

**Analysis of the European Commission's Proposal to Revise
the Tobacco Products Directive as it Pertains to Various
Design Restrictions for Tobacco Products**

Prepared by:

Professor Timothy M. Devinney

1 June 2013

Table of Contents

1.	REPORT INTRODUCTION	2
2.	CRITERIA FOR EVALUATION	5
3.	REVIEW OF THE STUDIES	16
	DETAILED ANALYSIS OF THE STUDIES	19
	Borland and Savvas (2012).....	19
	Mutti, et al. (2011).	23
	Moodie, Ford et al. (2011).	25
	Moodie and Ford (2011).	28
	Doxey & Hammond (2011).	30
	Hammond & Parkinson (2009).	33
	Hammond et al. (2009).....	35
4.	CONCLUSION	39
5.	EXHIBIT ONE – QUALIFICATIONS OF PROFESSOR DEVINNEY	42
6.	EXHIBIT TWO – RESUME	44
	Examples of Research Support Received	47
	Professional Consulting Experience (Selection)	49
7.	EXHIBIT THREE – SELECTED PUBLICATIONS	50
8.	EXHIBIT FOUR – OTHER MATERIALS CONSIDERED	56

1. REPORT INTRODUCTION

1.1 My full name is Professor Timothy M. Devinney. I am a Professor of Strategy at the University of Technology, Sydney (“UTS”) in Australia. In addition, I am a Conjoint Professor in the Faculty of Medicine at the University of New South Wales, Australia and a Visiting Professor at the Institute of Management at Humboldt, Berlin. I am an academic trained in the areas of Psychology, Public Policy, Economics, Statistics and Management. I have extensive experience in the conduct and evaluation of consumer research studies, both from an academic and commercial perspective. I have been involved most recently in an extensive set of research projects examining the degree to which social aspects of consumption influence behaviour.

1.2 Exhibits One to Three of this report set out in detail my professional qualifications, my current resume and a list of sample publications that I have written. However, in summary terms and amongst other areas of expertise, I am an expert in consumer survey research, experimental methods and associated statistical analysis. I have specialised knowledge in assessing the methodology of consumer survey research to determine the extent to which it provides credible, methodologically and empirically sound evidence (which I refer to in this report as “reliable evidence”) in support of stated conclusions. Although not limited to this area, I have extensive experience with these issues in the context of consumer goods. I have been requested to prepare this report for Japan Tobacco International (“JTI”) (I describe in further detail below the basis on which I have prepared this report).

1.3 I am the author of the following reports previously prepared for JTI:

- (a) “Analysis of Consumer Research Evidence on the Impact of Plain Packaging for Tobacco Products” dated 30 November 2010 (the “2010 Report”); and
- (b) “Analysis of Consumer Research Evidence on the Impact of Plain Packaging for Tobacco Products” (Updated to 2012) dated 29 June 2012 (the “2012 Report”).

1.4 The European Commission published its Proposal to Revise the Directive of the European Parliament and of the Council on the approximation of the laws, regulations and administrative provisions of the Member States concerning the manufacture, presentation and sale of tobacco and related products (the “Draft

Directive”) on December 19, 2012.¹ The following measures are included in the Draft Directive, which I will refer to together in this report as “various design restrictions”:

- (a) Requiring that cigarette packs must be cuboid, made of carton or soft material, have a flip-top lid, and contain at least 20 cigarettes;
- (b) Prohibiting the use of “misleading colours”; and
- (c) Prohibiting cigarettes with a diameter of less than 7.5mm (so-called “slim” cigarettes).

1.5 The European Commission also published an Impact Assessment (the “IA”) to accompany the Draft Directive that includes the evidence considered in respect of the various design restrictions. The IA makes the following conclusions in this regard:

- (a) *“Some of the current pack shapes make it difficult to effectively display health warnings affecting negatively the visibility and legibility of the warning. This is particularly the case for very narrow (including “lip-stick” shaped) packets which distorts text and picture warnings”;*²
- (b) *“Recent studies have demonstrated that packages have the potential to mislead smokers and potential consumers and present them with an erroneous comfort about the risk of smoking. For instance, FMC packets featuring the descriptors ‘slim’ or ‘extra slim’ were rated significantly more appealing than packets without those descriptors. In another study of young adults, so-called ‘super slim’ ‘perfume type’ FMC packages were associated with femininity, elegance, slimmness and reduced harm”;*³ and
- (c) *“Likewise, the shapes (e.g. slim) and colours (e.g. pink, black, denim blue) of individual FMC can mislead consumers by creating e.g. the impression of harmlessness. A study found that smokers of ‘slim’ FMC were more likely to believe that some FMC could be less harmful and that their own brand might be a little less harmful. A recent study in young Australian adults has shown*

¹ The Draft Directive is available at: http://ec.europa.eu/health/tobacco/docs/com_2012_788_en.pdf.

² See Page 31, Impact Assessment.

³ See page 32, Impact Assessment.

that the characteristics of the cigarette stick affect smokers' perceptions of the attributes of cigarettes".⁴

1.6 This report presents my evaluation of the consumer survey studies presented in the IA as they pertain to the various design restrictions (the "Studies") and to address the extent to which they provide reliable evidence that the various design restrictions would be effective in achieving the public policy goals of changing actual smoking behaviour, namely in:

- (a) reducing smoking uptake (also known as smoking initiation) among minors;
- (b) reducing smoking among minors and/or adults; or
- (c) increasing smoking cessation among minors and/or adults.

1.7 Two points are relevant in this regard:

- (a) A number of Studies that are considered in the IA are studies that I have previously reviewed in my 2010 and 2012 Reports. I have indicated where this is the case in respect of the individual studies discussed further in Section 3 below. While the conclusions in my 2010 and 2012 Reports were in respect of plain packaging, these conclusions were reached on the basis of an assessment of whether the methodology used was sound and credible using the evaluation criteria outlined below. I, therefore, confirm that the views in my 2010 and 2012 Reports in respect of these studies with remain the same in the context of the various design restrictions.
- (b) I have based my opinions and conclusions in this report on studies presented in the IA in support of the various design restrictions that presented original consumer research. Studies that did not generate or evaluate consumer evidence relating to the effectiveness of the various design restrictions in achieving the public policy in paragraph 1.6 above were not considered in the formulation of my opinions. For completeness, however, I have included a list of these studies in Exhibit Four.

1.8 In summary, and as discussed further in Section 3 of this report, it is my expert opinion that none of the Studies provides reliable evidence that the various design

⁴ See page 32, Impact Assessment.

restrictions would be effective in achieving the public policy goals of changing actual smoking behaviour.

1.9 For ease of exposition, I repeat the evaluation criteria employed in my 2010 and 2012 Reports and in Section 2 below.

2. CRITERIA FOR EVALUATION

2.1 In evaluating the Studies a number of factors come into play in determining their validity with respect to addressing the questions being set. As in my 2010 and 2012 Reports, I build on the criteria outlined by Dr. Keegan.⁵ For ease of exposition, I repeat the outline I presented in my 2010 Report as way of a summary of the criteria discussed in detail by Dr Keegan. This table also provides a brief definitional description of the criteria.

2.2 I would also note that in line with my 2010 and 2012 Reports and Dr Keegan's Reports I have applied the same professional standards for researchers as outlined in the various codes of conduct from groups such as the UK-based Market Research Society (MRS),⁶ the American Association for Public Opinion Research (AAPOR),⁷ the World Association for Public Opinion Research (WAPOR),⁸ the Council of American Research Organizations (CASRO),⁹ the International Statistical Institute

⁵ As discussed in paragraph 1.3 of my 2010 Report, I have reviewed the following documents prepared by Dr Warren Keegan, which review, amongst other things, publicly available consumer survey evidence cited in support of plain packaging for tobacco products: (i) "Analysis of Consumer Survey Evidence Relevant to the UK Department of Health Consultation on the Future of Tobacco Control" dated 2 September 2008; (ii) "Analysis of Consumer Survey Evidence Relevant to the UK Department of Health Consultation on the Future of Tobacco Control – a supplemental report" dated 19 June 2009; (iii) "Analysis of Consumer Survey Evidence Relevant to DG SANCO's Proposal to Increase the Size of Health Warnings on Tobacco Packaging" dated 24 November 2010. In addition, I have also reviewed the document prepared by Dr Keegan entitled "Analysis of Consumer Survey Evidence Relevant to the Display Ban Requirement in England" dated 28 April 2010. In this report, I refer to these four documents collectively as "the Reports".

⁶ Market Research Society (2005), *MRS Code of Conduct*. <http://www.mrs.org.uk/standards/downloads/code2005.pdf>.

⁷ American Association for Public Opinion Research (2005), *AAPOR Code of Professional Ethics & Practices*. <http://www.aapor.org/aaporcodeofethics>.

⁸ World Association for Public Opinion Research, *WAPOR Code of Professional Ethics and Practices*. <http://www.unl.edu/WAPOR/ethics.html>.

⁹ Council of American Research Organizations (2008), *Code of Ethics and Standards for Survey Research*. <http://www.casro.org/pdfs/CodeVertical-FINAL.pdf>.

(ISI),¹⁰ the Canadian Market Research and Intelligence Association (MRIA),¹¹ and the European Society for Opinion and Marketing Research (ESOMAR).¹²

2.3 Because many of the Studies utilise experimental or quasi-experimental approaches to consumer research, I believe that a number of additional criteria are relevant. The relevance of these additional criteria is due to the nature of the proposals for the introduction of the various design restrictions. These criteria were outlined in my 2010 and 2012 Reports and are repeated here for ease of exposition. Specifically:

- (a) Because the various design restrictions envisaged in the Draft Directive do not currently exist and consumers are being asked to state an ‘intention’ relating to purchasing circumstances that are not currently available, they are being forced to speculate about a specific behaviour that may or may not arise in new circumstances. Hence, researchers must be cognisant of the degree to which the experimental task creates outcomes that can be linked specifically to behaviour.
- (b) Because the various design features envisaged in the Draft Directive have embedded within them a social outcome – i.e., the sheer existence of the various design restrictions is based on the belief that it will make a product category less “attractive” and therefore will change actual smoking behaviour – individuals will most likely know the intent of the investigator. Hence, the researcher must be cognisant of the degree to which the study itself enhances artificially the salience of the factors being studied.
- (c) Ultimately, the goal of policy related research is to examine the efficacy of a change in policy on a change in behaviour. Changes in behaviour themselves require the affected individuals to either use different decision models or different criteria within their existing models. Hence, it is critical for

¹⁰ International Statistical Institute (1985), Declaration on Professional Ethics. <http://isi.cbs.nl/ethics.htm>.

¹¹ Market Research and Intelligence Association (2007), *Code of Conduct for Members*. <http://www.mria-arim.ca/STANDARDS/CODE2007.asp>.

¹² European Society for Opinion and Marketing Research, *ICC/ESOMAR International Code on Market and Social Research*. http://194.38.169.84/uploads/pdf/professional-standards/ICCESOMAR_Code_English_.pdf.

researchers to have as complete an understanding as possible of the parameters of consumers' decision models.

2.4 Points (a) and (b) in paragraph 2.3 relate specifically to what is known as the attitude-behaviour gap or the difference between 'stated' intentions (i.e. what respondents tell those conducting a survey that they intend to do) and 'revealed' or actual purchases.¹³ Good research will attempt to reduce this problem by focusing on three factors, which I will add to my list of criteria:¹⁴

(a) **Incentive compatibility.** Incentive compatibility addresses the extent to which the methodology used by the researcher allows (or makes) subjects to reveal their true behaviour (if they currently engage in an activity, such as purchasing or not purchasing a product that is currently available) or what that behaviour would be if they were given the opportunity (in situations where there is no opportunity to reveal that behaviour, such as in the case of a new product). For example, it is well understood that forcing individuals to make a trade-off or asking them to pay a real price creates incentives that align better with their actual purchasing behaviour. Incentive compatibility is a particularly critical issue when asking:

- a) 'intention' questions – e.g., “if faced with these alternatives, which would you choose?”;
- b) 'speculative' questions – e.g., “how do you think a person faced with these alternatives would behave?”; and
- c) most types of 'self-report' questions – e.g. 'how likely is it that you do/would engage in a behaviour?'.

¹³ This is an extensive area of research. Some representative publications include: Sheerana, P. (2002), “Intention-Behaviour Relations: A Conceptual and Empirical Review,” *European Review of Social Psychology*, 12: 1-36; Carrigan, M. and A. Attalla, (2001), “The Myth of the Ethical Consumer – Do Ethics Matter in Purchase Behaviour?” *Journal of Consumer Marketing*, 18: 560-578; Ajzen, I. and M. Fishbein (2002), “The Influence of Attitudes on Behaviour,” in Albarracín, D. Johnson, B. and M. Zanna (Eds.), *The Handbook of Attitudes*, Abingdon, UK: Routledge, 173-221.

¹⁴ An overview of the importance of these issues is given in Devinney, T., Auger, P. and G Eckhardt (2010), *The Myth of the Ethical Consumer*, Cambridge UK: Cambridge University Press, pp. 56-59.

- (b) **Inference of salience.** Inference of salience addresses the degree to which the sheer addition of a factor that would otherwise not be part of the consumer's decision is all of a sudden added into the mix. For example, asking individuals about newly added attributes to existing products – i.e., aspects of a product that consumers know do not currently exist in what is offered in the market – heightens the salience of the new information making it more likely that the consumers will over-react to the new aspects of the product.

- (c) **Context.** Context addresses the degree to which the decision individuals are being asked to make is outside the context in which it might normally be made. For example, it is quite common to find that individuals, when asked in a survey the degree to which they will act in a pro-social way – such as purchasing 'green' products, volunteering or donating to a charitable cause – will overstate very significantly the likelihood that they will do so. Part of this is related to salience and incentive compatibility but it is also the case that most social behaviours are context driven, meaning that it is the context that drives behaviour. Respondents know that, from society's perspective, it is 'good to be green', so they may overstate their intentions to behave in an environmentally friendly manner. The context of answering something in a survey is different from the context of opening one's wallet or sacrificing one's time.

2.5 Point (a) in paragraph 2.3 forces us to ask very specifically how an individual chooses to do (or not do) something. Given the policy goals outlined in paragraph 1.6 we can reframe this to read:

- (a) Do the Studies have an effective statistical decision model of smoking uptake (initiation) among minors?

- (b) Do the Studies have an effective statistical decision model that explains a reduction in smoking consumption among minors and/or adults?

- (c) Do the Studies have an effective statistical decision model of smoking cessation among minors and/or adults?

2.6 What these questions ask collectively is: do the Studies effectively tell us something about the cognitive process that people go through when making a decision with regard to smoking-related behaviour(s)?

2.7 The point of this criterion is the link between the structure and design of the study and the operative criteria that a consumer would be using in realistic purchasing circumstances. In other words, does the study **appropriately model decision making** when the decisions are smoking initiation, smoking reduction and smoking cessation?

2.8 The experimental studies examined in this report were attempting to determine what decision an individual would make in “what if” circumstances. What this implies is that studies that can most effectively mimic the decision making process and criteria used by the individual in realistic circumstances will be the most valid. Hence, an additional criterion that we need to consider is the degree to which the experimental approach was designed in a manner that allowed the researcher to model the decision making process consumers would be using in realistic purchasing environments.

2.9 When attempting to determine what an individual’s decision model is via the choices that they can make in experiments it is important that the decision model and the experimental structure are aligned. The most accepted method for doing this is via the application of what are known as discrete choice models (or its variant conjoint analysis). This approach requires that individuals: (a) choose amongst a set of alternatives, (b) rank a set of alternatives, and/or (c) rate a set of alternatives.

2.10 The set of alternatives – known as a ‘choice set’ – must be structured according to an experimental design that has certain specific properties that we also consider as part of our evaluative criteria.

(a) First, they must be **statistically efficient**. Statistical efficiency implies that the structure of the experiment allows the researchers to recreate the decision model in use, either by the individual or by a group of individuals.¹⁵ Statistical efficiency is important because most experiments cannot possibly have all individuals look at every possible combination of products features that might

¹⁵ Mathematically, efficiency is a comparison of the design used to the ‘optimal’ or best possible design (the one that ensures that the decision model is estimated with the greatest degree of

be potentially on offer in a market. Hence, the extent to which an experiment is efficient is the degree to which the choice sets allow the researcher to have confidence that they have enough information to say that their results are a realistic representation of how an individual or group of individuals would behave.

- (b) Second, it is important when looking at the choices that individuals are being asked to consider, that the attributes and features presented exhibit (as closely as possible) **orthogonality**.¹⁶ In product choice experiments the features of the product are broken into ‘attributes’ – such as brand, colour or price – and the attributes into ‘levels’. For example, the ‘levels’ of the attribute *price* might be €1, €3, €5, €7 and the attributes of *colour* might be red, green and blue. Orthogonality implies that the experiment examining *price* x *colour* be set up with $4 \times 3 = 12$ alternatives as shown in Table 1 and that the individual be presented with a series of price x colour alternatives – as represented by the cells A-L below – such that the effects of each combination can be evaluated independently.¹⁷ This is achieved by the fact that any combination of an attribute level in the experiment below will appear 1/12th of the time and Red, Green and Blue will appear in 1/3rd of the options presented.¹⁸ A lack of orthogonality makes the determination of the effect of specific attributes and

statistical precision). See, e.g., Street, D. and L. Burgess (2007). *The Construction of Optimal Stated Choice Experiments*, Hoboken, NJ: Wiley, 85-86.

¹⁶ Perfect orthogonality is sometimes difficult to achieve and there are classes of non-orthogonal experimental designs. However, these are normally reserved for quite complex experiments and none of the Studies examined here would, in my expert opinion, be classified as being so complex as to require anything other than a basic experimental approach. See, e.g., Kuhfeld, W., Tobias, R. and M. Garrett (1994), “Efficient Experimental Design with Marketing Research Applications,” *J. of Marketing Research*, 31: 545-557.

¹⁷ More specifically this is important when using linear regression models for estimation purposes. As noted by Kuhfeld W., Tobias, R., and M. Garrett (1994), “*a linear model is fit with an orthogonal design, the parameter estimates are uncorrelated, which means each estimate is independent of the other terms in the model. More importantly, orthogonality, usually implies that the coefficients will have minimum variance*” (p. 545). The first part of this statement implies that we can readily make general statements about the importance of an attribute independent of the other attributes. For example, we can talk of the effects of price independent of the effects of the different colours. The second part of the statement implies that the estimates of that importance are the ‘best’ estimates we can achieve.

¹⁸ Technically, orthogonality is achieved when “the joint occurrence of any two levels of different attributes appear in options with frequencies equal to the product of their marginal frequencies”. Hubert, I. and K. Zwerina (1996), “*The Importance of Utility Balance in Efficient Choice Designs*,” *J. Marketing Research*, 33: 307-317.

levels difficult and will also generally imply less than efficient statistical estimation.

		Colour		
		Red	Green	Blue
Price	€1	A	B	C
	€3	D	E	F
	€5	G	H	I
	€7	J	K	L

- (c) Third, it is important when looking at the choices that individuals are being asked to consider that the experiment exhibits **balance**.¹⁹ This means that the attribute levels are appearing across the choice sets an equal number of times. For example, in Table 1, it is possible that each individual would see the four prices 2 times each or three times each. However, an improper design would have them see €1 three times and €3, €5, and €7 only once. The lack of balance has two possible effects. First, the subject may notice the imbalance and focus on the levels (in this case the prices) that appear more/less frequently (we do not know in which direction the bias would go, just that it exists). Second, statistically we will have inefficient estimates of the effects of the levels because we have only one response for three of the prices and three responses for one of the prices.

¹⁹ Haaijer, R. and M. Wedel (2007), “Conjoint Choice Experiments: General Characteristics and Alternative Model Specifications,” in Gustafsson, A., Herrmann, A. and F. Huber (Eds), *Conjoint Measurement: Methods and Applications*, 4th Edition, Heidelberg, Germany: Springer. Note that sometimes balance and orthogonality are in conflict, particularly when some of the options presented might not make logical sense to the subject of the experiment. Such an example might be a car that had low price and lots of luxuries or very high engine power but also very high fuel efficiency. However, this again is not an issue relevant to the Studies being examined in this report, as there is no indication that the authors considered such factors when setting up their studies. In the case where such an issue exists, the research would generally use a non-orthogonal design of the types discussed by Kuhfeld, Tobias and Garrett (1994).

2.11 Point (a) in paragraph 2.3 relates also to intentionality. Morwitz, Steckel and Gupta²⁰ have shown that ‘intentions’ are most related to actual purchasing when:

- (a) They are for **existing products**;
- (b) They are for durable rather than **non-durable goods**;
- (c) They are for **short-term horizon** decisions rather than for long-term time horizon decisions;
- (d) Subjects are asked about purchase intentions for **specific brands** rather than for the product category in general;
- (e) Purchase intentions are measured as **‘trial’ rates amongst existing purchasers in the relevant product segment**, rather than being measured in terms of total market share; and
- (f) Purchase intentions are collected in a **comparative mode**, rather than monadically (e.g., a paired comparison versus asking the subject to evaluate a single alternative at a time).

2.12 As noted in paragraph 2.1 above, this report includes the evaluation criteria applied by Dr Keegan in his Reports. Table 2 provides a short summary of the criteria applied by Dr Keegan and highlights which are most relevant to this report. Those criteria that are not applied here are excluded only because they are not applicable to the research studies examined below. Specifically, the Studies are: (a) all relatively recent and hence do not suffer from ‘study age’ issues; and (b) are examinations within a single culture/country and hence do not suffer from issues of cross cultural bias.

²⁰ Morowitz, V., Steckel, J. and A. Gupta (2007) “When Do Purchase Intentions Predict Sales?” *International Journal of Forecasting*, 23: 347-364.

Table 2: A Summary of the Evaluative Criteria Outlined by Dr Keegan

Evaluative Criteria	Short Description of Criteria and Its Relevance
<i>1. Standards Compliance</i>	
Compliance with International Standards	Complies with the standards outlined in paragraph 2.2 above.
<i>2. Age of Study</i>	
Study age	Consumer research has a limited lifespan, the older the study, the less likely its current applicability. The study must be reflective of current market conditions and regulatory environment. This criterion is not applied to the Studies in this report.
<i>3. Field Administration Protocol</i>	
Question design	Ensuring proper question design is a requirement that is reflected across internationally accepted research standards. Questions should not cue responses, i.e., ‘beg the answer’. Questions should not make assumptions about a respondent’s knowledge or experiences. Respondents should be given the opportunity to give a “don’t know” or “no opinion” answer.
Interviewer response bias	To the extent that it may bias the results, neither respondents nor persons responsible for the data collection should be informed as to the sponsor or purpose of the study.
Researcher objectivity	A researcher, whatever his/her views or opinions on a topic, must ensure that the study design is

Evaluative Criteria	Short Description of Criteria and Its Relevance
	impartial and not designed to yield any particular result.
	To the extent that an author's advocacy influences the study design, the study's reliability and validity suffers.
Response reliability	<p>Observing what people do is a better predictor of behaviour than recording how people respond to questions about what they think they will do, or what they think others will do, or what they report they have done.</p> <p>In consumer research, the gold standard is to get as close as one can to observing behaviour. The gradient of research reliability, from most reliable to least reliable, is generally as follows:</p>

Reliability by Data Collection Method

Reliability	Data Collection Method	Research Type
Most reliable	Direct observation	Observed Behavioural
↓	Recent recall of behaviour	Observed Behavioural
	Recall of non-recent past behaviour	Self-Reported Behavioural
	Prediction of future behaviour	Opinion / Attitudinal
Least reliable	Prediction of others' future behaviour	Opinion / Attitudinal

4. Appropriateness of Sampling Frame

General appropriateness The sample should reflect the population relevant to the question at hand allowing for the greatest degree of generalisation.

Age of respondents Conducting research among minors presents particular issues that must be accounted for to

Evaluative Criteria	Short Description of Criteria and Its Relevance
---------------------	---

ensure the reliability of the data collected.²¹ Young respondents are more likely to feel pressured during an interview situation; such pressure can result in answers that are inaccurate.

It is much more difficult to ask a minor a difficult policy question and have an acceptable degree of confidence that the information collected will have any resemblance to the effect that would be observed if the policy were actually enacted. For example, asking a minor “*Will young people buy fewer bus passes if fares are increased?*” is unlikely to generate reliable data.

Focus groups

Focus group studies are exploratory. They generate hypotheses rather than findings that can be generalised to a wider population.²²

The reported findings of focus groups often have no statistical significance due to the small sample size and informal nature of the responses.

5. Analysis

Statistical significance

It is imperative that authors refrain from projecting results that are not statistically significant to general populations or markets.²³

It is widely recognised in the research community that statistical significance is a necessary pre-requisite in determining that a causal relationship is

²¹ Churchill, G., and D. Iacobucci (2005), *Marketing Research: Methodological Foundations*, 9th Edition. Orlando, FL: Dryden, 387-390.

²² Churchill, G., and D. Iacobucci (2005), *Marketing Research: Methodological Foundations*, 9th Edition. Orlando, FL: Dryden, 81-85.

Evaluative Criteria	Short Description of Criteria and Its Relevance
Unsupported results or conclusions	<p>an observed result and not caused by chance, error or other factors.²⁴</p> <p>In interpreting study results, authors sometimes make ‘leaps’ between the data yielded by the study and the conclusion the author puts forth. It is concerning when an author draws conclusions that are not supported by the research.²⁵</p>
False comparison	<p>It is imperative that authors refrain from generating comparisons and drawing conclusions from comparisons that are not reflective of actual real life conditions.</p>
Cross cultural applicability	<p>Cultures can have unique characteristics that must be accounted for when designing a study, and specifically, a questionnaire. Cultural differences exist both between countries and, indeed, in many cases, within different geographical regions of one country.</p> <p>This criterion is not applied to the Studies in this report.</p>

3. REVIEW OF THE STUDIES

3.1 My review in this report has focused on the primary research studies which were presented in the IA in support of the various design restrictions in the Draft Directive to determine whether they contain reliable evidence that the various design

²³ International Statistical Institute (1985), *Declaration on Professional Ethics*. <http://isi.cbs.nl/ethics.htm>, paragraphs 1.3 and 3.1.

²⁴ Thompson, B. (1994), “The Concept of Statistical Significance Testing,” *Practical Assessment, Research & Evaluation*, 4.5.

²⁵ American Association for Public Opinion Research (2005), *AAPOR Code of Professional Ethics & Practices*. <http://www.aapor.org/aaporcodeofethics>, paragraph I.A.2 and I.A.3.

restrictions will achieve the public policy goals set out in paragraph 1.6 above. The Studies are listed below, with a study-by-study analysis set out from paragraph 3.2 below:

- (a) Borland, R and S Savvas (2012). Effects of stick design features on perceptions of characteristics of cigarettes. *Tobacco Control*. Hereinafter referred to as “Borland and Savvas (2012).”²⁶
- (b) Mutti S., Hammond. D., Borland, R., Cummings, KM., O’Connor, RJ., Fong, GT. (2011). Beyond light and mild: cigarette brand descriptors and perceptions of risk in the International Tobacco Control (ITC) Four Country Survey. *Addiction*; 106: 1166 – 75. Hereinafter referred to as “Mutti et al. (2011).”²⁷
- (c) Moodie, C., Ford, A., Mackintosh, A. and G Hastings (2011). Young people’s perceptions of cigarette packaging and plain packaging: An online survey. *Nicotine and Tobacco Research*, Advance Access, publicised on October 24, 2011. Hereinafter referred to as “Moodie, Ford et al. (2011).”²⁸
- (d) Moodie C. and A. Ford (2011). Young adult smokers’ perceptions of cigarette pack innovation, pack colour and plain packaging. *Australian Marketing Journal*, 19: 174-180. Hereinafter referred to as “Moodie & Ford (2011).”²⁹
- (e) Doxey, J and D Hammond (2011). Deadly in Pink: The impact of female-oriented cigarette packaging among young women. *Tobacco Control*, 20 (5); 353 – 60. Hereinafter referred to as “Doxey and Hammond (2011).”³⁰
- (f) Hammond, D. and C. Parkinson (2009). The impact of cigarette pack design on perceptions of risk. *The Journal of Public Health*, Sep; 31 (3): 345-53. Hereinafter referred to as “Hammond & Parkinson (2009).”³¹

²⁶ Referred to at footnote 151 of the Impact Assessment.

²⁷ Referred to at footnotes 146, 148 and 150 of the Impact Assessment.

²⁸ Referred to at footnotes 146 and 148 of the Impact Assessment.

²⁹ Referred to at footnotes 146 and 148 of the Impact Assessment.

³⁰ Referred to at footnote 147 of the Impact Assessment.

³¹ Referred to at footnote 146 of the Impact Assessment.

- (g) Hammond D., Dockerell, M., Arnott, D., Lee, A. and McNeill (2009). Cigarette pack designs and perceptions of risk among UK adults and youth. The European Journal of Public Health, Volume 19, Issue 6, 631-637. Hereinafter referred to as “Hammond et al. (2009).”³²

³² Referred to at footnote 146 of the Impact Assessment.

DETAILED ANALYSIS OF THE STUDIES

Borland and Savvas (2012)

3.2 This is an internet survey examining the perceptions of smokers and non-smokers as to the characteristics of the dimensions and design of cigarette sticks on attractiveness, quality and taste perceptions. It concludes that “[c]haracteristics of the cigarette stick affect smokers’ perceptions of the attributes of those cigarettes and thus are a potential means by which product differentiation can occur. A comprehensive policy to eliminate promotional aspects of cigarette design and packaging needs to include rules about stick design.” (page 1). This study was not evaluated in my 2010 or 2012 Reports, and a short evaluation is given below.

Analysis

3.3 A total of 160 adults in Australia aged between 18 to 29 years participated in the study. Eight percent were smokers and twenty percent recent quitters. No information is given as to the distribution of the sample in terms of geographic location or other factors that may impact on the nature of the study. It is noted that 27.5 percent of the sample is unemployed, which is considerably above the unemployment rate for adults in this age category (even accounting for the fact than many of the unemployed could be also categorized as students).³³

3.4 This study is dominantly limited by its design and structure of the questions. Ultimately, however, the most comprehensive limitation of the study is that it attempts to make broad conclusions about behaviour without having estimated a behavioural model or examined actual behaviour.

3.5 Individuals were presented with three sets of computer images of various cigarette stick designs. From this the study participant is asked to rate them based on which is “most” and “least” attractive, “weakest” and “strongest” strength of taste, “highest” and “lowest” quality, and which you would be “most likely” or “least likely” to choose. Individuals were also given the option to say that all the cigarettes were the same. From this information, the researchers artificially scored the

³³ See, e.g., http://www.budget.gov.au/2011-12/content/glossy/skills/html/skills_overview_21.htm, accessed 20 April 2013.

“most/highest/strongest” response a 5 and the “least/lowest/weakest” response a 1. All the non-rated designs were rated a 3. This structure suffers from a number of serious flaws.

3.6 First, the most obvious issue is related to the use of images rather than actual cigarettes. This fact is of particular importance here as the study itself argues that it is most of the aspects of stick design are tactile and, hence, cannot be evaluated effectively with images. At a minimum some validity check needs to have been done where a subsample of individuals evaluated the stimuli using either both the actual cigarette sticks and the computer generated stimuli – which is superior since it allows evaluating the method both ‘between’ and ‘within’ subjects – or simply evaluated the actual cigarette sticks – which allows for ‘between’ subjects comparisons.

3.7 Second, querying individuals about the different designs will induce the individual to make cognitive attributions about what they are being asked – i.e., variants in stick design – must be related to what they are being asked to evaluate. The solution to this problem is rather simple and that is to add into the survey items that are unrelated to the purpose of the study – dummy questions – to see whether or not the individuals are actually responding in a manner that is representative of their beliefs rather than responding to the instrument. The very high correlations seen across the different evaluations (given in Table 5) imply that there are clearly such method bias issues. Rather than do this, the researchers attempt to justify this *post hoc* by arguing that while correlation amongst the ratings are high, the pattern does not imply that individuals are giving responses biased by the instrument itself. This, however, is not sufficient to justify conclusively that there is no response bias.

3.8 Third, the researchers used a best-worst evaluative structure but did not use an actual best-worst experimental design.³⁴ A simple best-worst design for 5 stimuli (as is used in this study for Sets A, B and C) would require individuals to evaluate 10 blocks of 3 stimuli. The importance of using such designs is that rather than leading to

³⁴ Such designs were discussed in my 2010 and 2012 reports. An addition recent overview of this approach is given in Louviere, J., Lings, I. Islam, T., Gudergan, S., & T. Flynn (2013) "An Introduction to the Application of (Case 1) Best-Worse Scaling in Marketing Research," *International Journal of Research in Marketing* (Forthcoming). Available at: http://portal.idc.ac.il/en/main/research/IJRM/Documents/IJRM_D-08-00204_Louviere%20et%20al_forthcoming.

arbitrary scoring systems as is seen in this study where “best” is given a ‘5’ and “worst” is given a ‘1’, one can easily run a multinomial logit model of the design against the responses. This provides the researcher with an immediate utility estimate of each stimuli and a complete and robust evaluation of the items being studied – in this case, stick design. The value in such a structure is that it not only allows the investigator to get pure utility measures that are based on a clear behavioural model,³⁵ but also that one can utilize the fact that they have six evaluations of five items from each individual, allowing them to get a more accurate picture of individual level variation in the evaluations.³⁶

3.9 In addition to the design issues there is also a methodological issue that arises from the poor design. The researchers evaluate each item and each breakdown (e.g., gender) one item at a time. This assumes no relationship amongst the various evaluations (which is clearly not true as evidenced by Table 5) and that the sampling was done in a way that effectively controls for sources of variation from the socio-demographic covariates (e.g., gender, age, education and so on) – which was clearly not the case as it was never indicated that this was how the sampling was done. I have criticized many studies in my 2010 and 2012 reports for using this approach, and this criticism applies here as well. Two aspects of this are worth noting:

- (a) The correlations in Table 5 imply that either there is a more complex structural relationship amongst the evaluations or that there is a method bias issue.
- (b) Had the study used an efficient best-worst design it would have been possible to estimate individual level models and hence relate the differences in the responses to the underlying covariates (e.g., gender, age, education, etc.) directly and simultaneously. Absent this, the investigators could have attempted to do a more sophisticated analysis using a multinomial logit model

³⁵ This was noted in my 2012 report as well and can be see in Marley, A.A.J and J. Louviere (2005). Some Probabilistic Models of Best, Worst, and Best-Worst Choices, *Journal of Mathematical Psychology*, 49(6): 464–480. Another alternative would be to use a ranking task (i.e., ranking the 5 items). There is a direct relationship between using a ranking task and a best-worst task, although the ranking task requires individuals to do only one set of evaluations and hence the research cannot see degree to which there is individual level variance (or certainty) in preferences.

³⁶ The optimal best-worst design for 5 stimuli is to have individuals evaluate 10 blocks of three items. This implies that each of the 5 stimuli will appear 6 times ($3 \times 10 / 5 = 6$).

on the responses they received. This would have been done by coding the “most/highest/strongest” response as ‘+1’, the “least/lowest/weakest” response as ‘-1’, and no response as ‘0’ and estimating a model that related all the covariates to the responses simultaneously.

3.10 The final major limitation of the study is in the results. There are two aspects to this:

- (a) First, Tables 2, 3, and 4 reveal one dominant fact – that the current design is preferred. All other designs are inferior. There is not one single case in which an alternative design is rated superior to the current design on the market. This is a common inertia finding that can imply only that people prefer what they know and alternative presented to them can lead to random or obvious inferences that may or may not be related to anything other than randomness.³⁷
- (b) It is unclear how the results of this study relate to those specifically examining the effects of various aspects of cigarette design and packaging on women’s choices and preferences (e.g., Doxey, 2009). This study shows quite mixed effects.

Conclusion

3.11 Overall, it would be erroneous to draw policy conclusions based on this study. The study argues for three conclusions: (a) “*that [the] characteristics of cigarette stick design can be used by consumers to make judgements about characteristics of the cigarettes*”; (b) “*they have the capacity to act as a form of promotion and should be regulated where public policy is to eliminate as far as possible promotion of tobacco products*”, [and] (c) “*it provides evidence supporting the decision of the Australian government to regulate stick design features as part of its Plain Packaging legislation.*” There is no empirical justification for these conclusions. Point (a) is speculative in that the authors never measure whether consumers make meaningful

³⁷ See, e.g., Dubé, J.-P., Hitsch, G. J. and P. E. Rossi (2010). State Dependence and Alternative Explanations for Consumer Inertia. *The RAND Journal of Economics*, 41(3): 417–445; Shin, S., Misra, S., and D. Horsky (2012). Disentangling Preferences and Learning in Brand Choice Models. *Marketing Science*, 31(1): 115–137.

judgments, only that they respond to a survey, where the dominant conclusion is that they exhibit inertia. Point (b) is invalid in that the study never examined “promotion” and provides no evidence of the effectiveness of the promotion of the stick design. Point (c) is an erroneous conclusion in that the point of the plain packaging legislation was ostensibly to reduce cigarette smoking and initiation, yet these were never tested, nor was there any evidence of a behavioural link between the responses in the surveys in the study and actual behaviour.

Mutti, et al. (2011).

3.12 This study uses data from the fifth wave of the ITC Four-country survey to examine smokers’ beliefs about the perceived risk related to certain types of cigarette brands. The conclusion is that “*despite current prohibitions on the words ‘light’ and ‘mild’, smokers in Western countries continue to falsely believe that some cigarettes brands may be less harmful than others. These beliefs are associated with descriptive words and elements of package design that have yet to be prohibited.*” (page 1). This study was not evaluated in my 2010 and 2012 Reports and a short evaluation is given below.

Analysis

3.13 The validity of the ITC Four-country survey is based mainly on its representativeness and the fact that it is longitudinal. In the case of the questions utilized in this study there are a number of worrying methodological issues and the conclusions drawn by the researchers as to the importance of the effects they find are overstated.

3.14 The first methodological issue arises in how ‘light/mild’ are measured. Respondents were asked to indicate whether they agreed or disagreed with the following statements: (a) “*light cigarettes are smoother on your throat and chest than regular-strength cigarettes*”; (b) “*light taste means less tar*”; and (c) “*harsh smoke is more dangerous*”.³⁸ The questions were preceded by “*a statement that the term ‘lights’ was being used to refer to cigarettes that were being promoted with terms such as light, mild or low in tar*”. The proximity of the questions to the “priming”

³⁸ See page 3, Mutti et al. (2011).

statement about the meaning of light cigarettes induces an enormous and fatal bias. If I am being told that products labeled light have these characteristics, then asked if these products have these characteristics it is very likely that I will respond that this is the case. Indeed, to call these beliefs “false” when they are beliefs that are being primed into the participant is inconsistent with best research practice. If the whole intent of the ITC survey was to gather “realistic” information such priming invalidates this.

3.15 Secondly, the questions themselves are directionally biased. In the three questions the key words are (a) “*smoother*”, (b) “*less tar*”, (c) “*more dangerous*”. All are pointing directionally to agreement with the statements. A more balanced approach would have been to use neutral statements such as “*light cigarettes are no more or less harmful than regular cigarettes*” or to have split the sample so that the questions were directionally different – with half the time individuals would be receiving a question such as “*light cigarettes are more harmful than regular cigarettes*” – with these three questions being rotated for direction (meaning that there were six versions of the survey for these questions where each individual has a 50:50 chance that each question was directionally positive or negative). Removing this bias is absolutely critical because the researchers define “false beliefs” based upon agreement with these statements. If the statements are not effective anchors then one cannot define what is “false” about a belief.

3.16 Further, the questions about tar values are arbitrary and, in many ways, meaningless. Respondents were asked: “*how closely, if at all, are the tar numbers on cigarette packs, related to the amount of tar that smokers take in their bodies?*”.³⁹ In short, asking individuals about how accurate the tar measures are is querying them about something they would not be expected to know.

3.17 There are a number of other methodological concerns with this study, which I have also raised in my analysis of studies such as Moodie, Ford, et al. (2011), Hammond & Parkinson (2010) and Doxey & Hammond (2011).

³⁹ See page 3, Mutti et al. (2011).

3.18 Firstly, rather than evaluate each question across its entire range of values, the researchers erroneously and inappropriately aggregate items to create binary variables. This is most egregious when the aggregation includes “Don’t Know” responses. As their measure of “false beliefs” is based on the scale that arises from this aggregation there is little that can be effectively gleaned from it.

3.19 Secondly, instead of conducting a full-scale multivariate analysis that accounts for the potential inter-relationships between their dependent variables the researchers conduct analyses one at a time. This leads to the classic “sunspot” bias where the real effects on the dependent variables may be being caused by effects that have little or nothing to do with the supposed antecedents. This can arise because of bias – common method bias – inter-relationships amongst the dependent variables, or unknown missing causal factors that can be influencing all of the dependent variables.

3.20 However, the most problematic aspect of this study is once again the lack of any behavioural model of “false beliefs”. This leads to attempts at ex post speculation as to why certain results are what they are. This lack of a model with any specific hypotheses or any basis in extant areas of research makes the interpretation of the results little more than opinion as to effects.

Conclusion

3.21 Overall it is my conclusion that it would be erroneous to draw policy conclusions based on this study. The methodology suffers from significant bias in the construction of the measure, creating a lack of validity in the conclusions. The supposition that the authors are measuring “false beliefs” is based entirely on a poorly constructed index.

Moodie, Ford et al. (2011).

3.22 This study relates to young people’s perceptions of cigarette packaging and plain packaging. It is based on an online survey of 658 young people aged between 10 and 17 years of age. A detailed evaluation of this study is given in sections 4.78-4.95 of my 2012 Report. For ease, I have summarized the key areas of concern with the study below.

3.23 The dominant overarching issue with this study is the lack of any effective behaviour measures and the lack of a structure that takes into account the attributes critical to choice. Hence, the conclusions drawn from the study that pack design and colour matter are based on a study design which has failed to consider key drivers of choice or purchasing behaviour (such as price, availability and peer influences). As noted throughout my 2010 and 2012 Reports, and as I have said above, to be able to effectively understand the role of packaging on choice it is important to have as complete a behavioural model as possible and to structure the empirical aspects of the study in a way that allow for the importance of all the key drivers to be considered in tandem. Focusing on isolated product attributes independently creates salience effects that lead to serious overstatement of importance as subjects respond. This is because the study focuses respondents on the issues that the researcher is studying, rather than the issues that are truly relevant to them. Failure to structure the studies correctly so that they are balanced and orthogonal across the key decision factors reduces the researchers' ability to estimate what the true impact of key decision drivers are.

3.24 Therefore, overall, one cannot take many conclusions from this study give its lack of interdependent behavioural validation. But even despite this there are conclusions that one can point to that call into question the conclusions drawn by the researchers. I will focus on two of these to illustrate the issue.

3.25 Firstly, the authors conclude that: *“when shown four differently colored completely plain cigarette packs, displaying only health warnings, many young people associated pack color with product strength and also level of harm...This is consistent with the meanings that people, often automatically, attach to color...Although past research has typically explored pack color in the presence of branding (including brand name or pack descriptors), this study suggests that color alone is sufficient to mislead consumers”*.⁴⁰

3.26 The first issue with this conclusion is the statement *“this study suggests that color alone is sufficient to mislead consumers”*.⁴¹ There is no evidence of this at all. Firstly, because the study examined each factor separately it cannot make any

⁴⁰ See page 6, Moodie, Ford et al. (2011).

⁴¹ Ibid.

statements about whether something alone mattered. This can only be done by excluding all other effects, as would be the case in a structural equations model.

3.27 Second, the conclusions about the influence of colour are based on a limited part of the sample. Suppose that one reframed the data in Table 3 (as is done in the table below which shows the percentages making any or no colour association). I concentrate only on the total results and smokers since their definition of “susceptibility” is questionable.⁴²

	All	Ever-smokers	Never-Smokers
Any Association	48.25	60.25	43.75
No Association	51.75	39.75	56.25

3.28 What we see is that overall there is a 50:50 chance that a respondent makes any association at all with colour. Even in the case of ever-smokers, this number is not that materially different in that 40 percent make no association and the ever-smokers are 28 percent of the sample. What we do not know, of course, is whether that association is related in any way to anything material. For example, I can make an association because I am knowledgeable about a relationship (hence ever-smokers might use information about the products they have used). Alternatively, I may make an association because I am being asked to make an association and am more susceptible to assuming that such an association exists when I have no knowledge. It is even possible that both of these factors are at play in the different groups. Hence, the fact that an association exists does not de facto imply that it related to misperceptions of pack features.

3.29 Second, it is argued that *“one in three thought that young people would be most likely to smoke either the plain slide or the super Slims pack. Of particular interest is the fact that susceptible never-smokers were almost three times more likely than nonsusceptible never-smokers to show a preference for the super Slims pack and*

⁴² The authors’ definition is given as *“nonsusceptible never-smokers were those who indicated that they would “definitely not” smoke a cigarette during the next year, with susceptible never-smokers those whose response was anything other than definitely not.”* (see page 3 of the study). As noted with other recordings this is dubious statistically as it makes an arbitrary assumption that can be easily addressed by effect coding the responses.

twice as likely to like the slide pack, which suggests that package design features such as shape and method of opening either appeals to, or perhaps reinforces, susceptibility to smoke".⁴³ There are two issues with this conclusion.

- (a) First, the statement that suggests pack design features such as shape and method of opening either appeal to, or perhaps reinforce, susceptibility to smoking is based on a specific definition of susceptibility that can be questioned, and makes an inference about appeal and reinforcement that is not tested in any way (as the conclusions are drawn by comparing across individuals). Without a behavioural model, no conclusion can be made on whether this is true or not.
- (b) Second, the preference for the slide-pack is conditional on individuals expressing an opinion at all and then the way in which the query is framed. Based on the data in Table 5 of the study, 57 percent express no preference at all. When they do express an opinion the slide-pack is found to be more "attractive" (25 percent to 10 percent) but when the question is reframed to what they believe others would do, this result changes to 23 percent and 21 percent in favour of the flip-top box. This brings up the basic concern that the responses given are related entirely to the structure of the question design.

3.30 Overall, this study attempts to draw causal conclusions about behaviour without having a causal behavioural model. This creates a dilemma where policy conclusions are being inferred rather than proven in any meaningful way. As noted above, the data suffers from methodological and empirical limitations that make its conclusions difficult to justify. Put simply, this study does not provide reliable evidence to suggest that the various design restrictions will influence the public policy goals outlined in paragraph 1.6 above.

Moodie and Ford (2011).

3.31 This study relates to a sample of 18-35 year old smokers' perceptions of cigarette pack innovation, pack colour and plain packaging. A total of 54 such

⁴³ See page 7, Moodie, Ford et al. (2011).

respondents in Glasgow, Scotland, participated in 90-minute focus groups with 6 or 7 participants. I have previously reviewed this study in detail in my 2012 Report (see paragraphs 4.108-4.113). For ease of exposition, I have summarised my key concerns with the study below.

3.32 The study is limited first and foremost by the use of focus groups. As I have noted in my 2010 and 2012 Reports, focus groups are non-representative and subject to significant research and social bias effects that render their applicability to significant policy and commercial decisions severely limited. At best, focus groups can be used to work through nascent ideas that serve as a basis of more rigorous investigation. However, in this case, the value of the conclusions must be viewed sceptically as the conclusions of the authors outlined in section 1.1 have no scientific basis and amount to little more than a “considered opinion” of the researchers.

3.33 In this study, the focus group discussions were motivated by examinations of pack types – where the pack differed by opening, and also a ‘perfume’ pack style – colour (red, light blue, green and white) and plain packaging. The only conclusion that appears to be drawn which is consistent in the study is that women reacted differently from men, with women being slightly more responsive to styling when discussing the alternative packs. Other than this conclusion, the remainder of the discussion was ad-hoc justifications of quotations and opinions of the various participants.

3.34 In conclusion, and consistent with my 2012 Report, this study is completely lacking in validity and, at best, can “*be considered as [a source for] potential hypotheses for more rigorous scientific studies*”.⁴⁴ The relevance of a small-scale focus group study to major policy decisions must be questioned. In addition, the sample is non-representative. Therefore, in my opinion, this study does not provide reliable evidence as to whether the various design restrictions would help achieve the public policy goals listed in paragraph 1.6 above.

⁴⁴ See paragraph 4.113 of my 2012 Report.

Doxey & Hammond (2011).

3.35 This study examines the impact of package and cigarette design on young women. I reviewed this study in detail in sections 4.38 to 4.47 of my 2010 Report in its original form (i.e. as a Masters Thesis). A total of 512 females between the ages of 18 and 25 were recruited throughout Canada through a market research service. I concluded in my report that this was one of the most methodologically flawed studies I examined: *“This study is methodologically flawed in terms of its basic design, execution and statistical analysis. The structure of the experimental tasks and questions is flawed and fails to meet the standards of good experimental research. The use of arbitrary aggregation without theoretical justification renders the conclusions invalid as they are subject to a plethora of other explanations. The statistical modeling is atheoretical and does not align with any behavioural model”*.⁴⁵

I have summarized the key limitations of the study below.

3.36 This study suffers from a number of serious methodological flaws, many of which follow directly from the fact that they repeat the same basic form of research method used by Hammond, et al. (2009) and Hammond and Parkinson (2009). Most obviously, the lack of actual behavioural outcomes or incentive compatible measures that represent how individuals would make choices in the broader contexts of purchasing are quite serious. They undermine the relevance of any conclusions drawn about perceptions.

3.37 This issue of the perceptions is made worse by the artificiality of the question design and measure construction. Questions relating to tar delivery – *“How much tar do you think these cigarettes would have compared to other cigarette brands?”* – and health risks – *“How would the health risks of these cigarettes compare to other cigarette brands?”* – are (a) assuming that the individual is competent to understand the meaning of “tar delivery” and (b) define “health risks” in a manner that is comparable between individuals (in other words my definition of “health risk” is comparable to yours). In addition, the comparator of “other cigarette brands” throws in an arbitrary base case against which the subject is supposed to compare. However, there is no guarantee that one person’s “other cigarette brands” is the same as another

persons, or that many individuals would know little or nothing about the tar yield of these “other cigarette brands” and what that might mean, if anything, in terms of comparative “health risks” of the different brands.

3.38 The research uses 5-point scales to measure her perception constructs – e.g., where the scales range from “a lot better” to “a lot worse”. However, these 5-point scales are then arbitrarily aggregated so that they were dichotomous (i.e., “1” and “0”). Such arbitrary aggregation is completely unacceptable based on the norms and standards of market research as it imposes *a priori* restrictions on the data for no other reason than the researcher wants to use a specific empirical approach and to create an index by aggregating up the 0-1 measures. As noted by Rossiter:⁴⁶ *“the formation of such an index requires that constructs of interest be conceptually defined (described) in terms of (1) the object, including its constituents or components, (2) the attribute, including its components, and (3) the rater entity. Failing this, the conceptual definition of the construct will be inadequate for indicating how the construct should be (operationally) measured.”* What this means specifically is that (a) it must be clear what is being rated – the ‘object’ must be clear to those being asked to do the rating, (b) what makes up the construct is well articulated – what the ‘attributes’ of the object are understood and valid, and (c) all of the raters are comparable – in other words, the raters are knowledgeable and relevant. The last statement implies that without this clear articulation then the measure is conceptually invalid.

3.39 The problems with this aggregation can be seen in a few examples which show that there is no theoretical or logical justification for such an approach:

- (a) “Brand appeal” ratings were determined by a question: *“In your opinion, how appealing would this brand of cigarettes be to young women your age compared to other brands on the market?”* The ratings asked were then aggregated so that “a lot more appealing” and “a little more appealing” received a score of “1” and “no difference”, “a lot less appealing”, “a little less appealing” and “don’t know” received a score of “0”. In essence, this measure

⁴⁵ See paragraph 4.47 of my 2010 Report.

⁴⁶ Rossiter, J. (2002), “the C-OAR-SE Procedure for Sale Development in Marketing”, *International Journal of Research in Marketing*, 19: 305-335.

treats someone who has no idea (i.e., don't know) in the same category as "a lot less appealing" and "no difference".

- (b) "Health risk" perception ratings were determined by a question: "*How would the health risks of these cigarettes compare to other cigarette brands?*" The ratings asked were then aggregated so that "a lot less risk" and "a little less risk" received a score of "1" and "no difference", "a lot more risk", "a little more risk" and "don't know" received a score of "0". In essence, this measure treats someone who "doesn't know" (and is willing to say so) in the same category as someone who views the risks as high!

3.40 These arbitrarily aggregated measures are then further confused by the creation of an index in which each package in the experiment is added up to create a 1-8 scale that is meant to create an overall "Brand Appeal" index, "Perceived Taste" index, "Tar Delivery" index, and "Health Risk" index. This creates what amounts to a theoretic formative measure for Appeal, Taste, Tar and Health that supposedly measure the overall perception of the 8 packs seen by the subject. However, what this measure means is totally unclear. How it relates conceptually to the choices being asked of the subject in the experiments is also unclear. To use Rossiter's well-established formulation described in paragraph 4.40 of my 2010 Report and paragraph 3.38 here, it is unclear what the components of this index are. For example, although it might be clearer what an "8" and "0" meant for such scales but intermediate measures are completely confusing. If someone rated the 8 brands as a 4 on "tar" this could arise because 4 of the brands were perceived to have a "lot less tar" and 4 of the brands were not rated "Don't know" or all 4 of the brands were perceived to have "a little more tar" and 4 "a little less tar". In other words, there is absolutely no way to determine what actual perceptions led to the intermediate scores. Mean statistics are provided for these indices in Doxey (2009)'s tables 4, 8, 12, and 16 and what we see is, as a matter of accepted norms for such assessment, concerning. For the most part these indices are skewed toward the lower end meaning that the vast majority of subjects end up rating most of the packs in a manner that leads to a "0". But a "0" is the most confused score as it confuses "a little" and "some" with "don't know" and "no difference".

3.41 As in Hammond, et al. (2009) the experimental conditions are confused. They do not allow for effective and efficient comparison of the package attributes, as the design is not efficient, orthogonal or balanced. For example, the brands appear different numbers of times and the “Male” brands that are meant as controls have completely different brand names, pack dimensions and colours. A properly designed study would control for brand effects, dimension effects, colour effects, price, and other package and product attributes and do so in a way that would have proper experimental design characteristics. This would, as noted earlier in paragraph 2.10 allow for the measurement of specific attribute and level effects (essentially the components of the packaging), along with the interaction between those factors and other components of the product.

3.42 This criticism of the experimental design means that there is no logical link between the design and the statistical model being estimated. In other words, the vast majority of the analyses are based upon pairwise comparisons (such as “standard” versus “plain”) where the scores are completely dependent upon the alternative against which they are being compared. What this implies is that, at best, the analysis only allows us to make a statement about that specific package in that specific experiment against that specific alternative. This is why the lack of orthogonality and balance is important. An orthogonal and balanced design allows the researcher to make statements about the components of the product independent of the alternatives. In this study, we are left with the result that we can say absolutely nothing generalisable about any of the results of the analysis.

3.43 In short, and given the above, this study does not provide reliable evidence that the various design restrictions would have any impact on changing actual smoking behaviour.

Hammond & Parkinson (2009).

3.44 This study relates to the impact of cigarette pack design on perceptions of risk. The authors recruited 603 respondents aged 18 years and over, smokers and non-smokers, between January and March 2007 from shopping malls in Ontario, Canada. I evaluated this study in detail sections 4.16-4.22 of my 2010 Report and concluded

that: *“this study provides no valid evidence to support the propositions that the packaging information is “misleading” and “deceptive” and that “these terms [health and taste descriptors] are equivalent in the minds of many smokers when used on packaging.”*⁴⁷ For ease, I have summarized the key limitations with the study below.

3.45 A severe limitation of this study is the bias induced by creating scales based on what the researchers believed should be true rather than any theoretical model or via pre-testing procedures that would justify their conclusions. As noted on page 3: *“prior to the study, one package from each pair was identified as the package most likely to be rated as higher tar, smoother taste and lower health risks, based on a priori hypotheses”*. Exactly what these a priori hypotheses were is never explained. This is a particularly worrying issue for two reasons. First, as noted by the researchers, they biased the presentation of the package alternatives by having what they considered to be the *“higher tar, smoother taste and lower health risk”* alternatives always being *“listed first for each pair”*. Proper designs would require this to be randomized or structured in a manner that does not potentially bias any responses. Second, the conclusions of the paper are based entirely on the correlations seen amongst these artificially created researcher-based scales. For example, the statement: *“These findings raise important questions about taste versus health descriptors in cigarette packaging”* (page 7) is based entirely on the correlations amongst the *a priori* scales for which there are no justifications given.

3.46 It is more worrying when one compares results between Hammond, et al. (2009) (discussed in detail in paragraphs 4.3 to 4.15 of my 2010 Report and in summary in paragraphs 3.48 to 3.56 below) and this study and a number of inconsistencies that appear to be sampling related. First, this study was conducted with adults and one conclusion is that *“smokers were significantly more likely to perceive differences in taste, tar delivery and health risk”* (page 6). However, Hammond, et al.’s (2009) table 2 discussed in paragraph 4.8 of my 2010 Report, shows no effects of difference perceptions from smoking intensity. This reinforces the conclusion that leaving out non-smoking adults from that sample makes the

⁴⁷ See paragraph 4.22 of my 2010 Report.

comparisons invalid. Second, the percentage of adults who indicate “no difference” in Hammond, et al. (2009) hovers around 60% over all the packages examined. However, in this study the numbers average less than 10%. It is difficult without any hypotheses to understand such large differences and why this sample would choose the “no difference” option so little. One possible conclusion is that the structure of the study is itself an ‘artifact’ that is influencing the choices being made by the subjects.

3.47 Put simply, this study does not provide reliable evidence of the type set out in paragraph 1.6 above.

Hammond et al. (2009).

3.48 This study reviews data obtained from 516 adult smokers and 806 minors aged between 11 and 17 years of age in the United Kingdom who participated in an online survey. Researchers assessed the degree to which adult smokers and minors perceived differences in tar, lower “health risk”, taste, and “attractiveness” of different packages. I have reviewed this study in detail at paragraphs 4.3 to 4.15 of my 2010 Report, but have summarized the key limitations with the study below.

3.49 Firstly, this study suffers from study design problems and empirical measurement issues, as well as a questionable sampling frame. It fails to account for the issue of prior knowledge and the degree to which a “false belief” is simply an artifact of the failure to sample non-smoking adults. Overall, while it purports to examine the intentions of individuals, it does not have either a theoretical or empirical model of how those decisions are made, either ideally or in the practical reality of the marketplace.

3.50 In this study there are 5 package ‘attributes’ – brand, colour, smoothness, “light” and size. Brand has six ‘levels’ – Marlboro, Mayfair, L&B, Richmond, Silk Cut, and B&H. Colour has 8 levels – Gold, Red, Silver, Purple, Dark Grey, Dark Red, Plain White and Plain Brown. Smooth – labeled Smooth or not – Size – Regular and King – and Light – labeled Light or not – have two levels each. This would imply a $6 \times 8 \times 2 \times 2 \times 2 = 384$ design, which in theory is workable. However, this implies a very large number of possible product alternatives that have to be incorporated into the experimental structure. The authors neither explain the design they have chosen, nor

do they justify it in any manner that is consistent with best practice experimental design science.⁴⁸ Nor, when they exclude potential options from the mix of alternatives they present to their subjects, do they justify why those alternatives are being excluded. Overall, giving the individual only 14 choices ensures that: (a) the design is very inefficient statistically and unlikely to create an estimateable decision model and (b) is unbalanced as the various attributes/levels appear an unequal number of times (as noted above). The implication is that we can say nothing definitive about the nature of the choices that would arise because we cannot isolate the specific effects of the components of the packaging. Nor without a properly structured design can one justify general statements as to the impact of specific colours, brands or labels.

3.51 The science underlying the application of experimental methods to consumer choice can be extremely complicated, but, putting it in simple terms to assist a lay reader of this report, there are two key points to note. First, for the reasons set out in paragraphs 3.38 and 3.39 above, the authors have not accounted for all the pack design variables they have introduced and which could influence decision making and would allow the researcher to make valid statements about the influence of the composition of the package on the decisions being made. For example, I note in paragraphs 3.38 and 3.39 that basic aspects of experimental design science – orthogonality and balance – are being ignored. Researchers can, in certain circumstances, apply what are known as non-orthogonal and non-balanced designs, which adds even more complexity to the study and requires that quite sophisticated statistical analyses be applied that account for the nature of the experimental design. This calls into question the plain packaging findings of the Hammond, et al. (2009) study as the results could be arising for any number of effects that the researchers failed to control for in the structure of their experiments. Secondly, all of the possible permutations that are relevant to meaningful conclusions should form part of the study design. In paragraph 3.50, I noted that there were 384 possible alternatives that could potentially arise from the mixture of ‘attributes’ and ‘levels’ in this study. This does not mean that each person needed to see all 384 alternatives or even that all of the 384 permutations needed to be in the experiment. However, the experimental design had

⁴⁸ Street, D. and L. Burgess (2007). *The Construction of Optimal Stated Choice Experiments*, Hoboken, NJ.

to be structured statistically so that the researcher was confident that they could get as much of the information that was possible given that there were 384 alternatives. This is what is known as the ‘efficiency’ of the experimental design. Very simply, it asks, how many and which of the 384 possible alternatives do people have to see so that the researcher gets enough information to make a statistically valid guess at the decision model being used.

3.52 There is also the issue of the appropriateness of the sample. It is unclear why only adults who smoked were the relevant comparator group to minors who both did and did not smoke. So a critical comparison is missing and that is between non-smoker adults and non-smoker teens. This is telling in that table 2 in the paper reveals that there is no difference in the number of cigarettes smoked a day and perceptions of package differences, while the only consistent finding for minors in table 3 in the paper is that smokers were more likely to say there was a difference. This could be arising simply due to the fact that smokers were more knowledgeable (or believed they need to make what appeared to be a ‘knowledgeable’ choice).

3.53 The questions ultimately do not represent actual decisions. For example, the conclusions that are drawn are based upon a belief that “lower tar”, “smoother taste” and “package attractiveness” are the operative decision making criteria, without any validation that that is indeed the case in reality (an incentive compatibility issue). Nor do we know the degree to which they serve as effective differentiators for consumers in their choice. In other words, we do not know the ordering of their importance in making a decision and how they might stand up against other salient attributes, such as price, brand, availability, peer group effects and so on. It is best practice in choice experiments to span the domain of all the relevant attributes the consumer considers in making a choice,⁴⁹ most notably price (which is excluded from this study) and also product availability. Otherwise, one risks biasing the results by making those factors that the researcher highlights as the most important in making a choice.⁵⁰

⁴⁹ See, for example, Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge, UK: Cambridge University Press, who notes that the choice sets presented to people should be “*exhaustive, in that all possible alternatives are included*” (p. 15).

⁵⁰ Hensher, D., Rose, J. and W. Greene (2005), *Applied Choice Analysis: A Primer*, Cambridge, UK: Cambridge University Press.

3.54 The choice questions do not represent actual choices, and as a result, force the subject into a situation where they are speculating about “trying to reduce” smoking and/or “making it easier to quit”. Again, we do not know anything about the hypothesized decision model related to smoking reduction or cessation and whether the questions are increasing the salience of the packaging as a driver of this behaviour (otherwise, why would the research be asking the subject the question?). We do not know what it is that the individual is actually thinking when attempting to link the act of “trying to reduce” smoking to the package; in other words what the cognitive process is that is driving the answer to the question. This, again, is an example where a “pseudo opinion” (discussed above at paragraph 3.52) could arise due to the fact that the individual is being asked to speculate on something for which there may be no relationship at all.

3.55 The statistical analysis is inappropriate. With paired comparison choice tasks the appropriate modeling structure is a logit.⁵¹ The dependent variable would simply be a binary or multinomial choice – which option(s) was(were) chosen – and the independent variables would be the experimental design. The individual level factors – such as gender, age, social status, smoking status and so on – would be incorporated as covariates.⁵² This would allow for a more immediate comparison of the influence of the package attributes and their levels and would also account for the heterogeneity amongst individuals. It would also allow for a direct modeling of the decision making structure of the individual and would hence alleviate the need for making inferences from indirect questioning.

3.56 In conclusion, this study applies an inadequately designed paired comparison choice study and an inappropriate statistical analysis. The researchers also do not have

⁵¹ A logit model estimates the probability that an option is chosen based upon the levels of the attributes that option possesses plus a random component. As noted by Train (2003), it is “by far the easiest and most widely used discrete choice model. ... Its popularity is due to the fact that the formula for the choice probabilities takes a closed form and is readily interpretable.” The structure, history and appropriateness of this model is discussed by McFadden in his 2000 Nobel Prize Lecture; McFadden, D. (2001), “Economic Choices,” *American Economic Review*, 91:351-378. For rating scales it is more appropriate to use a multinomial probit model (see, Haaijer, R. and M. Wedel (2007) for the difference in the statistical models).

⁵² How these ‘covariates’ are added into the model is will be based upon how they are hypothesized to influence a decision. Train (2003) and Henscher, et al. (2005) show how different statistical models need to be used to account for the different ways in which factors such as age, social economic status or other demographics are accounted for in choice models.

a sampling frame that allows them to draw the conclusions that they do as they are comparing knowledgeable adult smokers with both knowledgeable and unknowledgeable non-adult smokers and non-smokers. Accordingly, I do not consider this study reliable evidence as to whether the various design restrictions would help to achieve the public policy goals listed at paragraph 1.6 above.

4. CONCLUSION

4.1 Ultimately, my overall conclusion on the impact of the various design restrictions on actual smoking behaviour as outlined in paragraph 1.6 above is that the Studies fail to come to any consistent or valid conclusions. This is due not just to methodological issues but to a fundamental lack of a behavioural model that relates any aspects of attitudes and intentions as measured in Studies to actual behaviour seen in the market. The Studies' major flaw is that they assume behavioural effects exist rather than proving that they exist.

4.2 I have come to this conclusion based upon evaluations of each study on a series of criteria that determines the extent to which a study provides valid and generalisable conclusions that align with what an individual would do across a range of contexts, such as might exist in purchasing situations. My conclusions are that the Studies fail to account for significant factors that influence their validity and predictability and relevance to the policy questions stated. The major issues are:

- (a) **Salience and Incentive Compatibility.** The Studies should be generating results that represent realistic behaviour (incentive compatibility) without bias induced by making aspects of the product or the situation more salient than it would in reality. It is my conclusion that the Studies failed to meet reasonable incentive compatibility requirements. Because the Studies focus on packaging absent other salient attributes of the products (such as price) and other factors, we do not know the degree to which the Studies provide realistic information about the degree to which packaging matters.
- (b) **Context.** It is my conclusion that the Studies fail to provide experimental or situational contexts that created realistic scenarios in which the individual

would be applying the decision model that the individual used when making purchasing decisions or decisions that related to the policy goals.

(c) **Attitudes and Intentions \neq Behaviour.** The Studies concentrate almost entirely on stated preferences, attitudinal measures and focus group opinions, often about the likely behaviour of others (rather than the respondents themselves). One cannot assume any predictive accuracy with respect to actual purchasing behaviour or the intended policy goals.

(d) **Methodological Limitations.** I have also come to the conclusion that the Studies – in varying degrees and with varying mixtures – fail on a number of research quality dimensions. In particular they fail with respect to their: Field Administration Protocol, Appropriateness of the Sample Frame, and Analysis.

(i) **Field Administration Protocol** asks whether the **questions** were appropriately constructed and relevant, there was no **researcher bias**, the researcher was **objective**, and the **responses** received were relevant. In the case of all of the Studies, there are significant limitations in the structuring of questions.

(ii) The **sampling frame** in the Studies vary making comparisons between the Studies and drawing generalizable conclusions from the Studies nearly impossible.

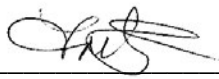
(iii) The empirical analysis in the Studies suffers from significant **statistical validity** issues. These include experimental design limitations, flaws in the coding of data, and invalid and significantly limited statistical procedures when considered in light of best practice.

4.3 It is my expert opinion that the Studies relied on in the IA in respect of the various design restrictions do not provide reliable evidence that such a measure would be effective in achieving the public policy goals of changing actual smoking behaviour, namely in:

(a) reducing smoking uptake (also known as initiation) among minors;

- (b) reducing smoking consumption among minors and/or adults; or
- (c) increasing smoking cessation among minors and/or adults.

4.4 I confirm that insofar as the facts stated in my report are within my own knowledge I have made clear which they are and I believe them to be true, and that the opinions I have expressed represent my true and complete professional opinion.

Signature  _____

Name: Professor Timothy M. Devinney

Date: 1 June 2013

5. EXHIBIT ONE – QUALIFICATIONS OF PROFESSOR DEVINNEY

Ex.1.1 I am a Professor of Strategy at the University of Technology, Sydney. In addition I am a Conjoint Professor in the Faculty of Medicine at the University of New South Wales and a Visiting Professor at the Institute of Management at Humboldt University – Berlin.

Ex.1.2 My educational background includes a B.Sc. (1977) in Psychology from Carnegie Mellon University, and M.A. (Public Policy, 1979), MBA (Economics and Statistics, 1981) and PhD (Economics, 1984) from the University of Chicago.

Ex.1.3 I have held academic positions at the University of Chicago (Lecturer), Vanderbilt University (Asst Professor), University of California – Los Angeles (Asst Professor), The Australian Graduate School of Management (AGSM) (Associate, Chaired Professor, Professorial Research Fellow) and the University of Technology – Sydney (Professor).

Ex.1.4 I have held visiting Professorships at the Universities of Trier, Frankfurt, Ulm, Hamburg and Humboldt University in Germany, London Business School in the UK, Copenhagen Business School in Denmark, Hong Kong University of Science and Technology and City University in Hong Kong.

Ex.1.5 I have taught MBA and doctoral courses at University level for over 25 years. I teach in the marketing, international business, strategic management, innovation and statistics/research methods areas. I was the Founding Director of the Executive MBA Program at the AGSM and have taught extensively on executive development programs around the world.

Ex.1.6 I am one of the leading researchers in the social sciences in Australia, having secured extensive research funding through the Australian Research Council and other external funding bodies.

Ex.1.7 I have published in the leading business journals in the field of social sciences including the *Journal of Marketing*, *Journal of International Business Studies*, *Management Science*, *Organization Science*, *Strategic Management Journal*, and many others. I am on the editorial board of 10 of the leading academic journals and serve as Co-Editor of the *Academy of Management Perspectives* and the *Advances in International Management* Series published by Emerald. I am also the author or editor of more than 6 books.

Ex.1.8 I am a Fellow of the Academy of International Business, an International Fellow of the Advanced Institute of Management (UK), a Fellow (Distinguished Member) of the Australia New Zealand Academy of Management, a Research Awardee of the Alexander von Humboldt Foundation (Germany), and a Bellagio Residence Fellow of the Rockefeller Foundation. My work has been recognised by numerous organisations including being awarded the Researcher of the Year award by the Australia New Zealand Academy of Marketing.

6. EXHIBIT TWO – RESUME

Education

B.Sc. (Psychology), with highest honors, Carnegie-Mellon University, 1977

M.A. (Public Policy Studies), University of Chicago, 1979

M.B.A. (Economics and Statistics), University of Chicago, 1981

Ph.D. (Business Economics), University of Chicago, 1984

Academic Experience (excluding visiting positions):

University Professor of Strategy, School of Business, University of Technology
Sydney, July 2009–present.

Professor (Conjoint), Faculty of Medicine, University of New South Wales, July
2009–present.

Professor of Management, Australian Graduate School of Management (now
Australian School of Business), University of New South Wales, 1993–2009
(June). Director Centre for Corporate Change, 1999–2006. AGSM
Professorial Research Fellow, 2006–2009.

Assistant Professor of Management, Anderson Graduate School of Management,
University of California, Los Angeles, 1990–1992.

Assistant Professor of Management, Owen Graduate School of Management,
Vanderbilt University, 1982–1990.

Lecturer in Mathematics, Graduate School of Business, University of Chicago,
1981–1982.

Academic Experience (examples of courses taught):

AT UTS: Philosophy of Science (PhD).

At AGSM (Recent MBA/EMBA): International Business in Asia (MBA on site
project course in China), Corporate Strategy (MBA/EMBA), International
Business Strategy (MBA/EMBA), Strategic Management of Intellectual
Property (MBA shortcourse), Philosophy of Social Science (PhD), Ph.D.
Seminars (one on Corporate Strategy and one on Innovation), Globalization
of the Knowledge Based Organization (MBA).

At AGSM (Executive Programs): Managing Competitive Strategy,*
Technology Management,* The International Manager's Program,*
Managing Intellectual Property,* the Accelerated Development Program,
the General Manager Program, and the Development Program for Managers
(* indicates program directorship and development).

At Vanderbilt: Pricing, Managerial Decision-making and EC92, Economics of
the Firm, International Business, Industrial Strategy and Organization,
Regulation and Antitrust Economics, Financial Institutions, Information
Economics (Ph.D. Level)

At UCLA: Marketing Strategy, Product Management, MBA Projects.

Professional Associations/University Affiliations (including awards/recognitions):

Professional Associations and Activities Therein (Including Awards): American
Economic Association, Econometric Society, INFORMS, The Product
Development Management Association, Academy of International Business,
Academy of Management, Australia New Zealand Academy of
Management, Australian New Zealand Marketing Academy

INFORMS: Organizing Committee, Marketing Science Conference, Nashville
1987; Program Coordinator, Euro XII/TIMS International XXXI, Helsinki
1992 and TIMS XXXIII, Singapore 1995; Organizing Committee,
Marketing Science Conference, Sydney 1995

Academy of International Business (AIB): Chair, Academy of International
Business, Annual Meeting, Sydney, 2001. Journal of International Business
Studies (editorial board, consulting editor), 2006 JIBS Decade Award
Committee Chair; Faculty 2006 AIB Doctoral Consortium. 2007 Program
Organizing Committee (Track Chair). 2009 Program Organizing Committee
(Track Chair). 2010 Program Organizing Committee (Track Chair). AIB
Fox Best Paper Award Committee, 2009–2012. Best Paper Award Finalist
(2005), AIB Fellow, Elected 2008

ANZAM (ANZ Academy of Management): Executive Committee, ANZAM (2005–2006), Best Paper Award (2001, 2006 & 2007), 2007 ANZAM Conference Organizing Committee, Distinguished Member (Fellow), 2008 ANZMAC (ANZ Marketing Academy): Researcher of the Year Award, 2007 Academy of Management (AOM): Executive Committee and Chair, International Management Division of the Academy of Management (2004–2009). Raised funding for the Booz & Co./strategy+business eminent scholar award in International Business (2004–2009). Organized IMD PDW program (2005). Organized IMD Conference Program (2006). Served as Division Chair (2007–2008). Received one of three AOM Enterprise Awards (2005) to develop multimedia delivery of IMD Program activities. Chair IMD Communications Committee (2008–2009). IMD Service Committee (2010–2012). BPS Junior Faculty Consortium Faculty (2006). New Doctoral Consortium Faculty (2008). Carolyn Dexter Award Nominee (Social Issues in Mgt Division)—Best Intl Paper (2007). Evidence-Based Management Collaborative (2007–2008), Founding Member. Academy of Management Perspectives, Co-Editor (2011-2015), Associate Editor (2006-2010)

Strategic Management Society (SMS): Strategic Management Journal (editorial board), Global Strategy Journal (editorial board), International Management Interest Group: Representative-at-large (2008–2009), Associate Program Chair (2011), Program Chair (2012). Conference Organizing Committee, Rio de Janiero (2010).

Editorial and Refereeing Duties (Formal):

Director, Social Science Research Network (SSRN), International Business Research Network and Editor, International Business Strategy & Structure, 2009–present

Co-Editor, *Advances in International Management*, Emerald (with T. Pederson and L. Tihanyi), 2009–present

Co-Editor, *Academy of Management Perspectives*, (with D.Siegel)
 2011–present, Associate Editor, 2006–2011
 Associate Editor, *Australian Journal of Management*, 1995–2005
 Associate Editor, *Management Science*, 1988–1990
 Consulting Editor, *Journal of International Business Studies*, 2011–present
 Editorial Board, *Strategic Management Journal*, 2007–present
 Editorial Board, *Journal of International Business Studies*, 2003–present
 Editorial Board, *Global Strategy Journal*, 2010–present. Co-Editor of Special
 Issue on Managing Global Stakeholders (2013)
 Editorial Board, *Strategic Organization*, 2006–present
 Editorial Board, *Corporate Governance: An International Review*, 2007–
 present. Co-Editor of Special Issue on CSR, Institutions and Governance
 (2012)
 Editorial Board, *Journal of Product Innovation Management*, 1991–present
 Editorial Board, *Journal of Strategy & Management*, 2008–present
 Editorial Board, *European Management Review*, 2005–present
 Editorial Board, *Asia Pacific Journal of Management*, 2003–present
 Editorial Board, *BuR–Business Research*, 2007–present
 Editorial Board, *Journal of Small Business Economics*, 1993–2002

Examples of Research Support Received

- 1993–1994 Andrus Foundation, Washington D.C., Age as a Rating Factor in
 Automobile Insurance Pricing (with B. Cooil, Vanderbilt
 University).
- 1999–2002 SRG and UCG, Hong Kong, Measuring the Utility Value of
 Ethical Consumerism (with Patrice Auger, City University HK,
 and Jordan Louviere, Sydney) (SRG) (ARC) (granted 1/6/99)
- 2002–2003 CRC for Smart Internet Technology (Linkage Grant), Measuring
 Customer Response to Radical Future Technologies (with
 J. Louviere, UTS and industry partners Westpac) in 2004.

- 2003–2005 Australian Research Council (Discovery Grant), Cross-Cultural Differences in Perceptions Of Consumption Ethics (with G. Eckhardt, AGSM, and R. Belk, Utah).
- 2003 Nokia/Telstra, Using Lead User Research to Determine the Demand for 3G Service Delivery .
- 2003–2005 Transurban, Discrete Choice Modeling of Infrequent Road Users (with J. Louviere, UTS).
- 2004 AIM Fellowship (ESRC UK), Performance of UK Firms (with G. Yip and G. Johnson).
- 2005–2007 Australian Research Council (Linkage Program), Patterns of Rural Segmentation (with J. Louviere and S. Gudergan, UTS) (Described at: <http://www.ruralchoice.com.au>).
- 2005–2007 Australian Research Council (Discovery Grant), Information Provision and the Valuation of Social Attributes (with P. Auger, MBS, A. Gunthorsdottir, AGSM, J. Louviere and M. King, UTS).
- 2006–2009 Australian Research Council (Linkage Program; Bluescope Steel), A Simulation Based Approach to Understanding Alternative Supply Chain Configurations (with T. Coltman, J. Gattorna and T. Spedding).
- 2007–2010 Australian Research Council (Linkage Program; ACT Health), An Action Research Project to Strengthen Inter-Professional Learning and Practice Across the ACT Health System (with J. Braithwaite, R. Iedema, J. Westbrook, R. Foxwell, R. Boyce, K. Murphy, M.-A. Ryall, J. Beutel, M. Budge, W. Ramsey).
- 2009–2012 Australian Research Council (Discovery Program), The Value of CSR to Close Stakeholders: A Discrete Choice Modelling Approach (with P. Auger, MBS, and G. Dowling).
- 2010–2014 Australian Research Council (Discovery Program), Extreme Values: The Anatomy of Civil Society Supporters and Protest

- Groups (with R. Belk, York U., J. Schwalbach, HU-Berlin, P. Auger, MBS and A. Gunnthorsdottir).
- 2010-2012 Alexander von Humboldt Foundation. Research Reward (renewed).
- 2012–2016 Australian Research Council (Linkage Program), Technology and Innovation Management in High Risk Situations (with T. Coltman, R. Sharma, S. Gudergan, B. Brooks and N. Lin).
- 2013-2015 Australian Research Council (Discovery Program), Culture and the Individual: Advancing the Measurement of Culture (with D. Caprar, E. Maitland, P. Auger, C. Eckert, B. Kirkman and T. Peterson).

Professional Consulting Experience (Selection)

Management consultant for various organisations including large corporations—United Press International, IMS/Dun & Bradstreet (London), Apple Computer, Martin-Marietta (Department of Energy), NationsBank, Dominion Bank, Nuturn Corporation, The Tennessee Valley Authority, LG (Seoul), Permanent General Insurance, Boral, AT Kearney (London), GEC-Alstom, AMP, Morgan & Banks/TMP, GM/Holden, CSR, Mobil, Koppers Industries, SAP, Rolls Royce (UK), SAS Institute, Telecom Austria, Hanimex/Rabbit Photo, Thomson Publishing, Transurban, Nokia, Telstra, Sabanci Holdings (Turkey), Borusan (Turkey), Anadolu (Turkey) and Westfield Holdings.

Government and non-profits – the State Council of the PRC (China), the Government of PNG, Amnesty International, Greenpeace, WWF, The Property Council of Australia, Invest Australia, Australian Manufacturing Council and the City of Sydney.

7. EXHIBIT THREE – SELECTED PUBLICATIONS

Ex.3.1 “Is Microfoundational Thinking Critical to Management Thought and Practice?” *Academy of Management Perspectives*, 27, 2, May 2013.

Ex.3.2 “How Much Does a Company’s Reputation Matter in Recruiting?” *MIT Sloan Management Review*, Spring, 2013 (with P. Auger, G. Dowling, C. Eckert & N. Lin).

Ex.3.3 “Taking the “Non-“ out of “Non-Market” Strategy,” *Global Strategy Journal*, 3, 2.

Ex.3.4 “Continuing Differences Between Health Professions’ Attitudes: The Saga of Accomplishing Systems-Wide Interprofessionalism,” *International Journal for Quality in Health Care*, 25, 1, February 2013 (with J. Braithwaite & 6 others).

Ex.3.5 “Can the Socially Responsible Consumer be Mainstream?” („Kann der gesellschaftlich verantwortliche Konsument zum Mainstream werden?“) *Zeitschrift für Wirtschafts- und Unternehmensethik (J. for Business, Economic and Ethics)*, 12, 3 (with P. Auger & G. Eckhardt).

Ex.3.6 Institutional Theory in International Business and Management, *Advances in International Management (Volume 25)*, Emerald, 2012 (co-editor) (with T. Pedersen and L. Tihanyi).

Ex.3.7 “Perspectives on the Art and Science of Management Scholarship,” *Academy of Management Perspectives*, 26, 1, February 2012 (with D. Seigel).

Ex.3.8 “Social Responsibility, Global Strategy and the Multinational Enterprise: Global Monitory Democracy and the Meaning of Place and Space,” *Global Strategy Journal*, 1, 3-42 (launch issue), 2011.

Ex.3.9 “Customer Relationship Management and Firm Performance,” *Journal of Information Technology*, 26, 3, 2011 (with T. Coltman & D. Midgley).

Ex.3.10 “Dynamics of Globalization: Location-Specific Advantages or Liability of Foreignness?” *Advances in International Management* (Volume 24), Emerald, 2011 (co-editor) (with C. G. Asmussen, T. Pedersen and L. Tihanyi).

Ex.3.11 “Using Frontier Analysis to Evaluate Company Performance,” *British Journal of Management*, 21, 4, December 2010 (with G. Yip & G. Johnson).

Ex.3.12 “The Myth of the Ethical Consumer,” Cambridge: Cambridge University Press, 2010 (with P. Auger and G. Eckhardt).

Ex.3.13 “The Past, Present and Future of International Business and Management,” *Advances in International Management* (Volume 23), Emerald, 2010 (co-editor) (with T. Pedersen and L. Tihanyi).

Ex.3.14 “The Importance of Intangible Social Attributes in Consumer Purchasing Decisions: A Multi Country Comparative Study,” *International Business Review*, 19, 2 (with P. Auger, J. Louviere & P. Burke). A variant of this paper was nominated for the Carolyn Dexter Award for the best international paper at the 2007 AOM Conference.

Ex.3.15 “Measuring Organizational Performance as a Dependent Variable: Towards Methodological Best Practice,” *Journal of Management*, 35, 3, June 2009 (with P. Richard, G. Yip & G. Johnson).

Ex.3.16 “Measuring Long Term Superior Performance: The UK’s Long-Term Financial Performers, 1983-2004,” *Long Range Planning*, 42, 3, June 2009 (with G. Yip & G. Johnson).

Ex.3.17 “Is The Socially Responsible Corporation a Myth? The Good, Bad and Ugly of Corporate Social Responsibility,” *Academy of Management Perspectives*, 23, 2, May 2009 (shortlisted for best paper of 2009).

Ex.3.18 “The Financial Times Business Schools Ranking: What Quality is This Signal of Quality?” *European Management Review*, 5, 4, Winter 2008 (with G. Dowling & N. Perm-Ajchariyawong). There are three additional commentaries on this paper in the same issue.

Ex.3.19 “Formative versus Reflective Measurement Models: Two Applications of Formative Measurement,” *J. Business Research*, 61, 12, December 2008 (with T. Coltman, D. Midgley & S. Venaik).

Ex.3.20 “Do Social Product Features Have Value to Consumers?” *International Journal of Research in Marketing*, 25, 3, September 2008 (with P. Auger, J. Louviere & P. Burke).

Ex.3.21 “Do Managers Behave as Theory Suggests? A Choice Theoretic Examination of Foreign Direct Investment Location Decision Making,” *Journal of International Business Studies*, 38, 7, December 2007 (with P. Buckley & J. Louviere) best paper finalist, AIB Conference, Quebec City, 2005.

Ex.3.22 “Do What Consumers Say Matter? The Misalignment of Preferences with Unconstrained Ethical Intentions,” *Journal of Business Ethics*, 76, 4, December 2007 (with P. Auger).

Ex.3.23 “Using Best-Worst Scaling Methodology to Investigate Consumer Ethical Beliefs Across Countries,” *Journal of Business Ethics*, 70, 3, February 2007 (with P. Auger & J. Louviere).

Ex.3.24 “The Other CSR,” *Stanford Social Innovation Review*, Fall 2006 (with P. Auger, G. Eckhardt & T. Birtchnell).

Ex.3.25 “Dual Paths to Performance: The Impact of Global Pressures on MNC Subsidiary Conduct and Performance,” *Journal of International Business Studies*, 36, 6, 2005 (with D. Midgley & S. Venaik).

Ex.3.26 “Modular Strategies: B2B Technology and Architectural Knowledge,” *California Management Review*, 47, 4, Summer 2005 (with P. Richard).

Ex.3.27 “Client and Agency Mental Models in Evaluating Advertising,” *International Journal of Advertising*, 24, 1, 2005 (with M. Collins & G. Dowling).

Ex.3.28 “A New Perspective on the Integration-Responsiveness Pressures Confronting Multinational Firms,” *Management International Review*, 44, SI1, 2004 (with D. Midgley & S. Venaik). Reprinted in: J.A. Krug and J.D. Daniels, *Multinational Enterprise Theory*, Thousand Oaks, CA, 2007.

Ex.3.29 “What Will Consumers Pay for Social Product Features?” (with P. Auger, J. Louviere & P. Burke), *Journal of Business Ethics*, 42, 3, 2003.

Ex.3.30 “Knowledge Management: Philosophy, Process, and Pitfalls,” *California Management Review*, 44, 4, Summer 2002 (with A. Deering, D. Midgley, & C. Soo). Best Paper, 2001 ANZAM Conference.

Ex.3.31 “Managing the Global Corporation: Case Studies in Strategy and Management,” 2nd Edition, New York: McGraw-Hill, 2001 (with José de la Torre & Yves Doz).

Ex.3.32 “E-Business: Revolution, Evolution or Hype?” *California Management Review*, 44, 1, Fall 2001 (with D. Midgley, T. Coltman & A. Latekefu).

Ex.3.33 “The Organisational Imperative and the Optimal Performance of the Global Firm: Formalising and Extending the Integration-Responsiveness Framework,” *Organization Science*, 11, 6, 2000 (with D. Midgley & S. Venaik).

Ex.3.34 “Understanding Institutional Designs Within Marketing Value Systems,” *Journal of Marketing*, 63, Special Issue, 1999 (with S. Carson, G. John & G. Dowling).

Ex.3.35 “Paying the Piper an Incentive to Play a Better Tune: Understanding and Resolving Advertiser-Agency Conflicts,” *Journal of Business-to-Business Marketing*, 6, 1, Spring 1999 (with G. Dowling).

Ex.3.36 “A Formal Model of Trust Based on Outcomes,” *Academy of Management Review*, 23, 3, July 1998 (with R. Bhattacharya & M. Pillutla).

Ex.3.37 “The Essence of Corporate Strategy: Theory for Modern Decision Making, Sydney,” Allen & Unwin, 1997 (with Jeremy Davis).

Ex.3.38 “How Well Do Patents Measure New Product Activity?” *Economic Letters*, 41, April 1993.

Ex.3.39 “New Products and Financial Risk Changes,” *Journal of Product Innovation Management*, 9, September 1992.

Ex.3.40 “New Product Innovations and Stock Price Performance,” *Journal of Business Finance & Accounting*, 19, September 1992 (with P. Chaney).

Ex.3.41 “The Return to Advertising Expenditure,” *Marketing Letters*, 3, May 1992, (with B. Cooil).

Ex.3.42 “European Markets After 1992: Implications for Business Strategy,” Lexington, MA: Lexington Books, 1991, (with William C. Hightower).

Ex.3.43 “The Impact of New Product Introductions on the Market Value of Firms”, *Journal of Business*, 64, October 1991, (with P. Chaney and R. Winer).

Ex.3.44 “New Products Over The Business Cycle,” *Journal of Product Innovation Management*, 7, December 1990.

Ex.3.45 “Diversification Strategy and Performance in Canadian Manufacturing Firms”, *Strategic Management Journal*, 11, September 1990 (with T. Nguyen and A. Seror).

Ex.3.46 “Rationally Determined Irrationality: An Extension of the Thesis of Rationality as Anti-Entropic,” *Journal of Economic Psychology*, 10, November 1989.

Ex.3.47 Numerous other articles in other journals, books, and magazines. Note that none of the above includes book chapters or other publications, patents or magazine/newspaper publications that have been excluded for space reasons.

8. EXHIBIT FOUR – OTHER MATERIALS CONSIDERED

Ex.4.1 In addition to the Studies, the IA identifies a number of other materials which it claims are relevant to the various design restrictions. These additional materials are studies and other materials which are related to the various design restrictions as a regulatory initiative (either by their conclusions or their content), but which do not generate any original consumer survey evidence in this regard.

Ex.4.2 Given that the scope of my report addresses the extent to which publicly available consumer surveys provide reliable evidence that the various design restrictions will achieve the public policy goals set out in my report at paragraph 1.6, I have not considered studies or other materials which fall into the above criteria in formulating my conclusions in this report.

Ex.4.3 For completeness, however, I have set out below a list of the studies which are listed in the IA and which do not generate any original consumer survey research in this regard:

- (a) Tan, YL., and K. Foong (2012). How the Malaysian tobacco industry exploits loopholes in pictorial health warnings. *Tobacco Control* 21 (1): 55-56.
- (b) Thrasher, J., Hammond, D., Arillo-Santillan, E.A. (2010). The alchemy of Marlboro: transforming “light” into “mild”. *Tobacco Control*; 19:342-343.
- (c) Ernster, VL., Lloyd, G., Norman, LA., McCarthy, A., and AL Pinto (2001). Women and Smoking: A Report of the Surgeon General. Executive Summary. Atlanta, GA: Centers for Disease Control and Prevention.
- (d) Eher, F., Gallopel-Morvan, K., Beguinot, E., and Y Martinet (2011). Smokers and non-smokers’ perceptions of current cigarettes versus plain cigarettes. Conference presentation, SRNT Europe.