taboo – when the attributes traded off against each other belong to different ‘spheres’. Taboo trade-off aversion has been observed in a wide variety of contexts, but so far the notion had not been systematically explored, and modelled, in a discrete choice context. This study is the first to

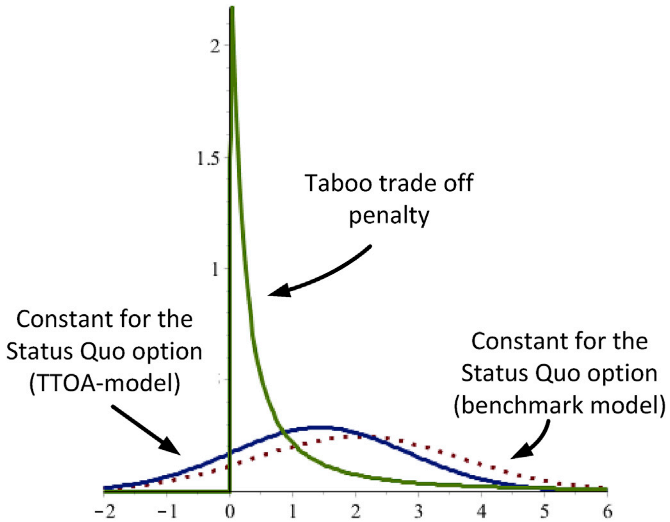

Fig. 3. Probability density functions for selected random parameters (from Table 6).

do so. To capture possible taboo trade-off aversion, we propose to extend the conventional linear RUM model by including so-called taboo trade-off penalties. Using several such model specifications, we then explore the presence (and size) of taboo trade-off aversion, using a data set specifically collected for this purpose. The choice experiment used for data collection concerned support for/opposition against a large scale transport policy which would affect several secular values (tax and travel time) as well as sacred ones (traffic injuries and fatalities). Results, based on estimation of a variety of (Mixed) Logit model specifications with and without taboo trade-off penalties, suggest that there is indeed a significant and sizeable taboo trade-off aversion underlying choice behaviour of respondents. We also show that ignoring this taboo trade-off aversion when it is present in the data, may lead to biased parameter estimates and market share predictions.

There is much work to be done before it can safely be concluded that taboo trade-off aversion is relevant for discrete choice analysis, and how it can best be modelled and studied. Our work should be considered a mere stepping stone for much needed follow up work along at least the following directions: first, it goes without saying that our results need to be replicated before they can be considered empirically robust. Preferably, such replication efforts should focus on choice contexts beyond transportation to study to what extent the notion of taboo trade-off aversion is relevant in different domains. For example, it seems likely that in areas such as health (e.g. de Bekker-Grob et al., 2012; Boeri et al., 2013), taboo trade-offs may play an even bigger role than in the relatively 'secular' transport context. Indeed, a recent study (Howard, 2017) on organ donation finds that offering potential donors a monetary reward in fact leads to a lower inclination to become donor – a finding which is fully in line with the notion that 'selling' one's organs for money is considered taboo by many.[8] Furthermore, in fields related to the environment and nature conversation (e.g., Czajkowski et al., 2009) taboo trade-off aversion is a potentially relevant behavioural concept (e.g. Stikvoort et al., 2016), because many people hold rights-based beliefs for flora and fauna and are reluctant to make trade-offs between environmental attributes and monetary attributes (e.g., Leamer and Lustig, 2017). In such follow up work, it would be interesting to explore which type of taboo trade-off aversion model works best under which conditions (also exploring choice sets with opt out options, and multinomial choice sets). Our finding that a model containing one generic taboo penalty performed best, may well be specific to our data.

Secondly, it is very important to note that – in line with the scope of this journal – our paper put forward the TTOA-model as a *behavioural* model that may be used for discrete choice analysis. The paper provides evidence that taboo trade-off aversion i) can be effectively modelled (in a way that allows estimation using conventional software), and ii) helps explain decision making behaviour in a discrete choice experiment. Much work needs to be done however, before we understand how the notion of TTOA should (not) be used for *policy* development. First, there is the very natural question how – if at all – our models connect with the axioms underlying (neoclassical) welfare economics, such as transitivity. This is something that should certainly be explored in future work. Put more broadly, the question how to use TTOA-models for policy making is very subtle indeed, and has to do a lot with the notion of 'agency', which is the capacity of the decision maker whose behaviour is modelled, to actually influence policy. Consider the situation we used in our experiments: our participants may experience disutility when choosing an alternative which embodied a trading off of human lives against taxes. But if the government simply imposes a policy that embodies such a taboo trade-off without consulting citizens, is it then meaningful to take such a disutility among citizens into account? Probably less so, because in that situation, citizens have no agency (i.e., they have no influence over the policy being implemented or not), and as such will likely not suffer from taboo trade-off aversion if the policy is being forced upon them. Situations and thought experiments like these are important to consider in the ongoing attempts to derive 'ethical' policies (e.g., Van Wee, 2011; Mouter et al., 2017). This study, and the modelling framework we have put forward, provide input to these vital debates; it would for example be interesting to study empirically to what extent the size of taboo trade off penalties differs when participants to the choice experiment are being put in either a citizen-frame (as in our study), or in a policy maker- or politician-frame.

Statement of contribution

The paper contributes to the literature by modelling and empirically analysing so-called taboo trade off aversion in a discrete choice context.

Taboo trade-off aversion implies that decision makers consider trade-offs morally problematic – or taboo – when the attributes traded off against each other belong to different 'spheres'. Specifically, it has been found that people dislike choosing an alternative which embodies an implicit trading off of a 'sacred' value such as a human life against a 'secular' value such as money. Taboo trade-off aversion has been observed in a wide variety of contexts, but so far the notion had not been systematically explored, nor modelled, in a discrete choice context.

This study is the first to do so. To capture possible taboo trade-off aversion, we propose to extend the conventional linear RUM model by including so-called taboo trade-off penalties. Using several such model specifications, we then explore the presence (and size) of taboo trade-off aversion, using a data set specifically collected for this purpose. The choice experiment which was designed for the purpose of testing the model concerned support for/opposition against a large scale transport policy which would affect several secular values (tax and travel time) as well as sacred ones (traffic injuries and fatalities). Results, based on estimation of a variety of (Mixed) Logit model specifications with and without taboo trade-off penalties, suggest that there is indeed a significant and sizeable taboo trade-off aversion underlying choice behaviour of respondents. We also show that ignoring this taboo trade-off aversion when it is present in the data, may lead to biased parameter estimates and market share predictions.

[8] Interestingly, that same study observed that when the monetary compensation was substantially increased, it did have a positive effect on inclination to donate. Apparently, each taboo has its limits.

Acknowledgements

The first author wishes to thank the European Research Council for financial support (ERC Consolidator grant BEHAVE – 724431). We thank three reviewers for providing a range of interesting comments on a previous version of this manuscript, and we are grateful for having had several fruitful discussions with participants at the ICMC in Cape Town, S-A.

References

Ben-Akiva, M.E., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. MIT press.
Bierlaire, M. (2003). BIOGEME: A free package for the estimation of discrete choice models, Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland.
Bierlaire, M., 2016. PythonBiogeme: a Short Introduction. Report TRANSP-OR 160706, Series on Biogeme. Transport and Mobility Laboratory. School of Architecture, Civil and Environmental Engineering, Ecole Polytechnique Fédérale de Lausanne, , Switzerland.
Bliemer, M.C., Rose, J.M., Chorus, C.G., 2017. Detecting dominance in stated choice data and accounting for dominance-based scale differences in logit models. Transp. Res. Part B Methodol. 102, 83–104.
Boeri, M., Longo, A., Grisolía, J.M., Hutchinson, W.G., Kee, F., 2013. The role of regret minimisation in lifestyle choices affecting the risk of coronary heart disease. J. health Econ. 32 (1), 253–260.
Chorus, C.G., 2014. Capturing alternative decision rules in travel choice models: a critical discussion. Chapter 13. In: Hess, Daly (Eds.), Handbook of Choice Modelling. Edward and Elgar, Cheltenham, UK.
Chorus, C.G., 2015. Models of moral decision making: literature review and research agenda for discrete choice analysis. J. choice Model. 16, 69–85.
Czajkowski, M., Buszko-Briggs, M., Hanley, N., 2009. Valuing changes in forest biodiversity. Ecol. Econ. 68 (12), 2910–2917.
Daw, T.M., Coulthard, S., Cheung, W.W., Brown, K., Abunge, C., Galafassi, D., …, Munyi, L., 2015. Evaluating taboo trade-offs in ecosystems services and human well-being. Proc. Natl. Acad. Sci. 112 (22), 6949–6954.
de Bekker-Grob, E.W., Ryan, M., Gerard, K., 2012. Discrete choice experiments in health economics: a review of the literature. Health Econ. 21 (2), 145–172.
Fiske, A.P., Tetlock, P.E., 1997. Taboo trade-offs: reactions to transactions that transgress the spheres of justice. Polit. Psychol. 18 (2), 255–297.
Gigerenzer, G., 2010. Moral satisficing: rethinking moral behaviour as bounded rationality. Top. Cognit. Sci. 2 (3), 528–554.
Harel, A., Porat, A., 2011. Commensurability and agency: two yet-to-be-met challenges for law and economics. Cornell Law Rev. 96 (4), 749–477.
Hauser, J.R., 1978. Testing the accuracy, usefulness, and significance of probabilistic choice models: an information-theoretic approach. Oper. Res. 26 (3), 406–421.
Hensher, D.A., Rose, J.M., Greene, W.H., 2015. Applied Choice Analysis: a Primer, second ed. Cambridge University Press.
Howard, K. (2017). Preferences For donating a family member's organs for transplantation: a discrete choice experiment. Presentation given at the *5th International Choice Modelling Conference*, Cape Town, South Africa
Keeney, R.L., Raiffa, H., 1976. Decisions with Multiple Objectives: Preferences and Value Trade-offs. Wiley.
Lancaster, K.J., 1966. A new approach to consumer theory. J. political Econ. 132–157.
Lancsar, E., Louviere, J., 2006. Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? Health Econ. 15 (8), 797–811.
Leamer, E., Lustig, J., 2017. Inferences from stated preference surveys when some respondents do not compare costs and benefits. Chapter 8. In: McFadden, D., Train, K. (Eds.), Contingent Valuation of Environmental Goods: a Comprehensive Critique. Edward Elgar, Cheltenham, UK.
Leong, W., Hensher, D.A., 2012. Embedding decision heuristics in discrete choice models: a review. Transp. Rev. 32 (3), 313–331.
McGraw, A.P., Tetlock, P.E., 2005. Taboo trade-offs, relational framing, and the acceptability of exchanges. J. Consum. Psychol. 15 (1), 2–15.
Mouter, N., van Cranenburgh, S., van Wee, B., 2017. An empirical assessment of Dutch citizens' preferences for spatial equality in the context of a national transport investment plan. J. Transp. Geogr. 60, 217–230.
Sælensminde, K., 2006. Causes and consequences of lexicographic choices in stated choice studies. Ecol. Econ. 59 (3), 331–340.
Stikvoort, B., Lindahl, T., Daw, T.M., 2016. Thou shalt not sell nature: how taboo trade-offs can make us act pro-environmentally, to clear our conscience. Ecol. Econ. 129, 252–259.
Tetlock, P.E., Kristel, O.V., Elson, S.B., Green, M.C., Lerner, J.S., 2000. The psychology of the unthinkable: taboo trade-offs, forbidden base rates, and heretical counterfactuals. J. Personal. Soc. Psychol. 78 (5), 853.
Tetlock, P.E., 2003. Thinking the unthinkable: sacred values and taboo cognitions. Trends Cognit. Sci. 7 (7), 320–324.
Train, K.E., 2009. Discrete Choice Methods with Simulation, second ed. Cambridge University Press.
Van Wee, B., 2011. Transport and Ethics: Ethics and the Evaluation of Transport Policies and Projects. Edward Elgar Publishing, Cheltenham, UK.
Zaal, M.P., Terwel, B.W., ter Mors, E., Daamen, D.D., 2014. Monetary compensation can increase public support for the siting of hazardous facilities. J. Environ. Psychol. 37, 21–30.