

An analysis of Timmermans' critique of Prospect Theory

Working paper

December 2012

Evert Jan van de Kaa

Delft University of Technology

Faculty of Technology, Policy and Management

Section Transport and Logistics

Preface and Summary

For a special issue of the *European Journal of Transport and Infrastructure Research (EJTIR)* Timmermans (2010) was asked to write a critical commentary on the suitability of Prospect Theory (PT) for travel behaviour research. When browsing through his article (from now on referred to as T) during a study about the transferability of PT's assumptions I came across a citation of an alleged inferior explanatory performance of PT with respect to people's choices in the TV game 'Deal or no Deal'. Out of curiosity I downloaded the cited working paper and found that the citation was not right. Successively I thoroughly reviewed the argumentations in T and several references that were advanced to support them. In this working paper my analyses and findings are documented extensively. A summary is published in *EJTIR*¹. Its editors had determined in advance that Timmermans would be given the opportunity to comment on my findings and that, if he did so, his response would close the discussion in *EJTIR*. Timmermans commented in a rejoinder². In this preface and summary I take these comments into consideration.

The underlying working paper was finalized in September 2012. It reviews the critiques in T in connection with the underlying arguments and referred findings. It shows the presence of errors and/or violations of good scientific practice, such as very inaccurate referring (Sect. 4.1), fallacious arguments (Sect. 4.7, Annex 4) and two texts that might suggest plagiarism (Sect. 4.5). In his rejoinder Timmermans admitted the inaccurate referring, but disregarded the fallacious-arguments-matter and denied any plagiarism. With respect to the latter, in my opinion he might be given the benefit of the doubt, though the rejoinder did not offer irrefutable evidence to the contrary like, for example, references to writings that explain both problematic passages in T. Another questionable assertion in T (p. 369) is the posit: 'Figure 1 gives an overview of dominant approaches and key issues that have been addressed and explored in the early years (1970-1980s). These are listed in the context of a general conceptual framework that summarizes the common elements of the various approaches (Timmermans, 1982)'. As neither Timmermans' 1982 article nor T does support these pretensions I considered them as fabricated (see Sect. 4.3 for an extensive underpinning). In his rejoinder Timmermans disregarded the reference to his 1982 article and advanced an alternative pretension of Figure 1, which I might not have classified as fabricated if these texts had been published in T instead.

Some more severe violations of good scientific practice were found in T:

- The statement that in a study by Blavatsky and Pogrebna (2007) 'remarkably, CPT never outperformed other decision theories, regardless of the assumed probabilistic choice rules' (T p. 376). Close reading of the cited paper revealed no information whatever that supports it. More telling, Blavatsky and Pogrebna (p. 10) explicitly stated that 'we do not estimate ... (cumulative) PT'. Apparently, Timmermans' assertion was fabricated (see Sect. 4.3 for my review).
- Several assertions that are weakening PT's credibility for travel choice modelling, such as: 'The most basic version of EUT ... is the expected value model ... Kahneman and Tversky questioned the validity of EUT ... *It should be noted that this position relates to the basic form of EUT*' (T p. 371-372, my emphasis). I spelled both articles in which Kahneman and Tversky posited PT and cumulative PT. Wherever they questioned the validity of EUT they consistently tested its standard form (with a concave utility function) and never the expected value model (with a linear utility function). The cursory statement in T was thus apparently fabricated. The same applies to several more assertions concerning PT (see Sect. 3 for my review).

In his rejoinder Timmermans did not pay any attention to these findings. This surprised me because such untrue statements should, in my opinion, be classified as scientific fraud if alternative explanations for their publication are missing.

¹ Van de Kaa, E.J. (2012). Timmermans' misleading critique of Prospect Theory actually supports its relevance for travel choice modelling. *European Journal of Transport and Infrastructure Research*, vol. 12, no. 4, pp. 440-458.

² Timmermans, H. (2012). Relevance of Prospect Theory to transport choice modeling: a rejoinder. *European Journal of Transport and Infrastructure Research*, vol. 12 (4), pp. 459-462.

After correcting for inaccuracies, errors, fallacies and untruthful citations, the remaining critical comments in T draw on dissimilarities between PT's assumptions and Timmermans' personal opinions about the human choice process (Sect. 5). At the actual state-of-the-art, human choice behaviour is an unobservable process and these opinions are thus not falsifiable. That is why, in my opinion, they do not provide scientific support for conclusions about the applicability of choice theories. In his rejoinder Timmermans pointed out that his critical comments in T were intended as his personal opinions indeed, and that he apparently disagrees with me about methodological matters, as I would consider that 'only a detailed historical account, based on solid theoretical or empirical support is academic, and anything else is inferior' (rejoinder p. 460). I believe that we disagree here indeed. Contrary to his suggestion in the rejoinder I am not opposed to personal opinions in well-balanced reflections about the state of the art and in formulating research agendas, for example, provided that they are explicitly advanced as such. In my opinion the substantiation of the comments in T with arguments and references to other work did not suggest that they were meant as mere personal opinions. However, as stated above, my concern is that most comments on the applicability of PT for transport choice modelling drew on dissimilarities with Timmermans' non-falsifiable opinions about transport choice processes. That is why, in my opinion, these comments do not offer scientifically valid insights in that applicability.

With respect to the supporting evidence of Timmermans' critiques on the applicability of PT this leaves his reviews of studies that compare observed and theoretically predicted choices. The empirical studies that were discussed in T seemed to demonstrate PT's alleged weak explanatory performance for transport choice behaviour. However, an examination of several of them revealed findings that corroborated PT's usefulness instead. A comparison with other reviews also suggests that the referred studies were chosen rather selectively. These other reviews demonstrated a good descriptive performance of PT for the choices observed in quite a number of empirical studies that were not covered in T (Sect. 6). In his rejoinder (p. 462) Timmermans repeated that 'in case of routine departure time and route choice, the question is whether these choice problems meet assumptions and reasoning behind prospect theory' without providing additional empirical support. However, since T was published several articles³ appeared that corroborate the earlier findings about the usefulness of PT for route choice modelling.

The final conclusion in T (p. 381-382) was: 'at the current state of development, PT lacks the rigor, scope, behavioural principles and mechanisms, and content validity to serve as a comprehensive theory of how individuals and households dynamically (re-)organize their activities and travel ... Applications of (C)PT to these types of choices represent an attempt to apply the theory in the wrong contexts.' As shown in this working paper, this conclusion drew on non-falsifiable statements and inferences, which in turn were for a large part based on questionable substantiations and arguments. It was thus based on scientifically invalid arguments. This does not mean that it is necessarily wrong. However, taking the findings from the empirical studies referred in T and those from other reviews into account I conclude that it is highly likely that an (extended) PT might serve as a rather comprehensive theory for transport choice explanation and prediction (Sect. 7). As discussed above Timmermans' rejoinder does, in my opinion, offer no clues to adjust this conclusion. His critiques on PT thus actually seem to support its applicability for transport choice modelling.

Hoeven, December 2012
Evert Jan van de Kaa

³ See e.g. Hjorth, K. and Ramjerdi, F. (2011), A Prospect Theory approach to travel time variability. *International Choice Modelling Conference*, Oulton Hall, Leeds; Ramos, G.d.M., Daamen, W. and Hoogendoorn, S. (2011), Expected Utility Theory, Prospect Theory and Regret Theory compared for prediction of route choice behavior. *Transportation Research Record*, vol. 2230, pp. 19-28.

Contents

| | |
|--|----|
| Preface and Summary | 2 |
| Contents | 4 |
| Keywords | 5 |
| 1. Introduction | 5 |
| 2. Interpretations of Prospect Theory | 7 |
| 3. Reproduction of the statements and underlying assumptions of PT | 9 |
| 4. Use of the relevant literature | 12 |
| 4.1 Characterization of T's bibliography | 12 |
| 4.2 A false account of other scientists' findings? | 13 |
| 4.3 A fabricated conceptual framework? | 13 |
| 4.4 Undeserved citations of one's own publications? | 15 |
| 4.5 Independently arriving at earlier published findings? | 16 |
| 4.6 Tendentious account of other scientists' findings? | 16 |
| 4.7 Selective, tendentious or false accounts of one's co-authored publications? | 17 |
| 5. Accuracy of the description of the actual choice process | 19 |
| 6. Use of empirical evidence | 22 |
| 7. Summary, recommendations and conclusions | 24 |
| Annex 1. Examination of the foundations of Figure 1 | 25 |
| Annex 2. Mental representation and cognitive environment | 28 |
| Annex 3: Probabilistic choice in risky settings or genuine errors? | 30 |
| Annex 4: Empirical evidence for what one cannot know | 32 |
| Annex 5: Relative utility, historical deposition and different frames of reference | 48 |
| References | 53 |
| Nabeschouwing | 57 |

An analysis of Timmermans' critique of Prospect Theory

Evert Jan van de Kaa

September 2012

Keywords

Good Scientific Practice, Scientific misconduct, Peer review, Affirming-the-consequent fallacy, Mental representation, Positive economics, Prospect Theory, Travel choice behaviour, Choice modelling.

1. Introduction

In my opinion readers and peer reviewers of a scientific journal like *EJTIR* are entitled to trust that the submitted articles offer an adequate explanation of the discussed concepts and to feel assured of the open-mindedness, solidity and, above all, truthfulness of citations, assertions and argumentations. The contributors should also have made a fair effort to find out whether the explanations, assertions or argumentations had previously been published and if so, should have credited the concerned authors by referring to them, thus avoiding any suspicion of plagiarism. If references are used to reduce the article length by introducing earlier explanations and/or substantiating argumentations, their relevance for those purposes should be beyond reasonable doubt. It may seem that referring to these elements of good scientific practice is stating the obvious. However, if a contributor does not behave according to these requirements even a solid peer review might not detect such scientific carelessness and/or misconduct⁴. I found suggestions of violations of several of these principles in a critical review of the relevance of Prospect Theory for travel choice modelling under uncertainty (Timmermans, 2010). Particularly, I uncovered the following errors and/or violations of good scientific practice:

- Misrepresentation of elements of PT (from now on used for referring to Prospect Theory) as advanced by Kahneman and Tversky (Sect. 3);
- Inaccurate referring (Sect. 4.1);
- Fabrication of empirical evidence questioning the suitability of PT for transport choice modelling (Sect. 4.2);
- Fabrication of findings from a literature survey that, if it existed, was not accounted for (Sect. 4.3);
- Unnecessary references to one's own publications (Sect. 4.4);
- Signs of plagiarism (Sect. 4.5);
- Misrepresentation of other scientists' findings (Sect. 4.6)
- Selective, tendentious and/or false accounts of findings from one's co-authored publications (Sect. 4.7)
- Positing and/or citing conclusions based on fallacious arguments (Sect. 4.7);
- Using personal opinions in argumentations without stating these as such (Sect. 5); and
- Selective and prejudiced use of empirical evidence supporting and/or undermining PT's credibility (Sect. 6).

Not knowing the intentions behind these errors and/or violations it is not for me to judge whether the boundaries of scientific integrity have been exceeded. In this working paper I intend to present my findings both carefully and extensively.

⁴ Contemporary interpretations of scientific misconduct are variations on 'Presentation to the scientific community of fabricated, falsified, or misappropriated observations or results and violation against good scientific practice' (National Research Ethics Council, Finland, 1998, in Nylenna *et al.*, 1999 p. 58).

My investigation into the appropriateness of the references in T (from now on used for referring to Timmermans, 2010) was straightforward: I consulted the publications referred to in order to check whether their content supported the corresponding citations and inferences in T. Where feasible I just browsed through the concerned texts but quite often a more thorough secondary analysis appeared required to arrive at solid conclusions. In my report hereafter I strive to offer quotations of corresponding texts from both T and the referred literature, thus allowing the reader to judge for himself.

With respect to the argumentations in T I agree that ‘it is critical that researchers ... systematically assess and compare the strengths and limitations of competing theories and models’ and that ‘fundamental discussions on the relevance of a particular approach are difficult to find in the transportation and travel behaviour literature’ (T p. 370). But for hardly if any of the few dozens of critical comments on PT was indicated whether or not this also applies to the nowadays dominant Random Utility Maximization (RUM) theory or to other theories and models that are currently applied in transport research, let alone that a systematic comparison was offered. As the comments in T thus apparently aimed at an absolute, rather than relative, assessment of the strengths and limitations of PT, one might expect that T would have applied some threshold or criterion to establish their (ir)relevance for modelling uncertainty in travel choices. I found no indication of this, which makes an unbiased appraisal of these critiques complicated.

Looking for a methodologically sound approach for such an unbiased appraisal I considered that the behavioural sciences have quite a history in discussions about the correctness and/or applicability of choice theories and models. Well-known examples are the ongoing criticism (e.g. Simon, 1955; Kahneman and Tversky, 1979) of neoclassical economics and other versions of Utility Theory (UT) and the rejection of both UT’s ‘unbounded rationality’ and the heuristics-and-biases concept of Tversky and Kahneman (1974) by the ‘ecological rationalists’ (e.g. Gigerenzer and Todd, 1999). Many comments in T and other publications that have been proposed against PT were similar to those considered in these discussions. Perhaps as a response to early criticism on neoclassical economics, Friedman (1953) wrote an authoritative article that sheds light on the value one might attach to such arguments. I will apply his line of reasoning in my review of the arguments in T. As it was published towards the end of 2010, no references used in this review will be from works that were published later.

In Sect. 2 I will characterize PT’s constituent assumptions and the different interpretations of it that one might encounter in transport research and the behavioural sciences at large. In Sect. 3 the reproduction of the assumptions of PT is analysed. Sect. 4 discusses the references in T. Sect. 5 is concerned with his comments about the accuracy of PT’s description of the human choice process and Sect. 6 evaluates the empirical evidence he advanced regarding the usefulness of PT for travel choice modelling. My review closes with a summary and some recommendations and conclusions (Sect. 7).

2. Interpretations of Prospect Theory

Before accounting for the methodology that I followed for the appraisal of the argumentations in T about the usefulness of PT it seems appropriate to examine the different interpretations, including mine, of that theory in the literature. KT (from now on used to refer to Kahneman and Tversky, 1979) posited PT as a number of basic assumptions about the choice behaviour of individuals. They presented these assumptions as generic textual descriptions. They posited, for example, that ‘gains and losses are defined relative to some neutral reference point (which) usually corresponds to the current asset position ... however, the location of the reference point ... can be effected by the formulation of the offered prospects and by the expectation of the decision maker’ (KT p. 274); ‘strictly speaking, value should be treated as a function of two arguments: the asset position that serves as reference point⁵, and the magnitude of the change from that reference point’ (KT p. 277); and ‘we hypothesize that the value function for changes in wealth is normally concave above the reference point and often convex below it’ (KT p. 278). Obviously, this leaves quite some room for person- and context-specific accentuation. However, in PT’s constituent publication KT (p. 275) only ‘discuss choice problems where it is reasonable to assume ... that the ... edited prospects can be specified without ambiguity’. These problems concerned several choices between gambles with probabilistic and certain outcomes. To explain the aggregate and/or the majority of choices in their experiments they considered just one out of several specific choice behaviour strategies, that are combinations of rules for framing, attribute appreciation, probability weighting, comparison of alternatives and choice criterion. Particularly, the strategy adopted in KT’s discussion of choice problems considered the current asset position as reference point and an attribute value function that was concave for gains, solely convex for losses and independent of the position/size of the reference point. Though Kahneman and Tversky did not exclude within-context interpersonal heterogeneity in choice behaviour strategies they did not consider this explicitly in KT or elsewhere

The difference in KT between the generic textual description of PT’s basic assumptions and their more accentuated application to risky choice settings gave rise to different interpretations of PT. Many researchers disregarded the textual descriptions and adopted the value and weighted-probability functions as used in the applications in KT. To distinguish this application, in which probability weighting is applied to given probabilities, from Cumulative Prospect Theory (CPT) (Tversky and Kahneman, 1992) in which it is applied to cumulative probabilities, some later authors (e.g. Li and Hensher, 2011) called it Original PT (OPT). Some authors, including Timmermans, narrowed PT’s interpretation further by constraining its applicability to choices under risk and uncertainty, while from Thaler (1980) onward other scientists categorize reference-dependent choice between alternatives with ‘certain’ outcomes under PT as well. Without wronging any of these views I conceive PT as defined by the assumptions that are articulated in the texts in KT, that is as a generic theory or paradigm, with accentuated versions like OPT, Reference-Dependent theory (Tversky and Kahneman, 1991) and CPT. This view is similar to a broad interpretation of UT as a coordinating theory with e.g. RUM, Neoclassical UT and Expected UT (EUT) as versions (e.g. Van de Kaa, 2010a).

I agree with Timmermans that the use of choice theories for descriptive analysis, which ‘is concerned with people’s decision making as it is, not as it should be’, is more relevant to transportation research than fundamental research concerned with normative theories (T p. 368). KT advanced PT as a descriptive theory following a critique of EUT, which they also conceived as a descriptive choice theory. Though Bernoulli (1738) might have meant to advance the latter in a prescriptive-normative sense, KT’s treatment agrees with a long-standing practice in economics and transport research. In that perspective

⁵ Note that several feasible framing rules might yield reference points different from the current asset position. Also note that Kahneman and Tversky (1984 p. 343 and following pages) used the more generic term ‘reference state’ for the same concept and in later publications often used point ‘point’, ‘state’, ‘value’ or ‘situation’ indiscriminately to indicate ‘the reference value to which current stimulation is compared (which) also reflects the history of adaptation to prior stimulation’ (Kahneman, 2002).

Friedman (1953 p. 4) would class both UT and PT under positive economics⁶, as they ‘deal with what is, not with what ought to be’.

Friedman (1953 p. 7) considers a theory as a mixture of a ‘language’ and ‘a body of substantive hypotheses’. He posits that ‘viewed as a language, theory has no substantive content ... Its function is to serve as a filing system organizing empirical material and facilitating our understanding of it; and the criteria by which it is to be judged are appropriate to a filing system’. Elsewhere I compared the completeness and consistency of the assumptions of PT and UT, from a functional perspective (van de Kaa 2008, 2010b). I found that the latter consisted of a complete, non-redundant set but the former had to be extended with assumptions about transferability, to allow for the use of PT for choice prediction⁷. Considering the rather constraint interpretations of PT that are commonly used I also accentuated the assumptions about reference state, reference state shifts, choice rules and heterogeneity in choice behaviour strategies in my posit of Extended PT (EPT). Except for this functional comparison of its basic assumptions with those of UT I found no systematic examination of the ‘language’ of PT in the transport literature.

With respect to the ‘body of substantive hypotheses’ Friedman (1953 p. 8) states that a ‘theory is to be judged by its predictive power for the class of phenomena which it is intended to explain. Only factual evidence can show whether it is right or wrong or, better, tentatively accepted as valid or rejected’. He elaborates this latter perspective in appraisals of several critiques on mid-20th century mainstream economics. Several similar argumentations about the validity of descriptive theories can be found in the social sciences. A sophisticated example is the position that Chater *et al.* (2003) took up in the debate on the (ir)rationality of UT and other theories, as initiated by Gigerenzer and Todd (1999).

Timmermans considered the critiques on PT ‘from the perspective of the development of choice modelling in transportation research’ (T p. 370). When Friedman wrote his essay, computers and programming were scarcely out of the egg. He did not examine the connection of theories and their implementation in mathematical models. In the actual review I consider the algorithms of choice models as mathematical implementations of theoretical assumptions about the choice behaviour of individuals. Compared to those of the more generic theory the assumptions of their mathematical model implementations are commonly much stronger, which eases their modelling. E.g. McFadden (2001) offered a nice overview of the accentuation of assumptions from neoclassical UT via Random Utility Maximization (RUM) to multinomial (MNL) discrete choice models with linear-additive utility specification.

A prerequisite for the suitability of choice models for predictions is that, in various contexts, their outcomes offer a fair approximation of the choices that people exhibit actually. It is, of course, not necessary that the considered people actually apply the same rules, make the same calculations, apply the same decision criteria and perform these in the same sequence as according to the theoretical assumptions, let alone as their mathematical implementation. It suffices if these model algorithms and theoretical assumptions offer a so-called paramorphic representation⁸ of the actual choice process (e.g. Hoffman, 1960; Swait, 2001). This agrees with the statement that ‘the relevant question to ask about the assumptions of a theory is not whether they are descriptively “realistic”, for they never are, but whether they are sufficiently good approximations for the purpose in hand’ (Friedman, 1953 p. 15). I will follow this line of thought in my evaluation of the critiques in T, and apply this to PT’s mathematical algorithms in the same way as the theoretical assumptions that they aim to simulate.

⁶ Friedman (1953 p. 4) characterizes positive economics by its task, which is ‘to provide a system of generalizations that can be used to make correct predictions about the consequences of any change in circumstances’. There has been some discussion of his essay among methodologists, e.g. about whether his concept of positive economics is an art or science, but this does not weaken the relevance of his argumentation for applied sciences like transport research.

⁷ I was incited to do so by the comment of Prof. em. Piet Bovy (pers. comm., 15-04-2004) that, while UT was well-equipped for model estimation and prediction, PT had proven its explanatory usefulness but was not a ‘complete’ theory as assumptions about its use for prediction were missing.

⁸ That is, distinct in form but analogous in the nature and product of their operations.

3. Reproduction of the statements and underlying assumptions of PT

In his thorough discourse of the kinds of evidence advanced as pros and cons for and against the validity of descriptive theories Friedman (1953 p. 34) elaborates a case where critique is forwarded against alleged in stead of actual statements of the theory. It concerns criticism on the theory of imperfect competition: ‘Marshall, it is said, assumed “perfect competition” ... The reader will search long and hard - and I predict unsuccessfully - to find in Marshall any explicit assumption about perfect competition or any assertion that in a descriptive sense the world is composed of atomistic firms engaged in perfect competition’. Similarly, Chater *et al.* (2003 p. 66) elaborate criticism in which rational calculations by the human mind are pinned unfoundedly on descriptive theories. I consider critiques based on such ‘straw man’ fallacies (see Annex 4) as erroneous and/or unfair.

T (p. 372) summarizes some core assumptions of PT in a concise overview. He aptly characterizes it briefly as ‘in PT, choice is based on transformed probabilities and outcomes as gains and loss’ (T p. 372). Next, he cites its generic formulae for the assessment of the overall utility of prospects and for the preference relation/choice criterion, and refers to the editing⁹ stage in which a reference point, gains and losses are established and the evaluation phase in which the alternatives are valued. For the latter, T mentions that the value function is concave for gains and convex for losses and that a monotonically increasing weighted-probability function transforms objective probabilities into subjective probabilities¹⁰. CPT is characterized as a different theory in which ‘cumulative probabilities rather than the probabilities are transformed’ and which allows for different weighting functions for gains and losses. Disregarding the minor flaws this overview offers, in my opinion, a fair though narrow reproduction of the essence of PT.

However, I found several notions in the article that are at odds with the premises of PT. For obvious reasons these were not supported by references to concrete texts, let alone quotations. I will discuss these in the order they appear in the text.

1. ‘KT questioned the validity of EUT ... *It should be noted that this position relates to the basic form of EUT*’ (T p. 372, my emphasis); previously T (p. 371) defined this form as ‘The most basic version of EUT ... is the expected value

model¹¹ ... it assumes that $u_i^n = \sum_{j=1}^J (p_j^n \cdot x_j^n)$... a deterministic decision rule is assumed in this classical case’ in which,

in turn, ‘each outcome j of the n^{th} risky prospect is defined by the values of a vector of observable attributes $X = \{ x_k; 1 \leq k \leq K \}$ ’ (T p. 370).

It is true that KT questioned the validity of UT. T’s cursory statement above, however, suggests that KT only considered a poor version of EUT which would weaken PT’s credibility. However, this *is not true*. Preceding their critique on EUT KT explicitly mentioned that, following most economic applications, they adopted the assumption of risk aversion/ a concave utility function $u(x)$. That is, they assumed

$u_i^n = \sum_{j=1}^J p_j^n \cdot u(x_j^n)$ in stead of expected value theory. I checked all comparisons between observed

choices from ‘different’ choice sets by Kahneman and Tversky and found that for all the considered choice experiments they consistently considered the expected utility formula with a concave utility $u(x)$. Moreover, with one exception the violations can also be found for any function $u(x)$ which is continuously increasing with x .

2. ‘PT assumes that decisions under risk and uncertainty are based on objective probabilities’ (T p. 373). This is not true, see KT (p. 274, 288): ‘The theory is developed for ... stated probabilities but it can be extended to more involved choices’; ‘where the probabilities of outcomes are not explicitly given ... decision weights must be attached to particular

⁹ Initially KT used the term ‘editing’ but from 1984 onward they systematically called the same concept ‘framing’, which I adopt here as a synonym.

¹⁰ Note that KT (p. 275, 280) used the term ‘decision weight’ instead and explicitly denied that this is a probability measure.

¹¹ I follow Timmermans’ definitions here. Actually, Bernoulli (1738) posited his concept of a concave utility function of money *against* the expected value concept which was developed in a correspondence between Pascal and Fermat (1654).

events rather than to stated probabilities, but they are expected to exhibit the essential properties that were ascribed to the weighting function.’

3. ‘It may be conceptually richer to distinguish between mental representation, cognitive environment¹², preference structure and choice rule to avoid any confounding as potentially done in PT’ (T p. 373). This critique apparently draws on the conceptual framework in Figure 1 that appeared to be fabricated (see Sect. 4.3). Kahneman and Tversky clearly considered, for example, a decision frame as ‘a representation of the act, outcomes and contingencies that are relevant to the decision maker’ (Tversky and Kahneman, 1992 p. 299) and according to Kahneman (2000, p. xiv) the ‘true objects of evaluation and choice (in PT) are neither objects in the real world nor verbal descriptions; they are mental representations’. More recently, Kahneman was even honoured¹³ by fellow scientists of the Federation of Behavioral and Brain Sciences for, among other things, his contribution (together with Tversky) to modelling ‘the interplay between the alternative framing¹⁴ of information, its mental representation as a function of the internal state of the decision maker, and the decisions based on that information.’

4. ‘learning models for decisions under uncertainty may have more to offer than non-dynamic models of decisions under uncertainty such as (C)PT’ (T p. 373-374); ‘loss aversion implies that travellers will likely experience that they could have done better. Repeatedly using updated reference points will then, ceteris paribus, lead to decisions and choices that deviate from the predictions of standard PT’ (T p. 378). Both learning models, (C)PT and UT are static choice models in the sense that, otherwise than dynamic models like e.g. Decision Field Theory (Busemeyer and Townsend, 1993), they do not consider changes in choice during a particular choice process. Just as learning models (C)PT rejects UT’s assumption that individuals have a complete preference ordering of all feasible alternatives, which does not depend on the actual circumstances and does not change between successive choices. As most applications of (C)PT to transport research assumed a constant choice frame during recurrent choices I explicitly included reference updating in my posit of EPT but this does definitely not conflict with the posit of PT. In KT even a whole section was devoted to the shifts of reference points that may occur between two successive choices in a sequence.

5. ‘In (C)PT risk attitude is nothing but a descriptive label of the curvature of the utility function and the weighted-probability function presumed to underlie travel choices ... It cannot be ruled out that the characteristic curvature can be caused by mechanisms other than risk attitudes’ (T p. 374). These comments seemingly discredit the validity of PT as a descriptive model of choice. In KT the reference-dependent framing of attribute levels is explained as a consequence of underlying psychophysical adaptation and perception mechanisms and in Kahneman and Tversky (1984) a more extensive psychophysical substantiation of the shapes of both the value and weighted-probability functions is given. Thus, PT’s fourfold pattern of risk attitude is indeed a descriptive label but this is founded on generally accepted underlying psychophysical mechanisms. If the credibility of a theory would be increased by considering risk attitude as more than a descriptive label PT should have been praised in T – but he the reverse happened.

6. ‘Because PT assumes a deterministic utility function and utility-maximizing behaviour, given the edited prospects, it implicitly assumes that individuals do take all information into account’ (T p. 376). Maybe this reasoning is based on the misconception that a compensatory, value-maximizing appraisal of attributes is only feasible in connection with a complete, context-independent preference ordering of all feasible alternatives as presumed in UT? KT only posited that PT’s edited-prospects-value-maximization algorithms should explain the choices of most individuals of the concerned population. Anyhow, the cited argument is specious, though insufficiently elaborated to diagnose it as circular reasoning viewed apart or in combination with other fallacies. In connection with T’s texts around it the conclusion was also superfluous, unless it was meant to weaken the credibility of PT.

7. ‘The literature on riskless choices has identified a series of effects influencing riskless choice behaviour, including effects of choice set, context, and taste variation to mention a few ... (C)PT does not take these effects into account’ (T p. 380).

¹² For an explanation of this term, see Annex 2.

¹³ see <http://www.fabbs.org/index.php?cID=169> (accessed 2012-06-04).

¹⁴ Note that several feasible framing rules might yield reference points different from the current asset position.

Obviously, it cannot be excluded that PT does not account for an unspecified effect implicated by T but posited in this way, the latter statement is not true. PT was proposed as a theory in which, different from (E)UT, an individual's judgments (or tastes) might differ dependent on the composition and presentation/perception of the choice set and the choice context, see for example KT (p. 275): 'the preference order between prospects need not be invariant across contexts, because the same offered prospect could be edited in different ways depending on the context in which it appears.'

4. Use of the relevant literature

4.1 Characterization of T's bibliography

In Table 1 I classified the entries in the bibliography of T according to content, field and document type.

Table 1 Characterization of the entries in the bibliography

| Direct theoretical or empirical evidence about the suitability of PT→ ↓ Scientific discipline and document type | Yes | No | Not referred |
|--|-------|-------|--------------|
| TRANSPORT RESEARCH | | | |
| Journal articles and book chapters | 5 | 9 | - |
| CD-ROM proceedings and working papers ¹ | 6 (4) | 7 (5) | 2 (1) |
| OTHER DISCIPLINES | | | |
| Journal articles and book chapters | 3 | 7 | - |
| CD-ROM proceedings and working papers ¹ | - | 1 (1) | - |

¹ Figures between brackets: papers that were republished as peer reviewed articles before T was proofed.

The occurrence of two entries that did not appear anywhere else in T seems questionable. For an article aiming at a critical review of the applicability of PT to transport research I also found the very frugal referencing to the large body of literature on PT from economics, decision theory and other social sciences conspicuously. The retrieval of 16 entries was hampered because, as grey literature, they were not accessible via Web of Knowledge or SCOPUS. For 10 of these this was unnecessary, as they were already published as peer-reviewed journal articles before T was proofed. By the courtesy of a colleague who attended several meetings at which the remaining were presented I was able to recover several of the other documents.

Closer inspection of T's references exposed several factual inaccuracies which complicated appraisal and/or retrieval:

- T p. 372: the reference to Kahneman and Tversky (1979) for the power functional form of the value function was wrong, this formula was not published before the posit of CPT by Tversky and Kahneman (1992 p. 309);
- T p. 375, 382: the journal article referred to as written by Camerer was actually written by Carbone;
- T p. 376, 382: the working paper of Blavatsky and Pogrebna (2007) was misdated as 2008;
- T p. 380: Habin and Miller should have been spelled Habib and Miller;
- T p. 380: the 2009 TRB conference paper of Chen *et al.* was misdated as 2008;
- T p. 382: the title of the 2005-article of Avineri and Prashker in *Transportation Research C*, 'Sensitivity to Uncertainty: The Need for a Paradigm Shift', was cited wrongly. It was apparently mixed up with the slightly different title of their 2003-publication in *Transportation Research Record* 1854, which reads: 'Sensitivity to Uncertainty: Need for a Paradigm Shift'. The correct title for their 2005-article is 'Sensitivity to travel time variability: Travelers' learning perspective';
- T p. 382: Borgers *et al.*'s paper did not cover pp. 221-234 in 'Advances in Culture, Tourism and Hospitality Research' but pp. 215-226 of its Volume 1, which was not published in Bangalore by MacMillan but in Oxford, UK by Elsevier.
- T p. 382: Apparently, Chen *et al.* contributed to the 88th Annual TRB Meeting, not the 85th.

In the following I aim to use the correct references, also in citations of T.

4.2 A false account of other scientists' findings?

To me, initially the most convincing argument that questioned the applicability of PT to travel behaviour research was the statement (T p. 376) that in a study by Blavatskyy and Pogrebna (2007) 'remarkably, CPT never outperformed other decision theories, regardless of the assumed probabilistic choice rules'. But this statement was apparently fabricated, as close reading of the working paper revealed that it contains no information whatever that supports it. Blavatskyy and Pogrebna examined the choices between a risky lottery and an 'opt-out' monetary bank offer for certain, in the UK and Italian versions of the TV game 'Deal or No Deal'. They considered different choice theories embedded in combinations of two datasets (Italy or UK) and in two versions (static and dynamic) of five stochastic choice models, yielding 20 loglikelihood listings. Before introducing the seven examined choice theories Blavatskyy and Pogrebna (2007 p. 10) explicitly stated that 'we do not estimate ... (cumulative) PT'. They mentioned that one might consider Rank-dependent utility theory (RDU), which they examined, as a version of PT in which the participants' assets at the beginning of the game act as reference state, which would rank all choice options as gains. This RDU offered the best loglikelihood in five of the 20 listings, the second best in six, the third best in eight and the fourth best in one. In any listing it thus performed better than at least three other theories. Even if RDU is considered to be a version of CPT the statement that CPT never outperformed other decision theories was thus invented. This incited me to start a thorough examination of the referring in T.

4.3 A fabricated conceptual framework?

Timmermans refers several times to Figure 1 (replicated below) as substantiating evidence for critiques on PT. Its title is 'Conceptual framework and key topics in seminal behavioural analyses in marketing, urban planning and transportation research' and it is introduced as follows: 'Especially outside of transportation research, a large number of studies, based on a variety of theories, concepts, and measurement approaches has been suggested to analyze and model individual and household decisions. Figure 1 gives an overview of dominant approaches and key issues that have been addressed and explored in the early years (1970-1980s). These are listed in the context of a general conceptual framework that

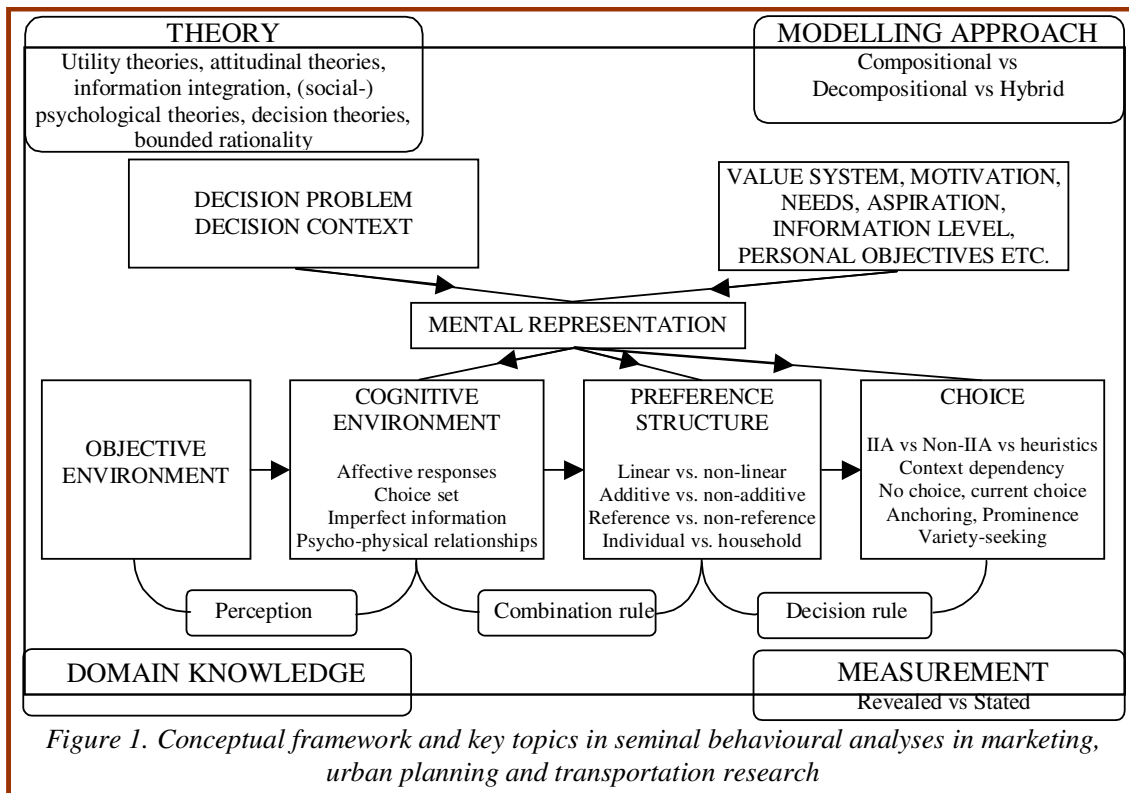


Figure 1. Conceptual framework and key topics in seminal behavioural analyses in marketing, urban planning and transportation research

Figure 1 Conceptual framework in Timmermans (2010)

summarizes the common elements of the various approaches (Timmermans, 1982)’ (T p. 369). In the remaining texts no further account is rendered of the origin and foundations of this diagram, nor of the purpose of its creation, nor of its entitlement to the claim that it offers an overview of early research and a general conceptual framework as is implied in the quotation above. Also, except for the 1982-article no further references that might provide such evidence are made. This suggests that Figure 1 was taken from the 1982 article.

Contrary to this expectation the 1982-article did not provide any account of the general conceptual framework nor an overview of the early research that Figure 1 claims to provide. It empirically investigated consumers’ ratings of some ‘objective’ attributes of shopping centres and contained a hypothetical process flowchart for destination choice (replicated and discussed in Annex 1). As any explicit support for the pretensions of Figure 1 is missing in both the 2010 and 1982 articles I examined whether the figure was self-explanatory and unambiguous (see Annex 1). This appeared not to be the case. For example, its keywords might be conceived as states, processes or functions; many research topics might be assigned to different keywords; the arrows might be conceived as relationships indicating exertion of control or time sequence or flow of information; etc. It is conspicuous that the keyword ‘cognitive environment’, which in the 1982 diagram apparently represented a set of perceived destinations with their attributes, was adopted in Figure 1 as a core element of a general choice behaviour concept. I found the prominent presence of the ‘mental representation¹⁵’ block even more confusing, in view of its position amidst several blocks that might accommodate specific forms of it, like ‘value system’, ‘cognitive environment’ and ‘combination rule’, to mention a few.

To grasp an idea of the degree to which the keywords in Figure 1 are nevertheless representative for the research effort that it claims to describe I did an ‘all fields, all document types’ search using Scopus (Table 1). The occurrence of most keywords in the concerned literature appeared very low. This indicates that these keywords and the attribution of topics to them are inappropriate for the provision of an overview of the dominant approaches and key issues that it claims to describe. In my opinion this shows that Figure 1 is hampering an overview of the concerned research rather than providing it.

Table 2 Occurrence of keywords in marketing, urban planning and transport literature

| Number of articles found in full text search on SCOPUS ¹ → which contained the choice-related concepts ↓ | All years | 1970-1989 | 1970-1989 |
|--|------------|------------|-----------------------|
| | All fields | All fields | 3 fields ² |
| “decision problem” OR “decision context” | 15,168 | 1,229 | 15 |
| “value system” OR motivation OR needs OR aspiration OR “information level” OR “personal objectives” | 2,300,562 | 234,878 | >>100 |
| “mental representation” | 14,257 | 232 | 5 |
| “objective environment” | 155 | 10 | 2 |
| “cognitive environment” | 272 | 8 | 0 |
| “preference structure” | 1,906 | 88 | 7 |
| choice | 973,867 | 86,915 | >>100 |

¹ accessed Februari 2012. ² “urban planning” OR marketing OR (transport OR travel).

As I was not able to retrieve any evidence that supports the claim that Figure 1 offers an overview of early research as well as a general conceptual framework for choice behaviour, the firm statement of these pretensions (T p.369) and the reference to Timmermans’ 1982 article are, in my opinion, scientifically unfounded if not misleading. In my review of the argumentations I consider the following statements that were underpinned by referring to Figure 1 as unsubstantiated personal opinions:

- ‘cognitive environment’, next to ‘mental representation’, is ‘one of the key issues that have been addressed and explored in the early years’ (T p.369);

¹⁵ For the meaning and interpretation of this concept, see Annex 2.

- it ‘has been convincingly demonstrated’ that, ‘as choices in certain real world environments do not necessarily reflect underlying preferences, observed choices in uncertain environments do not necessarily depict risk attitudes and corresponding decision styles and the various effects influencing the decision outcome’ (T p. 374);
- ‘PT lacks the behavioral concepts and may be too simple to avoid confounding of the various effects, shown in Figure 1, influencing the decision outcome’ (T p. 374).

4.4 Undeserved citations of one’s own publications?

On a total of 40 references the bibliography in T, in addition to Timmermans’ 1982-article, refers to 12 papers that were co-authored by him. As discussed above and in Annex 1, the once-only reference to the 1982 article does not supply evidence that compensates for the missing account of the claims based on Figure 1. The listing of the paper by Arentze and Timmermans (T p. 382), which is not referred to in the text, was obviously not appropriate. Each of the remaining 11 entries is referred to only once in the text. I will now consider whether each of these references is appropriate.

References to Chorus *et al.* (three articles) and Sun *et al.* appear on page 370, listed under ‘different theories and models of decision making under risk and uncertainty (that) have been applied in transportation research’. This was done under the self-imposed constraint of only referring to 21st century publications, which is, in my opinion, not a good reason to refrain from crediting prior publications. I would have found referring to the influential articles of Senna (1994, EUT in connection with a mean-standard deviation of travel time) and Bates *et al.* (2001, EUT in connection with schedule delay) more appropriate than the references to successively Sen *et al.* (2001) and Bos *et al.* (2004), without aiming to discredit these latter articles in any way. In addition to these applications three more theoretical concepts were listed, provided with references to the four co-authored articles. Obviously, the work of Chorus *et al.* (e.g. 2006) on Regret Theory deserved a reference here, but a rationale for referring to them more than once¹⁶ is missing. As far as I could retrieve the article in *Transportation* by Arentze and Timmermans (2005) was one of the first applications of Bayesian Belief networks in transport research. In my opinion Timmermans and colleagues might also be credited for bringing this topic to the attention of transport researchers as they considered it in several more articles from 2005 onward. A reference to this work appears thus well-deserved, though Sun *et al.*’s 2009-paper on CD-ROM might not have been the best way to do so.

Without agreeing with the concerned argumentations, the references to Zhu and Timmermans (T p. 375) and Han *et al.* (T p. 377, 378) seemed appropriate to me. However, on page 380 a lengthy argumentation led to the conclusion that ‘choosing PT because the researcher feels a reference point is necessary is not necessarily an adequate reason as several other utility-based alternative theories have been shown to offer the same mathematical functionality’. I wholeheartedly agree with this statement, as I consider it self-evident that the reference point/state of PT is nothing else but an elaboration of the asset position to which an individual considers him or herself to be entitled, which is at the heart of the, essentially relative, utility concept¹⁷ as posited by Bernoulli (1738). A researcher who feels that a reference point is necessary would thus behave quite ignorantly if he adopted this and all other assumptions of PT for that reason. T does not refer to researchers who actually stated that they had done so. Even if a researcher acted this way, I cannot see how this would provide any evidence against or in favour of the use of PT in transport research. The argumentation leading to the conclusion cited above is thus redundant. In my opinion this also makes the five references that were advanced as supporting evidence undeserved. This concerns four co-authored papers with Borgers, Chen, Zhang and Zhu as leading authors.

¹⁶ In the overview one reference is given for each choice theory and applications such as in a stochastic model for multi-attribute choice were not considered to make a difference, except for Regret Theory. Chorus *et al.* (e.g. 2008) extended Regret Theory to multi-attribute multi-alternative choice settings and called it Random Regret Minimization. Just one reference to Chorus *et al.* would thus have been appropriate.

¹⁷ For an elaboration of their and other relative utility concepts and their connection with reference states, see Annex 3.

To summarise, five of the 13 references to Timmermans' own (co)authored papers appear to be appropriate the remaining eight seem redundant.

4.5 *Independently arriving at earlier published findings?*

The brief overview of applications of PT to transportation research mentions that 'Avineri and Prashker (2004, 2005, 2006) applied PT in a route choice setting ... They found evidence of non-linear decision weights and loss aversion' (T p. 377). However, in Avineri and Prashker (2006) PT was not applied in the concerned route choice context. They even did not refer to that term in that article and also any reference in it to the work of Kahneman and Tversky is missing. During my PhD research¹⁸ the latter publication attracted my attention because one experiment described in it was published earlier (2003) in *Transportation Research Record* 1854, together with an exposition of CPT and Bayesian Learning models. Avineri and Prashker (2006) presented an additional experiment and did not refer to their earlier consideration of CPT. I made a secondary analysis of the choice observations presented in the 2006 article and found evidence of 'non-linear weighted probabilities' as well as 'reference-dependent framing and loss aversion' (Table 8 in Van de Kaa, 2008 p.177). In the same table the same inference was listed for the experiments described in the 2004 and 2005 articles of Avineri and Prashker. To my knowledge my dissertation, and the article in *Transport Reviews* (Van de Kaa, 2010a) that draws on it, are the only publications other than T in which the same inferences about the usefulness of PT are drawn from the 2004, 2005 and 2006 articles of Avineri and Prashker.

During my PhD research I re-examined several evaluations of the usefulness of PT for explaining recurrent choice between probabilistic alternatives, in which the participants received feedback about the outcomes of their previous choices in the sequence (e.g. Barron and Erev, 2003; Avineri and Prashker, 2005). In view of the attention that KT paid to reference shifts I was surprised that the different studies adopted the same reference state for all successive choices. I considered that feedback-based updating of the reference state and heterogeneity in choice behaviour strategies might offer a fair explanation of such recurrent choices. This view is explicitly articulated in the constituent assumptions of EPT and elaborated in several places in my dissertation, see e.g. my evaluation of Avineri and Prashker (2005): 'Considering the salience of the 'instant endowment' phenomenon it seems highly likely that the experienced outcomes of successive choices caused reference shifts ... Following the assumptions of EPT, the consistency of intrapersonal choice behaviour in such recurrent choice contexts could be studied by presuming idiosyncratic reference state updating ...' (Van de Kaa, 2008 p. 171). I was surprised to read that 'PT does not take such feedback and consequent learning and adaptation into account ... Repeatedly using updated reference points will then, ceteris paribus, lead to decisions and choices that deviate from the predictions of standard PT' (T p. 378) without reference to my earlier publication of this idea.

4.6 *Tendentious account of other scientists' findings?*

Sect. 3.3. in T addresses PT's assumption of a deterministic utility function. It is introduced by the statement that 'implicitly, PT assumes that when faced with replicated identical binary choices, subjects will make the same choice. There is overwhelming evidence to the contrary. Carbone (1997), Hey and Orme (1994) and Ballenger and Wilcox (1997) to name a few report switching behaviour between 20 and 30%, *fundamentally questioning the assumptions underlying PT*' (T p. 375, my emphasis).

At first sight the firm statement that PT assumes that people make the same choices from replicated identical sets seems a fair interpretation of PT. However, T did not mention that this is also the common interpretation of RUM's systematic utility function: 'Random utility models assume, as does the economic consumer theory, that the decision maker has a perfect discrimination capability. However, the analyst is assumed to have incomplete information and, therefore, uncertainty must be taken into account' (Ben-Akiva and Bierlaire, 1999 p. 7). Not mentioned either is that 'most recent work on the modeling of decision making under risk (and indeed under

¹⁸ Note that Timmermans commented on the final draft of my dissertation in April 2008 and received the printed book in the summer of that year.

uncertainty) has assumed that the preference functional of the decision maker is deterministic', which is the first sentence in Hey (1995 p. 633), who is cited in T on the same page. The cited switching behaviour does indeed yield unarguable evidence that subjects not always make the same choice from identical choice sets. However, PT was not examined in any of these three references, let alone that its underlying assumptions were fundamentally questioned by their authors. Other than the text in T might suggest, these authors did not infer therefore that their findings questioned PT's assumptions fundamentally, T inferred it from their articles.

In Annex 3 I examined whether these articles support the inference in T. It appears that genuine human error was the best explanation for people's choice of different alternatives from recurrent corresponding choice sets. Kahneman and Tversky did not explicitly discuss human error in connection with choice in agreement with PT. They share this with the foundations of most theories of choice under risk and uncertainty but I found no evidence that any of these theories excluded human error. Keeping this in mind I found no explanation why these articles incited to question fundamentally the assumptions underlying PT.

Next, I tried to understand the meaning of the remaining texts of Sect. 3.3 in T. I cannot grasp why adding a logistic term to PT's value function by Schwanen and Ettema, Avineri and Prashker and several more transport researchers, to account for genuine human error and/or fluctuating tastes, should not be convincing. In my opinion this is not in conflict with PT. Obviously, the estimated parameters of PT and any other choice theory may differ depending on the applied stochastic model in which it is embedded to account for errors. But if this effect is so strong that wrong behavioural conclusions are drawn this would apply to any choice theory.

Summarizing, in none of the five studies discussed above did I find evidence that might give rise to questioning PT's assumptions more than those of any other choice theory. That is why, in my opinion, referring to this line of research in connection with discrediting PT's assumptions seems unfounded and tendentious.

4.7 Selective, tendentious or false accounts of one's co-authored publications?

Questioning PT's so-called experiment-based foundations T (p. 375) stated that 'Zhu and Timmermans (2010c) argued that ideally the analysis of stated preference/choice data should include both a model of preference and choice behaviour, plus a process model of how subjects create a mental representation of the hypothetical choice problem'. I have carried out an extensive re-examination of the article referred to (Annex 4). In brief, Zhu and Timmermans proposed several strong assumptions about the choice process, developed a model drawing on these assumptions, estimated the parameters of their model for the responses to a stated choice experiment and apparently deemed the fit of their model with these responses fair enough for conclusions like 'our results showed that respondents seem to have applied extremely simple decision heuristics in the first stage' (Zhu and Timmermans, 2010c p. 779). However, in the same way that different mental choice processes might be approached by the same mathematical algorithms and different mathematical algorithms, like for example a RUM model, might describe the same choices, the inferences made by Zhu and Timmermans' were based on a combination of 'affirming-the-consequent' and 'begging the question' and/or 'non-cause as cause' fallacies¹⁹. Therefore, this also applies to their conclusion above, even though it was formulated with some restraint. If Timmermans would have recognized the specious character of these argumentations the firm statement about the desirability to supplement choice models with mental-representation-creation process models would have been deliberately misleading. Assuming he was not aware of it makes the statement just another fallacy, this time formulated without much restraint.

¹⁹ These fallacies are well documented in the scientific literature, starting with Aristotle, *On Sophistical Refutations* (ebooks.adelaide.edu.au/a/aristotle/sophistical/index.html). See Annex 4 for their definition and their occurrence in Zhu and Timmermans (2010c).

In T was referred to another paper by Zhu and Timmermans (2010a) who should ‘have argued that travelers may use multiple reference points. In their conceptualisation, however, reference points do not serve as anchors to distinguish between gains and losses, but rather as thresholds for accepting a decision strategy or not’ (T, p. 380). Scrutinizing their paper²⁰ I did not find the term ‘reference point’ anywhere. Except for the bibliography ‘reference’ occurred only as ‘reference alternative’. This was introduced as follows: ‘the individual compares the alternative with a given reference alternative to judge whether the alternative should be accepted or not ...the reference alternative is an instance of the personal value space’ (Zhu and Timmermans, 2010a p.7). Their empirical study concerned the choice between going-home and continuing-shopping during a shopping trip. The continuing-shopping option was denoted as the reference alternative. Its utility was treated as an unknown parameter of the going-home alternative to be estimated. Such a ‘reference-alternative’ concept is very dissimilar from PT’s reference point that draws on hedonic adaptation to people’s earlier experiences. Both in their theoretical sections and empirical case study Zhu and Timmermans considered only one reference alternative and did not use the term ‘multiple’ anywhere in connection with ‘reference’ in any meaning. The citation that Zhu and Timmermans (2010a) ‘have argued that travelers may use multiple reference points’ (T p. 380) is thus fabricated.

In T (p. 380, my emphasis) was remarked that ‘it seems that transportation researchers have primarily explored the applicability of (cumulative) PT to incorporate reference points in their models to differentiate between gains and losses ... the use of *reference points or thresholds* has a long history in modelling riskless choices to model ... relative utility theory (Zhang *et al.*, 2004), historical disposition (Chen *et al.*, 2008; Habib and Miller, 2009) and different frames of references as a function of accumulated experiences (Borgers *et al.*, 2007). Hence ... several other utility-based alternative theories have been shown to offer the same mathematical functionality’. Here, in passing and without substantiation, T lumped reference points, that are current or expected asset positions, with threshold levels, which separate rejected from accepted alternatives. He also advanced the listed theoretical concepts and references as offering the same mathematical functionality as PT. To the best of my knowledge, the mathematical functionality of PT’s reference point is to locate a kink and a convex-concave transition in the value function. Habib and Miller (2009 p. 92) presented a mixed-logit implementation of ‘the theoretical framework of PT for riskless choice’ in a ‘reference-dependent residential location choice model within a relocation context’ and compared it with a conventional RUM model. Adopting the characteristics of the current residence as reference-state levels they found that ‘the reference-dependent model performs better than a conventional location choice model in terms of model fit and provides important behavioral insights’. Obviously, their model offered PT’s functionality as it was an implementation of it but it was not utility-based as is meant by T. The three articles that were co-authored by Timmermans are discussed in Annex 5. They describe utility-based models but did not consider reference points or offer the same mathematical functionality as PT. More seriously, all three co-authored articles contain flaws and/or misleading references and/or fallacies. None of the four cited articles provides evidence that ‘several other utility-based alternative theories have been shown to offer the same mathematical functionality’ (T p. 380).

²⁰ Their conclusions and the arguments on which these were founded contained the same fallacies as listed in the previous paragraph (see Annex 4). Their article does thus neither corroborate nor reject their choice concept.

5. Accuracy of the description of the actual choice process

Many of the objections to theories of choice imply that their assumptions about the choice process do not correspond with ideas about people's real-life choice processes that the critic posits as true beyond reasonable doubt, without providing factually supporting evidence for it. They often use vague, subjective criteria such as 'face validity' or adjectives as 'well-known' to strengthen their views. As people's choice behaviour is a predominantly unconscious, covert process (e.g. Nisbett and Wilson, 1977; Dijksterhuis, 2004), verifying or falsifying the truth of such essentially personal opinions about the real-life choice process, and thus of the critiques that are built on it, is impossible. Sometimes cognitive limitations are advanced against descriptive-theoretical assumptions that seemingly require extensive calculations. The poor information processing capacity of the conscious mind may then be advanced to support such critique. Some examples from social sciences are the objections by Simon (1955) and Gigerenzer and Todd (1999) to the utility maximization assumptions of UT and, by the latter, also to PT's value-maximization assumptions. In transportation research such critiques have been advanced against CPT's weighted-probability assumption (e.g. Fujii and Kitamura, 2004; Avineri and Prashker, 2006). But the information processing capacity of the unconscious is huge (e.g. Dijksterhuis, 2004), which is evidenced by the phenomenal computational skills of several *idiots savants*. People who are not able to perform complex calculations unconsciously might use simplifying heuristics or their memory to achieve a similar result, as predicted by descriptive theories that are conceived as paramorphic models of actual choice processes.

Friedman (1953 p. 21) illustrated this latter view with the following example: 'Consider the problem of predicting the shots made by an expert billiard player. It seems not at all unreasonable that excellent predictions would be yielded by the hypothesis that the billiard player made his shots *as if* he knew the complicated mathematical formulas that would give the optimum directions of travel, could estimate accurately by eye the angles, etc., describing the location of the balls, could make lightning calculations from the formulas, and could then make the balls travel in the direction indicated by the formulas. Our confidence in this hypothesis is not based on the belief that billiard players, even expert ones, can or do go through the process described; it derives rather from the belief that, unless in some way or other they were capable of reaching essentially the same result, they would not in fact be *expert* billiard players'. In his conclusions Friedman (1953 p. 41) restates briefly that 'a theory cannot be tested by comparing its "assumptions" directly with "reality." Indeed, there is no meaningful way in which this can be done. Complete "realism" is clearly unattainable, and the question whether a theory is realistic "enough" can be settled only by seeing whether it yields predictions that are good enough for the purpose in hand or that are better than predictions from alternative theories'. This line of thought makes it irrelevant whether or not theoretical assumptions correspond with algorithms of actual human choice processes, be it conscious or unconscious. For objections about such a lack of correspondence holds that 'criticism of this type is largely beside the point unless supplemented by evidence that a hypothesis differing in one or another of these respects from the theory being criticized yields better predictions' (Friedman, 1953 p. 31).

Using this standard to judge the relevance of the comments in T on particular assumptions of PT I found a dozen or two that did not meet these qualifications. Several of these were based on the alleged authority of Figure 1, the fallacious arguments or the misrepresentations of PT discussed in the preceding chapters, others were advanced without underpinning and for none was solid empirical evidence advanced. Listed in the sequence in which they appeared these comments are:

- T p. 373: 'PT assumes that decisions under risk and uncertainty are based on objective probabilities ... However, in situations ... such as in gambles, individuals do not know these probabilities';
- T p. 373: 'it is not very realistic to assume that they' (people) 'first assign probabilities and then apply some weighting scheme';
- T p. 373: 'it may be conceptually richer to distinguish between mental representation, cognitive environment, preference structure and choice rule to avoid any confounding as potentially done in prospect theory';
- T p. 374: 'in (cumulative) prospect theory risk attitude is nothing but a descriptive label of the curvature of the utility function and the weighted probability function presumed to underlie travel choices';
- T p. 374: 'PT lacks the behavioral concepts and may be too simple to avoid confounding of the various effects, shown in Figure 1, influencing the decision outcome';

- T p. 374-375: ‘the experimental tasks used to test prospect theory typically look artificial ... It means that one cannot rule out the possibility that violations reflect incongruent mental representation and simple error as opposed to systematic bias’;
- T p. 375: ‘the analysis of stated preference/choice data should include both a model of preference and choice behaviour, plus a process model of how subjects create a mental representation of the hypothetical choice problem’;
- T p. 375: ‘some transportation researchers (e.g. Schwanen and Ettema, 2009) have added an error term to the value function and assumed a utility-maximizing decision rule to derive a logit-form model with a scale factor equal to 1. This set of assumptions is not very convincing. Not only is the use of an error term in conflict with the original theory, but assuming utility-maximization in the choice part and not in the valuing part seems inconsistent. Moreover, the results depend upon the assumed scale parameter of the utility function’;
- T p. 376: ‘probabilistic choice rules would only allow for error in the activation of the underlying utility/valuing function. It does not imply however that risk attitudes are only captured in the curvature of the valuing function and decisions weights and not in the choice rule. There does not seem to be any inherent reason to assume that risk attitudes do not play a role in the choice stage of a decision problem’;
- T p. 376: ‘However, many of the examined biases can also be explained by the alternative assumption that individuals demonstrate bounded rationality. They may filter attributes, set thresholds on attribute levels or may apply simplifying choice heuristics’;
- T p. 377: ‘The assumption of given probabilities is also incongruent with the typical decision problem in activity-based analysis. In general, travellers will not know the objective probability of an outcome ... Consequently, differentiation between decision weights and objective probabilities, as assumed by PT, may be impossible’;
- T p. 378: ‘any comprehensive theory of travel behaviour under uncertainty should include principles and mechanisms how travellers develop beliefs about the credibility of the information and information source, how they learn about possible underlying control strategies and how they dynamically respond to information and recommendation provided under these circumstances. Standard (C)PT does not satisfy this criterion’;
- T p. 378: ‘Loss aversion does not seem an effective coping mechanism against regret! One would expect that the value function becomes less curved as uncertainty is reduced, reflecting proportionally less concern with small gains and losses with larger change. This process may be captured at the level of the value function, but it may also involve ignoring extreme outcomes in the mental representation of the decision problems and updating of beliefs, implying that the value function will be activated for a small domain only in which its curvature is (almost) linear’;
- T p. 378: ‘loss aversion does not seem an effective behavioural mechanism in case of information exchange and common attitude formation in social networks’;
- T p. 379: ‘the question is whether loss aversion also plays a significant role in routine behaviour such as departure time, route and destination choice’;
- T p. 379: ‘In the context of departure time, this conceptualization (the loss aversion) seems less appropriate: even if travellers would view late arrival as a loss, the consequences can be easily remedied by calling ahead, working more efficiently or appealing to the largely accepted excuse of congestion for being late’;
- T p. 379: ‘in case of well-articulated beliefs about the distribution of travel times, it is not readily evident why travellers would not directly act on their context-dependent beliefs of travel times and risk attitudes, rather than first processing and valuing travel time variability against some endogenous reference point’;
- T p. 379: ‘It is even more likely that they have built up strong beliefs about the effectiveness of departure time strategies, skipping processing of travel time information altogether’;
- T p. 380: ‘travel time uncertainty is not restricted to a certain link of a path, but multiple sources of uncertainty may occur, some of which may be shared by different routes’;
- T p. 380: ‘departure time and route choice are just part of daily activity-travel scheduling processes and should be modeled accordingly’;
- T p. 381: ‘the conceptual richness, the congruence of assumed causal mechanisms and structures, and the content validity of these models’ (based on PT, vdK) ‘as a manifestation of a theory of travel behavior under uncertainty is relatively poor compared to competing theories of travel behavior under uncertainty, such as (Bayesian) network learning models, and regret-theoretical approaches. These competing approaches are not more or less direct applications of theories originally developed in other domains, but try to develop a domain-specific modeling approach based on the salient features and key underlying processes of activity-travel behavior under uncertainty. Making travel decisions under uncertainty is not even close to gambling for money!’

A telling example is the last quotation above. It precedes T’s final conclusions and draws on a posited better performance of learning models and regret theory. Timmermans is well-known with these concepts in transport research settings, as he has co-authored many articles in which they have been considered. Yet the only arguments elsewhere in T that support PT’s inferiority for transport research compared to Bayesian learning models concern its inability to update references between recurrent

choices (which was a misrepresentation, see Sect. 4 item 4 above). For the claim about regret-theoretical approaches nowhere else in the article any supporting evidence is advanced. Regret Theory (e.g. Loomes and Sugden, 1982) was proposed as an alternative for PT that explained the same violations of EUT as discussed in KT. To the best of my knowledge, before 2011 no comparisons were reported of the performance of PT and Regret Theory in transport research. Sticking to the literature to which T referred, Tversky and Kahneman (1992 p. 305) showed that ‘the violations of independence reported in tables 1 and 2’ (which were explained by CPT, vdK) ‘are also inconsistent with regret theory’ while in Hey and Orme (1994) Rank-Dependent Utility theory with the Quiggin (1982) function²¹ outperformed the best-fitting specification of Regret Theory in explaining the observed choices in their gamble experiment.

In sum, T’s statements above should, in the absence of solid evidence, be conceived as personal opinions that lack scientific underpinning.

²¹ This is the same weighted-probability function as adopted in CPT. As Hey and Orme considered gambles in which losses did not occur this rank-dependent utility model coincided with CPT.

6. Use of empirical evidence

In an earlier extensive literature search (van de Kaa, 2008) I found many studies in transportation research (van de Kaa, 2010a) and the social sciences at large (van de Kaa, 2010b) in which the assumptions of PT appeared to explain choice under risk, uncertainty or certainty better than the canonical UT paradigm while I encountered hardly any irrefutable evidence of the opposite case. Outside the transportation field Levy and Levy, for example, concluded that the outcomes of three choice experiments were inconsistent with PT but Wakker (2003) showed that their results would have fitted well with PT, had they not disregarded the weighted-probability function. In another publication Blavatsky and Pogrebna (2007 p. 12) stated that ‘Blavatsky and Pogrebna (2006) find no evidence of loss aversion in the Italian and British versions of the TV game *Deal or No Deal*’. However, their 2006 report showed that a significant majority of participants refused the considered swap offer. Their analysis does definitely not rule out that many participants exhibited loss aversion in doing so. From this perspective the comments in T about the poor empirical support for PT surprised me.

T (p. 375, 377) believed that ‘prospect theory is largely based on experiments ... Most empirical evidence supporting prospect theory is based on gambling experiments in which subjects are requested to choose one of two prospects, specifying the probability of associated outcomes.’ The idea that (C)PT is mainly concerned with gambles might rise from the articles in which PT and CPT were proposed, which were the only publications of PT to which T referred. Browsing through Kahneman and Tversky’s (2000) anthology ‘*Choices, values and frames*’ reveals a wealth of experiments and real-life observations for which the different assumptions of PT offered an explanation, including choice under uncertainty, context-dependent framing, endowment effects, investment decisions etc.

T (p. 374, 375, T’s emphasize) doubted the credibility of the empirical evidence that these experiments offered for the usefulness of PT: ‘the experimental tasks used to test prospect theory typically look artificial ... many examples seem designed to articulate and amplify known biases ... Experimental tasks often look like quizzes to test whether students understand expected utility theory. They require ... the *calculation* of losses and gains and overall payoff. Subjects ... certainly will make mistakes ... the basis of responses in case of the gambling experiments are given probabilities and decision outcomes, how unrealistic they may be. Subjects’ ... mental representation may differ from the constructed reality ... one cannot rule out the possibility that violations reflect incongruent mental representation and simple error ... Because prospect theory is largely based on experiments, evidence of risk aversion may have been confounded with errors introduced in understanding the experimental task, the framing of the task itself, limited information processing/bounded rationality in completing the task or any other process affecting the response-generating process.’ The gambles discussed in KT and Tversky and Kahneman (1992) concerned choices from simple binary choice sets that participants might well have made intuitively. Such experiments were replicated dozens of times in countries all over the world and the observed choice patterns were remarkably similar. The outcomes are generally used in economics and other social sciences to test and compare the usefulness of EUT and alternative theories for choice under risk. I found no other publications in which their credibility for that purpose was doubted. T (p. 373-374) suggested that ‘learning models for decisions under uncertainty may have more to offer than non-dynamic models of decisions under uncertainty such as (cumulative) prospect theory’, referring to the findings of Hertwig *et al.* (2004), who investigated choice from similar gambles with probabilities that were either given or had to be learned from experience. T (p. 375) also accepted the findings from this kind of simple gamble as ‘overwhelming evidence’ for a non-deterministic utility function (wrongly, see Annex 3) without disputing the credibility of these experiments. Furthermore, as discussed in the previous section T (p. 381) stated that the content validity of PT for travel behaviour under uncertainty was poor relative to regret-theoretical approaches. Regret Theory (e.g. Loomes and Sugden, 1982) was proposed as an alternative for PT as it also explained the choices in exactly the same gambles as discussed in KT.

As discussed in Sect. 4.1, the statement that to explain the UK and Italian versions of ‘Deal or No Deal’ ‘CPT never outperformed other decision theories, regardless of the assumed probabilistic choice rule’ (T p. 376) was fabricated. According to Blavatsky and Pogrebna (2007), who did not consider loss aversion, the overall best performing theory was EUT with an expo-power utility specification (EUT-exp) while RDU offered the second-best fit. For Dutch, German and USA versions of ‘Deal or no Deal’ Post *et al.* (2008) compared the performance of EUT-exp with a fully-fledged PT implementation with reference

updating after each bank offer. In all five considered settings PT outperformed EUT, offering an improvement in loglikelihood between 11% and 16%. The ‘Deal or No Deal’ game thus confirms rather than weakens the large body of empirical evidence from real-life and experimental settings that support the usefulness of EPT for choice modelling.

Remarkably, T did not refer to any transport research study in which an implementation of PT was outperformed by UT or any other choice theory. Though he questioned, for example, the appropriateness of PT’s loss-aversion concept for transport research on several places (e.g. T p. 379: ‘the question is whether loss aversion also plays a significant role in routine behaviour such as departure time, route and destination choice’) he did not refer to the many transport studies in which reference-dependent models outperformed loss-neutral UT models in explaining choice between alternatives with certain outcomes (Van de Kaa, 2010a). T’s overview of concrete applications of PT to empirical transport behaviour research under risk and uncertainty (e.g. Li and Hensher, 2011, for an overview) is restricted, both in the number of studies and depth of treatment. For the few studies referred to no comparison is offered between the performance of PT and other choice theories. An exception is the study of Schwanen and Ettema (2009), for which a statement that seemingly reduced their credibility was reiterated: ‘e.g. Schwanen and Ettema have added an error term to the value function and assumed a utility-maximizing decision rule ... Not only is the use of an error term in conflict with the original theory, but assuming utility-maximization in the choice part and not in the valuing part seems inconsistent’ (T p. 373) and ‘A deterministic choice rule was assumed. Implicitly, this means that the authors assumed that the utility function is stochastic, theoretically violating prospect theory.’ I demonstrated the inappropriateness of this comment in Sect. 4.6 above. The improved model fit of CPT compared to EUT as found in this study was also trivialised: ‘Overall, differences with expected utility theory seem modest at best’ (T p. 377).

In summary, T suggested that empirical evidence supporting the usefulness of PT as a descriptive choice theory is almost limited to simple gambling experiments while omitting references to much empirical research in social sciences at large; T casted doubt on the credibility of this kind of experiments and its usefulness for understanding transport choice under risk and uncertainty and at the same time presented inferences drawing on the same kind of gambling experiments without any comment on the nature of their empirical underpinning; T fabricated a poor performance of PT in explaining the British and Italian versions of the ‘Deal or No Deal’ game for which its performance was not assessed, its sibling RDU performed second-best to EUT-exp, and PT outperformed EUT-exp in the Dutch, German and USA versions; T disregarded the empirical evidence for the usefulness of PT’s reference-dependent-framing and loss-aversion assumptions in transport choice under certainty; and, except for one study for which PT’s good performance was played down, T did not discuss the relative performance of PT in empirical transport choice studies under risk or uncertainty. In my opinion this is selective use of the empirical evidence and I consider it to be misleading. It might explain why in T’s final conclusion-and-discussion section empirical evidence undermining or supporting PT’s (ir)relevance for transport research is missing.

7. Summary, recommendations and conclusions

During an extensive examination of the foundations of the comments on PT in T I uncovered the following errors and/or violations of good scientific practice:

- Misrepresentation of elements of PT as advanced by Kahneman and Tversky/ straw-man fallacies (Sect. 3);
- Fabrication of empirical evidence questioning the suitability of PT for transport choice modelling (Sect. 4.1);
- Inaccurate referring (Sect. 4.2);
- Fabrication of findings from a literature survey that, if it existed, was not accounted for (Sect. 4.3);
- Underserved referring to one's own publications (Sect. 4.4);
- Calling suspicion upon oneself of plagiarism (Sect. 4.5);
- Misrepresenting other scientist's findings (Sect. 4.6)
- Selective, tendentious or false accounts of findings from one's co-authored publications (Sect. 4.7)
- Positing and/or citing conclusions based on fallacious arguments (Sect. 4.7);
- Using personal opinions in argumentations without stating these as such (Sect. 5); and
- Selective and prejudiced use of empirical evidence supporting and/or undermining PT's credibility (Sect. 6).

With respect to the uncovered errors and/or violations of good scientific practice one might wonder how these could creep into *EJTIR*. I must confess, however, that when I was reading the article for the first time I only found that most objections against PT's process assumptions were of a general character and applied to EUT and other familiar choice theories as well. If I had been a peer reviewer of T – which I was not – I am not sure at all that I would have found more drastic objections than that. It was only when I accidentally found the fabricated finding from the 'Deal or No Deal' game that my mind-set changed, from 'taking the trustworthiness of the article for granted' into 'being on the alert for violations of good scientific practice'. Uncovering these in T required much more effort than one might ask of even the most devoted peer reviewer.

I have the feeling that a thorough reading of this paper might be helpful for editors and peer reviewers in uncovering bad scientific practice. To my mind, a more than superficial screening of the bibliography might often be revealing. The presence of the following topics in an article should warrant a very critical attitude of editors and peer reviewers:

- Any arguments and evidence yielding firm conclusions about mental processes; and
- Any 'empirical evidence' for the 'absolute' applicability or validity of theories/models and their assumptions, instead of their 'relative' applicability, based on *ceteris paribus* model fits, compared with a well-known alternative choice concept.

The final conclusion in T (p. 381-382) is: 'at the current state of development, it (PT) lacks the rigor, scope, behavioural principles and mechanisms, and content validity to serve as a comprehensive theory of how individuals and households dynamically (re-)organize their activities and travel (departure, route choice, destination, transport mode decisions) along multiple horizons in uncertain, non-stationary environments in a ubiquitous information society, enforcing a diversity of travel control strategies, for which they can rely on past experiences. Applications of (C)PT to these types of choices represent an attempt to apply the theory in the wrong contexts.' This conclusion apparently draws on the statements and inferences without solid ground as were discussed in Sect. 5. These were for a large part based on the invalid substantiations and arguments uncovered in Sect. 3 and Sect. 4. T's final conclusion is thus based on scientifically invalid arguments and lacks solid theoretical and empirical substantiation, which reduces it to a personal opinion rather than a scientifically sound inference. This does not mean that its content is necessarily wrong. However, taking the empirical findings from other studies into account makes it, in my opinion, highly likely that an (extended) PT might serve as a comprehensive theory for transport choice explanation and prediction.

Annex 1. Examination of the foundations of Figure 1

A1-1 Introduction

After discovering the discrepancy between the quotation of Blavatsky and Pogrebna (2007) and the underlying working paper in T I got a different outlook on its Figure 1 (replicated in main text above). Initially, I had taken it for granted that it ‘gives an overview of dominant approaches and key issues that have been addressed and explored in the early years (1970-1980s). These are listed in the context of a general conceptual framework that summarizes the common elements of the various approaches’ (T p. 369). However, a more thorough investigation yielded hardly, if any, evidence that supported these claims. I will first consider the account that is given for the ‘general conceptual framework’ posit and next discuss the justification for the claim that it gives an overview of the most relevant research approaches and issues in marketing, urban planning and transportation research, in the 1970s and 1980s.

A1-2 A general conceptual framework of choice behaviour?

Superficial examination of Figure 1 shows that a legend for the ideograms, as used to symbolize the different concepts and their relationships in the conceptual framework, is missing. In search for self-explanation I considered the meaning of the ideograms in Figure 1. This was hampered by the absence of a statement of the purpose of its creation or its inclusion in the article. The rounded rectangles at the corner and in the lower part of the diagram suggest that it is meant as an UML-activity diagram²² describing a workflow of stepwise actions or activities. However, in such diagrams plain rectangles are not used and the rounded rectangles should be connected with arrows indicating the sequence of the activities. Rounded rectangles are also used in process flowcharts. They denote either the start or the end of the process, indicated by an outgoing and incoming arrow respectively. Such flowcharts comprise at least some rectangles, symbolizing generic processing steps, and arrows, indicating the flow of control. Also this convention is not followed consistently in Figure 1. It also does not comply with the conventions for data flow diagrams or functional flow block diagrams, for example.

Of course, one does not necessarily have to follow one particular convention in depicting a contextual framework but a tailor-made design without a statement of its purpose or a key to symbols makes the interpretation of the diagram’s entities ambiguous. One might, for example, conceive the keywords in the rectangles in Figure 1 as either denoting states, processes or functions and the arrows as indicating exertion of control, time sequence, flow of information and so on. A more thorough investigation of Timmermans’ texts revealed no definitions nor explanations for the keywords in the diagram and the issues in lower case that are assigned to them, nor a justification for their selection. To me, from a linguistic perspective most keywords seem to indicate states but most lower-case issues listed under them rather seem to apply to transformation processes. Whatever is the character of the keywords, I found no rationale for the assignment of the issues to them. For example, most disjunctive concepts listed under the heading ‘preference structure’ are to my best knowledge more commonly considered as judgment or utility/value specification issues while I would expect that the idiosyncratic-preference-order concept would appear in the value-system-etcetera block. Another example is the listing of context-dependency, as well as anchoring and prominence, under choice, and not under mental representation, perception or cognitive environment. More ambiguities arise from the unspecified use of the keyword ‘mental representation’, without an incoming arrow from ‘objective environment’: many other entities in the diagram, like value system, preference structure, combination rule, etc, might be conceived as specified mental representations – see Annex 2 for an overview of the meanings of ‘mental representation’ and ‘cognitive environment’.

As Figure 1 is not explained in T and is not self-explaining I went through the 1982- article of Timmermans that was referred to as its source. This describes a study of consumers’ choices between different shopping centres. It contains a ‘conceptual model’ (replicated as Figure 2 below) which ‘clearly

²² I consulted Wikipedia (accessed 03-02-2012) for the definitions of different types of diagrams and their symbols.

illustrates that spatial choice behaviour is considered to be the outcome of an individual decision-making process' (Timmermans, 1982 p. 172). There is no legend nor explanation for the meaning of the ideograms in the diagram. Apparently, the convention for process flowcharts is not followed and I found no other convention which explains the meaning of all symbols. I guess that the arrows indicate temporal sequences as well as processing steps, the diamonds only processing steps, and the rectangles states and/or inputs/outputs.

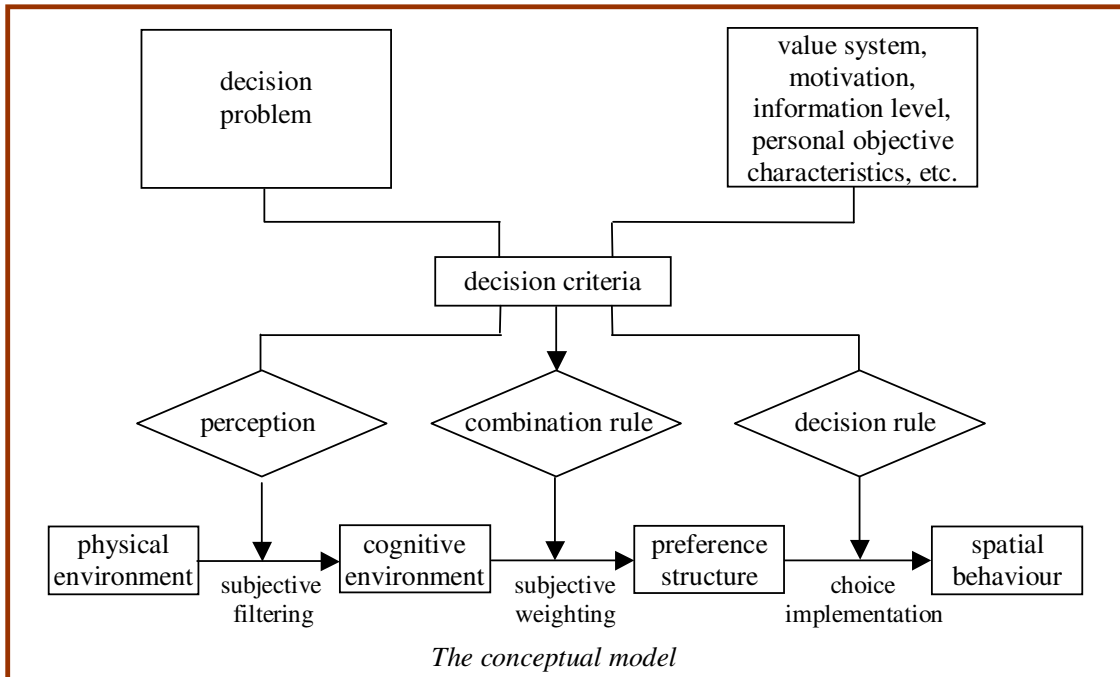


Figure 2 Conceptual model in Timmermans (1982)

Interestingly, the term 'cognitive environment' is neither defined nor mentioned elsewhere in the text. Its position in Figure 2 indicates that it is a mental phenomenon, contrary to its more common use as a descriptive label of the 'objective environment' (see Annex 2). It is apparently meant as synonymous to 'cognitive space', which is roughly defined in the sentence 'this perspective act typically involves a subjective filtering based upon imperfect information, the result of which is a *cognitive space*, consistent of n^1 ($n^1 < n$) destinations and m^1 ($m^1 < m$) attributes' (Timmermans, 1982 p. 173, his emphasize). This again is a rather constraint interpretation of spatial knowledge that people might employ in destination choice and a far cry from Norberg-Schulz's definition as cited by Golledge and Spector (1978 p. 407, their emphasis): '*The cognitive space* of the physical world involving imagery specifically related to perceived relationships. Within this mode the individual is able to think about spatial relations', to which was not referred by Timmermans. There can be no doubt that both interpretations of cognitive space conceive it as mental representations and, more specifically, as mental pictures, maps or images. I found no explanation why Timmermans did not use the same term in text and diagram for this mental representation, and why he did not adopt one that was more common in the contemporary literature, like 'spatial cognition' or 'mental map' (e.g. Evans *et al.*, 1981; Young and Richardson, 1979). In agreement with the diagram he treated it as the individual's subjective consideration choice set of shopping centre destinations. He surveyed the evaluations on a good-bad scale of shopping opportunities and accessibility attributes of several shopping centres by consumers who were familiar with them. In my opinion such judgments are more similar to part-worth utilities of the three concerned attributes of the 'physical environment' than to the corresponding actual levels of the 'cognitive space/environment', let alone that they provide a reasonable complete picture of that mental representation. Anyhow, Timmermans (1982) did not claim nor supply any evidence that his conceptual model as depicted in Figure 2 holds for other contexts than spatial choice.

The 1982-article did thus not pretend to provide a general conceptual framework. Also, it offered no clue for the attribution of the mental representation rectangle to T's Figure 1 nor for the other key topics in it. Nevertheless, several keywords in Figure 1 are apparently taken from the conceptual model for spatial choice (Figure 2), while neither in Timmermans (1982) nor in T a rationale was given that supports their applicability in a general conceptual framework of choice behaviour. If, for example, the constrained interpretation of 'cognitive environment' according to the 1982 article should be adopted, this would increase the incredibility of the framework – if not, it would increase its ambiguity. All over, Figure 1 differs too much from the conceptual model for destination choice to accept this reference as a proper account for its pretensions. As it is neither explained in the 2010 article nor self-explaining and the differences with the 1982 diagram are not accounted for nor referred to another document, the firm statement that it is 'a general framework that summarizes the common elements of the various approaches' (T p.369) is, in my opinion, unfounded if not misleading.

A1-3 Overview of issues in marketing, urban planning and transportation research?

Regarding for the posit that Figure 1 gives an overview of the most relevant research approaches and issues in marketing, urban planning and transportation research, in the 1970s and 1980s one should consider that Timmermans (1982 p. 172) stated that 'in recent years behavioural geography has provided a number of conceptualizations of spatial choice behaviour ... Fig. 1' (replicated above as Figure 2) 'summarizes the basic elements of these conceptualizations and adds some new elements and interpretations to them'. He did not indicate explicitly which new elements and interpretations he added. Also, he did not discuss the differences of his scheme with other process models of choice behaviour nor with a paramorphic interpretation of utility theory. He provided no positive evidence that it offers a more suitable transport choice modeling concept than these other theories. The reference to Timmermans (1982) does thus not offer any support for the overview-pretensions of Figure 1 in T.

Considering the entities in its rectangles seem at a first glance to justify the statement that Figure 1 offers the concerned overview. However, verifying or falsifying that statement is hampered as not any reference was given to the primary literature from the 1970s or 1980s that was allegedly overviewed. Also references to meta-analyses, reviews or anthologies of that primary literature were missing. Among the topics that are printed in lower case I missed many research items to which a multiple of empirical studies in psychology and decision theories were devoted, such as: framing of the choice context (e.g. Levin *et al.*, 1998, review); heuristic judgment and associated biases (e.g. Gilovich *et al.*, 2002, anthology); outcome-oriented v process-tracing approaches in choice process elicitation (e.g. Montgomery and Svenson, 1989, anthology); and loss appraisal v loss- neutral valuation (Kahneman and Tversky, 2000, anthology), to mention a few. Positive evidence for omitting these topics in Figure 1 was not found but my overview of the marketing and urban planning literature is too limited to dispute that it is justified.

To grasp an idea about the degree to which the keywords in Figure 1 are representative for the research effort that it claims to describe I did an 'all fields, all document types' search on Scopus (Table 1 in the main text). The occurrence of the denominations of most keywords in the literature to which the topics are assigned appears very low. This indicates that these terms and the attribution of topics to them are inappropriate. For example, the terms 'preference', 'judgment' and 'utility function' as a covering designation for 'linear vs. non-linear', 'additive vs. non-additive', 'reference vs. non-reference' and 'individual vs. household' each yielded over a hundred hits in the concerned literature against just seven for 'preference structure'. Also when disregarding the ambiguities discussed above, the selection of its keywords makes that Figure 1 apparently hampers an overview of the concerned research rather than providing it.

Annex 2. Mental representation and cognitive environment

A2-1 Introduction

One of the critiques in T (p. 373) is that ‘It may be conceptually richer to distinguish between mental representation, cognitive environment, preference structure and choice rule to avoid any confounding as potentially done in PT’. In this annex I examine the meaning of the ‘mental representation’ and ‘cognitive environment’ notions.

A2-2 Mental representation

‘Mental representation’ is a concept from cognitive psychology that is widely used in the social sciences including transportation research. The definitions that I retrieved from the literature are variations on the one proposed by Smith (1998 p. 391): ‘an encoding of some information, which an individual can construct, retain in memory, access, and use in various ways’. It is thus a broad concept, within which a main distinction between pictorial representations (or e.g. mental images), and nonpictorial (propositional or linguistic) representations might be discerned (Sober, 1976; Mani and Johnson-Laird, 1982). Several more specified categories of mental representations were hypothesized in psychology, which often are more or less overlapping and may or may not have been attributed to the pictorial or linguistic classes. Searching for it one will find categories like ‘environmental cognition’ (e.g. Sober, 1976); ‘cognitive maps’ (e.g. Evans *et al.*, 1981) or ‘mental maps’ (e.g. Dingemans *et al.*, 1986); ‘visual image’ (e.g. Mani and Johnson-Laird, 1982); ‘value image’ (Beach, 1990); and so on.

The schema-concept, which dates back to Plato, is the category of mental representations that has the longest history. Schemata may be defined as ‘cognitive structures stored in memory that are abstract representations of events, objects and relationships in the real world’ (Atkinson *et al.*, 1983 p. 214). Some more definitions of them and a recent application to transport research may be found in Plant and Stanton (2012). Within the schema-concept several subcategories may be discerned. One example is a stereotype, which is ‘a preconceived, standardised, and oversimplified expression of the characteristics which typify a person, situation, etcetera, often shared by all members of a society or certain sociological groups’ (SOED, 2002); ‘Stereotypes are forms of information and, as such, are thought to be stored in memory in a dormant state until they are activated for use’ (Gilbert and Hixon, 1991 p. 510). Distantly related are scripts, ‘embodying knowledge of stereotyped event sequences’ (Abelson, 1981 p. 715). These are considered to be deployed in habitual behaviour (e.g. Fujii and Gärling, 2003). Both might be the outcome of previous thinking, which might be defined as ‘the creation of mental representations of what is *not* in the environment’ (Dawes, 1988 p. 3, his emphasis).

In applied sciences, the term ‘mental representation’ is mostly used in connection with a more specific event or phenomenon in the real world. The cognitive neurologist Damasio (2001 p. 781), for example, defined ‘feelings’ as ‘the mental representation of the physiological changes that characterize emotions’. In a departure time choice context Fujii and Kitamura (2004) supplemented ‘mental representation of’ with successively ‘travel time’, ‘uncertainty’, ‘subjective travel time’, ‘numerical values’ and ‘uncertain travel time’. Searching on Web of Science²³ for publications in which the term is mentioned I found five articles co-authored by Timmermans. One was still in press and not accessible and in two the term was just mentioned in passing. A well-wrought discussion of the term is given in Arentze *et al.* (2008 p. 844-845). They propose ‘a mental representation of the decision problem that explicates the variables judged for evaluating choice alternatives’. This is conceived as a mental model and they consider that ‘to use a mental model in an active mode, it should be held in working memory. The limited capacity of working memory severely restricts the amount of information that can be represented and, therefore, only the aspects of the system perceived most relevant for the task and situation are taken into account. This implies that a mental model necessarily is tailored to a task and situation’. They thus consider choice behaviour, for example in a stated choice experiment, as consistently implementing a purpose-built script-like mental representation. Arentze *et al.* (p 866) further ‘argued that mental representations of decision problems include decision variables, a causal network, and utility judgments. The causal network links decision and situational variables to outcome variables (on attribute and benefit level) and utilities represent the person’s preferences for outcomes. We showed how these structures can be modeled as a decision network and used in an

²³ Accessed Spring 2012

active mode to make inferences, evaluate decision alternatives, and, when combined with a decision rule, to make decisions'. Their decision network simulates a strictly causal-sequential process of subjective consideration choice set formation, followed by attribute size, probability and expected utility assessment, taking situational variables like weather into account. Combined with a compensatory utility maximizing decision rule this simulates a choice process that hardly differs from a process in agreement with EUT. Obviously, EUT might also be considered as a script-type mental representation, but one that is not purpose-built for each choice context and not necessarily consciously implemented. Zhu and Timmermans (2010c p 778) adopted the same interpretation of mental representations of decision problems but 'to model both outcome and process of a decision, this paper proposed a modeling approach which incorporates attribute thresholds as the basic mechanism to model attribute selection and attribute representation.' Again, the assumed use of these and other non-compensatory decision rules might be conceived as implementation of a script-like mental representation that might most often not have been purposely built and consciously applied.

The previous overview of the definitions of mental representations and their occurrence in the literature might underline three important notions. Firstly, they are mental phenomena that are not part of the subject's 'objective environment'. Secondly, mental representations are states, which clearly distinguish them from other mental phenomena. Thirdly, it are psychological constructs whose existence and content is hypothesized, but cannot be confirmed or rejected by means of 'objective measurement' nor by introspection by the subject to whom they concern, due to the covertness of most mental phenomena (e.g. Nisbett and Wilson, 1977; Dijksterhuis, 2004). The first two notions became clear to me from the way in which Kahneman explained the fundamental difference between a subject's consideration choice set and the universal set of feasible alternatives: 'true objects of evaluation and choice are neither objects in the real world nor verbal descriptions; they are mental representations' (Kahneman, 2000 p. xiv, in a retrospective of his research on PT and judgment and choice at large). In search for a generic, systems-theoretical concept of human choice behaviour it brought me to define choice behaviour as 'a mental process that transforms mental representations of several courses of action and their expected outcomes into a choice ... It has mental representations, based on perception, as inputs and choices with respect to actions as outputs', where mental perception 'transforms the essentially physical sensory information into mental representations and arranges for the storage in and retrieval from memory' (Van de Kaa, 2008 p. 5, 8, 13). Though mental representations played a crucial role in this system-theoretical perspective, as inputs and outputs of separate constituent processes, I had to accept that assumptions about their existence and content are not falsifiable due to the covertness of most mental processes.

A2-3 Cognitive environment

In his 1982 article Timmermans mentioned a 'cognitive environment' concept in a diagram (see Annex 1). The interpretation of that term in social sciences leaves some doubt whether it refers to a characterization of someone's 'objective environment' or to a mental representation of it. The oldest mention that I found of it was in an article that investigated people's recognition of words. The words that had to be remembered were first presented together with different 'context words' and later their recognition was tested by presenting them together with the same or different context words. According to the authors (Tulving and Thomson, 1971 p. 123) 'the presence or absence of context words in our experiment affected the cognitive environment of to be remembered and tested words, and hence their encoded traces'. In my opinion this leaves some doubt whether 'cognitive environment' relates to the mind or the 'real' world. In education and human information transfer it is most often conceived in the latter sense. In transport research and urban planning the term is only incidentally used. In Timmermans' diagram (Figure 2) it was apparently meant as a mental representation, but the term was not found in the accompanying texts. Stern (2004 p. 71) also considered it, in passing, as a mental representation: 'The subjective map created from this direct and indirect information provides the cognitive environment within which the driver makes decisions.' Kwan (2000), Callois (2008) and Sunley *et al.* (2008) used it, all in passing, as a characterization of the objective environment. In view of the ambiguity of its meaning and the availability of plenty of synonyms the term should, in my opinion, better be avoided in transport research.

Annex 3: Probabilistic choice in risky settings or genuine errors?

One of T's critiques on PT is that it assumes a deterministic utility function while several researchers should have found that people use stochastic choice rules: 'implicitly, PT assumes that when faced with replicated identical binary choices, subjects will make the same choice. There is overwhelming evidence to the contrary. Carbone (1997), Hey and Orme (1994) and Ballenger and Wilcox (1997) to name a few report switching behaviour between 20 and 30%, *fundamentally questioning the assumptions underlying PT*' (T p. 375, my emphasis). In this annex I examine whether these articles support this inference.

Hey and Orme (1994) analysed the stated choices of 80 persons from two sets of 100 pairs of prospects. Each prospect was characterized by two outcomes (either zero or a gain in dollars) with given probabilities. In both sets the same 100 pairs of prospects were submitted, but the sequence in which this was done was randomized and the corresponding choice questions were not identical as the position of the prospects on the screen (left versus right) was switched between the sets. The authors reported an average switching frequency of 25% and compared the degree to which 10 different choice theories could explain the observed choices of each individual. These concerned, among more, EUT with linear (or risk neutral) utility specification; Generic EUT, with any continuously increasing utility function; two versions of Regret Theory; and two versions of Rank Dependent Utility Theory. PT was not considered. Hey and Orme (1994 p. 1300-1301, their emphasize) reasoned that 'the theories themselves ... are all theories of deterministic choice ... There *are* theories of stochastic choice, but these are not the concern of this paper ... This implies ... that the subject must be stating his or her preference *with some error*. Such error may arise from a variety of sources ... For rather obvious reasons, we confine attention to what we might term "genuine" error – mistakes, carelessness, slips, inattentiveness, etc'. They adopted a normal distribution with zero mean for this error, and thus analyzed the data with probit models. Considering the same data Hey (1995) found that a heteroscedastic model offered a better fit than the probit model. Nowhere in both articles I found factual evidence that attributing the choice-switching behaviour to anything but genuine error would offer a better explanation.

Carbone (1997) re-examined the choice data reported in Carbone and Hey (1995). These concerned 94 choices between two prospects, each characterized by two outcomes in dollars with given probabilities, made by 40 participants. Eight choices were between dominant and dominated alternatives, the remaining consisted of two corresponding sets of 43 prospects with left-right reversal and sequence order as sole differences between the sets. Carbone (1997) did not mention any switching percentage for the choice experiments that she considered but Carbone and Hey (1995) reported an average switching rate of 12% for their subjects. Following the generic EUT concept Carbone compared three different explanations for it: constant error, normally distributed error and several specifications of stochastic preference theory (Loomes and Sugden, 1995). PT was not considered. The probit model yielded the best fit for the choices of 20 respondents, the best fitting stochastic preference theory for 14 persons and constant error for six. Overall, genuine error thus appeared a better explanation for the observed choice switching than a stochastic choice model.

Ballenger and Wilcox (1997) did a similar choice experiment as Hey and Orme (1994) and Carbone and Hey (1995), which only differed in the number of two-times submitted prospect pairs (25) and the number of participants (120). The average switching rate for the 25 corresponding choice sets ranged between 11% and 30%, with an overall-average of 21%. They did not discuss whether genuine errors or probabilistic choice offered a better explanation for choice switching between corresponding choice sets. My interpretation of their conclusions about heterogeneity in choice behaviour and the (un)suitability of probit and logit models for its assessment is that they apply to all choice theories that are nowadays in use in transport sciences.

The cited switching behaviour does indeed yield unarguable evidence that subjects do not always make the same choice from identical choice sets. The predominant explanation for choosing different alternatives from recurrent corresponding choice sets that was found in these articles is genuine error, with a heteroscedastic rather than normal or logistic character. PT was not examined in any of these

three references, let alone that its underlying assumptions were fundamentally questioned by their authors. Other than the text in T might suggest, these authors did not infer therefore that their findings questioned PT's assumptions fundamentally.

Annex 4: Empirical evidence for what one cannot know

A4-1 Introduction

Questioning PT's so-called experiment-based foundations T (p. 375) stated that 'Zhu and Timmermans (2010c) argued that ideally the analysis of stated preference/choice data should include both a model of preference and choice behaviour, plus a process model of how subjects create a mental representation of the hypothetical choice problem'. This annex offers a thorough review of Zhu and Timmermans' (2010c) article (in this annex from now on abbreviated to ZT for short) and some other articles that were co-authored by Timmermans. It reveals several fallacies. In Sect. A4-2 I introduce the fallacy concept and some of its manifestations more extensively. Sect. A4-3 summarizes the bounded-rational process theory of choice behaviour that was proposed by Zhu and Timmermans (2010c) and Sect. A4-4 does so for its implementation in a discrete choice model and the estimation of its parameters on the outcomes of a stated mode-choice experiment. Sect. A4-5 examines their inferences and sheds light upon some fallacies. Sect. A4-6 exhibits similar fallacies that might have been made in some other transport research publications by the same authors. Sect. A4-7 offers a discussion and some conclusions and recommendations.

A4-2 Fallacies

According to the Free Dictionary, in Philosophy and Logic a fallacy is defined as 'an error in reasoning that renders an argument logically invalid' (www.thefreedictionary.com/fallacy, accessed May 2012). An inference that follows from a fallacious argumentation is not necessarily wrong. Considering it as such just for that reason would be a so-called 'argumentation from fallacy', which itself is a fallacy. However, if the specious character of the argumentation that underpins a conclusion had been recognized by the authors, its publication would have been a violation of good scientific practice. Assuming this was not the case such a conclusion might, from a scientific point of view, better be considered as a scientifically unsubstantiated personal opinion.

The first scientific treatment of fallacies and their exposure is attributed to Aristotle (384-322 BCE). Since then fallacies and their exposure has been a persistent if uncommon research topic in most social sciences. Though Aristotle discussed them in several of his works a concise overview and classification can be found in *On Sophistical Refutations* (Aristotle, \approx 350 BCE). Woods (1999) gave a nice exegesis of this work. He explains that it has to some extent the character of a practical guide for the argumentative contests that were popular in Aristotle's age, helping 'one participant (who) seeks to refute the thesis of an opponent by asking questions the answers to which constitute a *syllogism*. A syllogism is a sequence of propositions the last member of which (the conclusion) follows of necessity from the prior members (the premisses) ... A syllogism is a refutation of a thesis T just in case its conclusion is the contradictory of T, that is, the proposition not-T. On the other hand, an argument would seem to qualify as a *sophistical* refutation when it appears to be a refutation, but is not in fact' (Woods, 1999 p. 204-205). In this annex I will disregard this discourse context and examine whether the conclusions in the reviewed writings follow from the premisses on which they are based or not and if the latter is the case then in my opinion they are based on a fallacy. The fallacies that I have discovered in this research might be conceived as complying with Aristotle's Fallacies of 'Consequent', 'Stating as cause what is not the cause' and 'Begging the Question' and with the 'straw man fallacy' that was proposed later. Though many more fallacies were defined from Aristotle onwards these still appear in most contemporary overviews.

The Fallacy of Consequent 'arises because people suppose that the relation of consequence is convertible. For whenever, suppose A is, B necessarily is, they then suppose also that if B is, A necessarily is' (Aristotle, \approx 350 BCE, Sect. I.5) He successively illustrated this with examples of everyday-layman reasoning: 'people often suppose bile to be honey because honey is attended by a yellow colour: also, since after rain the ground is wet in consequence, we suppose that if the ground is wet, it has been raining; whereas that does not necessarily follow.' There is still some academic discussion about how Aristotle distinguished the Fallacy of Consequent from the other versions of the more generic Fallacy of Accident (e.g. Botting, 2012). According to Woods (1999, p. 211, his emphasis) the former 'is an early version of what has come to be known as the fallacy of affirming the consequent. In present day treatments it is the mistake of concluding that P on the basis of the two premisses, "If P then Q", and Q. Where P is

the *antecedent* and Q the *consequent* of the first premiss, the fallacy is that of affirming P on the basis of having affirmed the consequent Q'. Formulating the wet ground example in this way yields:

- A: If it rains, the ground gets wet (conditional antecedent-consequent premise).
- B: The ground is wet (nonconditional affirmation of the consequent).
- C: Therefore, it rained (conclusion of affirming-the-consequent fallacy).

I have adopted the Affirming the Consequent interpretation of Aristotle's Fallacy of Consequent for this annex and will denote it hereafter for short as the AtC fallacy.

Aristotle (\approx 350 BCE, Sect. I.5) defined the fallacy of 'Stating as cause what is not the cause' as occurring 'whenever what is not a cause is inserted in the argument'. It is now commonly named the 'Non-Cause as Cause', which I will further call the NCaC fallacy for short. According to Woods (1999) Aristotle gave a different interpretation of this fallacy in his *Rhetoric*. I have adopted the *Sophistical Refutations* interpretation²⁴ here, which was nicely exemplified by Woods (1999, p. 211): 'Suppose that: $P; Q; \rightarrow not-T$, is a refutation of the thesis T. Then the argument in question is a syllogism, hence a valid argument. As any reader of modern logic knows, if the argument at hand is valid, so too is the second argument: $R; P; Q; \rightarrow not-T$, no matter what premiss R expresses. But it is not a syllogism since a proper subset of its premisses, namely, $\{P, Q\}$ also entails its conclusion. Hence our second argument cannot be a refutation. This matters in the following way. Aristotle thinks of the premisses of a refutation as reasons for ('cause of') its conclusion. But since our second argument is *not* a refutation of T, R cannot be a reason for not-T.' In other words, if we would assume that the argumentation above leading to the conclusion C ('therefore, it rained') is valid, we could add any premise A1 like, for example, 'if the wind is west it rains' or 'if the wind is west it does not rain'. Any conclusion concerning wind direction in connection with raining or any premise A1 based on an argumentation extended in this way would then be logically invalid because of being the outcome of an NCaC fallacy. It is nowadays often denoted as the Fallacy of False Cause but, as the 'superfluous' premise is not necessarily wrong as such, I prefer Aristotle's naming.

Aristotle (\approx 350 BCE, Sect. I.5) described 'begging the question' (BtQ fallacy for short) as follows: 'Those (argumentations) that depend on the assumption of the original point to be proved, occur in the same way, and in as many ways, as it is possible to beg the original point'. Woods (1999 p. 212) exemplified that 'it is a flat-out violation of the definition of "syllogism". If what is to be proved is also assumed as a premiss, then that premiss is repeated as the conclusion, and the argument in question fails to be a syllogism'. He also discussed Aristotle's interpretations of BtQ in other works but I adopt the interpretation as quoted above²⁵. Circular reasoning might be considered as a version of this fallacy. BtQ is closely related to NCaC, as for 'those that depend upon the assumption of the original point and upon stating as the cause what is not the cause ... the conclusion ought to come about "because these things are so", and this does not happen where the premisses are not causes of it: and again it should come about without taking into account the original point, and this is not the case with those arguments which depend upon begging the original point' (Aristotle, Sect. I.6). As in the literature of applied sciences the argumentations that substantiate the conclusions may sometimes be fragmentary and/or partly implicit it is not always possible to classify a fallacious argumentation under NCaC or BtQ for certain.

One of many other fallacies defined after Aristotle's age is the 'Straw man' fallacy. It is an argumentation in which, for example in a political debate, a misrepresentation of the opponent's position is refuted. I was not able to retrieve the author who introduced it into scientific literature but the term was already well-known by Friedman (1953) who convincingly denounced and refuted the straw man in publications of critics of Marshall's *Principles of economics*. Here I conceive it more generally, including the part of an argumentation in a scientific article in which, without proper explanation, an interpretation of a concept is used that differs from how most readers might understand it. It is a manifestation of Aristotle's *Ignoratio elenchi* fallacy, 'which results from violating any of the conditions on what constitutes a proper refutation' (Woods, 1999 p. 209).

²⁴ According to Woods (1999) Aristotle gave a different interpretation of this fallacy in his *Rhetoric*.

²⁵ I will use this designation though its meaning in modern English is not self-evident, a phrase like 'assuming the conclusion' might be more appropriate (e.g. Liberman, Begging the question, languagelog.ldc.upenn.edu/nll/?p=2290, accessed May 2012).

Inferences that are based on any of these fallacies are not logically valid. This implies that the advanced argumentation does not offer the scientific underpinning that is suggested. In scientific papers such fallacies should thus be avoided or, if published, revealed, because they are misleading.

A4-3 Recall and discussion of the proposed bounded-rational mental choice process

ZT departed from the assumption that ‘In case of conjoint experiments ... respondents will try to figure out what the researcher is expecting and what the experiment is trying to accomplish, and dependent on these contextual attributes will construct a mental representation of the decision problem/experimental task’ (p. 766). They also assumed that respondents develop this script-type mental representation deliberately to reduce their mental effort while tailoring them to the stated choice task at hand. Giving a broad outline, they further assumed that the mental representation consists of an attribute-level-range elimination phase followed by utility assessment rules for the remaining attribute levels and satisficing rules for rejection/acceptance of the stated choice alternatives. Successively, this script is assumed to be applied consistently in two stages to each of the stated choice questions. In the first stage they choose whether to stick to the ‘status quo’ or reject it in favour of one or more alternatives. Only in the latter case they enter the second stage, to select a particular alternative: ‘respondents need to judge whether the experimentally varied alternatives may be satisfactory compared to their current alternative ... If new alternatives pass this screening test, we assume that the respondent compares the new alternatives in the second stage’ (p. 766).

ZT did not discuss the mental effort required for the development of mental representations that are tailor-made to arrive at a compromise between the respondent’s and researcher’s interests. They did not consider that their mental-representation-construction process might result in other mental models such as compensatory utility maximization or strong-lexicographic alternative comparisons. Also, they did not discuss alternative choice-process assumptions (see e.g. Van de Kaa, 2010b, for an anthology). Just as an illustration, my personal memory of stated choice surveys is that I took the researcher’s suggested choice context and his/her explanation of the stated choice game for granted and, after reading them, immediately started answering the choice questions, in the submitted order, by marking without conscious calculation or deliberation the alternative that seemed the best. The behavioural and naturalistic decision-making literature suggests that such a largely unconscious choice process might have been ‘recognition-primed’ (e.g. Klein, 1993), by retrieving a convenient mental ‘choice-behaviour-strategy’ script from an ‘adaptive toolbox’ (e.g. Payne et al., 1993) that is stored in memory as a consequence of accumulated real-life experiences with similar choices. A retrieved script might imply unconsciously performing a full compensatory utility maximization of all submitted attributes and/or of just those that are felt as most relevant, or applying any conceivable non-compensatory choice process like, for example, the algorithms proposed by ZT. Such a process would presumably be cognitively less demanding than the ‘tailor-made-construction-on-the-spot’ of that latter mental representation. Anyhow, without further empirical evidence there seems to be no reason for rejecting the choice concept of ZT, whether it is interpreted as an isomorphic or a paramorphic choice model, nor for preferring it over any of the earlier proposed models.

A4-4 Mathematical implementation of the choice process and empirical research

ZT implemented their choice process assumptions in a set of sophisticated mathematical models. Giving a broad outline, the overall-utility function of alternatives was specified as the sum of their attribute utilities which, in turn, were specified as step-functions of the attribute levels with attribute-threshold levels separating the intervals. A first-stage model simulated an indifference band around an idiosyncratic threshold-level in the overall-utility of each alternative in the choice set, which separated combinations of its attribute-level ranges that did merit renouncing the current alternative from those combinations that did not. If all alternatives of a choice set were rejected the current alternative was considered to be chosen in the first-stage process and the choice set was not considered in the second-stage model. If all but one of the alternatives were rejected, the current alternative was considered as rejected and the non-rejected alternative as chosen in the first-stage process. Only if two or more alternatives in a choice set passed the first stage were their overall utilities compared in the second-stage model.

A4-4.1 Mathematical implementation

Dedicated software was developed for parameter-estimation, including an incremental process of selection of only those attribute-threshold levels for which the corresponding parameters improved the model-fit. ZT also presented probabilities of preference structures among the respondents and/or the corresponding heuristics as should have been applied in the first and second stages of the choice process. The description of the software employed for the parameter-estimation process was too frugal to figure out whether or not respondent-specific preference structures were imposed to be consistently applied during successive choices. Individual-specific preference structures might also have been assessed afterwards, by combining the estimated model parameters and corresponding thresholds that, if applied consistently, explained the individual's choice sequence. In both cases the preference structures for which ZT presented probabilities of occurrence should be considered as invariant during the stated choice sequences of each individual. It implies that the probabilities of occurrence, whether estimated or statistically determined, are equal to the frequency of their employment over the survey population.

A4-4.2 The stated mode choice experiment

The empirical data for which the parameters were estimated were from a stated choice experiment, held in 2002 among citizens of Eindhoven, The Netherlands. The respondents had to choose one out of three trips, two hypothetical trips with a novel 'TURE' public transport system and a corresponding trip with their 'current transport mode'. Each choice had to be made for four different trip motives (work/study; daily shopping; non-daily shopping; and social/leisure activities). The different TURE trips were characterized by the levels of 8 attributes ((i) access time; (ii) egress time; (iii) waiting time; (iv) vehicle speed; (v) price; (vi) shelter-availability at stop; (vii) platform height; and (viii) number of seats). Four levels were submitted for attributes (i) ... (iii) and (v) and two for the other attributes. An orthogonal fraction of 32 TURE profiles was selected from the huge number that follows from a full factorial design. The degree to which 'extreme' combinations of time and cost attributes were submitted was not reported. The respondents were divided into two approximately equal groups, to which the TURE-attribute levels were presented in different sequences. Sequence 1 (hereafter S1) ran from (i) to (viii) and Sequence 2 (S2) went from (v) to (viii) followed by (i) to (iv).

The age, gender and education statistics of the 356 respondents were considered as largely consistent with those of the town's concurrent population. The current mode was chosen from 69% of the choice sets. The respondents were apparently not asked to report their 'current transport mode' (walking, train, car driver and so on) and the corresponding 'current' attribute levels for each motive. The utility of the current mode could thus not be assessed as a function of 'independent' attribute levels. Instead, 'it is suitably modeled as a probabilistic distribution λ ' (p. 773) which roughly implies that it is equated with an overall-threshold of TURE, above which the overall-utility of TURE gives rise to rejection of the current transport mode.

A4-4.3 Goodness-of-fit

ZT did not discuss the goodness-of-fit of their models, for example in terms of Rho-squared values. They only presented the loglikelihoods and Consistent-Akaike-information-criteria (CAIC) for their models. In addition to the implementations of their choice concept 'A conventional multinomial logit model (MNL) is also estimated for comparison' (ZT p. 765). This contained a linear-additive systematic utility specification of the TURE attribute levels and only motive-specific constants for the current-mode alternatives. This MNL-model was a far cry from the 'conventional MNL model' in travel choice research which, to the best of my knowledge, is a version of random utility maximization (RUM) in which the systematic utilities of all the considered alternatives are specified in the same way. This prevents a *ceteris paribus* comparison of the goodness-of-fit of these models and the inferences that rely on it: they concern two different behavioural choice theories embedded in two different regression models. This is correctly mentioned by ZT (p. 778) but in the same breath they compare the goodness of fit of both models: 'Direct comparison between the MNL models and the proposed models is difficult, if not impossible, since their theoretical assumptions are different. In terms of goodness-of-fit, the combined log-likelihoods of the proposed

model under both sequences are ... better than ... in MNL models ... (and) both CAICs are better²⁶ than those in MNL models'. Though this comparison is phrased in literally correct words I nevertheless consider it as a straw man fallacy because for all but the most suspicious readers it seems to suggest that ZT's proposed models outperformed the corresponding RUM implementations.

A4-4.4 Mode-use, trip-length and trip expenses in 2002 in The Netherlands

To grasp an idea of the degree to which their bounded-rational models would, for example, outperform a conventional RUM-model when taking revealed current-mode attribute levels into account I used 2002-statistics for transport-mode-use, trip-length and expenses for the Dutch population (RWS, 2003) and applied them to the Eindhoven transport system, to approximate the respondents' current-mode attribute levels. In Figure 3 some key data are presented and compared with those of the TURE-alternatives as submitted in the stated choice experiment²⁷. It appears that the time and money expenses were far from identically distributed over the different trip motives, which might be considered as a prerequisite for a meaningful substitution by a constant. This suggests that, if a conventional RUM model would have been applied to the submitted attribute levels of TURE and the corresponding levels of the actual travel modes, its goodness-of-fit would have outperformed ZT's MNL model and, even more, their bounded rational model.

A4-5 A critical re-examination of the inferences about the mental choice process

Apparently, ZT deemed the fit of their models with the TURE-survey data fair enough to arrive at strong conclusions about the mental choice process. These conclusions concerned the influences of the sequence of information presentation, the nature of the decision problem and the decision context on the decision heuristics that the respondents applied. I will re-examine the argumentations that supported their conclusions and consider the extent to which they are based on AtC fallacies.

A4-5.1 The sequence of information presentation

In their conclusions ZT (p. 779) stated that 'estimating the models under different attribute presentation sequences showed that the influence of information presentation is substantial on attribute selection and processing sequence.' Successively they summarize the empirical inferences leading to this conclusion. These inferences are re-examined first.

The first underlying inference was that 'not all attributes provided in the experiment are influential for the decision, suggesting that some attributes were completely ignored' (p.779). This inference could follow from a reasoning such as:

D: If an attribute has no influence on the decision, adding its estimated parameter to our model does not significantly improve the goodness-of-fit.

E: Adding the estimated parameters of some attributes provided in the experiment to our model did not significantly improve the goodness-of-fit.

F: Therefore, not all attributes provided in the experiment are influential for the decision.

The vehicle-speed, shelter-availability, platform-height and seat-number attribute levels did indeed fail to yield significant parameters in any of ZT's first-stage, second-stage and MNL models. This concerned both sequences. However, the fact that the models did not yield significant parameters for these attributes neither proves nor rules out that these attributes are influential for the decision. Respondents might have attached a value to them that might perhaps have yielded significant parameters in a conventional RUM model if, in agreement with the state-of-practice in stated travel

²⁶ For the CAIC calculation the estimated attribute-threshold levels were 'not counted as free parameters because only their weights are effective' (ZT p. 774). These levels in connection with their weights define the set of considered stepwise utility/value functions and/or preference structures. Therefore, I did not understand this reasoning. If these had been accounted for as free parameters, the CAIC estimates for their combined first-and second-stage models would have been much higher though still lower than for their MNL model.

²⁷ The Dutch 2002-statistics concerned all ages. Zhu and Timmermans surveyed adults only. In Figure 1 I disregarded study trips because by far most of these are made by minors. For the same reason I disregarded the car passenger mode.

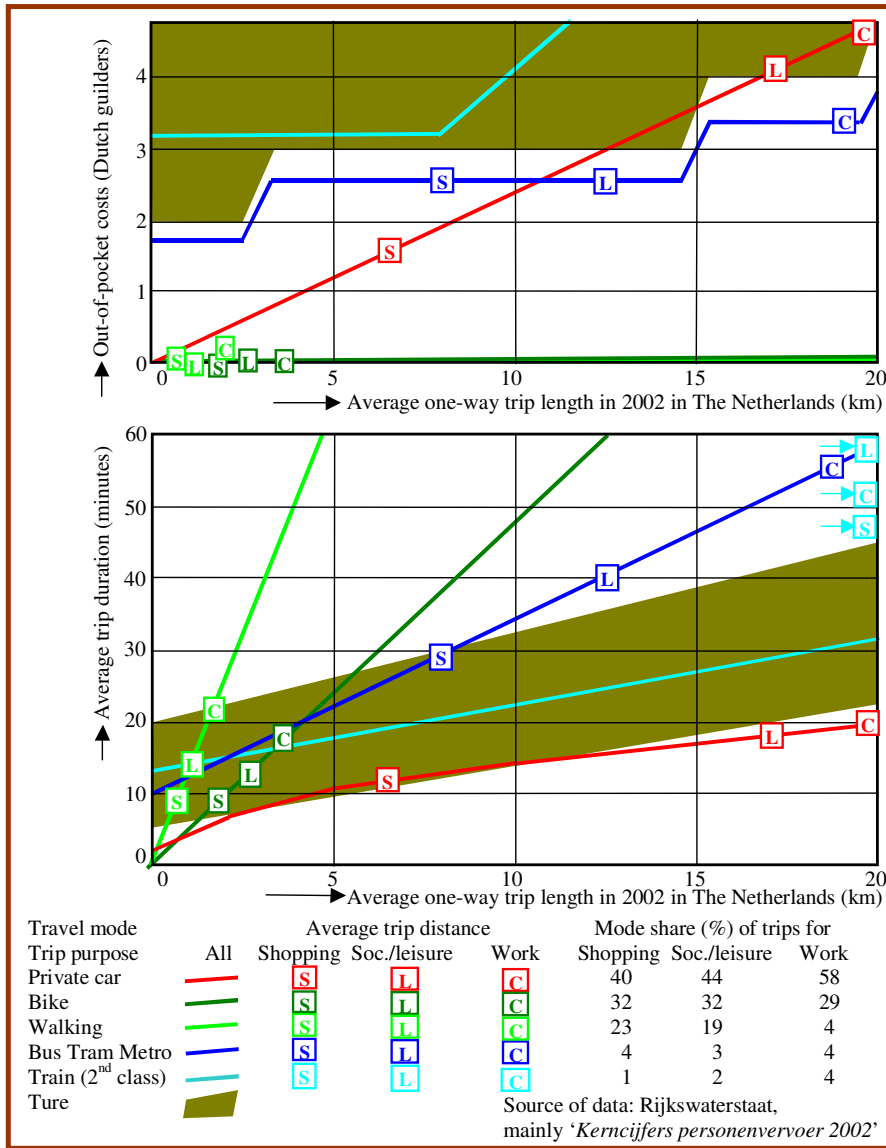


Figure 3 Key statistics for the Dutch 2002-trip characteristics, travel mode and purpose

choice research in those days (e.g. Burge et al., 2004), the respondents' current-mode attribute levels had been elicited and compared with the TURE attribute levels. As the antecedent in premise D is not the **only** sufficient condition for consequent E to occur, conclusion F is based on an AtC fallacy. Moreover, even if these attributes were not influential for the decision, this does not necessarily imply that the respondents did not consider them. Compared to the 'utilitarian' time and money attributes they might just as well have attached a small value to them, after full deliberate consideration or intuitively. The suggestion that they were completely ignored might thus have been caused by another AtC fallacy. Anyhow, whether or not some TURE attributes were not influential and/or ignored, their lack of significance was not influenced by the sequence in which they were presented and thus does not support the overall-conclusion above.

ZT (p.779) successively stated that 'the influence of the presentation sequence on the first-stage decision is about the selection of attributes. Because saving effort is the major consideration in heuristic choice, it is much easier for the respondent to passively use the attributes following the provided sequence rather than to actively locate the desired attributes. As a result, price was included in the decision when it was presented first, while it was excluded when it was presented last.' The

empirical evidence that they provided is that the first-stage model for S1 yielded relevant thresholds for access time (2 levels), egress time and waiting time while for S2 it found one threshold for each of these attributes and another one for price. As in S2 price was presented at the top of the attribute-levels list while in S1 it appeared in fifth place, ‘this suggests that information presentation format influenced the respondents’ decision strategies’ (p. 773). The following argumentation might lead to such a finding:

G: If the information presentation format influences the respondents’ decision strategies such that they consider the attributes in the provided sequence, the chance that an attribute’s estimated parameter improves the goodness-of-fit of our first-stage model significantly increases if it is provided earlier.

H: The estimated parameter for the price attribute improved the goodness-of-fit of our first-stage model significantly when it was presented first and not so when it was in fifth place.

I: Therefore, this suggests that the information presentation format influenced the respondents’ decision strategies such that they considered the attributes in the presented sequence.

Differences in the distributions of current-mode-users and/or value-of-travel-time-savings (VTTS) between the S1 and S2 subpopulations (both 178 respondents) might offer an alternative explanation: I figured out that if the respondents would have applied a compensatory-utility-maximizing appraisal, the majority of car drivers by far and many cyclists and pedestrians would have found out that all their TURE alternatives were dominated by their current mode (see Figure 3 and the next subsection). Together and averaged over the trip motives they made up the majority of respondents. For many bus passengers, however, the current mode would be dominated by all TURE alternatives.. It implies that, for example, a few more bus passengers in the S2- subpopulation might have caused the difference in significance of the price attribute. As this and other alternative conditions that might also be sufficient for the consequent H to occur²⁸ were not ruled out in the argumentation as summarized above this was an AtC fallacy.

Conclusion I is also weakened rather than supported by other circumstantial information. For example, if the provided attribute sequence had an important effect on the mental choice process one might expect additional, unambiguous evidence for it from the second-stage and MNL models but these all yielded significant parameters for access and egress time (both 1 or 2 levels), waiting time (2 or 3 levels) **and** price (always 2 levels). In the second-stage models the price parameters for S2 were even less negative than those for S1, which might have counterbalanced their higher significance in the first-stage models to some extent. Also, both the first- and second-stage models yielded very similar probability distributions of the preference structures (see below) for both S1 and S2. Anyhow, neither its foundation on an AtC fallacy nor its lack of circumstantial support means that the conclusion about the suggested influence of the information presentation format, which was formulated with some constraint, is wrong, but its substantiation is invalid.

One should consider that ZT (p. 779) included several causal relationships in their conclusion about the influence of attribute presentation sequences, like ‘Because saving effort is the major consideration in heuristic choice, it is much easier for the respondent to passively use the attributes following the provided sequence rather than to actively locate the desired attributes.’ As it is part of ZT’s empirics-based conclusions one might conceive it as an additional premise that they considered as affirmed by consequent H. As the antecedent formulated in premise G is a sufficient condition for consequent H to occur these additional premises might be substituted by any other assumption about the mental choice process without changing conclusion I. This would classify the effort-saving attribute-following conclusion as based on the NCaC fallacy. Another interpretation is that reducing mental effort was a core premise in ZT’s heuristic choice concept (see Sect. A4-3), and that they assumed beyond reasonable doubt that many respondents would therefore pay more attention to attributes that were presented earlier in the questionnaire. If they

²⁸ I could not rule out another possible explanation. That would arise if ZT adopted different attribute sequences in the parameter estimation process of S1 and S2. They did not report whether they did so or followed, for example, the same sequence when estimating the S1 and S2 models. Their thumb-nail sketch of the incremental parameter estimation software did not reveal if the sequence in which the significance of parameters was tested might influence their significance and selection.

therefore reworded these assumptions in their conclusions without further proof this would classify as a BtQ fallacy. In any case the argumentations above leading to these conclusions would not be valid. I was not able to retrieve another explicit or implicit reasoning in ZT that yielded a logically sound substantiation for these conclusions.

ZT (p.779) further advanced that ‘the influence of presentation sequence on the second stage decision is mainly about attribute processing sequence. Because the respondents preferred extensive information search in general, the following problem is to have an optimal sequence in order to save effort. If the presented sequence is consistent with the desired sequence in which more important attributes are processed earlier, it will be ideal; but if the presented sequence shuffles the desired sequence, it will cost the respondent more effort to reorganize the desired sequence or follow the provided sequence starting from minor attributes. In the latter case, the respondent could have expected more effort and gave more weight on effort in heuristic choice, resulting in more random choices.’ Considering this sophisticated argumentation one would consider that it is advanced to explain the assumed influence of the presentation sequence given the empirical finding that ‘the shape of the distribution under S2 is very similar to that under S1. The only apparent difference is that the probability of pure random choice is much lower’ (p. 777). Implicitly, ZT might have assumed that second-stage choosers consider the price attribute as most important and prefer sequence S2 (price at the top of the attribute list) over S1 (price in fifth place). ZT’s inference could be the result of an argumentation like:

- J: If the price attribute, which is of major importance, appears earlier in the questionnaire, less respondents will (try to save the effort required for adopting an appropriate attribute processing sequence etc., and consequently) apply pure random choice and our second-stage models will estimate a much lower probability of pure random choice than if the less important attributes appeared earlier.
- K: Our second-stage models estimate a much lower probability of pure random choice in S2, where attributes of major importance appeared earlier in the questionnaire, than in S1.
- L: Therefore, there is a presentation-sequence influence on the second stage decision (driven by respondents’ propensity to save mental effort etc.) that is mainly about attribute processing sequence.

This is a very simplified representation of ZT’s argumentation in which, in addition to those between brackets, several more assumptions about the mental choice process might have been considered to be substantiated. As ZT apparently considered the difference in sequence of less and more important attributes as a sufficient condition for consequent K to occur, the additional assumptions (indicated in L between brackets) are non-causes for it. The inclusion of these mental phenomena in the argumentation and its conclusion would then have been logically invalid, being based on the NCaC fallacy. If ZT assumed these mental phenomena to be true beyond reasonable doubt without further consideration in the argumentation, then their inclusion in the conclusions without further substantiation would classify as a BtQ fallacy. In any case, the argumentations leading to these conclusions would not be logically valid.

After disregarding the ‘between brackets’ parts, the argumentation above is still embedded in the, as yet not empirically substantiated, two-stage choice process, in which the current alternative is either chosen or rejected in the first stage. While, for example, the access time to the TURE stop was by far the most important attribute in ZT’s first-stage models, both the waiting-time and price attributes appeared to be more important in their second-stage models. If ZT had assumed that respondents kept their first-stage importance order in the second stage, consequent K would violate premise J instead of following it. In my opinion it would have been appropriate if ZT had paid some attention in their conclusions to this conditional ‘shift of preference’ between respondents’ first- and second-stage choices.

Looking for other conditions that might be sufficient for consequent K to occur one might consider that ZT assumed that random choice was applied when both TURE alternatives had only a small rank difference within the considered preference structure. This is quite different from the random choice rule of behavioural decision theory. Within ZT’s mental choice concept one might expect that some respondents would apply the latter random choice rule, as this would have required the least mental effort. As their first-stage models could not elicit this heuristic its probability could only be estimated

with their second-stage models, in which only 31% of all choice decisions were considered. The estimated second-stage random choice probabilities were 14% for S1 and 6% for S2. The elicited random choices might then have been made by just 8 out of all 178 S1-respondents and 4 of the 178 S2 respondents. It implies that the alleged sequence-effect would not have been found if, by chance, the number of random choosers in both S1 and S2 had been 6. If a conventional one-stage compensatory-maximizing choice behaviour was assumed it goes without saying that such small differences might also have been caused by, for example, coincidental differences in the current-mode-attribute levels between the S1 and S2 respondents. It is thus easy to see that the argumentation above is another AtC fallacy: If respondents applied their two-stage mental choice process and if they did so for the reasons listed on p. 779 and if this was modelled realistically this is a sufficient condition to explain the model results but there is no evidence that it is the **only** sufficient condition.

The inferences about the ignorance of some attributes, the first-stage selection of attributes, saving effort, the passive following of the provided sequence rather than actively locating the desired attributes, etc, were thus apparently based on AtC and NCaC and/or BtQ fallacies. ZT provided no solid empirical evidence for their conclusion (p. 779) that the influence of information presentation on attribute selection is substantial.

A4-5.2 The decision context

The trip motive was the sole characterization of the choice context that was considered by ZT (p. 779) who concluded that ‘the general trend is that the more value of the travel motive, the less strict will the decisions heuristic be, so TURE will be more probable to be accepted The sequence reverses for the rejection probabilities.’ The underlying empirical evidence was that all four motive-specific parameters estimated for the first-stage models appeared negative and significant, ‘meaning that the value of the current transport mode was higher than that of TURE for the sample of respondents in general’ (p. 773). These parameters varied going from the social/leisure motive ($\beta \approx -0.4$) over non-daily shopping ($\beta \approx -0.8$) and work/study ($\beta \approx -1.8$) to daily shopping ($\beta \approx -2.5$). Accordingly, the probability that the current alternative was chosen increased and the probability that TURE was chosen decreased in the same sequence. Also the (positive) motive-specific parameters of the current alternative that were estimated with the MNL model increased in the same sequence. In the second-stage models the motive-specific parameters appeared not significant.

To explain the negative parameters ZT (p. 773) assumed that ‘respondents were conservative about the new transit system’ and next proposed two explanations for the differences in it between the motives. One was that the assumed conservatism might decrease with the importance of the activity: ‘among the motive parameters, the one for daily shopping is the lowest. This is probably because daily shopping is the least important activity’. The other explanation is that people might not consider to switch to a novel transport mode due to strong habits as a consequence of the frequency of activities: ‘daily shopping is a very established activity ... The acceptance of TURE is higher for work (and) even higher for non-daily shopping and leisure, which suggests that TURE is more attractive for less frequent activities.’ This would follow from an argumentation in the vein of:

- M: If respondents are conservative about a new transport system the parameters for the trip motive, as estimated with our first-stage models, are negative and become more negative when the importance of the concerned activity decreases and/or its frequency increases.
- N: The parameters as estimated with our first-stage models were most negative for daily shopping, which is the least important and most frequent activity, and those estimated for work, non-daily shopping and leisure, which are less frequent, are increasingly less negative.
- O: Therefore, respondents are conservative about a new transport system and become more conservative when the importance of the concerned activity decreases and/or its frequency increases.

Without further substantiation one might wonder why getting one’s daily bread would be less important than buying fancy goods in the next big city or going to the cinema. One might also consider that in the Netherlands the shares of the work/study and social/leisure motives in the total number of trips (both app. 35%) were higher than for both shopping motives together (30%), let alone for the daily-shopping

motive alone. Both explanations for the motive-dependence of a conservative attitude to TURE are thus not very convincing.

An alternative explanation for consequent N might be that most respondents' trip modes as well as lengths, and their associated cost and duration, might be very different for different motives. Assuming that the distribution of the current trip characteristics of the survey population was similar to that of the Dutch population, the order-of-magnitude of the outcomes of a motive- and mode-specific total-travel-time versus cost trade-off in agreement with utility theory can be estimated. In doing so one should consider that in The Netherlands about 40% of the cost of all commuting were completely reimbursed by the employer and another 40% partially, while in general most social/leisure and shopping trips were not²⁹. Figure 3 shows that for all motives TURE might have offered interesting alternatives for current public transport users but that its share was only between 5 and 8%. Commuting and social/leisure trips by car were on average faster than the fastest TURE trip over the same distance, and the running costs were about the same as the cheapest TURE trip. The current mode would then have been the dominant alternative for almost all commutes by car drivers and for the greatest number of social/leisure trips by car. Shopping trips by car were on average slightly slower than the fastest TURE trip but the running costs, which in general were at the shopper's expense, were far below the cheapest TURE alternative, thus for most shopping trips sticking to the car would have been the dominant alternative even without considering its convenience in transporting the acquired goods. Figure 3 also shows that for cycling and walking trips of increasing duration TURE alternatives became more attractive. This particularly held for commutes, even more so if taking into account the frequency of cost reimbursement. One might thus expect that many cyclists and pedestrians would have preferred at least one TURE alternative in all choice sets over their current commute mode. For social/leisure trips the maximum time gain for a shift from a slow mode to TURE would have been smaller and the costs not reimbursed, thus only a minority would have preferred the fastest TURE alternative. For the average slow-mode shopping trips the time gain of TURE was smaller still and thus also the propensity to switch to it. Taking into account the current mode shares for each motive, the propensity to stick to the current mode would then have been highest for shopping, lowest for social/leisure and in-between for commuting.

The 2002 Dutch statistics did not discern between daily and non-daily shopping but the average shopping trip length for all modes was 4.2 km from Monday till Friday, 5.6 km on Saturdays and 12 km on Sundays (RWS, 2003). Considering that a significant part of the weekend trips concerned daily shopping the average length of the respondents' non-daily shopping trips might have been almost twice as much as for daily shopping. This suggests that most shopping trips by public transport concern non-daily purchases and that the share of slow-mode use for non-daily purchases is much lower than for daily shopping. In combination the longer length of non-daily shopping trips suggests a much higher overall-propensity to switch to TURE for non-daily shopping compared to daily shopping. A conventional compensatory-utility-maximizing choice process would thus presumably explain the same increasing propensity to choose TURE in the sequence from social/leisure over non-daily shopping and commuting to daily shopping.

The argumentation above leading to conclusion O, that people were conservative about the new transit system and became less conservative when the value of the travel motive increased, is thus an AtC fallacy: such attitudes are indeed sufficient to explain the differences in respondents' choices between the motives but as demonstrated above a common utility-maximizing appraisal, for example, could explain the same trend. No solid empirical evidence was provided for this and ZT's (p. 779) conclusion that 'the contextual effect of heuristic choice was ... significant for the first stage. The general trend is that the more value of the travel motive, the less strict will the decisions heuristic be'.

²⁹ The occurrence of travellers' cost reimbursement was retrieved from the data files of the Dutch 1997 VTTS survey (HCG 1998).

A4-5.3 The nature of the decision problem

The argumentations in ZT (p. 779) with respect to the nature of the decision problem were based on the probabilities of respondents' deployment of different preference structures, that are specific combinations of attribute thresholds, in the assumed two-stage choice process of 'comparing TURE alternatives with their current transport mode, and then comparing the alternatives if both of them are better than the current mode.' The main conclusion was that 'respondents seem to have applied extremely simple decision heuristics in the first stage ... At this stage, the respondent's major purpose is to reduce mental effort ... in the second decision stage ... heuristics implying extensive information search were applied more frequently.'

The preference structures with the highest probability as estimated for the first-stage choice model varied from 45% for social/leisure, 52% for non-daily shopping and 67% for work/study to 75% for daily shopping. Depending on the way in which they were assessed (see Sect. A 4-4) these percentages are identical to or model estimates of the percentages of the respondents that chose the current mode from all choice sets. The overall-average for all motives was 60%. Obviously, this is well below the 69 % of choices for the current alternative reported from the raw stated choice data. ZT (p. 774) concluded that this finding 'means the respondent's overall judgment was so strict that both TURE profiles would be rejected without consulting any information'. This and its contribution to the overall conclusion above might have been inferred from an argumentation like:

- P: A respondent seems to have applied an extremely simple decision heuristic if his overall judgment was so strict that both TURE profiles were rejected without consulting any information.
- Q: If a respondent's overall judgment is so strict that both TURE profiles are rejected (without consulting any information) he will choose the current alternative from all the submitted choice sets.
- R: Depending on trip motive, between 45% and 75% of all respondents chose the current alternative systematically from all the submitted choice sets.
- S: Therefore, many respondents (seem to have applied an extremely simple decision heuristic because they) had an overall judgment that was so strict that both TURE profiles were rejected (without consulting any information).

As judgment without consulting any information suggests a random choice, ZT apparently assumed that these respondents exhibited some strong inertia or status-quo preference. If that is why 75% of the respondents did not consider the TURE attributes for daily shopping trips one might wonder why many of them did so for the other motives. Anyhow, it is easy to see that the antecedent in Q is a sufficient condition for R to occur. Premise P in itself seems reasonable, particularly as it is formulated with some constraint. However, an alternative premise P₁ like 'A respondent seems to have applied an extremely complex choice behaviour strategy if his overall judgment was so strict that both TURE profiles were rejected after meticulously searching, consulting and processing all available information' seems just as reasonable. After adapting the 'between brackets' parts of premise Q accordingly the same argumentation would yield an alternative 'between brackets' part of conclusion S that is almost contrary to the one above. If the original 'between brackets' part of conclusion S is conceived as inferred from empirical findings it is thus based on the NCaC fallacy. Otherwise it may be considered as a BtQ fallacy. It is not supported by empirical evidence anyway.

After removing premise P and the between-brackets phrases in Q and S an argumentation remains that would be valid if the antecedent in Q was the only sufficient condition for R to occur. ZT did not demonstrate this. Looking for alternative explanations for the shares of respondents who systematically chose the current mode I considered that the respondents might have applied some straightforward compensatory-utility-maximizing trip-duration versus cost trade-off as discussed in the previous subsection. Using the motive-specific Dutch population statistics as approximation of their current-mode characteristics (see Figure 3) I estimated the percentages of the respondents for which that current mode might have dominated all TURE alternatives. This yielded roughly 75% for all kinds of shopping trips, 70% for commuting and 55% for social/leisure. Obviously, the distribution of current-mode characteristics over the survey population differed in many respects to the Dutch average. As, for example, Eindhoven is the seat of several institutes of higher education it has a relatively large population of students who either during the week and/or at weekends were entitled to free public transport. Taking this into consideration and adding study to the commute motive would reduce all estimated percentages by about 5%. One should consider that almost any time-cost trade-off concept

would yield similar percentages. The observed sequences of systematic choice for the current alternative might just as well have been the result of a first-stage selection of the 'best' time-for-money TURE alternative followed by its rejection in a second stage in favour of the current mode, to mention just another bounded-rational choice strategy. The same holds for a reference-dependent, loss-averse value maximizing as well as a loss-neutral, linear-additive utility concept, whether applied simultaneously to all alternatives or sequentially in two stages, the latter by first selecting the best TURE alternative followed by comparing it with the current alternative as well as vice versa. The percentages assessed according to all such choice concepts would become remarkably close to those found by ZT. As other mental choice concepts than the one assumed by ZT might cause consequent R to occur the reduced argumentation above is an AtC fallacy. As such it does not substantiate the statement that many respondents had an overall judgment that was so strict that both TURE profiles were rejected nor that they seem to have applied an extremely simple decision heuristic without consulting any information.

According to ZT the preference structure with the second highest estimated probability (on average 17%) implied 'that the respondent accepted both TURE profiles regardless of any information' (ZT p. 774). The probability decreased from 30% for social/leisure over 23% for non-daily shopping and 11% for work/study to 5% for daily shopping. These percentages apparently concerned the respondents who, for the concerned motives, chose a TURE alternative from all the submitted choice sets. The thread of the whole article suggests that ZT might implicitly have applied an argumentation like:

- T: If in the first decision stage, for a particular trip motive, a respondent rejects the current alternative in each choice set, (we assume that) he exhibits 'the most relaxed preference structure, representing the case that all overall values will be accepted' (ZT p. 768).
- U: If in the first decision stage, for a particular trip motive, a respondent rejects the current alternative in each choice set he accepts both TURE profiles regardless of any information and directly enters the second decision stage.
- V: Depending on trip motive, between 5% and 30% of all respondents rejected the current alternative in each choice set.
- W: Therefore, there is a 5 - 30% probability, depending on trip motive, that the respondent exhibited the most relaxed preference structure by accepting both TURE profiles regardless of any information and directly entering the second decision stage.

It is obvious that the conclusion that part of the respondents exhibited the most relaxed preference structure is based on an NCaC fallacy. However, even disregarding antecedent T and its corresponding conclusion of a relaxed preference structure the argumentation is based on a genuine logical error: it is very possible that the concerned respondent compared the current alternative successively with each TURE alternative in the first stage, either superficially or in-depth, and rejected the current alternative from each choice set in favour of at least one, rather than both, of the TURE alternatives. The argumentation above is therefore an 'arguing in a circle' version of the BtQ fallacy. Not knowing the submitted choice sets, I was unable to calculate the corresponding percentages of current-alternative rejection that would follow from a compensatory-utility-maximizing trade-off of the time and money attributes by assuming that those for the current mode were representative of the Dutch population. However, from Figure 3 it is easy to see that these would exhibit a similar motive-dependency. Even if the antecedent in U had been sufficient for V to occur – which it was not, see above – other conditions might also be sufficient, thus in the argumentation above an AtC fallacy was embedded in the BtQ fallacy. What remains is that there is a 5-30% probability that the respondent chose at least one TURE alternative from each choice set in the first decision stage if in that stage he either chose or rejected the current alternative. This does not substantiate ZT's overall-conclusion that 'respondents seem to have applied extremely simple decision heuristics in the first stage ... At this stage, the respondent's major purpose is to reduce mental effort rather than to avoid risk' (p. 779).

The other 10 to 14 first-stage decision heuristics were elicited when the same respondent chose the current alternative from one or more stated choice sets and one or both TURE alternative from the other sets. This completed the 69% first-stage choices for the current mode as reported above and left 31% of

the choices for the second-stage models. It appeared that the number of relevant attribute thresholds was higher than that found for the first-stage models, which defines a much higher number of 36 to 54 preference structures. For the probability of these preference structures ZT (p. 776) found that ‘the concentration is around the discriminant thresholds that imply low-risk³⁰ heuristics ... Small discriminant thresholds were less favoured ... Large discriminant thresholds were not favoured, either’. Again taking the validity of their two-stage choice concept for granted ZT (p. 779) advanced these results in their conclusions as showing that ‘risk perception turned out to be the determinant factor of heuristic choice in the second decision stage. As a result, heuristics implying extensive information search were applied more frequently than simple strategies.’ This suggests an underlying argumentation such as:

X: If (we assume that) respondents apply a two-stage bounded-rational choice process drawing on attribute-threshold levels and have rejected the current alternative in the first stage, risk perception will incite most respondents to refrain in the second stage from selecting a decision heuristic with a high or low overall threshold, and to select a low-risk heuristic with moderate threshold levels, even though this would imply an extensive information search, because most people are risk-averse.

Y: Our second-stage model determined that low-risk heuristics were applied more frequently than simple strategies and that heuristics with small and large discriminant thresholds were less favoured.

Z: Therefore, risk perception turned out to be the determinant factor of heuristic choice in the second decision stage.

In their explanation of the high probability of low-risk heuristics ZT (p. 776) remarked: ‘That means information search was preferred in general by the respondents. In this sense, their decision processes approach the assumption of rational choice.’ These decision processes would indeed take the price levels and several time attributes of both TURE alternatives into consideration. As discussed extensively in the previous subsections, a one-stage compensatory-maximizing cost-time trade-off might explain similar choices from the considered second-stage choice sets. As any evidence that the respondents actually applied the concerned choice process is missing the argumentation above is another AtC fallacy. Obviously, the argumentation leading to the conclusions about the underlying mental processes might be classified as either NCaC or BtQ fallacies.

A4-5.4 Summary

ZT presented several conclusions about the mental processes that the participants in a stated mode-choice survey went through. They founded these conclusions on a comparison of their assumptions about those mental processes (overview in Sect. A4-3 above) and the empirical outcomes of mathematical models, in which their assumptions of these mental processes were implemented and the parameters were estimated for the outcomes of the concerned stated choice survey (overview in Sect. A4-4). In this section I have re-examined the argumentations leading to these conclusions. For expository reasons I have summarized these as ‘logical’ combinations of theoretical antecedent-consequent premises, empirical consequent, and conclusion, which by coincidence ran from A to Z. To safeguard my argument from the straw-man fallacy I followed ZT’s formulations closely. As it appeared that compensatory-utility-maximization and alternative choice concepts/models might also explain the empirical findings all these argumentations contained the AtC fallacy. The conclusions concerning the underlying mental processes also drew on the BtQ and/or NCaC fallacies. I was not able to retrieve an alternative explicit or implicit reasoning in ZT that yielded a logically sound substantiation for any of the fallacious conclusions that were discussed in this section.

³⁰ ZT (p. 769) assumed that ‘avoiding extreme situations seems to be a human nature. Assume that this rule also applies to respondents selecting a decision heuristic. By selecting a high or low overall threshold, the respondent will correspondingly have a high probability of rejecting or accepting an alternative (and thus may be liable to) the expected regret resulting from a potential false rejection or false acceptance.’ Applying a low-risk heuristic is thus similar to exhibiting a preference structure with a set of moderate attribute thresholds.

Although the contents of ZT's conclusions are not based on a logically valid substantiation and the available circumstantial information is not very supportive either they are not necessarily wrong. From a scientific point of view, however, they might be better considered as unsubstantiated personal opinions.

A4-6 Fallacious conclusions in other publications co-authored by Timmermans

During this research I encountered fallacious argumentations similar to those in ZT in five other transport research publications that were (co)authored by Zhu and/or Timmermans (Zhang *et al.*, 2004; Borgers *et al.*, 2007; Chen *et al.*, 2009; Zhu and Timmermans, 2010a;b). I will now give a short summary of my findings, starting with Zhu and Timmermans' papers.

In what was apparently the earliest³¹ submitted paper in a series of very similar papers Zhu and Timmermans (2010b) elaborated essentially the same strong assumptions about the mental choice process as in ZT. The model implementations differed only slightly from the second-stage model in ZT. The empirical studies concerned pedestrians' choice between going-home and continuing-shopping in a shopping street in Beijing, China. The utility of going-home was essentially modeled as a combination of step functions of two attributes, elapsed shopping time and the considered hour of the day. It was not considered as dependent on characteristics of the person and/or the environment like hour-of-the-day-dependent travel characteristics such as travel time or transit frequency. 'Additionally, two discrete choice models were estimated for comparison' (Zhu and Timmermans, 2010b, p. 68). These concerned MNL and mixed logit models. In both discrete choice models the systematic utility was specified as the sum of an alternative-specific constant and logarithmic functions of elapsed shopping time and hour-of-the-day. A utility-theoretical implementation would rather consider the time spent shopping as a utility and thus subject to the decreasing marginal utility principle, implying a concave-increasing function of shopping time, while postponing the homebound trip would hamper other activities, thus implying a convex-increasing disutility function of the hour-of-the-day. It is telling that even though a concave-increasing disutility function was enforced on the hour-of-the-day attribute the CAIC of the discrete choice models (MNL: 2197; Mixed logit: 2211) models approached those of the proposed model (2181³²) so closely. In addition to a RUM model with a convex function of the hour-of-the-day attribute several alternative discrete choice models, such as a schedule-delay or a reference-dependent home-going choice model, would in my opinion be appropriate to explain their empirical data as well. Therefore, I consider Zhu and Timmermans' (2010b, p. 72-73, my emphasis) conclusion that 'The overall model fit outperformed that of two *typical* discrete choice models' as the result of a straw man fallacy. Most other conclusions were logically valid findings about the characteristics of their proposed model but some, such as that with their model 'evidence of rational choice behaviour would be obtained if estimated thresholds are such that all factors are taken into account' were apparently based on a combination of the AtC and the BtQ and/or NCaC fallacies like those extensively discussed in the previous section.

Zhu and Timmermans (2010a) described exactly the same choice concept and implementation and its estimation for the same data set, including estimations of two more data sets from a shopping street in Shanghai. For one of the additional datasets the CAIC of the discrete choice models outperformed that of the proposed model, for the other the proposed model offered the best fit. Zhu and Timmermans (2010a, p. 17-18) presented many more conclusions about the mental choice processes and the ability of their model to elicit these, for example: 'Pedestrians' habit of using typical time points as judgment thresholds were reflected well by the model estimations', and 'the threshold-based representation is not only a better realization of human cognitive characteristics, but a better explanation for the heterogeneity of non-compensatory decision rules that bounded rational humans actually use as decisions mechanisms.' These are easy to denounce as the result of a similar combination of AtC and BtQ/NCaC fallacies as found in ZT. Their conclusion that 'the proposed model was also shown to be able to better capture behavioral heterogeneity for the large samples in our data than the multinomial and mixed logit models' are the result of a straw man fallacy, for the same reasons as explained in the preceding paragraph.

³¹ First received by *Environment and Planning B* in April 2008.

³² In this article the attribute thresholds were not counted as free parameters while in the later 2010a paper they were. I adopted the CAIC as listed there for the same dataset, see my comments in Sect. A4-3.3.

Zhang *et al.* (2004) proposed a choice concept that was remotely similar to Regret Theory. In their models, essentially the *differences* in standard utilities between the target alternative and all other alternatives were modeled instead of the target alternative's standard utility itself. They estimated parameters for a stated activity-trip-schedule-choice survey and found a moderate goodness-of-fit (Rho-squared app. 0.12). They denoted this as 'statistically high' while refraining from mentioning that in an earlier publication (Wang *et al.*, 2000), to which all but one of the authors contributed, Rho-squared values above 0.4 were found for common RUM models estimated on the same data. I consider this as a straw man fallacy if not scientific misconduct. Not surprisingly some conclusions about the usefulness of this choice model are based on AtQ fallacies. Annex 5 offers a more extensive review of this article.

Borgers *et al.* (2007) investigated whether the residential district influenced people's choice of recreational areas. Their small stated choice survey (111 respondents) was not customized to the respondents' relevant experiences, the goodness-of-fit of the estimated linear-in-parameters MNL models (Rho-squared ≈ 0.08) was poor and no alternative utility specifications were considered. I found several alternative explanations for their model outcomes, implying that some of the main conclusions were based on AtC fallacies. Annex 5 offers a more extensive review of this article.

Chen *et al.* (2009) proposed a residential choice model in which parameters for the distance-to-work and other accessibility attributes were specified as partially dependent on the pre-relocation levels. I found that for the distance-to-work they imposed a utility function that violated the diminishing marginal utility principle. They did not consider any alternative utility specification nor did they provide other information that allows estimation of the goodness-of-fit – or lack of it – of their model. Clearly taking the applicability of their estimated model for granted they arrived at several conclusions about the choice process that apparently relied on AtC fallacies in connection with NCaC or BtQ fallacies. Annex 5 offers a more extensive review of this article.

A4-7 Discussion, conclusions and recommendations

This annex disclosed several fallacies in six papers that sneaked their way into scientific transport-research publications unnoticed. There are some interesting characteristics that these six papers share: they all started from particular assumptions about people's mental transport choice processes; all implemented them in choice models; all estimated these models from empirical data sets; and all offered no opportunities for an unbiased appraisal of the goodness-of-fit. In four papers an alleged better goodness-of-fit compared to so-called 'conventional' or similarly denoted models was advanced as supporting evidence, which I conceived as straw man fallacies. In all the papers the AtC fallacy substantiated conclusions that claimed the usefulness of the alternative choice model. Where conclusions were drawn about the mental choice process these were also founded on the BtQ and/or NCaC fallacies. The content of the inferences based on these fallacies and of the recommendations arising from them is not necessarily wrong but as they are not based on solid ground they should be considered as the authors' personal opinions without a scientifically valid substantiation.

At the root of these fallacies might be that the concerned authors exhibited a strong belief in the trustworthiness of their theoretical assumptions about the mental choice process and how these should be implemented in choice models. They started from developing strong assumptions about the individuals' mental representations and processes that could arrange for the considered transport-related choices. What's more, in doing so some statements sneaked their way into their argumentations that were posited as true beyond reasonable doubt, see e.g. ZT (p. 766): 'Respondents will try to figure out what the researcher is expecting and what the experiment is trying to accomplish, and dependent on these contextual attributes will construct a mental representation of the decision problem/experimental task.' They incidentally used vague, subjective criteria such as 'face validity' to strengthen their views see e.g. Zhu and Timmermans, 2010b p. 73: 'the approach on pedestrians' go-home decision showed evidence of face validity'. Most authors also seemed to be convinced that the algorithms in mathematical models should approach the assumed actual mental choice processes as closely as possible, see e.g. Zhu and Timmermans (2010a, p. 3): 'In real life, however,

shifting between different degrees of decision criteria in different decision contexts is common ... Therefore, it will be a mis-specification to model such decision process based on the random utility paradigm.’

This approach to choice modelling is based on a misunderstanding of the role of theoretical assumptions about the human choice process in choice modelling. There is abundant evidence from psychology that people’s choice behaviour is a predominantly unconscious, covert process (e.g. Zajonc, 1980; Dijksterhuis *et al.*, 2006; Van de Kaa, 2008 Sect. 2, for an overview). This is well-substantiated in neuroscience (e.g. Bechara and Damasio, 2005). As a consequence, even the causal explanations provided afterwards by individuals themselves for their choices appeared notably unreliable (e.g. Nisbett and Wilson, 1977). This covertness might explain the large array of process theories about human choice behaviour that were proposed in the social sciences without much success. Verifying or falsifying the truth of the underlying assumptions is impossible as due to their covertness the concepts conceived to describe them were not measurable. That seems why none of these was widely accepted. For many years most scientists have dropped the idea of true-to-nature modelling of mental choice processes and become convinced that a paramorphic representation would suffice (Hoffman, 1960; Swait, 2001). Such a representation may be distinct in form from the real choice process, but analogous in the nature and product of its operations.

Friedman (1953 p. 21) illustrated this latter view with the following example: ‘Consider the problem of predicting the shots made by an expert billiard player. It seems not at all unreasonable that excellent predictions would be yielded by the hypothesis that the billiard player made his shots *as if* he knew the complicated mathematical formulas that would give the optimum directions of travel, could estimate accurately by eye the angles, etc., describing the location of the balls, could make lightning calculations from the formulas, and could then make the balls travel in the direction indicated by the formulas. Our confidence in this hypothesis is not based on the belief that billiard players, even expert ones, can or do go through the process described; it derives rather from the belief that, unless in some way or other they were capable of reaching essentially the same result, they would not in fact be *expert* billiard players’. This line of thought makes it irrelevant whether or not theoretical assumptions correspond with actual mental choice processes, be it conscious or unconscious. It would have made the many fallacious argumentations that were advanced to substantiate theoretical assumptions redundant. However, Friedman (1953: 41) concluded that ‘the question whether a theory is realistic “enough” can be settled only by seeing whether it yields predictions that are good enough for the purpose in hand or that are better than predictions from alternative theories.’ This demands empirical evidence that a proposed novel choice model outperforms at least one commonly used choice model in a solid *ceteris-paribus* comparison of predictive and/or descriptive ability (see for a useful methodology e.g. Van de Kaa, 2010a). For Zhu and Timmermans’ models this would, for example, have been feasible by estimating them for the outcomes of the state-of-the-art Dutch 1997 national value-of-travel-time study, by disregarding the individuals’ revealed reference-trip-attribute levels while taking them into account in the RUM model and/or a more advanced choice theory like, for example, Prospect Theory. This would have made the straw man comparison unnecessary.

Annex 5: Relative utility, historical disposition and different frames of reference

A5-1 Introduction

One of the comments in T (p. 380, my emphasis) was that PT's reference-point concept might be redundant as several utility-based theories offered the same mathematical functionality: 'it seems that transportation researchers have primarily explored the applicability of (cumulative) PT to incorporate reference points in their models to differentiate between gains and losses ... the use of reference points or thresholds has a long history in modelling riskless choices to model ... relative utility theory (Zhang *et al.*, 2004), historical disposition (Chen *et al.*, 2008; Habib and Miller, 2009) and different frames of references as a function of accumulated experiences (Borgers *et al.*, 2007). Hence ... several other utility-based alternative theories have been shown to offer the same mathematical functionality'. In this annex I examine the extent to which the articles of Zhang *et al.*, Chen *et al.* and Borgers *et al.* substantiate this claim.

A5-2 Relative utility theory?

The 'use of reference points or thresholds ... (in) relative utility theory (Zhang *et al.*, 2004)' was advanced as one of the 'several utility-based alternative theories that have been shown to offer the same mathematical functionality' as the use of reference points in PT (T p. 380). This section examines the extent to which the model and/or relative-utility-concept in Zhang *et al.*'s article substantiate this claim.

The term 'relative utility theory' does not occur in Zhang *et al.* (2004) but they proposed a choice model that claimed to draw on the concepts of relative utility and relative interest. In micro-economics, the utility of alternatives is commonly conceived as relative to people's current personal circumstances, which convention can be retraced to Bernoulli (1738 p. 24-25, his emphasis): 'The determination of the *value* of an item must not be based on its price, but rather on the *utility* it yields ... *Any increase in wealth ... will always result in an increase in utility which is inversely proportional to the quantity of goods already possessed*'. His concave utility function took this relative character of utility into account. The diminishing sensitivity assumption of PT is similar to Bernoulli's (relative) marginal utility concept, though applied to the changes in assets relative to a reference point/state rather than to their absolute size. Since then, the idea of some neutral reference state, which moves with hedonic adaptation following experienced choice outcomes, was firmly established in psychophysics, see e.g. Frederick and Loewenstein (1999). Many social scientists assume that people's aspirations are also influenced by social comparison (e.g. Schwarz and Strack, 1999), a phenomenon that in economics is sometimes popularised as 'keeping up with the Joneses'. The utility of alternatives is then conceived to depend on social-comparison driven reference levels (e.g. a neighbour's or peer's asset position). In economics the 'relative utility' concept is commonly conceived as in Van de Stadt *et al.* (1985 p. 179): 'utility is a completely relative concept, that is, an individual evaluates a bundle of consumption goods by comparing it to the consumption bundles of others, or perhaps to the bundles the individual has consumed in the past.'

Though Zhang *et al.* (p. 218, my emphasis) referred for its meaning to Van de Stadt *et al.* they successively extended the relative utility concept as follows: 'it is necessary to define the concept of relative utility. We argue that *an individual evaluates an alternative by comparing it to other alternatives*, or perhaps to the alternatives the individual chose in the past, or to alternatives chosen by other individuals'. They posited this extension without any further foundation, motivation or reference. They elaborated this addition by positing that the utility of each alternative is relative to the utilities of all other alternatives in the choice set. Successively, they mentioned that they would furtheron disregard the original interpretations of relative utility.

Zhang *et al.* (p. 216, 221) apparently considered that 'the concept of relative utility assumes that utility is meaningful only relative to some reference point(s)', 'the utility is a relative concept and meaningful only in the presence of some reference point(s)'. Note that the term 'reference point' is hardly ever used in the relative utility literature where terms like reference group and reference level abound. The term 'reference point' in connection with human judgment was published before PT was posited. However, Kahneman and Tversky (1979) were, to my best knowledge, the first who defined it in a choice context. On p. 217

Zhang *et al.* referred to this interpretation: ‘Tversky and Kahneman (1991) argued that choice behavior is dependent on status quo or reference point(s) and empirically confirmed that change of reference point might lead to preference reversal’, without referring to the loss-aversion concept. Apparently, this definition was not useful for their relative utility concept. Instead, Zhang *et al.* (p. 219) suggested that they adopted another earlier published definition of the term: ‘ U_j regards the standard utilities of all other alternatives except for the alternative j of interest as the reference points, which is suggested by Tversky and Simonson (1993)’. However, Tversky and Simonson (1993 p. 1183, my emphases) deliberately distinguished between ‘the principle of *loss aversion*, according to which losses loom larger than the corresponding gains. Gains and losses are defined relative to a *neutral reference point* that generally corresponds to the decision maker’s status quo or current endowment’, and the ‘*extremeness aversion hypothesis*’, which would apply ‘in some situations (where) decision makers may evaluate *options* in terms of their advantages and disadvantages, *defined relative to each other*’, in which case ‘a natural extension of loss aversion suggests that disadvantages loom larger than the corresponding advantages’. Tversky and Simonson thus explicitly avoided to denote the attribute levels of the alternatives as reference points and did definitely not assume that they could be valued as standard utilities. Zhang *et al.*’s suggestion that they did so is thus misleading.

As both PT and the articles of Kahneman, Tversky and Simonson were discussing loss aversion extensively it is conspicuous that Zhang *et al.* nowhere referred to it. Essentially, the models that they developed differ from similar RUM specifications by (i) considering the *differences* in standard (non-relative!) utilities between the target alternative and all other alternatives, instead of the target alternative’s standard utility itself, and (ii) the application of ‘relative-interest’ coefficients to different destination-stop-pattern combinations. The standard utilities of alternatives were specified as the sum of linear and quadratic functions of the attributes and did not depend on the attribute sizes of their ‘reference points’. Their models did not offer the mathematical functionality of PT’s reference state, which locates a kink and a convex-concave transition in the value function.

Zhang *et al.* estimated MNL and NL (nested logit) models for the stated choices between different destination-activity-travel schedules, each with four estimated relative-interest parameters and with these four parameters fixed at unity. Best fitting was an NL model with estimated relative-interest parameters. It had a moderate adjusted-Rho-squared-value of 0.123, slightly better than the value of 0.122 found for the same model with equal relative-interest parameters. The MNL models had slightly lower adjusted-Rho-squared values (0.115 and 0.112, respectively). The same stated choice data were analysed before (Wang *et al.*, 2000). They found much better adjusted-Rho-squared-values of 0.386 and 0.399 for estimated MNL and NL models that were similar, but with just standard utilities instead of standard-utility differences and without relative-interest terms. Zhang *et al.* were well-known with this work as they referred to it for details about the stated choice experiment and as Zhang’s co-authors Borgers, Timmermans and Wang were among its authors. They should have realised that referring to and explaining of the better fit of the standard RUM models would have facilitated a proper appraisal of their new choice model but refrained from mentioning this factual evidence. Instead, Zhang *et al.* (p. 225) called their Rho-squared values ‘statistically high goodness-of-fit indices’.

Summarizing, Zhang *et al.* devised a choice model that considered differences in standard utilities. They called this a relative utility concept though they knew that this term was already commonly used to denote utility assessment in social comparison settings. They considered that they had to define reference point(s), apparently to justify the use of the term ‘relative utility’. They misleadingly suggested that Tversky and Simonson (1993) conceived ‘the standard utilities of all other alternatives except for the alternative of interest as reference points’ while these scientists deliberately avoided that naming. Though Zhang *et al.* knew that the fits of their ‘relative utility/relative interest’ models with the data set for which they established it were previously outperformed by far by conventional RUM-models they omitted mentioning this. Their reference points did not offer the mathematical functionality of locating a kink and/or a convex-concave transition in the utility/value function.

A5-3 Historical disposition?

Chen *et al.* (2009) should have demonstrated that utility-based models in connection with ‘historical disposition’³³, offered the same mathematical functionality as PT’s reference point (T p. 380). This section examines the extent to which the model and/or historical-deposition-concept in Chen *et al.*’s article substantiate this claim.

Though Chen *et al.* once-only mentioned PT’s reference point they did not identify the pre-relocation attributes as such. Instead, they developed a GEV-type RUM model in which parameters for the distance-to-work, shopping-accessibility, open-space-accessibility and recreational-opportunities attributes were partly dependent on the relevant ‘absolute’ pre-relocation levels. This was done by specifying these parameters as $((\alpha_1 + \alpha_2/(x+1))$, in which α_1 and α_2 are constants to be estimated and x is the pre-relocation attribute level. Chen *et al.* (p. 2763-2764) said they did follow ‘the law of diminishing marginal utility’ in specifying logarithmic functions for the part-worth utilities of the accessibility attributes. Obviously, this law applies to the utility of ‘goods’, that are things that people want to have. The shopping, open-space and recreational attributes were indeed defined as ‘goods’: an increase in their levels implied a better accessibility. The concave logarithmic function thus holds for them. However, when commute distance increases people’s monetary expenses and travel time will increase. It is thus a ‘bad’ or disutility, for which Chen *et al.* found negative parameter values indeed. If we take a decrease in commute distance as argument their utility function transforms into a continuously increasing, convex utility function which implies *increasing* marginal utility and thus violates the law of diminishing marginal utility. That is why, in my opinion, it does not classify as a utility-based model.

Chen *et al.* estimated their model for relocations of 566 household between 1989 and 2002, in the Puget Sound region. Taking their size and attribute level ranges into account the part-worth commute-distance utility appeared at least an order-of-magnitude higher than those of shopping, open-space and recreation accessibility. The parameters for commute distance were negative indeed, and in view of its convex utility specification one might expect a poor model fit. The authors listed just one loglikelihood value (of their full model?), which does not allow a pseudo-R-squared assessment. They did not consider any alternative utility specification nor did they provide other information that allows esteeming the goodness-of-fit –or lack of it – of their model. Chen *et al.* (p. 2771-2) apparently took the applicability of their estimated model for granted in presenting findings like ‘For commute distance we observe that, as prior commute distance increase, the estimate for commute distance becomes less negative. Or, people are less sensitive to long commute distances if they have lived with long commutes before’. It is easy to see that the monotonously decreasing contribution of the pre-relocation attribute parameters would have the effect mentioned in the first cited sentence, whatever part-worth utility function would be specified. Obviously, as prior- and post-relocation attribute levels will commonly be correlated the logarithmic utility function will increase this effect. Of course, one might specify a commute distance parameter that increases with the pre-relocation attribute level which, in connection with a convenient utility-functional form, would impose that as commute distance increases the estimate for commute distance becomes more negative. The second sentence in the citation above is thus clearly the outcome of an ‘affirming the consequent’ fallacy. The same holds for conclusions in Chen *et al.* (p. 2773) like ‘the results of our study appear to support the adaptive human behavior hypothesis. In particular, we found that people have become more tolerant to long commute distances and viewed attributes like retail, open space, and retail opportunities as more valuable after having been exposed to them in a large quantity in the past’ (p. 2773), which are based on this fallacy in combination with the ‘begging the question’ and/or ‘non-cause as cause’ fallacies.

Taking the commute distance prior to relocation as reference point Chen *et al.*’s model estimate can be expressed into a reference-dependent value function. This yields a somewhat higher appraisal of a decrease/gain in commute distance compared to an increase/loss of equivalent size, a convex value

³³ Apparently, ‘disposition’ is a misspelling: Chen *et al.* (p. 2760) hypothesized ‘a historical deposition effect for residence location choices, which states that people’s location preferences are likely to be a function of the attributes of where they lived before.’

function in the loss as well as gain domain and no change in marginal value at the reference point. Obviously, their model does not offer the same mathematical functionality as PT.

A5-4 Different frames of reference?

T (p. 380) mentioned that Borgers *et al.* (2007) had demonstrated that utility-based models in connection with ‘different frames of references as a function of accumulated experiences’ offered the same mathematical functionality as PT. This section examines the extent to which the model and/or frame-of-reference concept in Borgers *et al.*’s article³⁴ substantiate this claim.

Borgers *et al.* (p. 225) wrote: ‘this study assumes that individual’s utility for spatial choice alternatives may vary as a function of the location of individuals. We assume that individuals’ expectations are formed as a function of their experiences, and that these experiences may differ depending on where they live’ and ‘the findings of the model estimation supported the premise underlying this study. Estimated parameters suggest that the utility people derive from park attributes reflect their daily way of life. In addition, there is evidence of compensation: some attributes that are less or not available in the direct living environment are appreciated more’. Borgers *et al.* tested their hypothesis by performing a small-scale (111 respondents) conjoint choice experiment in Eindhoven, The Netherlands, in which people had to make up their minds about different parks characterized by 22 accessibility and level-of-service attributes. Obviously, a full factorial design of the conjoint choice experiment will not have been feasible. If the small fraction of it that was submitted was discussed this was on the missing pages. The same holds for questions about personal circumstances like dog-ownership and type of residence, and/or revealed preference information like frequency, aim and travel mode of earlier park visits that might have been posed – in the paper only some information was found about age, gender and having children.

Two full-fledged MNL-RUM models were estimated, both with a linear-additive specification of park attributes. In one model an additional city-resident-specific parameter was added, in the other the parameter of each attribute was split in an average and a city-resident specific parameter. The only person-specific attribute that was considered in these models was thus the geographical location of the residence, which was attributed to either the (central) town or the suburban belt. The estimation of both models showed that utility decreases with distance and increases with park size. All significant parameters were of the expected sign, except for accessibility by public transport. Borgers *et al.* (p. 223) suggested that this might ‘reflect the fact that public transport in the city is rarely used for visits to parks, and only by particular segments of the population’. The conclusion that individuals’ utility varies as a function of the location of their residence was based on the second model. This had a poor goodness-of-fit (Rho-squared 0.08). Its estimated parameters for the town-residents-specific attributes of children-playground-availability, dog-walking-opportunity, wonderful-view-opportunity and public-transport-accessibility were significant at the 5% level. Also the ‘distance decay effect as measured by the distance to urban park attribute is steeper ... (and) the range of utility for the type and size of the park is larger for the people living in the city’ (p. 225).

Obviously, the distribution of the types of residence in an area is strongly correlated with the distribution of family types. Flats, for example, accounted for about 30% of the residences in Eindhoven compared to about 15% in the surrounding communities, while the shares of households with children were 28% and 40%, respectively. Borgers *et al.* did not consider the possibility of a relative under-representation of parents with children in their town-residents sub-sample, which seems an obvious explanation for the observed underrating of playground-availability. Likewise, they did not consider the correlation between household- and/or residence-type and dog-ownership. In the Netherlands, less than one-third of the families keep a dog and there is a negative influence of population density on it. Again, a relative under-representation of dog-owners in their town-residents subsample seems an obvious explanation for the observed underrating of dog-walking-opportunities. It might also be the case that people with children and/or dogs are less inclined to visit a park just for a ‘wonderful view’ than others,

³⁴ The book in which this paper was published was not available in my university library but it was accessible on Google Books, with just 3 pages of the paper missing.

which might offer an alternative explanation for the relatively higher average utility of that attribute among town residents. The estimated town-resident-specific public-transport-accessibility parameter is much more negative than the average for all respondents. Taking Borgers *et al.*'s comment above into consideration it is hard to imagine that this is caused by area-related preferences, as home-side public-transport-availability is better for town-residents and these might well have been over-represented in the small fraction of the respondents that might have had an interest in it. It is easy to see that people visiting a park for walking their dog or accompanying their children to their playground will exhibit a strong aversion to an increasing distance while they might consider the smallest submitted park size as less appropriate for that purpose. Indirectly, the probable relative over-representation of these categories in the suburban population might therefore also explain the differences in the parameters for the distance and park-size attributes between town- and suburban-residents sub-samples.

The differences between the town- and suburban-residents subsamples in the estimated parameters might possibly also have been caused by confounded attribute interactions. In view of the large number of attributes, the linear-additive utility specification and the small number of possible combinations that might have been submitted in the conjoint experiment I missed a discussion of this in the article. Anyhow, the small survey population, the even smaller number of people who might have an interest in attributes like dog-walking or public-transport-accessibility and their uneven distribution over city and suburbs, in connection with the low goodness-of-fit of the model does definitely not exclude other explanations for the observed stated preferences, such as the type of residence independent of its distance to the city centre. Though the main conclusion of Borgers *et al.* (p. 225) is not necessarily wrong or right the successive conclusions that 'evidence for spatially dependent utility functions was found ... there is evidence of compensation: some attributes that are less or not available in the direct living environment are appreciated more' are clearly the result of an affirming-the-consequent fallacy.

According to T Borgers *et al.*'s utility-based models should, in connection with 'different frames of references as a function of accumulated experiences', have offered the same mathematical functionality as PT. I did not find the wording 'frames of reference' in the text, nor another definition of that concept in the sense as commonly used in decision theory. One might guess that T interpreted this concept as: adaptation to experiences undergone by living in a town results in preferences that differ from those following from living in its suburbs, all else being the same. Conceived as such, the re-examination above showed that Borgers *et al.* did not provide convincing evidence for such a frame of reference. Accounting for the spatial distribution of respondent-specific characteristics as, for example, type-of-residence, dog-ownership and having children might have yielded an improved model fit while reducing or annihilating the 'town-resident-location' effects. An even better model fit might be attained with a discrete choice model that allows accounting for the individuals' adaptation of a reference state following experienced accessibility and use of actually visited parks, in connection with the successive reference-dependent re-framing and valuation of the submitted park-attribute-levels.

References

- Abelson, R.P. (1981). Psychological status of the script concept. *American Psychologist*, vol. **36**, no. 7, pp. 715–729.
- AMR (1999). *The Value of Travel Time on UK Roads – 1994*. The Hague, Accent Marketing and Research and Hague Consulting Group.
- Arentze, T.A., Dellaert, B.C.G., and Timmermans, H.J.P. (2008). Modeling and measuring individuals' mental representations of complex spatio-temporal decision problems. *Environment and Behavior*, vol. **40**, no. 6, pp. 843–869.
- Arentze, T.A. and Timmermans, H.J.P. (2005). Representing mental maps and cognitive learning in micro-simulation models of activity-travel choice dynamics. *Transportation*, vol. **32**, pp. 321–340.
- Aristotle (≈350 BCE). *On Sophistical Refutations*. English translation by W.A. Pickard-Cambridge. University of Adelaide, eBooks@Adelaide, <ebooks.adelaide.edu.au/a/aristotle/sophistical/index.html>.
- Atkinson, R.L., Atkinson, R.C. and Hilgard, E.R. (1983). *Introduction to Psychology*, Eighth Edition. New York, Harcourt Brace Jovanovich.
- Avineri, E. and Prashker, J.N. (2004). Violations of Expected Utility Theory in Route-Choice Stated Preferences: The Certainty Effect and Inflating of Small Probabilities. *Transportation Research Record*, vol. **1894**, pp. 222-229.
- Avineri, E. and Prashker, J.N. (2005). Sensitivity to travel time variability: Travelers' learning perspective. *Transportation Research C*, vol. **13**, pp. 157-183.
- Avineri, E. and Prashker, J.N. (2006). The Impact of Travel Time Information on Travellers' Learning under Uncertainty. *Transportation*, vol. **33**, pp. 393-408.
- Barron, G. and Erev, I. (2003). Small Feedback-based Decisions and Their Limited Correspondence to Description-based Decisions. *Journal of Behavioral Decision Making*, vol. **16**, pp. 215-233.
- Ballinger, T.P. and Wilcox, N.T. (1997). Decisions, error and heterogeneity. *The Economic Journal*, vol. 107, pp. 1090-1105.
- Bates, J., Polak, J., Jones, P. and Cook, A. (2001). The valuation of reliability for personal travel. *Transportation Research E*, vol. **37**, pp. 191-229.
- Beach, L.R. (1990). *Image Theory: Decision Making in Personal and organizational Contexts*. Chichester, John Wiley & Son.
- Bechara, A. and Damasio, A.R. (2005). The somatic marker hypothesis: A neural theory of economic decision. *Games and Economic Behavior*, vol. **52**, no. 2, pp. 336-372.
- Ben-Akiva, M. and M. Bierlaire (1999). Discrete Choice Methods and Their Applications to Short Term Travel Decisions. In: R. W. Hall (ed), *Handbook of Transportation Science*, pp. 5-33. Norwell, MA., Kluwer Academic Publishers.
- Ben-Elia, E. and Shifitan, Y. (2010). Which road do I take? A learning-based model of route-choice behavior with real-time information. *Transportation Research A*, vol. **44**, pp. 249-264.
- Bernoulli, D. (1738). Exposition of a new theory on the measurement of risk (English translation from Latin, republished 1954). *Econometrica*, vol. **22**, pp. 23-36.
- Blavatskyy, P. and Pogrebna, G. (2006). Loss Aversion? Not with Half-a-Million on the Table! Working Paper Series no. 319. Institute for Empirical Research in Economics, University of Zürich.
- Blavatskyy, P. and Pogrebna, G. (2007). Models of stochastic choice and decision theories: why both are important for analyzing decisions. Working Paper Series no. 319. Institute for Empirical Research in Economics, University of Zürich.
- Borgers, A.W.J., Timmermans, H.J.P. and van der Waerden, P.J.H.J. (2007). Spatially differentiated utility functions for urban greenspace: a test based on conjoint choice experiments. In: A.G. Woodside (ed.), *Advances in Culture, Tourism and Hospitality Research*, Vol. 1, pp. 215-226. Oxford, UK, Elsevier.
- Bos, I., Ettema, D.F. and Molin, E.J.E. (2004). Modeling the Effect of Travel Time Uncertainty and Traffic Information on the Use of Park and Ride Facilities, *Transportation Research Record*, vol. **1898**, pp. 37-44.
- Botting, D. (2012). Fallacies of Accident. *Argumentation*, vol. **26**, pp. 267–289.
- Burge, P., Rohr, C., Vuk, G. and Bates, J. (2004). Review of international experience in VOT study design. *European Transport Conference*, Strasbourg.
- Busemeyer, J.R. and Townsend, J.T. (1993). Decision Field Theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, vol. **100**, iss. 3, pp. 432-459.
- Callois, J.-M. (2008). The two sides of proximity in industrial clusters: The trade-off between process and product innovation. *Journal of Urban Economics*, vol. **63**, pp. 146–162.

- Carbone, E. (1997). Investigation of stochastic preference theory using experimental data. *Economic Letters*, vol. 57, pp. 305-311.
- Carbone, E. and Hey, J.D. (1995). A comparison of the estimates of Expected Utility and Non-Expected-Utility Preference Functionals. *The Geneva Papers on Risk and Insurance Theory*, vol. 20, pp. 111-133.
- Chater, N., Oaksford, M., Nakisa, R. and Redington, M. (2003). Fast, frugal and rational: How rational norms explain behavior. *Organizational Behavior and Human Decision Processes*, vol. 90, pp. 63-86.
- Chen, C., Chen, J. and Timmermans, H.J.P. (2009). Historical deposition influence in residential location decisions: a distance-based GEV model for spatial correlation. *Environment and Planning A*, vol. 41, no. 11, pp. 2760-2777.
- Chorus, C.G., Arentze, T.A., Molin, E.J.E., Timmermans, H.J.P. and van Wee, G.P. (2006). The Value of Travel Information: Decision-Strategy Specific Conceptualizations and Numerical Examples, *Transportation Research B*, vol. 40, pp. 504-519.
- Clark, A.E. and Oswald, A.J. (1996). Satisfaction and comparison income. *Journal of Public Economics*, vol. 61, pp. 359-381.
- Damasio, A.R. (2001). Fundamental Feelings. *Nature*, vol. 413, pp. 781-781.
- Dawes, R.M. (1988). *Rational choice in an uncertain world*. Orlando, USA, Harcourt Brace College Publ.
- Dijksterhuis, A. (2004). Think different: the merits of unconscious thought in preference development and decision making. *Journal of Personality and Social Psychology*, vol. 87, no. 5, pp. 586-598.
- Dijksterhuis, A., Bos, M.W., Nordgren, L.F. and Van Baaren, R.B. (2006). On making the right choice: The deliberation-without-attention effect. *Science*, vol. 311, no. 5763, pp. 1005-1007.
- Dingemans, D., Sperling, D. and Kitamura, R. (1986). Mental maps and the refueling behavior of vehicle drivers. *Transportation Research Record*, vol. 1092, pp.1-10.
- Evans, G.W., Marrero, D.G. and Butler, P.A. (1981). Environmental learning and cognitive mapping. *Environment and Behavior*, vol. 13, no. 1, pp. 83-104.
- Frederick, S. and Loewenstein, G. (1999). Hedonic Adaptation. In: D. Kahneman, E. Diener and N. Schwarz (eds), *Well-being: The Foundations of Hedonic Psychology*, pp. 302-329. New York, Russell Sage Foundation.
- Friedman, M. (1953). The methodology of positive economics. In: M. Friedman (ed), *Essays in Positive Economics*, pp. 3-43. Chicago, University of Chicago Press.
- Fujii, S. and Gärling, T. (2003). Development of script-based travel mode choice after forced change. *Transportation Research F*, vol. 6, pp. 117-124
- Fujii, S. and Kitamura, R. (2004). Drivers' Mental Representation of Travel Time and Departure Time Choice in Uncertain Traffic Network Conditions. *Networks and Spatial Economics*, vol. 4, no. 3, pp. 243-256.
- Gigerenzer, G. and Todd, P.M. (1999). Fast and frugal heuristics: The adaptive toolbox. In: G. Gigerenzer, P.M. Todd and ABC Group (eds), *Simple heuristics that make us smart*, pp. 3-34. Oxford, UK, Oxford University Press.
- Gilbert, D.T., and Hixon, J.G. (1991). The trouble of thinking: activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, vol. 60, pp. 509-517
- Gilovich, T., Griffin, D. and Kahneman, D. (eds) (2002). *Heuristics and Biases*. New York, Cambridge University Press.
- Golledge, R.G. and Spector, A.N. (1978). Comprehending the Urban Environment: Theory and Practice. *Geographical Analysis*, vol. 10, no. 4, pp. 403-426.
- Habib, M.A. and Miller, E.J. (2009). Reference Dependent Residential Location Choice Model within a Relocation Context. *Transportation Research Record*, vol. 2133, pp. 92-99.
- Harless, D.W. and Camerer, C.F. (1994). The predictive utility of generalized Expected Utility theories. *Econometrica*, vol. 62, pp. 1251-1289.
- HCG (1998). *Value of Dutch Travel Time Savings in 1997 – Final report*. The Hague, Hague Consulting Group.
- Hertwig, R., Barron, G., Weber, E.U. and Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, vol. 15, pp. 534-539.
- Hey, J.D. (1995). Experimental investigations of errors in decision making under risk. *European Economic Review*, vol. 39, pp. 633-640.
- Hey, J.D. and Orme, C. (1994). Investigating generalizations of Expected Utility Theory using experimental data. *Econometrica*, vol. 62, pp. 1291-1326.
- Hoffman, P.J. (1960). The paramorphic representation of clinical judgment. *Psychological Bulletin*, vol. 57, pp. 116-131.
- Kahneman, D. (2000). Preface. In Kahneman, D. and Tversky, A. (eds). *Choices, Values and Frames*, pp. ix-xvii. New York, Russell Sage Foundation.

- Kahneman, D. (2002). Maps of Bounded Rationality: A Perspective on Intuitive Judgment and Choice. In Frängsmyr, T. (ed) *Les prix Nobel 2002*, pp. 449-489. Stockholm, Nobel Foundation.
- Kahneman, D., Knetsch, J.L. and Thaler, R.H. (1991). Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. *Journal of Economic Perspectives*, vol. **5**, no.1, pp. 193-206.
- Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, vol. **47**, pp. 263-291.
- Kahneman, D. and Tversky, A. (1984). Choices, Values, and Frames. *American Psychologist*, vol. **39**, no. 4, pp. 341-350.
- Kahneman, D. and Tversky, A. (eds) (2000). *Choices, Values, and Frames*. New York, Russell Sage Foundation.
- Klein, G.A. (1993). A Recognition-Primed Decision (RPD) model of rapid decision making. In: Klein, G.A. et al. (eds), *Decision Making in Action: Models and Methods*, pp. 138-147. Norwood, CT, Ablex Publishing Corporation.
- Kwan, M.-P. (2000). Analysis of human spatial behavior in a GIS environment: Recent developments and future prospects. *Journal of Geographical Systems*, vol. **2**, pp. 85-90
- Levin, I.P., Schneider, S.L. and Gaeth, G.J. (1998). All Frames Are Not Created Equal: A Typology and Critical Analysis of Framing Effects. *Organizational Behavior and Human Decision Processes*, vol. **76**, no. 2, pp. 149-188.
- Li, Z. and Hensher, D. (2011). Prospect theoretic contributions in understanding traveller behaviour: a review and some comments. *Transport Reviews*, vol. **31**, pp. 97-115.
- Loomes, G. and Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, vol. **92**, pp. 805-824.
- Loomes, G. and Sugden, R. (1995). Incorporating a stochastic element into decision theories. *European Economic Review*, vol. **39**, pp. 641-648.
- Mahmassani, H.S. (1990). Dynamic models of commuter behavior: Experimental investigation and application to the analysis of planned traffic disruptions. *Transportation Research A*, vol. **24**, pp 465- 484.
- Mani, K. and Johnson-Laird, P.N. (1982), The mental representation of spatial descriptions. *Memory & Cognition*, vol. **10**, no.2, pp.181-187.
- McFadden, D.L. (2001). Economic Choices (revised version of Nobel Prize Lecture). *American Economic Review*, vol. **91**, pp. 351-378.
- Michea, A. and Polak, J. (2006). Modelling Risky Choice Behaviour: Evaluating Alternatives to Expected Utility Theory. *Paper presented at the 11th International Conference on Travel Behaviour Research*, Kyoto.
- Montgomery, H. and Svenson, O. (eds) (1989). *Process and Structure in Human Decision Making*. Chichester, John Wiley & Sons.
- Nisbett, R.E. and Wilson, T.D. (1977). Telling More Than We Can Know: Verbal Reports on Mental Processes. *Psychological Review*, vol. **84**, no. 3, pp. 231-259.
- Nylenna, M., Andersen, D., Dahlquist, G., Sarvas, M. and Aakvaag, A. (1999). Handling of scientific dishonesty in the Nordic countries. *Lancet*, vol. **354**, no. 9172, pp. 57-61.
- Payne, J.W., Bettman, J.R. and Johnson, E.J. (1993). *The adaptive decision maker*. New York, Cambridge University Press.
- Plant, K.L. and Stanton, N.A. (2012). Why did the pilots shut down the wrong engine? Explaining errors in context using Schema Theory and the Perceptual Cycle Model. *Safety Science*, vol. **50**, pp. 300-315.
- Post, T., Van den Assem, M.J., Baltussen, G. and Thaler, R.H. (2008). Deal or No Deal? Decision making under risk in a large-payoff game show. *American Economic Review*, vol. **98**, pp. 38-71.
- Quiggin, J. (1982). A Theory of Anticipated Utility. *Journal of Economic Behavior and Organization*, vol. **3**, pp. 323-343.
- RWS (2003). *Kerncijfers personenvervoer 2002* (in Dutch). Rotterdam, Rijkswaterstaat - AVV.
- Schwanen, T. and Ettema, D. (2009). Coping with unreliable transportation when collecting children: Examining parents' behavior with Cumulative Prospect Theory. *Transportation research A*, vol. **43**, pp. 511-525
- Schwarz, N. and Strack, F. (1999). Reports of Subjective Well-Being: Judgmental Processes and Their Methodological Implications. In: D. Kahneman, E. Diener and N. Schwarz (eds), *Well-Being: The Foundations of Hedonic Psychology*, pp. 61-84. New York, Russell Sage Foundation.
- Sen, S., Pillai, R., Joshi, S. and Rathi, A.K. (2001). A Mean-Variance Model for Route Guidance in Advanced Traveler Information Systems. *Transportation Science*, vol. **35**, pp. 37-49.
- Senna, L.A.D.S. (1994). The influence of travel time variability on the value of time. *Transportation*, vol. **21**, pp. 203-228.
- Simon, H.A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, vol. **69**, pp. 99-118.

- Simon, H.A. (1978). Rational decision-making in business organizations. In: A. Lindbeck (ed), *Nobel Lectures in Economic Sciences 1969-1980*, pp. 343-370. Singapore, World Scientific Publ. Comp.
- Smith, E.R. (1998). Mental representation and memory. In: D. T. Gilbert, S. T. Fiske and G. Lindzey (eds), *Handbook of Social Psychology* (4th ed., vol. 2), pp. 391-445. New York, McGraw Hill.
- Sober, E. (1976). Mental representations. *Synthese*, vol. **33**, pp. 101-148.
- SOED (2002). *Shorter Oxford English Dictionary on CD-ROM*. Version 2.0. Oxford (UK), Oxford University Press.
- Stern, E. (2004). Spatio-temporal patterns of subjectively reported congestion in Tel Aviv metropolitan area. *Journal of Transport Geography*, vol. **12**, pp. 63–71.
- Sunley, P., Pinch, S., Reimer, S. and Macmillan, J. (2008). Innovation in a creative production system: the case of design. *Journal of Economic Geography*, vol. **8**, pp. 675–698
- Swait, J. (2001). A non-compensatory choice model incorporating attribute cutoffs. *Transportation Research B*, vol. **35**, pp. 903-928.
- Thaler, R.H. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization*, vol. **1**, pp. 39-60.
- Timmermans, H. (1982). Consumer choice of shopping centre: an information integration approach. *Regional Studies*, vol. **16**, no.3, pp. 171-182
- Timmermans, H. (2010). On the (ir)relevance of Prospect Theory in modelling uncertainty in travel decisions. *European Journal of Transport and Infrastructure Research*, vol. **10**, pp. 368-384.
- Tulving, E. and Thomson, D.M. (1971). Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology*, vol. **87**, no. 1, pp. 116-124.
- Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, vol. **185**, pp. 1124-1131.
- Tversky, A. and Kahneman, D. (1991). Loss aversion in riskless choice: a reference-dependent model. *Quarterly Journal of Economics*, vol. **106**, pp. 1039-1061.
- Tversky, A. and Kahneman, D. (1992). Advances in Prospect Theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, vol. **5**, pp. 297-323.
- Tversky, A. and Simonson, I. (1993). Context-dependent Preferences. *Management Science*, vol. **39**, no. 10, pp. 1179-1189.
- Van de Kaa, E.J. (2008). *Extended Prospect Theory. Findings on Choice Behaviour from Economics and the Behavioural Sciences and their Relevance for Travel Behaviour*. TRAIL Research School, Delft.
- Van de Kaa, E.J. (2010a). Applicability of an Extended Prospect Theory to travel behaviour research: a meta-analysis. *Transport Reviews*, vol. **30**, pp. 771-804.
- Van de Kaa, E.J. (2010b). Prospect Theory and choice behaviour strategies: review and synthesis of concepts from social and transport sciences. *European Journal of Transport and Infrastructure Research*, vol. **10**, pp. 299-329.
- Van de Stadt, H., Kapteyn, A. and Van de Geer, S. (1985). The relativity of utility: Evidence from panel data. *The Review of Economics and Statistics*, vol. **67**, no. 2, pp. 179-187.
- Wang, D., Borgers, A., Oppewal, H. and Timmermans, H. (2000). A stated choice approach to developing multi-faceted models of activity behavior. *Transportation Research Part A*, vol. **34**, pp. 625-643.
- Wakker, P.P. (2003). The data of Levy and Levy (2002) “Prospect Theory: much ado about nothing?” actually support Prospect Theory. *Management Science*, vol. **49**, pp. 979-981.
- Woods, J. (1999). Files of Fallacies, Aristotle (384-322 BC). *Argumentation*, vol. **13**, pp. 203-220.
- Young, W. and Richardson, A.J. (1979). Residential area location preference surfaces. *Transportation Research Record*, vol. **707**, pp. 39–47.
- Zajonc, R.B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, vol. **35**, no. 2, pp. 151-175.
- Zhang, J., Timmermans, H., Borgers, A. and Wang, D. (2004). Modelling traveler choice behavior using the concepts of relative utility and relative interest. *Transportation Research B*, vol. **38**, no. 6, pp. 215-234.
- Zhu, W. and Timmermans, H. (2010a). A decision making model incorporating choice of decision criteria. In: *Proceedings of the 89th TRB Annual Meeting*, Washington, D.C.
- Zhu, W. and Timmermans, H. (2010b). Cognitive process model of individual choice behaviour incorporating principles of bounded rationality and heterogeneous decision heuristics. *Environment and Planning Part B*, vol. **37**, no. 1, pp. 59-74.
- Zhu, W. and Timmermans, H. (2010c). Modeling simplifying information processing strategies in conjoint experiments. *Transportation Research B*, vol. **44**, no. 6, pp. 764–780.

Nabeschuwing

Voor een speciale uitgave van het *European Journal of Transport and Infrastructure Research (EJTIR)* heeft Timmermans desgevraagd een kritische beschouwing geschreven over de toepasbaarheid van Prospect Theory (PT) voor transportonderzoek (Timmermans, 2010; hierna aangeduid met T). Tijdens een onderzoek naar de bruikbaarheid van PT voor verkeerspredicties heb ik zijn artikel doorgenomen. Daarbij kwam ik een verwijzing tegen naar een paper waarin zou staan dat PT het keuzegedrag van deelnemers aan de TV show “Spel zonder Grenzen” slecht benaderde. Uit nieuwsgierigheid heb ik die paper gedownload. Toen ik hem doorlas viel me op dat de bewering in T niet klopte. Daarna heb ik de beweringen en redeneringen in T en in een aantal daarin gerefereerde publicaties bestudeerd. Mijn analyses en bevindingen heb ik gedocumenteerd in dit werkdocument. Een uitgebreide samenvatting ervan is gepubliceerd in *EJTIR*³⁵. De hoofdredacteur daarvan had vooraf bepaald dat Timmermans de gelegenheid zou krijgen om hierop éénmalig te reageren. Daarmee zou de discussie in het tijdschrift worden gesloten. Timmermans schreef inderdaad een repliek voor *EJTIR*³⁶. In het voorwoord van dit werkdocument heb ik de inhoud ervan samengevat waarbij ik ook mijn reactie op Timmermans’ repliek meenam. Deze nabeschuwing is een Nederlandse bewerking van dat voorwoord.

Afgezien van voorwoord en nabeschuwing werd dit werkdocument afgerond in september 2012. Het beschrijft mijn analyse en interpretatie van de kritieken en onderliggende beweringen, argumenten en literatuurverwijzingen in T. Het toont diverse fouten en/of schendingen van wetenschappelijke integriteit aan, zoals zeer onzorgvuldige literatuurverwijzingen (sectie 4.1), drogredenen (sectie 4.7) en teksten die de suggestie van plagiaat kunnen wekken (sectie 4.5). In zijn repliek geeft Timmermans de onzorgvuldige literatuurverwijzingen toe, maar gaat niet in op de drogredenen en ontkent iedere vorm van plagiaat. Wat betreft het laatste verdient hij m.i. het voordeel van de twijfel, hoewel hij in zijn repliek geen concreet tegenbewijs levert door b.v. naar eigen papers te verwijzen waaruit de beide twijfelachtige teksten worden verklaard. Een andere discutabele bewering in T (p. 369) was: ‘Figure 1 gives an overview of dominant approaches and key issues that have been addressed and explored in the early years (1970-1980s). These are listed in the context of a general conceptual framework that summarizes the common elements of the various approaches (Timmermans, 1982)’. Omdat deze pretenties noch in T, noch in Timmermans’ artikel uit 1982 werden onderbouwd concludeerde ik dat Figuur 1 gefabriceerd was (zie sectie 4.3). In zijn repliek laat Timmermans de verwijzing naar zijn artikel uit 1982 buiten beschouwing en beschrijft andere pretenties van Figuur 1. Als hij dat in T ook op deze wijze had verwoord en als zodanig had gebruikt in de daaropvolgende argumentaties was er m.i. geen sprake geweest van fabriceren, maar dat was niet zo.

Dit werkdocument signaleert ook enkele ernstiger overtredingen van goed wetenschappelijk gedrag:

- De bewering dat in een studie van Blavatsky and Pogrebna (2007) ‘remarkably, Cumulative PT never outperformed other decision theories, regardless of the assumed probabilistic choice rules’ (T p. 376). Nauwgezet doorspitten van de geciteerde paper leverde geen enkele informatie op die deze bewering verklaart. Integendeel, Blavatsky and Pogrebna (p. 10) schreven expliciet dat ‘we do not estimate ... (cumulative) PT’. Het lijkt er dus op dat Timmermans’ bewering was gefingeerd (zie sectie 4.3).
- Meerdere beweringen die de geloofwaardigheid van PT voor de modellering van transportkeuze gedrag geen goed doen, zoals: ‘The most basic version of EUT ... is the expected value model ... Kahneman and Tversky questioned the validity of EUT ... It should be noted that this position relates to the basic form of EUT’ (T p. 371-372, mijn cursivering). Ik heb beide artikelen waarin Kahneman en Tversky PT en Cumulative PT voorstelden meerdere malen doorgenomen. Waar ze daarin de geldigheid van EUT onderzochten deden ze dat consequent voor de volwaardige versie, d.w.z. met een concave nutsfunctie. Nergens deden ze dat voor wat Timmermans “de basale versie van EUT” noemt, d.w.z. die met een lineaire nutsfunctie. De gecursiveerde bewering lijkt dus ook gefingeerd.

³⁵ Van de Kaa, E.J. (2012). Timmermans’ misleading critique of Prospect Theory actually supports its relevance for travel choice modelling. *European Journal of Transport and Infrastructure Research*, vol. 12, no. 4, pp. 440-458.

³⁶ Timmermans, H. (2012). Relevance of Prospect Theory to Transport Choice Modeling: a Rejoinder. *European Journal of Transport and Infrastructure Research*, vol. 12 (4), pp. 459-462.

Hetzelfde geldt voor een aantal andere min of meer negatief getoonzette beweringen over PT (zie sectie 3).

In zijn repliek besteedde Timmermans geen aandacht aan deze bevindingen. Dat verbaasde me, omdat m.i. het doen van zulke onware beweringen wetenschapsfraude is, tenzij er een overtuigende alternatieve verklaring voor wordt gegeven.

Na correctie van onnauwkeurigheden, drogredenen, fouten en/of onware beweringen blijken de meeste kritieken in T gebaseerd op verschillen tussen de aannamen van PT en Timmermans' meningen over het keuzeproces van mensen. Het verloop van dat keuzeproces is, bij de huidige stand der kennis, niet waarneembaar. Daardoor zijn meningen erover niet falsifieerbaar en leveren ze geen wetenschappelijke onderbouwing voor conclusies over de toepasbaarheid van theorieën die keuzegedrag willen modelleren (zie sectie 5). In zijn repliek merkt Timmermans op dat zijn kritische commentaren in T zijn persoonlijke meningen waren en dat we verschillend aankijken tegen methodologische zaken, waarbij ik zou vinden dat 'only a detailed historical account, based on solid theoretical or empirical support is academic, and anything else is inferior' (repliek p. 460). Ik denk dat we hier inderdaad fundamenteel van mening verschillen. Anders dan Timmermans' repliek suggereert ben ik niet tegen het gebruik van persoonlijke meningen bij b.v. het formuleren van een onderzoekagenda of in een state-of-the-art artikel, maar m.i. moeten die dan wel expliciet als zodanig worden verwoord. De onderbouwing van de commentaren in T met argumenten en literatuurverwijzingen wekte bij mij niet de indruk dat ze als persoonlijke meningen bedoeld waren. Maar mijn bezwaar was vooral dat de toepasbaarheid van PT voor transportkeuze-modellering werd afgemeten aan de mate waarin PT's aannamen overeenkomen met Timmermans' niet-falsifieerbare eigen meningen over de onderliggende processen. Daardoor leveren deze commentaren m.i. geen wetenschappelijk verantwoorde inzichten op over de toepasbaarheid van PT.

Wanneer de kritieken in T die zijn gebaseerd op meningen over het keuzeproces buiten beschouwing blijven resteren verwijzingen naar studies waarin waargenomen en theoretisch voorspelde keuzen worden vergeleken. De empirische studies die in T werden besproken lijken te wijzen op een gebrekkige geschiktheid van PT voor transportkeuze onderzoek. Maar een nadere analyse ervan toonde aan dat meerdere studies de geschiktheid van PT eerder ondersteunden dan in twijfel trokken. Bovendien suggereerde een vergelijking met uitgebreidere reviews dat de keuze van de in T besproken empirische studies nogal selectief was. Die andere reviews leverden nogal wat andere empirische studies op die juist een goede geschiktheid van PT suggereren (zie Sectie 6). Desondanks wordt in Timmermans' repliek (p. 462) zonder aanvullende onderbouwing herhaald dat 'in case of routine departure time and route choice, the question is whether these choice problems meet assumptions and reasoning behind prospect theory'. Ik vond dat opvallend omdat sinds de publicatie van T er weer verschillende artikelen zijn verschenen³⁷ die aantonen dat PT goed toepasbaar is voor de modellering van het onderzochte routekeuzegedrag.

De slotconclusie in T (p. 381-382) was: 'at the current state of development, PT lacks the rigor, scope, behavioural principles and mechanisms, and content validity to serve as a comprehensive theory of how individuals and households dynamically (re-)organize their activities and travel ... Applications of (C)PT to these types of choices represent an attempt to apply the theory in the wrong contexts.' Zoals in dit werkdocument is aangetoond berust deze stelling op niet-falsifieerbare meningen, deels onderbouwd met ongeldige en/of onware beweringen en argumenten. Daarmee is de stelling zelf niet noodzakelijkerwijs onjuist. Maar gezien de resultaten van mijn heranalyse van de in T besproken empirische studies en die van de eerdere reviews concludeerde ik dat PT wel degelijk een breed toepasbare theorie biedt voor het verklaren en voorspellen van transportkeuzen (zie sectie 7). De repliek van Timmermans geeft m.i. geen aanleiding om deze conclusie aan te passen.

Hoeven, December 2012
Evert Jan van de Kaa

³⁷ Zie bijvoorbeeld Hjorth, K. and Ramjerdi, F. (2011), A Prospect Theory approach to travel time variability. International Choice Modelling Conference, Oulton Hall, Leeds; en Ramos, G.d.M., Daamen, W. and Hoogendoorn, S. (2011), Expected Utility Theory, Prospect Theory and Regret Theory compared for prediction of route choice behavior. *Transportation Research Record*, vol. 2230, pp. 19-28.