Too Good To Be True? An Assessment of the Melbourne Travel Behaviour Modification Pilot

Anthony Morton[1] and Paul Mees[2]
[1] Monash University, Melbourne, Vic, Australia
[2] University of Melbourne, Melbourne, Vic, Australia

1 Introduction

Governments, planners and analysts across Australia agree that mode shift from the automobile to walking, cycling and public transport is desirable for environmental, social and health reasons, but in all our major cities trends are heading in the opposite direction. Various remedies have been proposed, but all have their drawbacks. Road pricing, for example, is widely supported by transport planners, but is extremely unpopular with the public. Curtailing road expansion in favour of increased investment in public transport would be popular with environmentalists and many community groups, but is strongly resisted by road authorities and motoring organizations. Wouldn't it be nice if there was an inexpensive, pain-free way of moving towards sustainable urban transport?

Enter Travel Behaviour Modification (TBM), represented in Australia by TravelSmart (an adaptation of the IndiMark® concept) and Travel Blending®. TBM uses individualised marketing to change public perceptions of the attractiveness of more sustainable modes, with the aim of changing travel behaviour. Before we consider the effectiveness of TBM, it is worth reflecting on why this approach might be expected to gain wide support.

1.1 A political economy of Travel Behaviour Modification

The attractiveness of TBM to governments is obvious: it is much cheaper than investing in infrastructure or services, and less controversial than schemes like road pricing. At the other end of the scale, the source of attraction is indicated by the trade-mark symbols attached to products like Indymark®: TBM programs tend to be the property of consultants who, naturally, believe in their product and wish to sell it.

There is another, potentially more controversial, reason TBM might be popular. Programmes like TravelSmart can shift responsibility for solving transport problems from governments and transport planners to the public. Instead of asking why public transport is so poor or why pedestrians are marginalised to benefit motorised traffic, the approach seems to be telling the public that things are not as bad as they seem and to 'change its attitude'.

TBM therefore has the potential to be conscripted as part of what Beder (2000) calls 'greenwash' campaigns, in which governments and businesses responsible for environmental damage seek to deflect attention from themselves to individual members of the community. Indeed, Beder's description of the unofficial motto of greenwash campaigns is similar to that of TBM consultants: "It is easier and less costly to change the way people think about reality than it is to change reality" (Beder 2000: 109).

A similar point is made less 'politically' in a review of TBM techniques by the UK Department for Transport:

> It is clear that the techniques will only work 'on their own' where there is a large gap in perception between what exists and what people believe exists. For public transport, where services and travel quality [are] much higher than perceived, personalised

> approaches can have very large effects, but where such a gap does not exist the travel behaviour effects could be negligible… It would seem that they need to be thought of as an integral part of a strategy rather than as some form of 'public relations' exercise, when nothing substantive is being done to address strategic transport priorities (DfT 2002: 8.3).

1.2 TravelSmart in Australia

Commencing in Perth in 1997, TravelSmart has spread to all mainland capital cities, assisted by its adoption in 2003 by the Federal government. TravelSmart received $18.3 million from the Federal Greenhouse Gas Abatement Program, and remains the only transport project to have been funded under the program. Local government has become involved with TBM, through a Federally-funded initiative of the International Council for Local Environmental Initiatives called Cities for Climate Change™.

The enthusiasm of the Federal government for TravelSmart contrasts starkly with that government's hostility to other initiatives that might promote sustainable urban transport, from a carbon tax to funding for urban public transport. Numerous other local and state governments that are backing the programme are similarly notorious for their lack of interest in real change to transport priorities.

One such government is that of Victoria, which in 2002 released the *Melbourne 2030* metropolitan strategy, proposing to increase the share of motorised trips by public transport from 9 per cent to 20 per cent by 2020. Two years later, responding to criticism that no funding or other serious initiatives had been provided to meet this ambitious target, the government released *Linking Melbourne*, a "comprehensive and integrated transport plan". *Linking Melbourne* contained no commitment of funding either, but it did endorse TravelSmart in glowing terms, reporting that a pilot programme on the Alamein corridor in 2003 had reduced car driver trips by 10 per cent and increased public transport, cycling and walking by 27, 23 and 26 per cent respectively (DOI 2004: 35).

The Victorian government's support for TravelSmart appears to be a classic case of 'greenwash' as defined by Beder (and DfT, UK), since it is a substitute for, and diversion from, 'hard' policy changes. But this should not be allowed to obscure the potential of the technique: if the results reported for the Alamein trial are valid, TravelSmart has the potential to make dramatic changes to mode share at modest cost. While 'greenwash' and the large sums of public funding now potentially available to TravelSmart consultants have the potential to encourage wishful thinking about the program's potential, these factors do not automatically compel a negative verdict. Rather, they point to the need to examine very closely the credibility of the reported changes in travel behaviour.

1.3 Debating the effectiveness of TBM

There was a spirited debate about the effectiveness of TravelSmart at the 2003 ATRF. Stopher and Bullock (2003) raised a series of concerns about the reliability of reported behavioural changes and associated environmental benefits, focussing on a review of the results reported for TravelSmart/IndiMark® programs in Perth. A rebuttal was presented in a multi-author paper (Roth et al 2003), which in our view answered some, but not all, of the concerns raised in the original paper.

In particular, Stopher and Bullock commented on bus boarding data showing a large increase in patronage in South Perth in 2000, coinciding with a TravelSmart program there. This increase had been relied on as confirmation of mode shift changes reported in surveys

(Socialdata 2003: 13-14). The problem is that the increase in bus patronage appears to have occurred in February, while the individual marketing component of TravelSmart ran from March to June (Stopher & Bullock 2003: 10-11). The UK Department for Transport review expressed similar concerns about the Perth bus patronage data (DfT 2002: 5.26).

Roth et al respond (2003: 9) that "[t]he installation and commencement of the IndyMark® intervention began at the end of January 2000. Stopher's assertion that it began in March 2000 is incorrect." This response is not convincing. It is true that aspects of the South Perth programme began at the end of January, but the main activity at this time was the installation of improved signage at bus stops (DoT 2000: 4). The individualised marketing component did not commence until March, and one key component, home visits from bus drivers, did not start until April (DoT 2000: 7).

So not only must something else have caused patronage to rise in February, but there appears to be no discernible effect during or following the actual TBM programme. In fact, the most likely cause of the patronage increase was not individualised marketing, but improved bus services introduced in South Perth in January 2000 (Socialdata 2003: 14).

Further doubt on the validity, or at least the durability, of the behaviour changes brought about by TravelSmart in South Perth comes from census data on mode share for the journey to work. The 2001 census was conducted a year after the conclusion of the TravelSmart project, which covered the whole municipality of South Perth. Moriarty and Kennedy (2004) report the following census results for South Perth and the adjacent Central Metropolitan and Victoria Park districts (table 1), where no TBM programmes took place.

Table 1 **Mode share for journey to work, 1996 and 2001 (per cent)**
(source: Moriarty & Kennedy 2004)

	Central metro		Victoria Park		South Perth	
Year	1996	2001	1996	2001	1996	2001
Car	78.2	75.5	76.1	75.9	83.7	81.9
Public transport	11.9	13.2	15.8	16.8	10.7	11.2
Walk/cycle	8.1	9.5	5.9	5.6	3.9	4.8

The mode share for walking, cycling and public transport increased across inner Perth, as it did over the same period in inner Melbourne and Sydney. The most likely explanation is an increase in the share of workers employed in the central business district, due to the combined effects of central employment growth and gentrification of the inner city. It is difficult to see any impact from the previous year's TravelSmart programme in South Perth, since the change was similar to that recorded in adjacent districts.

1.4 The central issue: reliability of survey results

These results are difficult to reconcile with the survey results reported in the South Perth study, which reported similar large changes to those claimed for Melbourne's Alamein corridor in 2003. This highlights a concern about the mode share changes reported from TBM projects which was not raised by Stopher and Bullock (2003), but which has been noted by Moriarty and Kennedy (2004), namely the influence of errors arising from survey 'artifacts'.

2 A social psychology of Travel Behaviour Modification

To obtain credible results, a TBM evaluation study must be designed to control for any foreseeable source of bias or systematic error. Those conducting the evaluation will therefore take sensible precautions such as comparing the outcomes for the target group with those for an appropriately selected control group (members of which are excluded from TBM interventions), and endeavouring to survey the same individuals before and after the TBM intervention.

However, any study of the effectiveness of TBM is necessarily a study of human behaviour. The fact that the experimental subjects are people (rather than inanimate objects) introduces a set of subtle yet potentially significant sources of systematic error, usually called 'artifacts' in the psychology literature where they are extensively studied.

> We have learned that much of the complexity of human behaviour is inherent, but we have also learned that some of this complexity may result from uncontrolled aspects of the research situation, especially from the interaction between the researcher and the participant....
>
> Artifacts are not simply inconsequential effects in a research design; they may actually jeopardise the validity of the researcher's inferences from his or her results. Another way of saying this is that artifacts are unintended or uncontrolled human aspects of the research situation that confound the investigator's conclusions about what went on in the study (Rosnow and Rosenthal 1997: 2–3).

2.1 The expectancy effect

One example of an experimental artifact is *expectancy bias*. This occurs when experimenters' expectations about the result of a study inadvertently become a self-fulfulling prophecy due to feedback effects between experimenter and participant.

> Behavioral researchers, like other scientists generally, conduct research specifically to test hypotheses or expectations about the nature of things. When the researcher's hypothesis or expectation leads unintentionally to behavior toward the research participants that increases the likelihood that the researcher's expectation will be confirmed, we call this an *expectancy effect* (Rosnow and Rosenthal 1997: 42–43).

It is important to understand that expectancy effects and other artifacts are problematic in the *evaluation* phase of a TBM project, even where they present no difficulty for the TBM intervention itself. If a positive attitude on the part of TBM campaigners contributes to a greater shift in travel behaviour than a neutral or negative attitude, one is entitled to draw a positive conclusion to that effect. But if a researcher testing the effectiveness of a TBM campaign behaves in a way that induces participants to *say* they changed their behaviour when in fact they did not, the result of the entire study is thrown into question. The former is a useful marketing tactic; the latter is an expectancy artifact.

In other situations, expectancy effects are controlled through the familiar 'blind experimenter' approach, where experiments are designed to ensure that the data collector does not know whether the experimental target group or a control group is being observed. In TBM evaluation studies, although travel mode shifts are typically tested against those in a control group, there is scant indication that the target group is treated relative to the control group in a properly 'blind' fashion. As will be seen below, expectancy bias can intrude even in the statistical analysis of data after it has been collected.

2.2 The 'good subject' effect, and the behavioural guilt trap

Many artifacts identified in behavioural research arise from the experimental participants themselves. Among the most fundamental is what psychologist Martin Orne (1962, 1970) dubbed the 'good subject' effect, which arises from the participant's desire to please the researcher:

> [Orne] noted that, at the conclusion of many of his experiments, the participants often asked questions such as 'Did I ruin the study?' After postexperimental interviews with his participants, Orne deduced that what they had meant was 'Did I perform well in my role as experimental subject?' or 'Did my behaviour demonstrate what the study was designed to show?' (Rosnow and Rosenthal 1997: 64)

Because the usual method of evaluating TBM programs is through self-reporting surveys rather than direct observation of travel behaviour, TBM studies are particularly vulnerable to participant-related artifacts. They are also more vulnerable to 'good subject' effects when the experimental target is a group that self-selects as wanting to change their behaviour. The effect can also arise in a different but related form, where the motive is not so much to please, as to avoid being evaluated negatively by the researcher. Social psychologist Martin Rosenberg (1969) provided early experimental evidence of this effect, which he called *evaluation apprehension*.

> [Rosenberg] said that typical participants approached the psychology experiment anticipating that the experimenter would evaluate their psychological competence. Not surprisingly, most participants became apprehensive about being evaluated negatively (or at least not positively), and they developed their own hypotheses about how to win approval and to avoid disapproval (Rosnow and Rosenthal 1997: 68).

While Rosenberg's observations were in a psychological context, a similar effect occurs whenever participants' behaviour is evaluated against a socially desirable norm, for example when studying the effectiveness of programmes aimed at getting people to quit smoking or to increase their fitness. This kind of *behavioural guilt* is also to be expected in TBM studies, particularly when the target group self-identifies as seeking behaviour change.

Another way to view the psychology of the TBM participant is using the concept of an 'investment trap' described by Plous (1993: 243–244). When people are selected, or select themselves, into a program where they receive incentives such as free tickets, travel advice, and home visits by behaviour-change consultants, they become conscious of the ever-growing investment being made in their personal travel behaviour. When the time finally comes to complete the 'after' survey, there is an understandable reluctance among those whose behaviour has not changed to admit this fact.

Rosnow and Rosenthal (1997: 81–85) discuss strategies for minimising the occurrence of 'good subject' and related artifacts, for example by making the study objectives deliberately vague; by introducing bogus objectives; or by using strategies that encourage honest reporting, such as subject anonymity. Unfortunately, typical TBM studies display all the characteristics that lead to questionable results in this regard: they use before-and-after survey techniques, spaced relatively close together in time, on volunteer participants who are identified to the researchers, told exactly what behaviour is 'desirable', are themselves motivated to report a desirable change in behaviour, and whose own reports are relied on by the researchers to obtain their results.

2.3 Sampling and nonresponse bias

Darrell Huff's classic work on common errors in statistical reasoning (Huff 1954) begins with a warning on the inherent bias in a self-selecting sample. There is now a very large body of research pointing to systematic ways in which volunteers differ in aggregate from the general population. Rosnow and Rosenthal (1997: 97–104) collect the results of several hundred studies and conclude that several systematic differences exist: for example, volunteers tend to be more educated, be more seeking of approval, and have higher social status than nonvolunteers. When recruited for specific studies, volunteers tend to be those who are particularly interested in the topic under investigation and have an expectation of being favourably evaluated.

The distinction reveals itself also when surveys are used to obtain data for studies, as is usual practice in TBM. The systematic error that arises due to the difference between respondents and nonrespondents is termed *nonresponse bias* in the literature. In general one cannot quantify the extent to which nonresponse bias alters a result (save that the effect is likely to be small in the fortunate case where the response rate is high). However, knowledge of the general characteristics of volunteers and the particular characteristics of the study situation usually allows the researcher to predict the direction in which the bias is likely to occur.

In the case of TBM surveys there are two important predictions that follow from what is known about volunteering behaviour. First, those who have a high interest in transport issues and are strongly motivated to change their travel behaviour are more likely both to respond at the outset and to continue responding throughout the study. Second, those who are caught in the behavioural guilt trap of the previous section have a strong incentive not to respond to the 'after' survey; this enables them to avoid being unfavourably evaluated on the one hand, and to avoid giving a false report on the other.

3 A case study: TravelSmart in Alamein

By way of illustrating the principles of the previous section and the manner in which they may bias a real study, we now turn to a detailed analysis of the results of the Alamein TravelSmart pilot study referred to in the Introduction.

Our analysis is necessarily limited to the publicly available results presented in the final report for the TravelSmart Alamein project (SocialData / ITIR 2004). This report is the source for the claim in *Linking Melbourne* that the project produced increases in walking, cycling and public transport mode share of up to 27 per cent, and reductions in car mode share in the order of 10 per cent. These behaviour shifts are supposed to have resulted solely from the IndiMark® brand of TBM conducted over two months in 2003, without any alterations in the quality of transport service. If true, this would constitute significant evidence in favour of the efficacy of TBM.

(Note: for the sake of clarity in what follows, all figures for changes in transport mode share are *relative* changes except where specified otherwise. Thus, if public transport used to represent 10 per cent of trips and now represents 11 per cent of trips, the change in mode share is 10 per cent, not 1 per cent.)

3.1 When 27 per cent really means 12 per cent

In fact, at least half of the reported mode shift results from a basic error of interpretation. The target population for the study was a group of 6,465 households in the vicinity of the Alamein

train line in Melbourne. Also surveyed was a separate 'control' group of 413 households from the same geographical area. The mode share results for the target group and the control group are reproduced as Table 2 below, taken from Table 5.6.3 of the Socialdata report:

Table 2 **Reported mode shares before and after TBM (per cent)**
(Source: SocialData / ITIR 2004: 43)

Mode	Target Group			Control Group		
	Before	After	Change (%)	Before	After	Change (%)
Walking	12	14	+17	13	12	-8
Bicycle	1	1	+18	1	1	-3
Motorbike	0	0	0	0	0	+1
Car as driver	52	49	-10	54	56	+3
Car as passenger	26	25	-6	24	24	+1
Public transport	9	11	+12	8	7	-12

It should be noted that in order to obtain accurate figures for the percentage change in mode share, we have had to use other tables in the report (Table 5.6.7 for the target group, and Table 5.6.4 for the control group). Throughout the report, all percentages are reported only to the nearest integer—hence the reporting of the motorcycle mode shares as zero. This means that using just those figures reported in Table 5.6.3, for example, the actual change in public transport mode share in the target group could be as much as 35 per cent (8.5 per cent to just under 11.5 per cent) or as little as 10 per cent (just under 9.5 per cent to 10.5 per cent). Fortunately, Table 5.6.7 presents the before-and-after statistics as absolute numbers of trips rather than just percentage mode shares, and it is evident from this table that the change in public transport mode share resulting from SocialData's analysis is in fact 12 per cent.

So why did *Linking Melbourne* report the mode share change as 27 per cent? Because in the discussion following Table 5.6.3, the SocialData report explains that the mode share change should be 'corrected' for the 'control group effect'. Essentially the researchers have noticed that the mode shares in the control group changed between the 'before' and 'after' studies, have assumed that this is evidence for an actual mode shift (due to "seasonal and external influences") rather than a sampling artifact, and have inferred that, but for the TravelSmart intervention, the target group's mode shares would have changed by the same amount. So, they explain, the public transport mode share of the target group would actually have *fallen* by 12 per cent without the marketing action, but instead rose by 12 per cent; the real effect of the intervention is therefore 1.12 divided by 0.88, or a 27 per cent increase.

There are several problems with this reasoning. First, the influence of trends or seasonal influences on transport mode shares at the population level over a period as short as six months can be discounted; as Seethaler and Richardson (2003: 25) explain, the available evidence indicates that no such effects currently exist:

> For public transport... the report of Booz Allen Hamilton… does not provide results that would indicate any long-term trend or strong seasonality.... All in all, the analysis of VATS 1994–1999 data indicates no specific need to adjust for seasonality when conducting 'before' and 'after' studies related to a TravelSmart policy implementation. However, attention needs to be paid to choosing the same weekdays for the 'before'

and the 'after' surveys to control for the existing patterns with regards to the day of week.

Second, the reported 12 per cent fall in public transport mode share (from 8 per cent to 7 per cent) in the control group does not in any case appear to be statistically significant. The appropriate significance test here is a standard difference-of-means test with binary outcomes (use or non-use of public transport) with N = 413 and a 1 per cent difference in sample means. (The sample size N is based on households rather than persons or trips, because it is factually inaccurate to regard trips by members of the same household, or trips by the same person at different times, as logically independent events.) The test relies on one statistic not provided in the report, namely the actual number of households that changed from use to non-use of public transport, or vice versa, between the two surveys. But even making the worst-case assumption that all the change was from use to non-use, and none in the other direction, the null hypothesis that the underlying means are equal is barely rejected at the 95 per cent significance level (and thus not rejected at any higher level of significance).

One also cannot completely discount the psychological factors operating within the control group, members of whom could not have failed to notice that their travel habits were surveyed in detail once, then once again after six months, a sure signal that behaviour change research is taking place. It is entirely possible that if individuals in the control group knew (or had simply inferred) that others in the area were taking part in a TravelSmart campaign, they might manifest a 'good subject' effect of their own by (consciously or otherwise) under-reporting their public transport, walking and cycling and/or over-reporting their car use to ensure that the researchers got the comparative result they wanted.

The 'control group effect' should therefore be presumed to be an artifact resulting from chance variations, particularly as on prior evidence one does not expect any such systematic change within the control group over such a brief period. Thus the 'real' figure for public transport mode shift is the 12 per cent actually obtained for the target group—assuming that this figure represents a correct interpretation of the survey data. As we show next, even this cannot be taken for granted.

3.2 When 12 per cent might mean zero

While it is debatable whether the control group would have received strong enough cues to be prompted into 'good subject' behaviour, there can be no doubt that among the participants within the target group, all factors leading to the expression of participant-mediated artifacts are in operation. That is: the study objectives are clear to the participants; the participants are motivated to report behaviour in line with those objectives (whether or not they actually did behave in this way); and they are given the opportunity to influence the results via the self-reporting nature of the surveys.

To have a significant misleading effect on the final results, the magnitude of the errors resulting from operation of behavioural guilt traps and other such artifacts need not be large. The study report itself is candid about the fact that even small changes in reported behaviour from the surveys can greatly affect the evaluation results:

> It is worth noting that a significant reduction in 'car as driver' trips needed a change, on average about one trip per week. Therefore, *a very small change in household behaviour is significant in aggregate results!* (Socialdata / IRL 2004: 33, emphasis in original)

Based on the discussion in Section 2, the study is vulnerable to the following effects in

particular:
• nonresponse bias arising from reluctance of those caught in the 'behavioural guilt trap' either to report a failure to change behaviour or to make a false report; and
• systematic under-reporting of car use or over-reporting of public transport use (whether intentional or not) in the 'after' survey resulting from the 'good subject' effect.

The response rates for the target group are noted in a table on page 41 of the Socialdata report. In all, 636 households responded to the 'before' survey, which represents 73 per cent of the total sample. When these same 636 households were contacted for the 'after' survey, 530 responded, giving a response rate of 83 per cent. Put another way, 27 per cent of households failed to respond to the 'before' survey, and of those who did respond, 17 per cent failed to answer the 'after' survey. Interestingly, if one considers total persons rather than households the response rate for the 'after' survey drops to 79 per cent; this indicates that larger households were somewhat less likely than smaller households to respond at this final stage.

Nonresponse bias can therefore enter at two stages—'before' and 'after'—and in each case biases the result to the extent that there is a systematic link between travel behaviour and likelihood of nonresponse. Because a matched sample is used to measure the change in behaviour, nonresponse to *either* survey alters the composition of the sample in *both* survey results. Thus, the final result will be biased whenever the actual change of behaviour differs significantly between respondents and nonrespondents to either survey, and in each case will exaggerate the reported behaviour change if the actual change was lesser among the nonrespondents.

Intuitively, one might expect that nonresponse to the 'before' survey is an indicator of non-interest in the TravelSmart process as a whole, and therefore that the 27 per cent of households who did not respond at this stage would, if surveyed, display less behaviour change as a group than the 73 per cent who did respond. The bias resulting from narrowing the survey sample to just this latter group is then likely to be significant.

The case of those who failed to respond to the 'after' survey is quite different, as these people *were* motivated to respond to the 'before' survey. But as we noted above, a significant factor in nonresponse to the 'after' survey is likely to be 'guilt effect' from program participants who failed to change their behaviour. So in both the 'before' and 'after' cases there are sound reasons to expect nonrespondents to have changed their behaviour less than respondents, and therefore to expect the nonresponse bias to exaggerate the apparent behaviour change. The likely bias from reporting errors due to the 'good subject' effect will compound this effect.

3.3 Weighting of results or amplification of artifacts?

In an attempt to compensate for nonresponse bias, the report's authors have applied a 'weighting' procedure to the survey figures. When carefully examined the effect of this procedure is to *amplify* the most likely source of bias, not correct it. The procedure is described as follows:

> For the target group there was a weighting of the after data (which was not needed for the before data).... the distribution of the IndiMark® groups ('I', 'R' and 'N') in the survey sample was corrected accordingly to the one in the IndiMark® campaign to balance differences in the response behaviour of the single groups. People not participating in the IndiMark® campaign were considered as occurring [*sic*] no changes in their travel behaviour (Socialdata / IRL 2004: 42).

In the TravelSmart exercise itself, households were classified into four groups: those interested in increasing their use of sustainable transport ('I'); those who already use sustainable transport regularly but would like further support ('R with'); users of sustainable transport who require no further support ('R without'); and those not interested in further contact ('N'). The 'I' and 'R with' groups received the full marketing treatment, so we may for simplicity speak of them collectively as the 'participants' and the remaining groups (as well as those not successfully contacted) as the 'non-participants'. The participants accounted for 3,505 of the 6,465 households in the target area, or 54 per cent.

In simplified form, the 'weighting' procedure (as inferred by us from the above statement) scales the results by appropriate factors to reconstruct what would presumably have been observed, had the proportion of participants in the final 530-household survey sample matched the 54 per cent proportion in the population. This is done on the basis that the entire observed behaviour change is attributable to the participants, and none to the non-participants. For example, if the proportion of participants in the sample is only 27 per cent rather than 54 per cent, all differences between the 'before' and 'after' figures should be doubled to obtain the 'true' result across the entire population.

It is, of course, the objective of this evaluation exercise to demonstrate that TravelSmart is an effective way to increase the use of sustainable transport. To introduce this as an assumption during the analysis provides a convenient short cut to such a demonstration, but is logically questionable. In fact this leads to a form of expectancy artifact, albeit one resulting from the optimistic treatment of data rather than of people. In the presence of 'good subject' artifacts, the two mutually reinforce and lead to a larger error than either acting in isolation—*even when the assumption is true*.

We illustrate this point with a hypothetical example. Suppose that the 636 households in the 'before' survey are a perfectly representative sample, and 343 of them (54 per cent) become full participants in the TravelSmart campaign. Among the entire participant population, one-third make a minor change from 10 per cent to 11 per cent public transport use while the other two-thirds make no change at all. Again let the survey sample be perfectly representative, so that 114 of the 343 participants (one third) increase their public transport use while the rest do not. In keeping with Socialdata's assumptions, the non-participants also make no changes in their public transport use.

The 'after' survey collects just 530 responses out of 636; we suppose this is due to 106 households (all participants who did not change their behaviour) being caught in a behavioural guilt trap and failing to respond. The final sample of 530 includes 237 participants (114 of whom changed their behaviour) plus 293 non-participants; thus the proportion of participants in the final sample is 45 per cent rather than 54 per cent.

In this example, the average absolute mode shift to public transport in the target population is 0.18 per cent (1 per cent, times one-third, times 54 per cent). In the raw survey result, the 'good subject' artifact raises the average mode shift to 0.22 per cent. To this the weighting procedure applies a scaling factor of 1.2 (54% divided by 45%) resulting in a reported mode shift of 0.26 per cent (in absolute terms). Rather than correct for the nonresponse bias, the weighting has amplified it, making the result look 43 per cent better than it is.

(In a very similar hypothetical example, the change from 10 per cent to 11 per cent public transport use is actually the result of misreporting by 114 respondents keen to play the role of good experimental subjects. In this case there is in fact no mode shift in the population, but the procedure leads to the reporting of a mode shift of 0.26 per cent regardless.)

3.4 Evidence for artifacts

It may be argued that to this point, our case for an exaggerated mode shift result has been speculative rather than empirical. Is there any statistical evidence in the reported figures supporting the hypothesis that artifacts are present? The answer is yes.

A crucial test is whether the proportion of programme participants is significantly less in the 'after' survey than in the 'before' survey. If this is the case it indicates a systematic tendency for participants to be less willing than non-participants to respond to the 'after' survey. Since non-participants have no special motivation to respond at higher rates than participants, while participants have already expressed a motivation to cooperate in a behaviour-change experiment, this in turn suggests that some participants are being deterred from responding, most likely as a result of having no 'good news' to report. In any event, a statistically significant difference in the proportion of participants between a survey sample and the population should call into question any result drawn from the sample, simply because of its manifestly unrepresentative nature.

The Socialdata report provides no explicit figures on the numbers of each IndiMark® group present in the survey sample. Nonetheless, Table 5.6.2 in the Socialdata report compares the mode share figures for the target group before and after the weighting procedure described above. Since the effect of weighting is to increase the public transport mode share at the expense of car trips, it can be inferred that the proportion of participants in the survey sample is indeed less than the 54 per cent in the population.

To test the significance of the discrepancy, we again have recourse to Table 5.6.7, which implies that the reported mode share for 'car as driver' after weighting is 48.8 per cent to three significant figures. From Table 5.6.2, the mode share before weighting is 50 per cent to the nearest integer. Therefore, the weighting procedure has reduced the 'car as driver' mode share by at least 0.65 per cent in absolute terms.

Again from Table 5.6.7, the mode share for 'car as driver' in the 'before' survey is 52.3 per cent, to three significant figures. The absolute reduction in mode share indicated by the *raw* survey result is thus no less than 1.8 per cent (to 50.5 per cent) and no greater than 2.8 per cent (to 49.5 per cent). It follows that the mode shift has been scaled by an amount no less than 1.23 (2.8 plus 0.65, divided by 2.8), and so the proportion of participants in the survey sample can be no greater than 54% divided by 1.23, or 44 per cent.

Thus, despite the fact that the report provides no direct information about the number of participants and non-participants in the survey of 503 households, we have been able to show that the proportion of participants cannot be greater than 44 per cent. Again applying a difference-of-means test with binary data, the difference between 44 and 54 per cent with a sample size of 530 is shown to be statistically significant at levels beyond 99.9 per cent (the Z-statistic is 4.62). It is thus extremely unlikely that such a discrepancy could have arisen by chance.

So the survey results themselves reveal that a significant number of participants in the TravelSmart programme did not respond to the follow-up survey, suggesting that they had not changed their behaviour. Although the results do not permit direct testing for the opposite 'artifact'—i.e. people exaggerating the extent to which their travel behaviour did change—it is likely that this phenomenon is also present.

4 Conclusion

TBM programs such as TravelSmart carry the promise of 'something for nothing'. They appeal to policy makers because they claim to increase the use of sustainable transport without requiring significant changes to transport policies. Whether motivated by 'greenwash' or a genuine desire to improve the environment, such policy makers have a strong incentive to want TravelSmart trials to succeed. TravelSmart consultants also want trials to succeed, because this will vindicate their belief in their own product (as well as ensuring future consultancies). Members of the public participating in the trials also have incentives, arising from 'good subject' effects and associated processes, favouring a positive outcome. The combined effect of these 'political economy' and 'social psychology' influences is to create a high likelihood that changes in travel behaviour apparently caused by TravelSmart are in fact wholly or largely artifacts.

Careful analysis of TBM evaluation studies, such as the South Perth programme and the Alamein pilot project, confirm the presence of artifacts and suggest that the promise of something for nothing may be too good to be true. Most such TBM studies have ignored the very significant human factors, well-documented in the psychology literature, that can introduce systematic errors into the research results. Typically, studies are designed in a way that exacerbates these factors and magnifies the resulting errors, as we have shown in the case of the Alamein study.

Unfortunately, the nature of the study objectives makes it difficult to avoid the intrusion of artifacts such as the 'good subject' effect, the behavioural guilt trap, and nonresponse bias. The main problems arise from the use of a self-reporting survey framework, that provides the opportunity for participants to act on a motivation to report desirable behaviour, or to avoid reporting undesirable behaviour by not responding.

We recommend that the following changes be made to the evaluation of Travelsmart and other TBM programmes:

1. The evaluation of TBM programmes should be carried out by parties completely independent of the consultants conducting the TBM intervention itself—as is occurring, for example, with the TravelSmart trial in the Melbourne municipality of Darebin. Independent evaluation will not of itself overcome the problem of artifacts, but it is a good start.

2. Assessment of changes in travel behaviour should not be carried out primarily through self-reporting-based surveys. Instead, information sources that are less likely to be contaminated by artifacts should be preferred, such as census data, counts of public transport boardings and pedestrian flows, odometer readings and other measures of VKT, possibly even GPS data. A good discussion of possible methodologies appears in the paper at this conference by Stopher, Greaves, Xu and Lauer.

3. It may be that in order to obtain credible results, researchers will have to resort to different survey methodologies whereby the researchers themselves observe and report on travel behaviour by the target group. Ideally this should be done in a 'blind' fashion, where the observers are denied knowledge of whether the people being observed are experimental participants or not.

In the meantime, researchers and policy makers should be skeptical of the claims made on behalf of Travel Behaviour Modification techniques such as Travelsmart. At present, we have no reliable evidence that they produce real changes in travel behaviour.

5 References

Beder, S (2000) *Global Spin: The corporate assault on environmentalism* Melbourne: Scribe.

DfT (Department for Transport) (2002) *A Review of the Effectiveness of Personalised Journey Planning Techniques* London: DfT.
www.local-transport.dft.gov.uk/travelplans/pjourney

DOI (Department of Infrastructure) (2004) *Linking Melbourne: Metropolitan Transport Plan* Melbourne: Victorian Government.

DoT (Department of Transport, WA) (2000) *TravelSmart: South Perth Large-Scale Individualised Marketing Project* Perth: DoT.

Huff D. (1954) *How to Lie with Statistics* London: Victor Gollancz.

Moriarty, P and Kennedy, D (2004) Voluntary travel change: an Australian case study, presented at *Third International Conference on Traffic and Transport Psychology* Nottingham, UK.

Orne M.T. (1962) On the Social Psychology of the Psychological Experiment: With Particular Reference to Demand Characteristics and Their Implications *American Psychologist* vol. 17, 776–783.

Orne M.T. (1970) Hypnosis, Motivation, and the Ecological Validity of the Psychological Experiment, pp. 187-265 of *Nebraska Symposium on Motivation* Lincoln: University of Nebraska Press.

Plous S. (1993) *The Psychology of Judgment and Decision Making* New York: McGraw-Hill.

Rosenberg M.J. (1969) The Conditions and Consequences of Evaluation Apprehension, pp. 279-349 of Rosenthal, R. and Rosnow, R.L. (eds) *Artifact in Behavioral Research* New York: Academic Press.

Rosnow R.L. and Rosenthal R. (1997) *People Studying People: Artifacts and Ethics in Behavioral Research* New York: Freeman.

Roth, M et al (2003) A Dialogue on Individualised Marketing: Addressing misperceptions, CD ROM *Papers of the 26th Australasian Transport Research Forum* Wellington: ATRF.

Seethaler R.K. and Richardson A.J. (2003) *Seasonality in Daily Travel Patterns* Melbourne: The Urban Transport Institute, Report 24–2003.

Socialdata Australia Pty Ltd and Institute for Transport and Infrastructure Research Ltd. (2004) *TravelSMART Alamein Line Public Transport Project: Final Report* Melbourne: Department of Infrastructure Tender Number TS091.

Socialdata Australia Pty Ltd (2003) *Technical Appendix: Travel Behaviour Change Program for the City of South Perth under the TravelSmart Program* (submitted to Department for Planning and Infrastructure, WA) Fremantle: Socialdata.

Stopher, P and Bullock, P (2003) Travel Behaviour Modification: A critical appraisal, CD ROM *Papers of the 26th Australasian Transport Research Forum* Wellington: ATRF.