4th International Conference of the Academy of Wine Business Research, Siena 17-19th July 2008
Refereed paper

How important is wine packaging for consumers?

On the reliability of measuring attribute importance with direct verbal versus indirect visual methods

Dr Simone Mueller
Research Fellow Wine Marketing
Ehrenberg-Bass Institute for Marketing Science, University of South Australia
GPO Box 2471 Adelaide, South Australia 5001
Telephone: (61 8) 8302 0906
Facsimile: (61 8) 8302 0042
Email: simone.mueller@unisa.edu.au

Prof Larry Lockshin
Professor of Wine Marketing
Ehrenberg-Bass Institute for Marketing Science, University of South Australia
GPO Box 2471 Adelaide, South Australia 5001
Telephone: (61 8) 8302 0261
Facsimile: (61 8) 8302 0042
Email: larry.lockshin@unisa.edu.au

How important is wine packaging for consumers?

On the reliability of measuring attribute importance with direct verbal versus indirect visual methods

Abstract

Wine packaging design has received a growing research interest in the last few years. Nevertheless different approaches to measure the relative importance of packaging compared to other extrinsic cues like brand name, origin and price yield deviating results. Verbal methods directly asking consumers about wine packaging relevance usually result in low packaging importance ratings contradictory to what we know from the market place. We review previous research in the measurement of packaging attribute importance and discuss psychological differences found between direct verbal and indirect visual methods. We compare the results of two methods to measure wine attribute importance: a direct verbal Best Worst Scaling (BWS) experiment versus an indirect visual discrete choice experiment (DCE). With BWS all visual extrinsic cues are not only measured as less important than verbal cues but also show a smaller variance between respondents, signalling a strong respondent agreement on their non-importance. Contrary, the DCE combining label and packaging attributes in wine bottle graphics in a shelf-like setting reveals a higher average importance and strong consumer preference heterogeneity of wine packaging design attributes and levels.

Our results imply that stimuli which are visually perceived by consumers cannot be reliably measured with verbal methods. This results in strongly biased results. Choice experiments with packaging graphics, which simulate consumers' real purchase behaviour in front of the shelf are a powerful tool for marketing practitioners. It allows them to efficiently measure the relative importance of design attributes for different consumer segments and to cost-efficiently test consumer acceptance of newly developed wine packaging in the market place.

Keywords: Wine packaging, verbal versus visual representation, research methodology, discrete choice analysis, Best Worst Scaling

Introduction

Wine packaging has received increasing research attention in the last few years (Barber, Almanza, & Donovan, 2006; Boudreaux & Palmer, 2007; Orth & Malkewitz, 2006 and 2008; Rocchi& Stefani, 2005; Szolnoki, 2007). Appearance and packaging of food products and wine play an important role in influencing consumer perception and subsequent acceptance (Imram, 1999).The first taste is almost always with the eye. Extrinsic packaging attributes provide consumers with social and aesthetic utility and strongly influence expectations of sensory perception (Deliza & MacFie, 1996; Gianluca, Donato, & Cavicchi, 2006; Sara R. Jaeger, 2006; Lange, Martin, Chabanet, Combris, & Issanchou, 2002). Those expectations have been shown to be very robust against later disconfirmation when consumers actually taste the product (Cardello & Sawyer, 1992). Despite what we know about the underlying psychological influence packaging exerts on product evaluation, contradictory findings were found on the relative importance of wine packaging compared to other extrinsic product cues as brand name, origin and price.

Several studies directly measuring the importance of attributes conclude that wine packaging design is rather unimportant (Goodman, Lockshin, & Cohen, 2005, 2006, 2007; Mueller, Lockshin, Louviere, & Hackman, 2007). Other studies find that strong consumer impressions are evoked by wine packaging design elements (Boudreaux & Palmer, 2007; Orth & Malkewitz, 2006) and that during in depth focus groups consumers reveal they consider packaging design features when making purchase decisions (Rocchi & Stefani, 2005; Szolnoki, 2007). A first indicative study including a relatively small subset of packaging attributes without considering product price by Szolnoki (2007) reveals that the importance of wine packaging designs differ when measured directly and indirectly.

It further can be expected that the importance of wine packaging design and preferred attribute levels differ for different wine consumers as empirically confirmed for other food products (Deliza, MacFie, & Hedderley, 2003; Silayoi & Speece, 2007). Nevertheless, the majority of previous wine packaging studies did not consider consumer preference heterogeneity, which is managerially important to target different consumer segments.

Thus, a major unresolved research question is how wine packaging preference and importance can be reliably and validly measured. To answer this question we will first discuss prior findings on different psychological processes initiated by visual and verbal information and review previous empirical studies comparing direct verbal and indirect visual attribute importance measurement. To test our two research propositions we compare wine packaging design importance and importance variance in two experimental settings – a direct verbal Best Worst Scaling study with an indirect graphical discrete choice experiment. We will discuss the validity and reliability of both methods and conclude how graphical choice experiments can provide the wine industry with extremely valuable advice for product development and consumer targeting.

Literature review

Different psychological processes initiated by visual and verbal information

Different types of information processing induced by verbal and visual information are found to be responsible for conceptual differences between verbal and graphical product representation. Paivio (1971) proposed a dual-coding hypothesis, implying that pictures tend to be processed simultaneously in an imagery system, whereas verbal representations are processed sequentially in an independent system. Findings by Allan Paivio & Csapo(1973) confirm that imagery can be substantially better recalled than verbal information and that image and verbal memory are independent and additive in their effect on recall, confirming the dual encoding hypothesis. Similar explanations, although from different perspectives, have also been brought forward by Das, Kirby, & Jarman (1975), Mandler & Johnson(1976) and Pick & Saltzman(1978). These findings assert that verbal descriptions are very likely to generate different connotations than the corresponding pictorial representations of the same product.

Other research focused on humans' processing ability and preference for verbal and visual information (Childers, Houston, & Heckler, 1985; Sojka & Giese, 2001). MacInnis & Price (1987) and Jarvenpaa (1989) showed that there are lower cognitive elaboration costs and higher benefits for graphical relative to verbal information. Furthermore, visual representations can enhance problem-solving capabilities without overloading decision makers (Lohse, 1997). Kosslyn (1994) discussed how humans have developed visual and

spatial skills and better retrieve information with visual cues. Lurie & Mason(2007) compared the context of visual versus verbal information, which showed vividness, evaluability and framing increased with visual information; that is, "a picture is worth a thousands words".

Empirical differences between verbal, visual and real product presentation

A review of the early literature comparing effects of verbal and visual product presentation of the 1980 can be found in Vriens, Loosschilder, Rosbergen, &Wittink (1998). Ambiguous findings of those early studies are very likely partially caused by very simple and unrealistic graphical representation techniques like line drawings. Whereas Holbrook & Moore (1981) found stronger effects for visual than for verbal sweater descriptions a replication of this study by Domzal & Unger (1985) for watches did not result in significant differences. Similarly, Louviere, Schroeder, Louviere, & Woodworth(1987) comparing verbal descriptions and visual photographic representations of state parks in choice experiments found only a few differences in part-worth between representation modes. Smead, Wilcox, & Wilkes (1981) compared real coffee makers and their verbal presentation and found more eye movement and more significant preference determining attributes for real products.

Vriens *et al.*(1998) compared the relative importance of different design attributes of car stereo equipment with verbal and visual representations in a conjoint analysis. Pictorial representation produced slightly higher relative importance for two of three design attributes and a somewhat greater heterogeneity among respondents. Despite a higher degree of task realism for photographic representations, they concluded based on hold-out tasks that verbal representation facilitated judgement and had higher predictive accuracy.

Dahan & Srinivasan (2000) compared verbal, visual and physical product presentation of bicycle pumps for a conjoint analysis-based product concept test and found strong between differences verbal and visual, but only minor differences between visual web animated and physical product presentation. They concluded that that Internet visual presentation and costlier real prototype experiments produced a close match. But they only used full concepts and price, no design was used to measure and combine several attributes.

Silayoi & Speece (2007) used a rating based conjoint study for packaged ready-to-eat products in Thailand and found a strong overall importance of packaging shape, packaging colour and packaging graphics. They also confirmed strong consumer heterogeneity with distinct segments focused either on visual aesthetics or verbal product information on the label.

Jaeger, Hedderley, & MacFie (2001) compared photographs and real products in conjoint analysis and found that apple photographs conveyed information about apple varieties equally well compared to prototype apple packages. Because of equal validity and lower application costs, they recommended using designed graphics for conjoint analysis.

Sethuraman, Kerin, & Cron (2005) findings supported the use of internet technology for conjoint analysis data collection. Online data collection was judged superior to a traditional offline (paper-and-pencil) method. The differences were explained by greater participation attention and involvement especially because of visual enhancement of the pictorial objects possible in web-based tasks.

Szolnoki (2007) is the only study known to the authors utilising wine. He compared the relative importance of wine packaging elicited with rating and rating-based conjoint analysis for German wine consumers. Using verbal direct measurement, packaging design was rated as second least important after wine flavour and origin, but surprisingly before brand. The rather weak discrimination between the items could have been caused by the usage of a five point rating scale (Cohen & Neira, 2003). For the rating based conjoint analysis he combined four attributes origin/grape variety, label layout, bottle form and bottle colour with either two or three levels in photographic bottle representations. Label style showed the overall largest effect with an importance of 40%, followed by origin/grape variety (30%), bottle colour (19%) and bottle form (12%). Neither brand nor prices were included in the conjoint design. Szoloniki (2007) found three heterogeneous consumer segments using hierarchical cluster analysis on conjoint attribute weights, thereby loosing statistical efficiency in a two-step procedure.

Differences direct versus indirect attribute importance measurement

Most recent research suggests that there are also fundamental differences between direct and indirect importance measurement methods, independent of whether they apply verbal or graphical stimuli. Van Ittersum, Pennings, Wansink, & van Trijp (2007) found in their meta analysis that methods measuring different dimensions of attribute importance usually have a lower correlation than methods analysing the same dimension. Direct methods as rating or BWS measure the underlying dimension of attribute relevance, determined by personal values and desires. On the other hand, indirect methods such as conjoint or discrete choice analysis measure the determinance of an attribute - its relevance in judgement and choice – which is seen as most important from a managerial perspective (Pennings & Smidts, 2003).

Louviere & Islam (2007) found context effects and the degree of definition of a reference frame were responsible for differences between directly and indirectly derived importance measures. They argued that the importance of price depends on the ranges of price values a respondent has previously experienced, expects to experience or as provided by the researcher. Directly asking for importance of price is only meaningful if all subjects use the same frame of reference (e.g. $7.99-$22.99 for a bottle of wine). How this reference can best be defined by a researcher (Huffman, 1997) and connects again to the difference of verbal and visual stimuli. While a attributes such as price, brand, and region can unambiguously be defined verbally this is not the case for visual stimuli. Imagine the number of different shades of red respondents could refer to if the reference is set verbally. This exemplifies how important graphical methods are to be able to define the same basis of reference for respondents.

Research Propositions

Drawing from prior findings on differences between direct and indirect attribute importance measurement, and verbal and visual product presentation we will analyse two research propositions:

1) The relative importance of wine packaging attributes will be lower under verbal representation in direct importance measures than under visual representation in indirect attribute importance analysis.

2) Respondent heterogeneity in inferred relative attribute importance for packaging attributes will be higher under visual than under verbal representation.

Research Method

1) Direct verbal attribute importance measurement

For direct verbal extrinsic wine attribute importance measurement we use Best-Worst Scaling developed by Finn & Louviere(1992) which has shown to be a powerful method for preference measurement in social sciences and marketing (Auger, Devinney, & Louviere, 2007; Cohen & Orme, 2004; Goodman *et al.*, 2006; J. Louviere & Islam, 2007; Marley & Louviere, 2005). BWS uses respondents' choices of the best (most important) and worst (least important) item in set to create a ratio-based scale and overcomes several biases resulting from scores or ratings. This results in better discrimination between attributes (Cohen & Neira, 2003; Marley & Louviere, 2005). Despite the fact that we use BWS here as a direct verbal method - as it also has mainly been used in the past - it has to be emphasised that BWS is not limited to verbal attributes but can equally be applied to graphical concepts, as will later be shown.

Based on previous studies (Orth & Malkewitz, 2006; Rocchi & Stefani, 2005) and in store analysis in Australian retail stores we selected a total number of 16 extrinsic wine attributes (see list in Table 3). We thereby limited the potential detail of the bottle and label attributes as analysed by Orth & Malkewitz (2006) to a few more aggregated attributes like label style, label form, bottle shape and bottle colour along with standard verbal attributes like brand and price. We assigned all 16 attributes to a Youden design with 24 choice sets and choice set size of 6.

740 regular wine consumers (purchasing and drinking wine at least twice a month) from around Australia, recruited in March 2007 via a panel provider, completed an online questionnaire. The sample is very similar to the population of Australian wine drinkers with a slightly larger share of younger consumers (see Table 1). Respondents were asked to state within each set of wine characteristics the ones that are most and least important for their purchase decision of a bottle of Shiraz wine in a retail store.

2) Indirect visual attribute importance measurement

Discrete choice analysis or choice-based-conjoint are now predominating the measurement of attribute importance according to trade announcements of the commercial market leader in conjoint software, Sawtooth Software. Discrete Choice Experiments (DCE), use experimental designs to combine attribute levels into bundles. Respondents are forced to make tradeoffs when choosing bundles or product concepts. This method has been shown to be more valid in predicting actual choice (Louviere, Hensher, & Swait, 2000).

For a first proof of concept we had to limit the number of extrinsic wine attributes in the DCE because of the exponential growth of design complexity with a linear increase in attributes and levels (Street, Burgess, & Louviere, 2005). From the total of 16 attributes used for the direct verbal BWS we selected three verbal (brand, price and region scored highest in BWS) and three visual attributes (label style, label colour and bottle shape, which widely vary in the Australian wine market). The limitation on six attributes with either two or four levels (listed in Table 2) allowed us to have every respondent complete a full choice design and subsequently analyse respondent heterogeneity. A complete comparison between the BWS and DCE importance weights will not be possible as the relative attribute importance in choice experiments depends on the presence and absence of other choice relevant attributes (Islam, Louviere, & Burke, 2007).

Table 1: Comparison of experiment samples to total population of Australian wine consumers (Roy Morgan Single Source Australia: Jan 2006 – Dec 2006)

		Roy Morgan (total wine consumers)	Best Worst Experiment (n=740)	Discrete Choice Experiment (n=244)
State	NSW	34.3%	30.9%	34.2%
	Victoria	25.7%	25.5%	22.3%
	Queensland	18.4%	17.4%	19.7%
	South Australia	7.7%	10.3%	7.8%
	Western Australia	10.8%	12.7%	7.8%
	Tasmania	2.3%	2.8%	5.2%
	Northern Territories	0.6%	0.4%	3.1%
Area	Capital Cities	65.3%	74.6%	64.2%
	Country Area	34.7%	25.4%	35.8%
Gender	female	52.2%	53.5%	51.2%
	male	47.8%	46.5%	48.8%
Age	18-24	8.2%	14.9%	13.4%
	25-34	16.1%	32.7%	20.5%
	35-49	31.4%	34.9%	32.0%
	>50	44.3%	17.5%	34.1%
Marital status	single	30.7%	37.3%	31.1%
	married/ de facto	69.3%	62.7%	69.9%
Children in household	yes	31.8%	43.1%	46.6%
	no	68.2%	56.9%	53.4%
Number of children	1	13.3%	18.0%	13.6%
	2	12.7%	14.1%	16.8%
	3+	5.7%	10.0%	9.9%
Personal monthly income (AUD)	under $20,000	18.1%	21.8%	25.9%
	$20,000 - $39,999	24.8%	28.8%	28.2%
	$40,000 - $69,000	32.6%	33.6%	26.8%
	$70,000 or more	24.7%	15.8%	19.1%
Education	Some Secondary/Tech.	14.6%	8.5%	10.1%
	Certificate	16.5%	19.5%	19.0%
	High School	17.7%	17.1%	16.4%
	Degree or Diploma	51.3%	54.9%	54.5%
Employment	full time work	47.7%	58.3%	49.2%
	part time work	20.3%	19.6%	20.4%
	not employed	32.0%	22.1%	30.4%
Home ownership	Own Home	76.0%	58.2%	67.0%
	Rent Home	24.0%	41.8%	33.0%

Table 2: Attribute and levels for visual Discrete Choice Experiment

	Attribute	Levels	1	2	3	4
1	Price	4	$7.99	$12.99	$17.99	$22.99
2	Label style	4	traditional	chateau	graphic	minimalistic
3	Label colour	4	whitish	yellowish	orange	dark grey
4	Brand	2	Jinks Creek	McWilliams		
5	Region	2	Henty	McLaren Vale		
6	Bottle shape	2	Bordeaux	Burgundy		

Price levels were chosen to cover the commercially most relevant price range for Australian wine. Based on a content analysis of several store checks four types of label styles – traditional, chateau, graphic, and minimalistic – were found to compose the vast majority of different wine labels. A quantitative analysis of wine label colours in several Adelaide retail outlets revealed the four colours off-white, yellowish, orange/red and grey/black to be most dominant. Brand and region levels were chosen to represent a well known and an unknown example of each. Bordeaux and Burgundy are the two most available bottle shape types in Australia.

Figure 1: Sample discrete choice experiment with graphical bottle representations

It is well known that the range of attribute variation and number of levels used in DCE designs influences the inferred attribute importance (Wittink, Krishnamurthi, & Reibstein,

1990). When comparing attribute importance between verbal BWS and visual DCE we therefore have to consider that the importance of attributes with only two levels (brand, region, and bottle shape) can be negatively biased compared to attributes with four levels.

Attributes and levels were assigned according a 2^3x4^3 orthogonal main-effect plan in 16 choice sets with choice set size 6, its statistical efficiency is about 91%. Graphical designers developed graphical bottle representations of all attribute levels with prices given below, typical for a retail environment. Respondents were asked to choose the wine from the 'shelf' they most and least prefer and stated if they realistically would purchase the most preferred wine (see Figure 1). While typical choice based conjoint experiments only ask respondents to choose the best option, asking them for the best and worst in each set provides significantly more choice information (see Louviere, Eagle, & Cohen (2005) and references given there).

244 regular wine consumers (purchasing and drinking wine at least twice a month) from around Australia, recruited via a panel provider, completed the online experiment. As for BWS, the sample is very similar to the total population of Australian wine consumers (see Table 1).

Analysis and Results

1) Direct verbal attribute importance measurement

For BWS we counted the number of times an attribute was chosen as most important (best) and least important (worst) on aggregated level. (Marley & Louviere, 2005) Calculating the square root of the ratio of best to worst frequency counts for each attribute results in a bias free measure of attribute importance on a ratio scale (Marley & Louviere, 2005). This ratio scale can be standardised to a maximum value of 100 to result in a probabilistic scale (Mueller, Francis, & Lockshin, 2007). The relative importance of each attribute can then easily be compared by its coefficient to the most important attribute. For example country of origin is about half as likely to be chosen most important as brand.

According to the standardised importance measure in Table 3 verbal attributes such as brand, price and region are most important for respondents' purchase decision of a bottle of wine. Other verbal extrinsic attributes like medals/awards, country of origin and alcohol level follow in the middle. By contrast, all visual wine characteristics are consistently found as least important. Verbal Best Worst Scaling results imply that characteristics like bottle shape and colour, and label shape and colour only are five percent as important as brand. If these results are valid wine marketers could stop spending money on label design and fancy bottle shapes but instead sell their wine in brown paper bags.

The differentiation between verbal and visual extrinsic wine cues becomes even more prominent if we consider consumer heterogeneity. The standard deviation of the average best minus worst counts per attribute indicates how much attribute importance deviates over the total sample (Mueller, Rungie, Goodman, Lockshin, & Cohen, 2008). The relationship between attribute importance and importance heterogeneity is depicted in Figure 2.

Table 3: Verbal Best Worst Scaling results (n=740)

	Best	Worst	Sqrt(B/W)	Sqrt stand.	B-W Mean	Stdev
Brand	3052	145	4.59	100.0	3.93	2.94
Midpriced wine	2392	203	3.43	74.8	2.96	3.07
Promotional pricing	2577	302	2.92	63.7	3.07	3.35
Region of origin	2433	317	2.77	60.4	2.86	3.18
Medals awards	2321	386	2.45	53.4	2.61	3.50
Country of origin	1911	324	2.43	52.9	2.14	2.91
Bottle size	564	631	0.95	20.6	-0.09	1.99
Alcohol level	718	905	0.89	19.4	-0.25	3.21
Closure material	369	960	0.62	13.5	-0.80	2.42
Organic	358	1348	0.52	11.2	-1.34	3.22
Capsule material	259	1288	0.45	9.8	-1.39	2.46
Label style	212	1839	0.34	7.4	-2.20	2.38
Bottle shape	166	1896	0.30	6.4	-2.34	2.23
Bottle colour	128	2075	0.25	5.4	-2.63	2.37
Label shape	166	2708	0.25	5.4	-3.44	2.78
Label colour	134	2433	0.23	5.1	-3.11	2.60

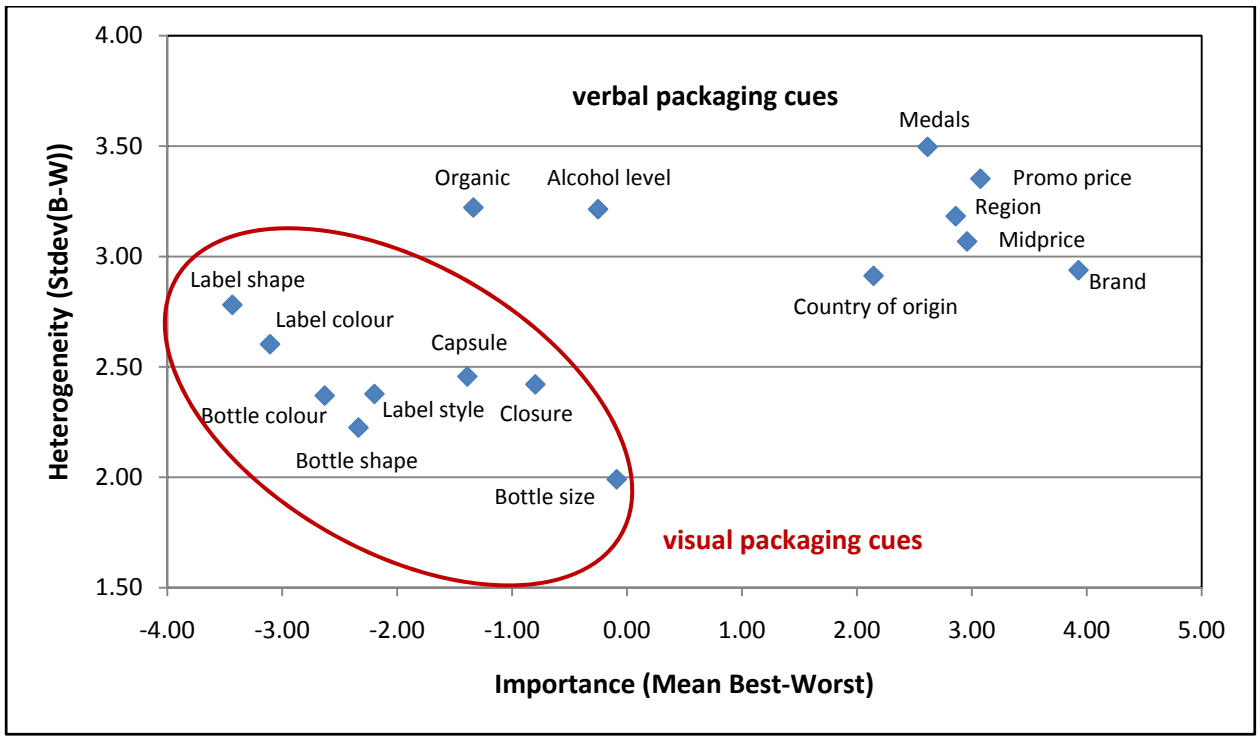

Figure 2: Relationship between attribute importance and heterogeneity

2) Indirect visual importance measurement

We used a scale extended latent class regression model to simultaneously estimate part worth utility parameters and class membership from our discrete choice experiment described above. Thereby individual-level Best-Worst scores for every attribute combination are

regressed against the effects coded attribute levels. We specify a linear regression model from the generalised linear modelling (GLM) family in which parameters (part worth utilities) differ across latent classes (Vermunt & Magidson, 2005). Our latent class model is defined by three components, the assumed probability structure (general mixture model probability structure), the distributional characteristic of the response variable (continuous B-W scores) and the linear scale extended utility regression function (Magidson & Vermunt, 2007). We estimated the model with Latent GOLD Choice 4.5 syntax module (Beta version).

For our model the best fit (lowest BIC value) was achieved with a model of K=5 classes and S=2 scale classes (λ_1=1, λ_2=0.39 (Wald=50.4, p=0.00), n_{s1}=191, n_{s2}=53). Utility part worth estimates for attribute levels for all five classes are given in Table 4. Wald statistics are significant for all attributes except for bottle form and indicate that attribute part worth utilities are significantly different between the classes, with the exception of bottle form, which is equally unimportant for all consumers.

Attribute importance is derived by calculating the range of estimated parameter values for each attribute and then normalising by dividing each attribute's range by the sum of all the attribute ranges. Attribute importance weights derived in this way can be slightly biased by different utility scales. Because of strong non-linearity in estimated price part-worth utilities (see Table 5), a priori standardisation of estimates by a linear price vector β_{price} was not possible as it would rather increase any potential bias. In the recent available beta-version of Latent Gold Choice (Statistical Innovations, Belmont, MA, USA) a derivation of attribute importance by the contribution of every attribute to the Log-Likelihood of the overall model as used by Louviere & Islam (2007) is not yet possible.

Table 4: Attribute importance weights for classes

	Class1	**Class2**	**Class3**	**Class4**	**Class5**	**Mean**
Class size	30%	23%	27%	10%	10%	100%
Brand	22%	16%	6%	13%	8%	14%
Region	8%	2%	2%	4%	5%	4%
Bottle form	4%	0%	0%	1%	0%	1%
Label style	4%	10%	63%	51%	84%	34%
Label colour	16%	6%	18%	20%	2%	13%
Price	47%	66%	10%	12%	2%	33%

The last column of Table 4 shows that in average over the total sample label style was most important, very closely followed by price. Brand and label colour were almost equally important as third and fourth most important, whereas region and bottle form followed as least important. For the attribute levels, all classes prefer the better known brand McWilliams over the made-up brand name Jinks Creek. Similarly all classes reveal a higher probability of choosing the well known region McLaren Vale over the rather unknown region Henty. This lower than expected attribute importance weights of brand and region are likely caused by only considering two attribute levels compared to four levels for all other attributes (Wittink *et al*., 1990). Future research with indirect visual importance measurement methods should include a similar number of attributes and levels of for all extrinsic attributes.

Table 5: Estimates of scale extended Latent Class choice model

		Class1	Class2	Class3	Class4	Class5	Mean	Std.Dev.	Wald	df	p
Class size		30%	23%	27%	10%	10%	100%				
		price + brand		label style + colour		label style					
		brand	price	flexible	chateau, graphic	minimalistic					
Predictors											
Brand	Jinks Creek	-0.835	-0.558	-0.195	-0.522	-0.250	-0.506	0.032	228.3	5	0.00
	McWilliams	0.835	0.558	0.195	0.522	0.250	0.506	0.032			
Region	Henty	-0.306	-0.072	-0.085	-0.173	-0.151	-0.163	0.022	33.1	5	0.00
	McLaren	0.306	0.072	0.085	0.173	0.151	0.163	0.022			
Bottle form	Bordeaux	0.145	-0.002	0.010	0.022	0.011	0.049	0.017	5.8	5	0.32
	Burgundy	-0.145	0.002	-0.010	-0.022	-0.011	-0.049	0.017			
Label style	traditional	0.101	0.029	0.910	-2.729	-2.118	-0.202	0.072	954.8	15	0.00
	chateau	0.168	-0.114	1.230	1.447	-0.497	0.461	0.073			
	graphic	-0.123	-0.303	1.129	1.443	-0.501	0.303	0.069			
	minimalistic	-0.145	0.388	-3.269	-0.161	3.116	-0.561	0.036			
Label colour	white	0.627	-0.073	0.223	0.648	0.010	0.297	0.066	172.5	15	0.00
	yellow	0.016	0.094	0.449	0.427	-0.050	0.188	0.065			
	orange	-0.016	0.183	0.161	-0.081	0.046	0.078	0.064			
	grey	-0.627	-0.204	-0.832	-0.994	-0.005	-0.563	0.029			
Price	$7.99	1.577	-2.612	-0.302	0.238	0.050	-0.188	0.073	883.7	15	0.00
	$12.99	-2.022	2.012	-0.367	-0.538	-0.028	-0.294	0.073			
	$17.99	-0.434	1.347	0.296	-0.149	-0.045	0.243	0.069			
	$22.99	0.879	-0.747	0.373	0.449	0.023	0.239	0.028			

R^2= 0.5325; LL =-8,048.99; BIC(LL) = 16,493.77, n = 244, #parameters = 72; Classification Error = 0.0857, 5 Classes and 2 Scale Classes

Nevertheless, the strength of visual extrinsic packaging cues, label style and label colour, stand in stark contrast to their importance measured by the direct verbal method above. The unimportance of bottle form for Australian wine consumers found in verbal only study indicates the reliability and power for discrimination of this indirect graphical measurement method – all three cues were almost equally unimportant when measured by the direct verbal method (Figure 2).

By jointly interpreting importance weights and attribute level utilities it becomes clear that the five classes can be grouped in two more general consumer types of almost equal size: consumers who consider price and brand as most important (class1 and class2 together form 53%), and respondents who mainly value label style and label colour (class3, class4 and class 5 amount to 47%).

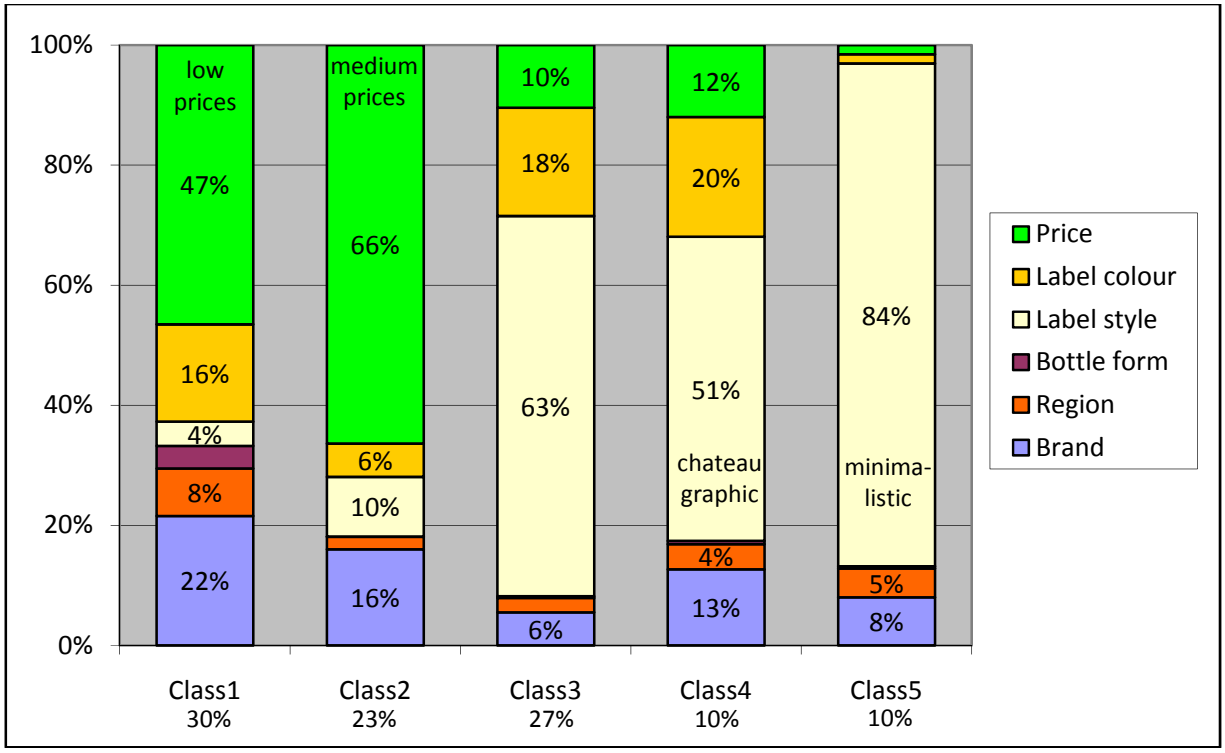

Figure 3: Importance weights for all classes

The first two classes base their wine choice mainly on price and brand, where class1 strongly prefers lower prices ($7.99), and class2 values medium prices ($12.99 and $17.99). Surprisingly consumers of class2 who favour medium prices and very likely perceive price as a quality indicator show a higher price importance than the low price choosing consumers of the first class. Label colour and label style of both price+brand classes are rather unimportant, but reveal different preferences. Whereas the low price class1 prefers traditional and chateau labels in white colour, the medium price class2 mostly likes minimalistic and traditional labels in yellow and

orange. Given only two regions were considered in our experiment, class1 values region with the highest importance of all clusters.

Despite the fact that class3 and class4reveal relatively similar attribute importance weights, they differ in the attribute levels most preferred. Class3 has a wide tolerance for all label styles and colours as long as they are neither minimalistic nor grey; both levels are very much disliked by this class. Both other label style oriented classes 4 and 5 have much more specific label and colour preferences. While the fourth class also dislikes grey, it has a narrower colour preference for white and yellow labels. The traditional label style is very disliked by class4, whereas chateau and graphical label styles are equally well liked. Of all classes class5 seems to apply the simplest decision heuristic when choosing wine, only accepting the minimalistic label style. From all other attributes only brand and region are very marginally important. Future research should endeavour to investigate how valid this respondent choice behaviour is for real market transactions.

Table 6: Sociodemographic differences between clusters

	Class1	**Class2**	**Class3**	**Class4**	**Class5**	Total
	30%	23%	27%	10%	10%	100%
average age	42.3	48.6 [b]	37.1 [a]	45.2	45.0	42.9
female %	47.2	41.1	50.7	72.0	66.7	51.2
male %	52.8 [b]	58.9	49.3	28.0 [a]	33.3	48.8
number of children in hh	1.04	1.00	1.02	1.23 [a]	0.78 [b]	1.02
home owner %	55 [a]	74	69	68	78 [b]	67
education	average	average	highest	lower	lower	
part time working%	12	19	13 [b]	27 [a]	22	17

In the next step we characterised respondents in the five different clusters by differences in their sociodemographics and wine behaviour. Surprisingly we found no significant differences in the wine purchase or consumption frequency, wine involvement or subjective wine knowledge between the five classes. In contrast, consumer segments were most strongly discriminated by sociodemographic characteristics such as age and gender, which could be caused by the overrepresentation of graphical attribute levels in the choice design. Both classes with high price and brand importance have a higher than average share of male respondents, implicating a very cognitive decision process, whereas especially the classes 4 and 5 with very specific preferences for label style are significantly dominated by female wine consumers. This could be a confirmation of previous studies exploring gender differences in decision making (Venkatesh, Morris and Ackerman, 2000; Powell and Ansic, 1997) that females tend to be more affective than cognitive decision makers when it comes to wine choice. These two classes also show the lowest share of under- and postgraduate education degrees and the highest share of part-time working. The third class with a high importance of label style but rather broad acceptance of vivid colours and label styles has the lowest average age and shows the highest education level of all classes. Contrary the oldest consumer class2 shows a strong preference for medium and higher prices. The cluster with the highest preference for low prices also has the lowest rate of home ownership, which is very highly correlated with available income in Australia.

Discussion

Our first proposition that the indirect graphical method utilising DCE results in a higher importance for visual extrinsic wine attributes was confirmed. Label style and label colour were on average the most (34%) and the forth most important attribute (13%) in the DCE. This stands in contrast to the direct verbal BWS method where label colour and label style are by far the least important attributes. Bottle form was not found to be an important choice driver for Australian wine consumers in the choice experiment (1%) or the verbal Best Worst task. Contrary to the direct verbal method, the graphical indirect DCE was much better able to discriminate the relative importance of visual cues.

Visual cues had a comparable or higher variance than verbal cues of the same number of attribute levels, e.g. price, in the DCE and were found to be important drivers of consumer segmentation. This stands opposite to the verbal BWS where visual extrinsic cues had a distinct lower heterogeneity than verbal packaging cues (Figure 2). Thus, our second proposition was also empirically confirmed.

As previously discussed in the literature review, BWS and DCE are not expected to result in identical attribute and importance weights because they measure different underlying dimensions (Van Ittersum *et al.*, 2007) and rely on different referent frames (Louviere & Islam, 2007). Our extremely contrary attribute importance findings for visual cues by both methods can hardly be explained with different underlying dimensions of attribute relevance and determinance. Though, graphical stimuli define an unambiguous reference frame in the DCE we would have expected a higher and not a lower variance of attribute importance in the direct BWS if different mental reference frames were the main reason for our observed differences between both methods.

Besides these explanations, at least two further reasons are responsible for explaining the observed differences between visual and verbal methods. The first are social demand characteristics, responsible for answers respondents believe are socially accepted and right (Cooley, 1983), which also operate unconsciously. It is surely socially more accepted to purchase a wine for its taste, quality, brand and regional reputation than for its packaging design, which could result in a lower directly measured importance. Secondly, visual cues are known to have subliminal effects which we are not aware of (Bornstein, Leone, & Galley, 1987; Monahan, Murphy, & Zajonc, 2000). If respondents do not consciously realise the impact of visual cues like colour and label design, they will not be able to report those effects in direct importance measures. Subliminal effects can then only be deciphered and quantified if respondents replicate their choice behaviour with visual cues in close to real choice settings like choice experiments.

We found mainly gender and age to be discriminating variables between consumer segments between those with high importance of cognitive cues (brand and price) compared to visual cues (label style and label colour). Wine involvement and wine consumption or purchase frequency were not found to be significantly different between consumer segments. Because the focus of this research was to test the concept of the validity and importance of the measurement of visual wine packaging and labelling cues, future studies should include a broader range of attributes and levels to further refine the description of different consumer segments.

Conclusion

The importance of visual wine packaging cannot be measured with direct verbal instruments. Instead indirect visual methods like conjoint analysis and DCE employing graphics are necessary to reliably capture those attributes. This also implies that the validity of attribute importance of wine packaging variables measured with direct verbal elicitation (e.g. rating) has to be doubted.

Besides those insights for research methodology, our findings also have high managerial relevance. Wine marketers can not only use DCE with graphically combined attribute labels for concept testing in new product development to find which packaging attributes are most important for their target consumer segment. Instead, photographically real labels, prototypes and innovative wine packaging (e.g. cans, tetra packs) can be included and tested for their relative performance compared to competitor products (Srinivasan, Lovejoy, & Beach, 1997). Only tactile experiences cannot (yet) be simulated with computer based experiments. But today's available graphical computer methods, high internet band width and representative online panels give the wine industry the chance to relatively inexpensively test and develop their product packaging in close to real life shelf settings compared to very expensive real market introductions with their high failure likelihood.

The relative attribute importance of price, brand, region, label style, label colour and bottle form included in the DCE was too limited to cover a complete picture of what drives Australian wine drinkers as only a subset of extrinsic attributes were considered in order to limit the complexity of the choice design. Research by (Louviere & Islam, 2007) has confirmed that attribute importance depends on the number of attributes and levels considered in a choice task. Future research should therefore include more (all relevant) attribute and levels in graphical DCE experiments. This also requires practical solutions to apply very complex choice designs in visual choice analysis.

Acknowledgements

The authors wish to thank Stewart Gluth, Head of Graphic Design Studio at South Australian School of Art, UniSA. This work was financially supported by Australia's grapegrowers and winemakers through their investment body the Grape and Wine Research and Development Corporation (GWRDC), with matching funds from the Australian Government.

References

Auger, P., Devinney, T., & Louviere, J. (2007). Using Best–Worst Scaling Methodology to Investigate Consumer Ethical Beliefs Across Countries. *Journal of Business Ethics, 70*, 299-326.

Barber, N., Almanza, B. A., & Donovan, J. (2006). Motivational factors of gender, income and age on selecting a bottle of wine. *International Journal of Wine Business Research, 18*(3), 218-232.

Bornstein, R. F., Leone, D. R., & Galley, D. J. (1987). The generalizability of subliminal mere exposure effects: influence of stimuli perceived without awareness on social behavior. *Journal of personality and social psychology, 53*(6), 1070-1079.

Boudreaux, C. A., & Palmer, S. (2007). A charming little Cabernet: Effects of wine label design on purchase intent and brand personality. *International Journal of Wine Business Research, 19*(3), 170-186.

Cardello, A. V., & Sawyer, F. M. (1992). Effects of disconfirmed consumer expectations on food acceptability. *Journal of Sensory Studies, 7*(4), 253-277.

Childers, T. L., Houston, M. J., & Heckler, S. E. (1985).Measurement of Individual Differences in Visual Versus Verbal Information Processing. *Journal of Consumer Research, 12*(2), 125-134.

Cohen, S., &Orme, B. (2004). What's your preference? *Marketing Research, 16*(2), 32-37.

Cohen, S. H., &Neira, L. (2003). *Measuring preference for product benefits across countries: Overcoming scale usage bias with Maximum Difference Scaling.* Paper presented at the ESOMAR 2003 Latin America Conference Proceedings, Uruguay.

Cooley, C. H. (1983).*Human nature and the social order*: Transaction Publishers.

Dahan, E., &Srinivasan, V. (2000).The predictive power of Internet-based product concept testing using visual depiction and animation. *Journal of Product Innovation Management, 17*(2), 99-109.

Das, J. P., Kirby, J., &Jarman, R. F. (1975). Simultaneous and successive synthesis: An alternative model for cognitive abilities. *Psychological Bulletin, 82*(1), 87-103.

Deliza, R., MacFie, H., &Hedderley, D. (2003). Use of computer-generated images and conjoint analysis to investigate sensory expectations. *Journal of Sensory Studies, 18*(6), 465-486.

Deliza, R., &MacFie, H. J. H. (1996). The generation of sensory expectation by external cues and its effect on sensory perception and hedonic ratings: A review. *Journal of Sensory Studies, 11*(2), 103-128.

Domzal, T. J., & Unger, L. S. (1985). Judgments of verbal versus pictorial presentations of a product with functional and aesthetic features. *Advances in Consumer Research, 12*, 268-272.

Finn, A., & Louviere, J. J. (1992). Determining the Appropriate Response to Evidence of Public Concern: The Case of Food Safety. *Journal of Public Policy & Marketing, 11*(2), 12-25.

Gianluca, S., Donato, R., &Cavicchi, A. (2006). Consumer expectations, liking and willingness to pay for specialty foods: Do sensory characteristics tell the whole story? *Food Quality and Preference, 17*(1-2), 53-62.

Goodman, S., Lockshin, L., & Cohen, E. (2005).*Best-worst scaling : a simple method to determine drinks and wine style preferences*. Paper presented at the Second Annual International Wine Marketing Symposium, Sonoma State University, July 8-9.

Goodman, S., Lockshin, L., & Cohen, E. (2006).*Using the Best-Worst method to examine market segments and identify different influences of consumer choice.* Paper presented at the 3rd International Wine Business and Marketing Research Conference, July 6-8, Montpellier.

Goodman, S., Lockshin, L., & Cohen, E. (2007).influencers of consumer choice in a retail setting - more international comparisons. *Wine Industry Journal, 22*(6), 42-48.

Holbrook, M. B., & Moore, W. L. (1981). Feature Interactions in Consumer Judgments of Verbal Versus Pictorial Presentations. *Journal of Consumer Research, 8*(1), 103-113.

Huffman, C. (1997). Elaboration on Experience: Effects on Attribute Importance. *Psychology & Marketing, 14*(5), 451-474.

Imram, N. (1999).The role of visual cues in consumer perception and acceptance of a food product. *Nutrition and Food Science, 99*(5).

Islam, T., Louviere, J. J., & Burke, P. F. (2007). Modeling the effects of including/excluding attributes in choice experiments on systematic and random components. *International Journal of Research in Marketing, 24*(4), 289-300.

Jaeger, S. R. (2006).Non-sensory factors in sensory science research. *Food Quality and Preference, 17*(1-2), 132-144.

Jaeger, S. R., Hedderley, D., &MacFie, H. J. H. (2001). Methodological issues in conjoint analysis: a case study. *European Journal of Marketing, 35*(11/12).

Jarvenpaa, S. L. (1989).The effect of task demands and graphical format on information processing strategies. *Management Science, 35*(3), 285-303.

Kosslyn, S. M. (1994).*Elements of Graph Design*. New York: W. H. Freeman.

Lange, C., Martin, C., Chabanet, C., Combris, P., & Issanchou, S. (2002). Impact of the information provided to consumers on their willingness to pay for Champagne: comparison with hedonic scores. *Food Quality and Preference, 13*(7-8), 597-608.

Lohse, G. L. (1997). The role of working memory on graphical information processing. *Behaviour & Information Technology, 16*(6), 297-308.

Louviere, J., & Islam, T. (2008). A comparison of importance weights/measures derived from choice-based conjoint, constant sum scales and best-worst scaling. *Journal of Business Research, forthcoming*.

Louviere, J. J., Hensher, D. A., &Swait, J. D. (2000). *Stated Choice Methods: Analysis and Application*. Cambridge: Cambridge University Press.

Louviere, J. J., Schroeder, H., Louviere, C. H., & Woodworth, G. G. (1987). Do the parameters of choice models depend on differences in stimulus presentation: visual versus verbal presentation? *Advances in Consumer Research, 14*(1), 79-82.

Louviere, J. J., Eagle, T.C., E., & Cohen, S. H. (2005). Conjoint Analysis: Methods, Myths and Much More. Unpublished Working Paper, Centre for the Study of Choice (CenSoC), UTS Sydney.

Lurie, N. H., & Mason, C. H. (2007). Visual Representation: Implications for Decision Making. *Journal of Marketing, 71*(1), 160-177.

MacInnis, D. J., & Price, L. L. (1987). The Role of Imagery in Information Processing: Review and Extensions. *The Journal of Consumer Research, 13*(4), 473-491.

Magidson, J., &Vermunt, J. K. (2007). *Removing the scale factor confound in multinominal logit choice models to obtain better estimates of preference.* Paper presented at the Sawtooth Symposium Conference Proceedings.

Mandler, J., & Johnson, N. (1976). Some of thousands words a picture is worth. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 529-540.

Marley, A. A. J., & Louviere, J. J. (2005). Some probabilistic models of best, worst, and best-worst choices.*Journal of Mathematical Psychology, 49*(6), 464-480.

Monahan, J. L., Murphy, S. T., &Zajonc, R. B. (2000). Subliminal Mere Exposure: Specific, General, and Diffuse Effects. *Psychological Science, 11*(6), 462-466.

Mueller, S., Francis, L., & Lockshin, L. (2007).Comparison of Best-Worst and Hedonic Scaling for the Measurement of Consumer Wine Preferences.Working Paper, Ehrenberg-Bass Institute for Marketing Science, University of South Australia.

Mueller, S., Lockshin, L., Louviere, J., & Hackman, D. (2007). Do respondents use extra information provided in online Best-Worst choice experiments? *Proceedings of the Xth Australian and New Zealand Marketing Academy (ANZMAC) Conference, Dunedin, New Zealand, 3-5 December 2007*, 3486-3493.

Mueller, S., Rungie, C., Goodman, S., Lockshin, L., & Cohen, E. (2008). Is there more information in Best Worst choice data? Using the variance-covariance matrix to consider respondent heterogeneity. Refereed paper, 4th *International Conference of the Academy of Wine Business Research*, Siena, 17-19thJuly, 2008.

Orth, U. R., &Malkewitz, K. (2006).Packaging design as resource for the construction of brand identity. Paper presented at the Proceedings of the *3rd Academy of Wine Business Research*, Montpellier.

Orth, U. R., &Malkewitz, K. (2008).Holistic packaging design and consumer brand impression. *Journal of Marketing*, 72, 64-81.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart and Winston, Inc.

Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology, 5*(2), 176-206.

Pennings, J. M. E., & Smidts, A. (2003). The Shape of Utility Functions and Organizational Behavior. *Management Science, 49*(9), 1251-1263.

Pick, H. L., & Saltzman, E. (1978). Modes of perceiving and processing information. In P. H. L & E. Saltzman (Eds.), *Modes of perceiving and processing information*. Hillsdale, NJ: Lawrence Erlbaum.

Powell, M., &Ansic, D. (1997), Gender differences in risk behaviour in financial decision-making: An experimental analysis, *Journal of Economic Psychology*, 18(6), 605-628.

Rocchi, B., & Stefani, G. (2005). Consumers' perception of wine packaging: a case study. *International Journal of Wine Marketing, 18*(1), 33-44.

Sethuraman, R., Kerin, R. A., & Cron, W. L. (2005). A field study comparing online and offline data collection methods for identifying product attribute preferences using conjoint analysis. *Journal of Business Research, 58*(5), 602-610.

Silayoi, P., &Speece, M. (2007). The importance off packaging attributes: a conjoint analysis approach. *European Journal of Marketing, 41*(11/12), 1495-1517.

Smead, R. J., Wilcox, J. B., & Wilkes, R. E. (1981). How Valid Are Product Descriptions and Protocols in Choice Experiments? *Journal of Consumer Research, 8*(1), 37-42.

Sojka, J., & Giese, J. (2001). The Influence of Personality Traits on the Processing of Visual and Verbal Information. *Marketing Letters, 12*(1), 91-106.

Srinivasan, V., Lovejoy, W. S., & Beach, D. (1997).Integrated Product Design for Marketability and Manufacturing. *Journal of Marketing Research (JMR), 34*(1), 154-163.

Street, D. J., Burgess, L., & Louviere, J. J. (2005). Quick and easy choice sets: Constructing optimal and nearly optimal stated choice experiments. *International Journal of Research in Marketing, 22*(4), 459-470.

Szolnoki, G. (2007). *Die Messung des Einflusses der aeusserenProduktgestaltung auf die Kaufbereitschaft: dargestellt am Beispiel Weisswein (Measurement of the influence of extrinsic product design on purchase intent for white wine).* Dissertation (PhD thesis) Justus-Liebig-UniversitaetGiesen, GeisenheimerBerichte Bd. 61, Geisenheim.

Van Ittersum, K., Pennings, J. M. E., Wansink, B., & van Trijp, H. C. M. (2007). The validity of attribute-importance measurement: A review. *Journal of Business Research, 60*(11), 1177-1190.

Venkatesh, V., Morris, M.G.,& Ackerman, P. (2000), A Longitudinal Field Investigation of Gender Differences in Individual Technology Adaptation Decision-Making Processes, *Organizational Behavior and Human Decision Processes*, 83(1), 33-60.

Vermunt, J. K., &Magidson, J. (2005). *Technical Guide for Latent GOLD 4.0: Basic and Advanced*. Belmont Massachusetts: Statistical Innovations Inc.

Vriens, M., Loosschilder, G. H., Rosbergen, E., &Wittink, D. R. (1998).Verbal versus Realistic Pictorial Representations in Conjoint Analysis with Design Attributes. *Journal of Product Innovation Management, 15*(5), 455-467.

Wittink, D. R., Krishnamurthi, L., &Reibstein, D. J. (1990). The Effect of Differences in the Number of Attribute Levels on Conjoint Results. *Marketing Letters, 1*(2), 113-123.