Journal of Marketing Research, Article Postprint, Volume XLVI

Brand Extension Strategy Planning:

Empirical Estimation of Brand-Category Personality

Fit and Atypicality

Rajeev Batra*

Peter Lenk

Michel Wedel

* All authors contributed equally; the order of names is alphabetical. Rajeev Batra is S.S. Kresge Professor of Marketing, Ross School of Business, University of Michigan, 701 Tappan Street, Ann Arbor, MI 48109-1234, USA, telephone 734 764 0118; fax 734 936 0279; email: rajeevba@umich.edu. Peter Lenk is Professor of Operations Management Science and Marketing, telephone 734 936 2619; fax 734 936 0279; email: plenk@umich.edu; Ross School of Business, University of Michigan, 701 Tappan Street, Ann Arbor, MI48109-1234, USA. Michel Wedel is Pepsico Professor of Consumer Science, Smith School of Business, University of Maryland, College Park, MD 20742, telephone 301 405 2162, fax 301 405 0146, email: mwedel@rhsmith.umd.edu.

** We gratefully acknowledge the data collection support of GfK and the Marketing Science Institute. We give special thanks to Raimund Wildner of GfK for his interest in this project. Helpful comments by Fred Feinberg and Zeynep Gurhan are also gratefully acknowledged.

Journal of Marketing Research, Article Postprint, Volume XLVI

Brand Extension Strategy Planning:

Empirical Estimation of Brand-Category Personality

Fit and Atypicality

Abstract

The vast majority of brand extensions reportedly fail, suggesting the need for methodologies that will allow for better strategic prediction of categories a brand should extend or license into. The prior literature suggests that brand extensions are likely to be more successful if a brand extends into another category into which its existing brand associations and imagery "fit" better, and that it may help as well if the extending brand is "atypical" (possessing associations and imagery that are broad and abstract rather than tied too closely to the brand's original product category). A methodology is developed, illustrated and validated in this study to estimate brand and category personality structures, using a Bayesian factor model that separates the two by using brand-level and category-level random effects. This methodology leads to measures of a brand's fit and atypicality. The authors illustrate and validate the model on two nationally-representative data-sets on brand personalities in three categories (Jeans, Magazines and Cars), and investigate the brand extension and licensing implications of the results obtained with the model.

Keywords: Brand Extensions; Marketing Strategy; Brand Personality; Category Personality; Marketing Research; Brand Management

Brand extensions – the use of an existing brand name on a new product in a new category, to benefit from the existing brand name's awareness and associations -- leverage the investments made in a company's existing brand names, and hedge against the risk of new product failures. The popularity of this strategy - over eighty percent of new product launches in a typical year are brand extensions (Brand Strategy 2005) - is due to the fact that it leads to higher consumer trial than the use of a new brand name, because of the awareness levels and association (imagery) equities of the brand name being leveraged (Keller 2003, p.582). Many companies today also seek to leverage their existing brand assets via licensing deals to other manufacturers in other categories (such as that of the Caterpillar brand into boots, made by Wolverine Footwear), or via co-branding arrangements (such as Harley-Davidson with Ford trucks). These "extendibility" advantages significantly contribute to a brand's financial value, since they raise the estimate of its future revenues (Keller 2003, p.499).

Not all brand extensions succeed, however, and there is a risk of failure backfiring on the image of the parent brand (Martinez and Pina 2003). In the United States, new products experience failure rates of between 80-90%, and brand extensions fail at a somewhat lower rate (Keller 2003, p.581-582). There has thus been a burgeoning academic research stream on the factors that promote or reduce the success of brand extensions. This academic research has highlighted the contributing role of the breadth and abstractness of the extending brand's associations and imagery, and the fit of these with the target category. This paper fills a void in the literature by proposing a methodology to estimate these constructs by separately measuring the association imagery of the extending brand, its "parent category," and that of the product category being extended into. Using a Bayesian factor-analytic model (cf., Ansari and Jedidi 2000, Ansari, Jedidi, and Dube 2002) for brand and category personality, we derive measures of a brand's atypicality with respect to its parent category and fit to the target category. In our empirical study, we estimate the brand personality model and brand atypicality and fit measures with a nationally-representative sample. We then validate these measures by predictively testing proposed brand extensions with a second, independent, nationally-representative sample. The methodology provides key strategic insights for generating, and assessing, successful brand extension, licensing and co-branding opportunities.

THE BRAND EXTENSIONS LITERATURE

A brand extension is the use of an existing brand name on a new product in a new category, to benefit from the existing brand name's attribute and imagery awareness and associations in gaining consumer trial, retailer distribution, etc., in the new category. For example, a cross-category brand extension could be one where a car brand such as Porsche extends into categories such as pens or eyeglasses. Presumably, consumers' favorable disposition towards Porsche, and its associations with prestige and exciting style, would extend to the new entries. Consequently, the literature argues that the extending brand name must first possess high awareness, and associations that are salient, strong, positive, relevant, and unique (Keller 2003, pp.600-601). These brand associations must then also "fit" the category being extended into, and be "broad and abstract" enough to accommodate the needs of that new category.

<u>Fit</u>

Extensions of a brand to a new category face the particular challenge of needing to "*fit*" (be seen as close to) the new product category being entered. Thus while the imagery surrounding the National Geographic brand name may fit the category requirements for travel clothing, travel shoes or binoculars, the fit would likely be poorer if Money magazine were to launch these same brand extensions. The necessity and basis for this "fit" have been the primary subjects of most academic research on brand extensions in the last fifteen years (Aaker and Keller 1990; for a review, see Keller 2003, pp.608-623). An existing brand name from another category fits a new product category if there appears to be a match at the level of concrete attributes (such as microprocessors, for Intel) or on the basis of abstract imagery or personality attributes (such as prestige and exciting style, for Porsche) (Park, Milburg, and Lawson 1991; Batra, Lehmann, and Singh 1993; Roedder-John and Loken 1993). The larger the number of salient, shared associations between the brand name and the new extension category, the greater the perception of fit. And, the higher the perceived fit, the greater the degree to which the perceptions and preference of the extending brand will be seen by consumers to "carry over" to its new product category. Fit at the level of imagery often is a greater determinant of brand extension success than the degree of favorable overall attitudes toward the extending brand, or the degree of physical similarity between the parent and entered-into product categories (Broniarczyk and Alba 1994).

Current methods of measuring fit simply ask consumers for their overall perceptual assessments, using direct rating scales such as "how well does the proposed extension fit with the parent brand" (Klink and Smith 2001; Keller 2003, p.604). There are at least three issues with the use of such measures. First, as shown by Klink and Smith (2001), consumers' answers to such direct questions are necessarily based on pre-conceived ideas of brand extendibility, and may be confounded with prior attitudes, yielding problematic estimates of fit. Second, these current 'overall fit' estimates do not offer any diagnostic insight into the "basis" of these fit assessments; understanding the specific associations that contribute to, or detract from, "fit" judgments could be important for brand extension opportunity identification and strategy planning. Third, the "concept testing" approach now prevalent does not allow the generation of new brand extension or licensing or co-branding ideas; it simply allows the testing of a limited number of already-generated concepts.

Addressing these limitations, we will present below a non-attitudinal method for empirically generating information about a brand extension's "fit" without asking for direct consumer judgments of fit and similarity between the extended brand and the new category, thus addressing some of the weaknesses of such approaches pointed out by Klink and Smith (2001). We will show through two validation studies from an independent sample of consumers that our approach, which does not rely on typically-used attitudinal measures of extension potential, nonetheless is able to predict them very well. Our approach will also enable strategic insights into the contributing sources of "extension fit" for a particular brand, as it explores which of many candidate brands and categories represents the highest-potential extension, licensing or co-branding opportunities, by studying their attribute and imagery associations. Importantly, our approach can be applied to generate, not merely to test, brand extension concepts, thus being of much greater strategic use than current methods.

Atypicality

In addition to possessing a high degree of fit with the category being extended into, brands with high "extension potential" also benefit from other qualities, according to prior literature. An important finding from brand extension research is that abstract associations are easier to extend than concrete associations – and a brand name that is too strongly identified with only its parent category, than with an abstract quality that spans multiple categories, can be more difficult to extend at all outside the category (Aaker and Keller 1990; Farquahar et al. 1992).

This finding is consistent with earlier research (Rosch et al. 1976; Johnson 1984) that "abstract" associations (such as "entertainment") are inherently more inclusive and super-ordinate (and thus broader, "fitting" into more product/service categories) than "concrete" associations (such as "television sets."). Brands that are marketed on the basis of inherently more abstract "lifestyle" associations (such as Ralph Lauren) have thus historically proved to be very extendible into many other seemingly-disparate product categories (such as table linen, sunglasses, even paint). Thus Heineken's strong association with the concrete "beer" category might make it less extendible than another beer such as Corona which also has a broader lifestyle association of a party-lover or beach-relaxation approach to living. Another instance is Clorox's failure to successfully extend into the detergent category in the late 1980s, apparently limited by its too-strong association with the "bleachness" quality of its "parent" bleach category. The literature therefore suggests that the degree to which a particular brand's associations and imagery are "atypical" of – and thus allow it to go beyond – the associations and imagery of its "parent" product category impacts the degree to which the brand possesses "abstractness" that enhances its ability to extend into other categories.

However, while this line of reasoning suggests that a brand whose imagery is just the standard, default imagery of its 'parent' product category should be limited in its extension potential beyond that category, this may not always be the case. As Tauber (1988) points out, for a brand extension to succeed its existing brand associations must also give the new entrant perceptions of being different and better than the existing brands in that new category, which he calls "competitive leverage" (powerful, unique and strongly-linked associations). An expanding brand could therefore sometimes benefit from, rather than be hurt by, its strong identification with the imagery of its "parent" product category. Thus while the concrete, category-linked “bleachness” association of Clorox may limit its extendability into categories where “strong cleaning power” is *not* valued, this same association may in fact be of great “leverage” advantage in an extended-into category where this association *is* highly valued. Brands may thus actually benefit from being closely linked to their parent product categories, rather than be hurt by it.

As a result of these two competing ideas, it is not entirely clear whether an extending brand should benefit from, or be hurt by, strong identification with the imagery of its parent product category. Nonetheless, it does appear that the degree of such identification might play an important role in determining a brand’s extension ability and power. It thus appears useful to obtain an empirical quantification of the degree to which an individual brand’s imagery overlaps with, or differs from, the imagery of the parent product category with which it is strongly identified. Naturally, the assessment of such “deviance from category baselines” depends on how the product category itself is defined: all cars, or just sports cars, or just family sedans? The Porsche brand may display some atypicality in a broader set of cars, but not if only sports cars alone are being studied. Despite these caveats, to our knowledge no study empirically assessing the quantitative degree of “atypicality” has yet appeared in the literature.

Our approach employs a factor analytic approach to modeling a consumer’s “net” brand associations as being shaped by both the overall associations of the category of which it is a part, as well as its own unique, idiosyncratic brand aspects. It seeks to provide an answer to the questions: To what extent, and for what reasons, is the parent brand seen as being too similar to – and thus possibly limited by – its parent category associations (“atypicality”), and to what extent, and for what reasons, is the personality of its parent category congruent with the imagery of the product categories to which it is potentially being extended (“fit”)?

To empirically assess the degree and type of fit and atypicality for a brand extension, we chose to work in the domain of "brand personality" brand associations, instead of associations about concrete brand attributes or benefits (although our approach is applicable to these as well), since these personality associations are inherently more abstract and therefore may be seen as more relevant across a wide set of product categories (Keller 2003, p.614). Indeed, the recent growth of "lifestyle mega-brands" (such as Ralph Lauren, Martha Stewart, even Nike) into multiple physically-unrelated product categories testifies to the extension-power of these abstract brand personality associations. We turn now a discussion of the literature on these "personality" associations, at the "brand" and "category" levels.

Brand Personality

Almost fifty years of research in marketing (Levy 1959; Martineau 1958) has shown that the perceptions and associations consumers have about brands go beyond their functional attributes and benefits, and include non-functional, symbolic qualities, often referred to as "brand image." Among these aspects of brand image are perceptions and associations about the brand's "personality," the "set of human-like characteristics associated with a brand" (Aaker 1997, p. 347). For instance, among soft drinks, Pepsi is often perceived by consumers as more "young," Coke as more "real and honest," Dr. Pepper as more "non-conformist and fun" (Aaker 1997, p. 348). As discussed earlier, these personality aspects of a brand extension often play a major role in consumer judgments of its fit and leverage in the new category into which it extends (Park, Milburg, and Lawson 1991; Batra, Lehmann, and Singh 1993).

Measuring Brand Personality

The appropriate measurement of existing brand personality imagery has been studied for over twenty years (Plummer 1984-85). Researchers have quite naturally sought to develop a valid and reliable measurement (survey) instrument of brand personality that is usable across various product categories and consumer segments, drawing on the extensive literature on human personality (Digman 1990; McCrae and Costa 1987), but going beyond it where necessary (Batra, Lehmann, and Singh 1993). The measurement instrument used most often recently is the one developed by J. Aaker (1997).

In her extensive development of this instrument, she sought to develop scales "generalizable across product categories" (Aaker 1997, p. 348), by having 631 respondents rate each of 37 brands on 114 personality traits - with these brands being carefully selected to represent a broad array of product/service categories, a few brands per category. She factor analyzed the between-brand variance after averaging the scores of each brand on each personality trait across multiple respondents. (In other words, the data matrix she factor-analyzed was based on pooled data from 37 brands across multiple product categories.) Using this aggregated category/brand matrix, she found five factors, labeled Sincerity (sample item: honest), Excitement (daring), Competence (reliable), Sophistication (upper-class), and Ruggedness (tough); her scale is described in more detail below.

It is widely acknowledged that "most of the research papers on brand personality are now based on Aaker's scale" (Azoulay and Kapferer 2003, p. 144), though there has been criticism of her scale on both conceptual grounds (Azoulay and Kapferer 2003; Caprara, Barbaranelli, and Guido 2001) and for its generalizability (Austin, Siguaw, and Mattila 2003).

Category Personality

There is also a prior literature which suggests that entire product categories (such as beverages), or sub-categories (such as beer or wine or milk), not only brands within them, also are perceived to possess a "personality." Levy (1986, p. 216-217) wrote "a primary source of meaning is the product (category) itself," pointing out that within the beverage category liqueurs connote discrimination, while wine symbolizes snobbism, beer sociability and democracy, soup tradition, and juices virtue. Coffee is seen as stronger and more masculine, tea as weaker and more feminine. Levy (1981, p. 55) also highlighted how user stereotypes - a common source of brand personality (Keller 1993) - differ for specific food categories: chunky peanut butter for boys, but smooth peanut butter for girls; lamb chops and salads for women, steaks for men. Other researchers such as Lautman (1991) have also noted that consumers appear to have a "schema" for different categories, clusters of inter-connected emotions, facts and perceptions stored in memory as a unit. Durgee and Stuart (1987) found that consumers associate "fun" with the entire ice-cream category. Batra and Homer (2004, p. 321) report finding potato chips rated more "fun" than expensive cookies, which were rated as more "sophisticated and classy."

Perhaps because of such background "category context", it has been shown that some brand personality measurement scale items appear, depending on the category, to pick up functional product category characteristics rather than brand personality ones. Thus in one study the brands rated highest on "energetic" were energizer drinks, while the item "sensuous" was most associated with ice cream brands (Romaniuk and Ehrenberg 2003). Given the well-known phenomenon of "concept × scale interaction" in the literature on measurement with scales such as semantic differentials (Osgood, Suci, and Tannenbaum 1957; Komorita and Bass 1967), it is not surprising that certain brand personality measurement items (in Aaker's scales, or others) might mean different things in different product categories (Caprara, Barbaranelli, and Guido 2001). Very importantly for our approach, these findings also suggest that the "net" personality perceptions of a brand are influenced by <u>both</u> the personality of the overall category of which it is a part (CP), <u>and</u> by its own idiosyncratic brand personality (BP) aspects.

We note here that despite the qualitative research evidence reviewed above that product categories as a whole have “personality” associations, there has not to our knowledge been any prior study that has empirically and quantitatively measured an entire category’s personality. Such measurement would be particularly important for the measurement of a particular brand’s atypicality within its parent category, as outlined above. Thus one of the challenges facing our proposed methodology is to first establish that a category’s personality can in fact be measured, and we will seek to do so as a basis for the measurement for the “fit” and “atypicality” assessment that is our main goal. We now describe the empirical data used in the analysis, and then the analytical approach being suggested, which utilizes a random effects hierarchical factor model to estimate brand and category personality. The subsequent sections present the results, followed by our validation studies. We conclude with a discussion.

METHOD

Data

We conducted two studies: the first collected data on brand personality perceptions, and the second study validated co-branding and extension concepts that were derived from the first study on a second sample of subjects. Both studies used a national sample of 200 subjects from an online national consumer panel, using a quota sample to match national proportions on gender, income, education, and location. The three product categories were Cars, Jeans and Magazines, with ten brands in each category (listed in our tables). These categories were selected because we thought they would each involve considerations of the user’s values and lifestyle, and because they would vary meaningfully on Ratchford's (1987) think-versus-feel dimensions (i.e, in the utilitarian, symbolic, and emotional benefits they provide). The collected data showed that they did indeed vary meaningfully on these dimensions, with cars being the highest on the utilitarian (useful, beneficial, practical) and highest on the symbolic (tells others about me, self image) benefits; jeans second-highest on the utilitarian and symbolic ones; and magazines highest on the hedonic (fun, pleasurable, enjoyable) attributes and lowest on the utilitarian and symbolic ones.

A major challenge in developing the questionnaire was its length and the resulting burden for respondents. Assessing a large number of brand personality items for 30 brands presents an insurmountable burden to respondents. Two solutions to this problem were adopted. First, a split questionnaire design was used (Adigüzel and Wedel 2008). The questions were split across three groups of respondents, with each respondent evaluating approximately a third of the brands for each category (with one common starting "anchor" brand for all). The splits were assigned at random. Second, for each of the brands, brand personality was assessed on the fifteen key items identified by Aaker (1997). She derived 5 factors, with a total of 15 facets, and used multiple (2-3) items for each facet. We used the first (highest loading) item she gave for each of her 15 facets (See Aaker 1997, p. 352). The items were selected to represent the five personality dimensions, as shown in Table 1.

[TABLE 1 ABOUT HERE]

Respondents responded to the personality items on a 9-point, ordinal scale with 1 described as "[brand] is not at all like this [personality item]" and 9 described as "[brand] is very much like this [personality item]." A screening question ensured that respondents were familiar with the categories and brands chosen for the study.

Analysis

Journal of Marketing Research, Article Postprint, Volume XLVI

Our approach involves a Bayesian factor analysis model in the spirit of factor models proposed by Ansari and Jedidi (2000), and Ansari, Jedidi, and Dube (2002), amongst others. It has the following features: a) It allows for different personality item loadings for each category. It does so to accommodate the fact that the items may have different meanings in the different categories. For example, the personality item "tough" may have different connotations for jeans and magazines. Category-specific loadings allow the personality factor "Ruggedness" to be differentially affected by the ratings for "tough" in different categories. b) Our conceptual view of the brand personality construct presumes that the brand-personality factor scores have a unique component due to the specific person providing the ratings, a component that is specific to the brand, and a component that is common to all brands within that category, i.e. a category specific component. Conceptually, this decomposition is related to the separation of overall brand associations and brand specific associations, proposed by Dillon, Madden and Mukherjee (2001). These components capture the notion that the personality perceptions derive from individual's views of both the brands and the category (Keller 1993; Batra and Homer 2004), and are crucial in our computations of brand "fit" and "atypicality" as explained below. Thus, the model captures both possible mechanisms – measurement, and category/brand schemata - for BP and CP structures. c) Our model is a confirmatory factor model, based on Aaker's brand personality measurement instrument. d) We accommodate the rank-order nature of the brand personality scales, and idiosyncratic usage of those scales by respondents, shown to be important by Rossi, Gilula, and Allenby (2001). e) We impute the data missing due to the split-questionnaire design of the brand personality survey at the same time as we estimate the model. Such imputation has been previously proposed by Ragunathan and Grizzle (1995), Lenk, Wedel, and Bockenholt (2006), and Adigüzel and Wedel (2008); here, we extend their approach to an ordinal data factor model. Thus all these features are integrated into a single model for the analysis of brand personality.

We use a confirmatory factor analysis approach (CFA) for Aaker's brand personality scales. For our data, brands are unique to categories, and "b|c" means that brand b is nested in category c. The P-dimensional personality factor scores $\beta_{i,b|c}$ for subject i and brand b nested within category c express respondent, brand and category influences on the latent personality dimensions. We represent the structure of the personality data through a random effects decomposition of the factor scores:

(1) $$\beta_{i,b|c} = \alpha_i + \alpha_{i,c} + \alpha_{i,b|c} \text{ for i =, 1…, n; b = 1,... B, and c = 1, …, C.}$$

The α_i are personality random effects that are unique to the person *i*, and common across the product categories and brands. The $\alpha_{i,c}$ are random effects that reflect personality traits common to all brands within category *c*, varying across categories: i.e. these are the category-personality scores. Lastly, $\alpha_{i,b|c}$ reflect personality traits that are specific to the brand *b* in product category *c* after adjusting for category specific personality traits: i.e. these are the ("pure") brand personality scores. The random effects in Equation (1) are mutually independent and normally distributed with mean of zero and the following standard ($P\times P$) covariance matrices:

(2) $$Var(\alpha_i) = \Lambda; \quad Var(\alpha_{i,c}) = \Lambda_c, \quad \text{and} \quad Var(\alpha_{i,b|c}) = \Lambda_{b|c}$$

The off-diagonal elements of all three matrices are zero, so that these personality random effects are uncorrelated. This assumption is standard in factor analysis, adopted here for parsimony. The random effects decomposition induces a correlation structure on the personality scores: 1) the covariance among brands in different categories is due to the common person effect; 2) the covariance among different brands in the same category is due to common person and category effects; and 3) the variance of a brand is due to person, category, and brand effects. Unlike traditional factor analysis, which assumes that the factor variances are one, we allow them to vary freely. (We set the variance of the first subject effect to a constant for identification.)

The brand personality factor scores $\{\beta_{i,b|c}\}$ determine the level for personality item j:

(3) $$U_{i,j,b|c} = \varphi_i + \mu_j + L'_{j,c}\beta_{i,b|c} + \varepsilon_{i,j,b|c} \text{ for j = 1, …, J}$$

where φ_i is a random effect that captures individual-level scale usage effects; μ_j is the grand mean for item j across all brands and subjects; $L_{j,c}$ is a vector of category-specific loadings, and $\varepsilon_{i,j,b|c}$ are normally distributed errors. In CFA each loading vector has only one free parameter, and the rest are zeros. If the loadings are not equal to one, φ_i and α_i are identified. The category specific loadings capture the possibly differential meaning of the items in every category, based on the theories of concept × scale interaction (Osgood et al 1957; Caprara, Barbaranelli, and Guido 2001) outlined above. Our formulation thus extends the original work of Aaker (1997), since there the personality structure in terms of the factor loadings is assumed invariant across categories.

The model can be summarized as:

(4) $$\begin{aligned} &U_{i,j,b|c} = \varphi_i + \mu_j + L'_{j,c}\left(\alpha_i + \alpha_{i,c} + \alpha_{i,b|c}\right) + \varepsilon_{i,j,b|c} \\ &\text{Var}(\varphi_i) = \tau^2 \text{ and } \text{Var}\left(\varepsilon_{i,j,b|c}\right) = \sigma^2_{j,b|c} \\ &\text{Var}(\alpha_i) = \text{diag}\left(\lambda^2_f\right); \text{Var}\left(\alpha_{i,c}\right) = \text{diag}\left(\lambda^2_{f,c}\right); \text{ and } \text{Var}\left(\alpha_{i,b|c}\right) = \text{diag}\left(\lambda^2_{f,b|c}\right) \end{aligned}$$

for subject i, personality item j, and brand b nested within category c. Further, respondent *i* selects scale point *k* for item *j* if the latent variable $U_{i,j,b|c}$ falls between two consecutive cut-points (c.f. Rossi, Gilula, and Allenby 2001) :

(5) $$Q_{i,j,b|c} = k \quad \textit{if and only if } \eta_i(k-1) < U_{i,j,b|c} \le \eta_i(k),$$

where there are *K* scale categories, $\eta_i(k)$ are cut-point parameters, and $Q_{i,j,b|c}$ is the observed data. To identify the model, we set the cut-point parameters $\eta_i(1)$ and $\eta_i(K-1)$ to –1 and 1,

Journal of Marketing Research, Article Postprint, Volume XLVI

respectively. There are K-3 unknown cut-points per respondent to be estimated from the data. Web-Appendix A summarizes the model, provides details about the cut-point model, distributional assumptions, identification constraints, the covariance structure of the model, and data imputation. Table A.1 details the induced covariances for $\{U_{i,j,b|c}\}$. Web-Appendix B provides a summary of the MCMC algorithm.

Fit and Atypicality Measures

At each iteration of the MCMC sampler, we compute brand locations and fit and atypicality measures. From these, we then compute means and standard deviations across iterations. Both the fit and atypicality measures are computed from each brand's location in brand perceptual space by using the factor scores $\beta_{i,b|c} = \alpha_i + \alpha_{i,c} + \alpha_{i,b|c}$. The computations only include factor scores for observed brands. To obtain the brand locations at each iteration of the MCMC, we first standardize $\{\beta_{i,b|c}\}$ within subject i across all brands in the study by subtracting off the means and dividing by the standard deviations:

(6)
$$\begin{aligned}
\psi_{i,b|c} &= \frac{\beta_{i,b|c} - \overline{\beta}_i}{s_{\beta,i}} \\
\overline{\beta}_i &= \frac{1}{n_i}\sum_{c=1}^{C}\sum_{b=1}^{B}\beta_{i,b|c}\,\chi(i\ evaluated\ b\,|\,c) \\
s_{\beta,i}^2 &= \frac{1}{n_i}\left[\sum_{c=1}^{C}\sum_{b=1}^{B}\beta_{i,b|c}^2\,\chi(i\ evaluated\ b\,|\,c) - n_i\overline{\beta}_i^{\,2}\right] \\
n_i &= \sum_{c=1}^{C}\sum_{b=1}^{B}\chi(i\ evaluated\ b\,|\,c)
\end{aligned}$$

where χ is the indicator function. Then, for brand b in category c, we averaged these scores to obtain the brand locations:

(7)
$$\xi_{b|c} = \frac{\sum_{i=1}^{n} \psi_{i,b|c}\, \chi(i \text{ evaluated } b \mid c)}{\sum_{i=1}^{n} \chi(i \text{ evaluated } b \mid c)}.$$

The brand personality space is five dimensional. The category personality ξ_c is the average of the brands in the category:

(8)
$$\xi_c = \frac{1}{B}\sum_{b=1}^{B} \xi_{b|c}$$

The *measure of fit* of brand b in category c to category c' is the root squared error (RSE) between $\xi_{b|c}$ and $\xi_{c'}$:

(9)
$$Fit(b \mid c, c') = \sqrt{\left(\xi_{b|c} - \xi_{c'}\right)'\left(\xi_{b|c} - \xi_{c'}\right)}$$

The smaller the RSE, the better the fit between brand b|c and category c'. The *atypicality measure* is the percent of variation in the brand personality that is not due to the category personality:

(10)
$$Atypicality(\,b \mid c\,) = \frac{\left(\xi_{b|c} - \xi_c\right)'\left(\xi_{b|c} - \xi_c\right)}{\xi_c'\xi_c + \left(\xi_{b|c} - \xi_c\right)'\left(\xi_{b|c} - \xi_c\right)}.$$

Atypicality ranges from zero to one. If the brand personality is completely inherited from the category personality, then atypicality is zero. Atypicality is one if the brand personality locations become infinitely far from the category locations.[1]

Model Comparisons

We estimated the confirmatory, hierarchical factor model on our data, which used Aaker's BP measurement instrument. In addition, to investigate the importance of the category-specific factor structure and the cut-point specification, we estimated two other models:

1. A model without the category specific factor structure, i.e. with one set of "pooled" loadings across the three categories. This model is comparable to the factor analysis of Aaker (1997), but goes beyond it to also accommodate heterogeneity across respondents, scale usage and missing data.
2. A model without the cut-point-specification that accommodates the ordered nature of the scales (the personality items were standardized within subjects to remove scale usage in a simpler fashion). This model has a confirmatory factor structure, a category-specific factor structure and accommodates heterogeneity across respondents and missing data.

The R-Square (correlation squared) between the observed ratings and the ones fitted by our model was 61.3%. Although it is not our purpose to directly compare our procedure with Aaker's, since our data are based on only a subset of her personality items, we believe that the model with the pooled-categories factor structure is a very strong benchmark to test our model against. The model with a category-specific factor structure has better fit (larger log marginal density) than the pooled-categories model: -53776 for the category specific structure versus -53941 for the common structure. Thus, there is evidence of a category influence on the structure of brand personality, which we will explore in more detail below. Finally, the correlation between the factor scores for the models with and without cut-points ranged from .92 to .98 (note that the LMDs are not comparable for these models because of a different data-distribution). This shows that the cut-point model results are robust and similar results would be obtained based on a simpler formulation. Yet, the cut-point model has been well established and is more appealing on theoretical grounds (Aitchison and Silvey 1957; McCullagh 1980; Gelfand, Smith, and Lee 1992; Rossi, Gilula, and Allenby 2001; and Johnson 2003).

Thus, we believe that the most critical components of our model are the confirmatory factor specification based on Aaker's (1997) scales, and the category-specific factor structure; the MCMC algorithm makes it feasible to bring these model specifications to the data. Other aspects, including the imputation of missing data due to the split-questionnaire design and the cut-point specification, are theoretically elegant and accommodate the data generating mechanism in the context of our application, but are perhaps less critical.

Brand and Category Personality Results

Table 1 displays the posterior means of the category-specific loadings for the brand personality items, as well as those from the analysis pooled across categories (see above), for comparison. The magnitudes of the loadings are different between categories, suggesting that many of these items may have different meanings in different categories, consistent with prior research (e.g., Austin, Sigauw, and Mattila 2003). For example, the loadings for the 'cheerful' item -- an indicator of the Sincerity brand personality dimension-- vary depending upon the product category. Its loading is largest for Cars, followed by Jeans, and lowest for Magazines. As another example, 'up-to-date' –- an indicator of the Excitement dimension in Aaker's model – loads highest on Jeans, followed by Cars and then Magazines. While these differences are not large, these results do suggest that the dimensions of brand personality may vary in their markers and measurement in different product categories, and that part of Aaker's "pooled" solution potentially confounds differences between categories. There are marked differences between the pooled analysis and those of the proposed approach. For example, the factor 'sophistication' is measured very differently by the category-specific loadings than by the pooled loadings: the loadings for upper-class are much lower for each of the categories than for the pooled solution, and the loadings for charming are higher for Jeans and Magazines than for the pooled solution, revealing that 'sophistication' has a very different meaning in each of the three categories. Further, tough loads much higher on 'ruggedness' for Jeans than for the pooled solution, while intelligent loads much higher on 'competence' for magazines than for the pooled solution, showing that the connotation of these variables is different for these categories. Given the previously cited literature on concept × scale interactions (Komorita and Bass 1967), these are expected, yet we believe important, findings.Table 2 shows the mean personality factor scores of the 30 brands on these five personality factors, where these factor scores reflect both the brand and the category personality determinants. We will not discuss the table in detail, but provide a few key examples to illustrate what, to us, seems to reflect the face validity of these results. Among cars, Porsche stands out for Exciting and Sophisticated, while it scores low on Sincere and Rugged. Chevy and VW are the most Sincere, and Chevy is Rugged while VW is not. It can be seen, in the jeans category, that on the Sincerity factor Levi's (and Lee) score relatively high, while the other brands, especially FUBU and Diesel, are low. Levi and Lee's are more Rugged than the cars, yet National Geographic scores even higher. Among magazines, Time, National Geographic, and Consumer Reports are the most Competent, while People, GQ, Cosmopolitan, and Rolling Stone are least Competent. Rolling Stone is judged to be Exciting, but not very Sophisticated. These personality perceptions are intuitively appealing and seem consistent with the widely held perceptions and communications of these brands.

[TABLE 2 ABOUT HERE]

The above results allow us to calculate one of the two key indices we seek, overall brand "atypicality." As described earlier, we define it as the percent of a brand's total personality score on each factor that is <u>not</u> attributable to the category average for that factor. That is, as in an Analysis of Variance, the variation in the factor score is decomposed into the category variation and the brand/person variation, and expressed as percentages. This score thus represents the "uniqueness" of each brand for each factor, i.e. that part of its overall personality that is "idiosyncratic" to the brand and not attributable to the category where it happens to be in presently. Our "atypicality" index allows us to see what percentage of a brand's score on each factor is "idiosyncratic" and not attributable to the category average (i.e., sum of squares due to brand, divided by sum of squares due to category plus brand). These statistics are presented in the second column of Table 3. We see that the Porsche, Mercedes and Jaguar brands are seen as the most atypical in the Cars category and Volvo the least; Levi and FUBU the most atypical (idiosyncratic) jeans brands, Gap the least; and National Geographic the most atypical magazine brand, Time the least. In other words, given our sample of respondents and brands, relatively speaking, Volvo most strongly evokes car-category imagery, Gap most strongly evokes jeans category imagery, and Time most strongly evokes magazine category imagery. As discussed earlier, brands (such as Heineken or Clorox) where the brand perception is too closely linked to a category perception tend perhaps not to possess broad "extendibility" beyond that category (although they could well benefit from their strong association with it in fewer, narrower, well-targeted extensions, via parent-category "leverage"). Thus the high parent-category influence on their brand personalities may make broad expansion relatively more difficult for the Volvo, Gap and Time brands. In contrast, the Porsche, Levi's and National Geographic brands are relatively more atypical of their categories, thus more abstract in meaning, thus potentially more broadly extendible into other categories. It testifies to the validity of these results that among these brands, the Porsche brand has in fact extended into many other non-automotive categories, including eyewear, briefcases, watches, even computer hard drives, all (we presume) on the basis of its reputation for style and design, not merely sports-car expertise (see: www.porsche-design.com). Similarly, National Geographic too has used its travel and explorer-type associations (very different from just 'magazine' imagery) to extend into travel clothing, travel shoes, binoculars, explorer boats and compasses, road atlases and globes, and many more (see: www.nationalgeographic.com/shop). (It is possible that these existing extensions may have influenced the perceived extendability of these two brands, but this is of no consequence in our procedure, since such extendability was not explicitly assessed by us.)

[TABLE 3 ABOUT HERE]

A complementary perspective on a brand's extendibility is that provided by our "fit" statistics, also provided in Table 3. These measures of "fit" of each brand with each of the three categories are computed as the root square error (RSE or distances) of a brand's own personality scores relative to those of the category in question. Thus, in reading these "fit" statistics, a lower number indicates a smaller brand to category distance and thus a closer "fit." We illustrate some of the insights that can potentially be gleaned from these statistics. We see, for example, that if a jeans brand were to extend into the car category (via a co-branding deal perhaps, as has been done by Eddie Bauer, Harley Davidson, L. L. Bean, Coach, etc.), Calvin Klein, Polo and Tommy Hilfiger would be viable brands (lowest distances of .804, .864 and .915). Among car-entering magazine entrants, GQ has the smallest distance (1.027). For an outside-category brand going into the jeans category, the (automotive) Pontiac brand seems to be the least distant (.265), as does the GQ magazine brand (.737) among magazines. A new magazine brand, if coming from cars or jeans, would be a good fit if it came from Honda, VW, Saturn (.490, .573, .583) or Gap (.740). Post-hoc reflection suggests that these particular clothing and automotive brands come closest to being "lifestyle" brands, representing particular attitudes towards life: GQ representing stylishness, Pontiac excitement, Saturn and Honda representing sobriety and moderation, VW a more youthful and engaged "drivers wanted" attitude to life, and Gap an attitude of classic casualness.

It can be seen that a few brands (Honda; Saturn; VW; Pontiac; Levi's; National Geographic) actually have lower distances in this five-dimensional personality space with categories other than their own. This means that in our data (given our sample of brands for each category) their location in this space is closer to the centroid of another category than to their own. There are two explanations for this result. First, from a substantive point of view, it shows (for example) that since magazines as a category are (directionally) less exciting, sophisticated and rugged than cars (see Table 2), but more competent, Honda's personality profile better matches magazines than cars: it is perceived to be less exciting and sophisticated than the average car, but more competent. Recall that we analyze ratings in a personality space, not in a functional attribute space, so this does not mean that Honda is seen as making better magazines than cars, only that its personality is closer to an average magazine-personality rather than to an average car-personality. The second explanation is statistical: perfect discrimination requires complete separation of the brands in personality space. In actuality, perfect discrimination does not occur, and there is overlap among brands in the personality space in our application, even though category personalities are different. Furthermore, an extending brand will need to only look at its distances from categories other than its own, so these own-category distances are not the ones of primary interest.

In a further analysis we computed the distance (RSE, defined similarly as the one above) between every pair of these 30 brands, to allow strategic ideation on partners for licensing or co-branding deals. Low distances (high co-branding potential) emerged for these cross-category pairs of brands: Porsche and Jaguar cars with A&F jeans and Cosmopolitan and People magazines; Mercedes and Lexus cars with Polo or CK jeans; Chevy or Pontiac cars with Lee or Levi jeans; Volvo cars with Consumer Reports and Parenting magazines; Rolling Stone magazine with FUBU jeans; GQ magazine with Diesel jeans; among many others. Figure 1 shows these distances graphically in a clustering dendrogram obtained with a hierarchical average linkage algorithm. While any explanation of these is necessarily post-hoc and subjective, the groupings in the tree do appear to us to be very reasonable, and revealing of the strategic, idea-generating power of the empirical and analytical approach we have presented here. We test these results in two follow-up validation experiments, the results of which are to be presented next.

[INSERT FIGURE 1 HERE]

VALIDATION TESTS

In the follow-up study we asked subjects to indicate, for each of the thirty brands in the previous study, which of the other brands in one of the other categories might make "a good partner" in their judgment, in the same way that Ford Explorer had partnered with Eddie Bauer, and Disney with Delta Airlines. Subjects were (as in the main study above) nationally representative adults familiar with the brands used in the brand extension ideas they were asked to rate. The sample was different than the sample in the main study, however. Each respondent filled in three tables, the first for our ten jeans brands (rows) potentially partnering with our car brands (columns), the second for our ten magazine brands (rows) partnering with our ten jeans brands (columns), and the third for our ten magazine brands partnering with our ten car brands (columns). In each table, the respondents were asked to think about each brand listed in the ten rows, and to put as few or as many check marks for the ten brands listed in the columns, with each check mark indicating "a good partner" in their judgment. We deliberately used the word "partner," since it left ambiguous the mode of partnership, potentially encompassing both licensing as well as "co-branding" deals.

Our hypothesis was that the data from this first validation study would show subjects indicating higher brand "partnership" potential in those pairs where our previous (main) study had estimated higher "fit" for the across-category pair of brands being rated, and for those brands estimated in the main study as having higher "atypicality" (regardless of the brand it was being paired with). We analyzed the results of the frequency tables with log-linear models, for the three combinations, Magazines-Jeans, Magazines-Cars and Jeans-Cars respectively, explaining the frequencies of partnering from the main effects of atypicality and fit. We controlled for the effects of prior attitudes to the brands in both categories by using average ratings of the attitudes towards these brands as covariates in the log-linear model. Note that the measures of atypicality and fit were derived from our proposed model and the first data set, then applied here to data on brand partnering evaluations in an independent sample of individuals. We used atypicality in the model, so that higher values of atypicality, and lower values of fit (distance), indicate brands that should be more abstract and brand pairs that have better fit.

The analyses reveal the following. Atypicality has no significant main effects in two out of the three category pairs, but it does have a significant effect for magazines-jeans (coefficient -.662, SE = .252, $p<.05$). The fit measure is highly predictive of partnering appeal in each of the three categories: increasing the fit (by decreasing the RSE) by .1 increases the odds of partnering by a factor of 1.47 for jeans-cars (b = .385, SE = .106, $p<.01$), 1.35 for magazines-jeans (b = .300, SE = .101, $p < .01$) and 1.44 for magazines-cars (b = .362, SE = .093, $p < .01$). These are substantial effects, which underline the impressive predictive performance of our model, especially given the fact that it is obtained on an independent sample of subjects and that entirely different measures were collected in these two samples (i.e. brand personality ratings in the first, pick-any co-branding data in the second). Thus, we conclude that the predictive performance of atypicality is less strong than that of fit, but that of fit is truly impressive.

We also validated the fit and atypicality measures from the main study by presenting hypothetical brand extensions to the second sample of subjects and obtaining ratings for them. See Table 4 for the rated items, which used 9-point semantic differential items (e.g., illogical/logical, uninteresting/interesting) similar to those previously used in the literature on brand extension acceptance, discussed in Keller (2003). The candidate brand extensions were purposefully picked to include combinations using brands that rated low and high on the fit and atypicality measures that were derived from the main study. A factor analysis of the ratings data (see Table 4) indicated that two components had eigenvalues greater than one and accounted for 82% of the total variance. The first component measures the 'Appropriateness' of the brand extension, with items such as logical, makes sense, appropriate, and fit (plus attitudinal items such as positive) having loadings between .8 to .9, and the second measures its 'Novelty,' with high loadings on original, new and special.

[INSERT TABLE 4 HERE]

We then constructed validation-study factor scores for each individual and brand extension, and regressed them onto the atypicality and fit measures from the main study. Because we have repeated measures (15 brand extensions) for each subject in this second validation study, we included a random effect for subject. We used control variables for prior brand attitudes and knowledge, of the extending brand and the new category (averaged over its brands), collected in this validation study. Both the atypicality and fit measures from our main study have a significant relationship with the first ratings factor, Appropriateness, from this validation study. (Similar results hold without the control variables.) In these regression results, brand extensions judged more "Appropriate" tend to possess imagery that fits well with the category personality of the *targeted* category (our fit measure, negative sign implying smaller distance), and have imagery that is *less* atypical (more typical) of the *parent* category (the negative sign for atypicality). Since standard attitudinal items (positive, favorable, good) also load highly on this Appropriate ratings factor, these results suggest that consumers rate a brand extension more positively if the imagery of the extending brand (1) fits that of the extended-into category and (2) is highly typical of the parent category. This latter result does not support the notion that category-atypical brands ought to be more successful extenders; instead it suggests that extensions are more favored if the extending brand seems to be operating within its parent-category competency, suggesting that such extensions may have more "leverage," as presented in our theoretical discussion earlier of this atypicality construct. The former result again provides evidence of the external validity and value of the proposed fit measure from our main study.

For the second ratings factor, Novelty, the effect of the atypicality measure from our main study was not significant, but that of fit was again significant, in a positive direction. Since in our fit statistics a lower number indicates a smaller brand to category distance and thus a closer "fit," the positive sign means that for a brand extension to be viewed as "Novel", it has to be viewed as different from the category it is entering. This makes intuitive sense. Thus our results suggest an interesting challenge for marketing managers in that an appropriate (high-fit) extension may lack novelty, which could lead to an extension without a distinctive personality in the target category. On the other hand, if the fit (appropriateness) to the target category is low, one risks having an extension that may be novel/distinctive, but is viewed as inappropriate. It is interesting that our attitudinal items (favorable, positive, good) load much more on Appropriateness than on Novelty, suggesting that (for our sample) Appropriateness is more important than Novelty in shaping overall attitudes to the brand extension.

DISCUSSION AND IMPLICATIONS

We have developed and estimated a random effects, hierarchical factor model, using consumer perception data on brand personality ratings, that (a) separates "category personality" from the brand's own "unique personality;" (b) computes the contribution of the latter to its total brand personality imagery, and thus the degree to which it is "atypical" of its original product category; and (c) quantifies the extent to which its unique brand personality imagery "fits" the personality imagery of several "candidate" product categories for purposes of extension, licensing, or co-branding. Our model accommodates separate brand and category personality effects, using a confirmatory model based on Aaker's brand personality measurement instrument, and allows for the ordered nature of the rating scales and missing data due to the split nature of the questionnaire.

The success of our validation studies in confirming the estimates of our main study, more for 'fit' than 'atypicality' (first validation test) and showing their impact on perceived extension appropriateness (second validation test), supports our model and the metrics derived from it, especially those of "fit." While our results for our measure of the extending brand's atypicality from its parent category do not provide strong support for the theoretical notion that atypical brands possess more abstract, thus more extendible, imagery, it may well be that these results are limited by our focus on personality attributes, or our small sample of categories, or the fact that we only modeled linear (but not quadratic) effects of atypicality. Such a relationship may yet exist on more functional attribute dimensions or in other categories; indeed, finding support for it may even depend on the specific categories being considered. Alternatively, there may be imperfections in our conceptualization or measurement of atypicality, which future research should examine further. Despite the need for more work on atypicality, our model capabilities should allow marketers greater ability than before to strategically generate ideas about which brands could potentially extend into which categories, or "partner" with which other brands in those categories (see Figure 1), based only on collected brand imagery data, rather than having to obtain explicit consumer reactions in advance to every possible concept or combination. Given the high failure rates in existing brand extensions, the burgeoning business in brand licensing deals, and the acknowledged limitations of existing academic methodologies to identify successful brand extension possibilities (Klink and Smith 2001), these would appear to be important contributions.

While the "attribute domain" of our illustration concerned only "brand personality" imagery and associations, it should be noted that the method presented here is more general, and is potentially extendible into the domain of more physical attribute and benefit associations as well. Our model may even allow cross-cultural market researchers to tease out the influence of culture (Aaker, Benet-Martinez, and Gariolera 2001) and culture-specific response scale usage (Ter Hofstede, Steenkamp, and Wedel 1999) on consumer perceptions of globally-marketed brands (Steenkamp, Batra, and Alden 2003). In the brand personality domain, our results demonstrate that the meaning of brand personality descriptors can be influenced by the category context. For example, a "reliable" magazine may have a different connotation than a "reliable" car. A criticism of existing inventories of brand personality descriptors and dimensions (such as that developed by Aaker 1997) is that they are often used, and interpreted, uniformly across very different products and categories. Our approach gives us a way to assess which particular brand personality traits may differ in meaning across categories. Further, while differences in "category personalities" had previously been qualitatively discussed by researchers such as Levy (1959), our method enables a level of empirical quantification heretofore missing from that literature.

An additional benefit of the proposed methodology is that it accounts for two important features of many brand personality datasets. The first is the rank order nature of the personality scales and idiosyncratic scale usage (Rossi, Gilula, and Allenby 2001). This disentangles respondents' behavior with respect to the measurement scales and their underlying brand personality perceptions. The second is the absence of entire blocks of data due the use of split questionnaires (Ragunathan and Grizzle 1995; Adigüzel and Wedel 2008), which seems common or even desirable in large questionnaires on brand personality to reduce respondent fatigue and boredom. Without (proper) data imputation, it would be close to impossible to analyze such a highly fractionated dataset. The method developed here does all of that, simultaneously.

It might be argued that by allowing the measurement items to have category-specific factor loadings, factor scores across brands in different categories are non-comparable. We acknowledge this concern, but believe our comparisons are still useful. If one specified a model forcing identical loadings for all categories (not the varying ones seen in Table 1), the estimated brand personality factor scores would be confounded with (biased by) the different connotations of the items in different categories, muddying the across-category comparisons of the brands on the latent, abstract, underlying dimensions. When consumers make a judgment about the appropriateness of a brand in one category extending into another, they clearly have to make comparison on abstract dimensions (such as ‘Sophistication’) that span categories (Johnson 1984). These abstract dimensions are measured in our model by our latent factors. By specifying category specific loadings, we allow the observed item responses to impact the latent brand factor scores in category-specific ways, thus enabling (we believe) more accurate comparisons of the brand associations across categories on these latent factors (abstract constructs). Note here that in our CFA model the latent factor scores determine the item ratings through the loadings, not conversely. That is, the confirmatory factor model is a "reflective model." In such a reflective model, the subjects' brand latent factor scores are the primitive construct. These scores are then expressed in the item ratings, which are the manifestation of the latent construct. We are allowing for the possibility that the manifest rating items are differentially impacted by the latent factor scores across categories; thus the same latent factors are being estimated across categories, while their manifest, observed items are being computed with differential loadings across categories. This allows the latent factor scores to be compared across categories (as consumers will seek to do), even if their factor loadings are different across categories.

Several limitations remain, beyond those concerning atypicality discussed above, suggesting other avenues for future research. The data used in our application came from just three product categories and thirty brands. Any category personality estimates obtained will obviously depend on the specific sample of brands and respondents included, and will reflect the category personality better (as is the case with all samples) if the sample is a random sample of all brands in that category, and as that sample gets larger.

References

Aaker, David A. and Kevin L. Keller (1990), "Consumer Evaluations of Brand Extensions," *Journal of Marketing*, 54 (January), 27–41.

Aaker, Jennifer L. (1997), "Dimensions of Brand Personality," *Journal of Marketing Research*, 34 (August), 347–56.

———, Veronica Benet-Martinez, and Jordi Gariolera (2001), "Consumption Symbols as Carriers of Culture: A Study of Japanese and Spanish Brand Personality Constructs," *Journal of Personality and Social Psychology*, 81 (September), 492–508.

Adigüzel, Feray and Michel Wedel (2008), "Split Questionnaire Design." *Journal of Marketing Research*, 25 (5), 608–617.

Aitchison J. and S.D. (1957), "The Generalization of Probit Analysis to the Case of multiple Responses," *Biometrika*, 44 (June), 131-150.

Ansari, Asim and Kamel Jedidi (2000), "Bayesian Factor Analysis for Multilevel Binary Observations," *Psychometrika*, 65 (December), 475–96.

———, Kamel Jedidi, and Laurette Dube (2002), "Heterogeneous Factor Analysis Models: A Bayesian Approach," *Psychometrika*, 67 (March), 49–78.

Austin, Jon R., Judy A. Siguaw, and Anna S. Mattila (2003), "A Re-Examination of the Generalizability of the Aaker Brand Personality Measurement Framework," *Journal of Strategic Marketing*, 11 (June), 77–92.

Azoulay, Audrey and Jean-Noel Kapferer (2003), "Do Brand Personality Scales Really Measure Brand Personality?" *Journal of Brand Management*, 11 (November), 143–55.

Batra, Rajeev and Pamela Miles Homer (2004), "The Situational Impact of Brand Image Beliefs," *Journal of Consumer Psychology*, 14 (3), 318–30.

———, Donald R. Lehmann, and Dipinder Singh (1993), "The Brand Personality Component of Brand Goodwill: Some Antecedents and Consequences," in *Brand Equity and Advertising*, David A. Aaker and Alexander Biel, eds., Hillsdale, NJ: Lawrence Erlbaum Associates.

Brand Strategy. London: Dec. 2004/Jan.2005, p.44

Broniarczyk, Susan and Joseph W. Alba (1994), "The Importance of the Brand in Brand Extension," *Journal of Marketing Research*, 31 (May), 214–28.

Caprara, Gian Vittorio, Claudio Barbaranelli, and Gianluigi Guido (2001), "Brand Personality: How to Make the Metaphor Fit?" *Journal of Economic Psychology*, 22 (June), 377–95.

Digman, John M. (1990), "Personality Structure: Emergence of the Five-Factor Model," *Annual*

Review of Psychology, 41 (1), 417–40.

Dillon, William R., Thomas J. Madden, and Soumen Mukherjee (2001). “Understanding What’s in a Brand Rating: A Model for Assessing Brand and Attibute Effects and Their Relationship to Brand Equity,” *Journal of Marketing Research*, 38 (November), 415–29.

Durgee, Jeffrey F. and Robert W. Stuart (1987), “Advertising Symbols and Brand Names that Best Represent Key Product Meanings,” *Journal of Consumer Marketing*, 4 (Summer), 15–24.

Farquahar, Peter, Julia Y. Han, Paul M. Herr, and Yuji Ijiri (1992), “Strategies for Leveraging Master Brands,” *Marketing Research*, September, 32–43.

Gelfand, Alan E. and Adrian F.M. Smith (1990), “Sampling Based Approaches to Calculating Marginal Densities,” *Journal of the American Statistical Association*, 85, 398–409.

———, Smith, A.F.M., and Lee, T.M. (1992), “Bayesian Analysis of cConstrained Parameter and Truncated Data Problems Using Gibbs Sampling,” *Journal of the American Statistical Association*, 87 (418) June, 523–32.

Johnson, Michael D. (1984), “Consumer Choice Strategies for Comparing Noncomparable Alternatives,” *Journal of Consumer Research*, 11 (December), 741–53.

Johnson, Timothy R. (2003), “On the Use of Heterogeneous Thresholds Ordinal Regression

Models to Account for Individual Differences in Response Style,” *Psychometrika*, 68 (4), 563–83.

Keller, Kevin L. (1993), “Conceptualizing, Measuring, and Managing Customer-Based Brand Equity,” *Journal of Marketing*, 57 (January), 1–22.

——— (2003), *Strategic Brand Management*, 2d Ed. Upper Saddle River, N.J. Prentice-Hall.

Klink, Richard R. and Daniel C. Smith (2001), “Threats to the External Validity of Brand Extension Research,” *Journal of Marketing Research*, 38 (August), 326–35.

Komorita, S.S. and Alan R. Bass (1967), “Attitude Differentiation and Evaluative Scales of the Semantic Differential,” *Journal of Personality and Social Psychology*, 6, 241–44.

Lautman, Martin R. (1991), “End-Benefit Segmentation and Prototypical Bonding,” *Journal of Advertising Research*, 31 (June/July), 9–18.

Lenk, Peter, Michel Wedel, and Ulf Bockenholt (2006), “Bayesian Estimation of Circumplex Models subject to Prior Theory Constraints and Scale-Usage Bias," *Psychometrika*, 71 (1), 33–55.

Levy, Sidney J. (1959), “Symbols for Sale,” *Harvard Business Review*, 37 (July/August), 117–24.

——— (1981), "Interpreting Consumer Mythology: A Structural Approach to Consumer Behavior," *Journal of Marketing*, 45, 49–61.

——— (1986), "Meanings in Advertising Stimuli," in *Advertising and Consumer Psychology*, Jerry Olson and Keith Sentis, eds. Westport, CT: Praeger, 214–26.

Martineau, Pierre (1958), "The Personality of the Retail Store," *Harvard Business Review*, (January/February), 47–55.

Martinez, Eva and Jose M. Pina (2003), "The Negative Impact of Brand Extensions on Parent Brand Image," *Journal of Product and Brand Management*, 12 (7), 432–48.

McCrae, Robert R. and Paul T. Costa (1987), "Validation of the Five-Factor Model of Personality Across Instruments and Observers," *Journal of Personality and Social Psychology*, 52 (January), 81–90.

McCullagh, Peter (1980), "Regression Models for Ordinal Data (with Discussion)," *Journal of the Royal Statistical Society, Series B*, 42, 109–142.

Osgood, Charles E., George J. Suci, and Percy H. Tannenbaum (1957), *The Measurement of Meaning*. Urbana: University of Illinois Press.

Park, C. Whan, Sandra Milburg, and Robert Lawson (1991), "Evaluation of Brand Extensions: The Role of Product Level Similarity and Brand Concept Consistency," *Journal of Consumer Research*, 18 (September), 185–93.

Plummer, Joseph T. (1984/85), "How Personality Makes a Difference," *Journal of Advertising Research*, 24 (December 1984/January 1985), 27–31.

Raghunathan, Trivellore. E. and James E. Grizzle (1995), "A Split Questionnaire Survey Design," *Journal of American Statistical Association*, 90 (March), 54–63.

Ratchford, Brian (1987), "New Insights About The FCB Grid," *Journal of Advertising Research*, 27 (August/September), 24–26.

Roedder-John, Deborah and Barbara Loken (1993), "Diluting Brand Equity: The Impact of Brand Extensions," *Journal of Marketing*, 57 (July), 71–84.

Rosch, Eleanor, Carolyn B. Mervis, Wayne D. Gray, David M. Johnson, and Penny Boyes-Braem (1976), "Basic Objects in Natural Categories," *Cognitive Psychology*, 8, 382–439.

Romaniuk, J. and Andrew Ehrenberg (2003), "Do Brands Lack Personality?" Marketing Science Centre Research Report No. 14, May, University of South Australia.

Rossi, Peter. E., Zvi Gilula, and Greg M. Allenby (2001), "Overcoming Scale Usage

Heterogeneity: A Bayesian Hierarchical Approach," *Journal of the American Statistical Association*, 96 (March), 20–31.

Song, Xin-Yuan and Sik-Yum Lee (2001), "Bayesian Estimation and Test for Factor Analysis Model with Continuous and Polytomous Data in Several Populations," *British Journal of Mathematical and Statistical Psychology*, 54, 237–63.

——— (2002), "Bayesian Estimation and Model Selection of Multivariate Linear Model with Polytomous Variables," *Multivariate Behavioral Research*, 37 (4), 453–77.

Steenkamp, Jan-Benedict E.M., Rajeev Batra, and Dana L. Alden (2003), "How Perceived Brand Globalness Creates Brand Value," *Journal of International Business Studies*, 34, 53–65.

Tauber, Edward M. (1988), "Brand Leverage: Strategy for Growth in a Cost Controlled World," *Journal of Advertising Research*, (August/September), 26–30.

Ter Hofstede, Frenkel, Jan-Benedict E.M. Steenkamp, and Michel Wedel (1999), "International Market Segmentation Based on Consumer-Product Relations, *Journal of Marketing Research*, 36 (February), 1–17.

FOOTNOTE

1. We also tested two other measures of Atypicality, brand to category RSE (distances) and the cosine between brand and category vectors with broadly similar results.

Journal of Marketing Research, Article Postprint, Volume XLVI

TABLE 1

VARIATIONS IN CFA FACTOR LOADINGS BY CATEGORY FOR FIVE-FACTOR PERSONALITY MODEL

		Posterior Means				Posterior Standard Devi		
Dimension	Item	Cars	Jeans	Mags.	Pooled	Cars	Jeans	Mags.
Sincerity	Down-to-Earth	1.925	2.068	1.837	1.907	0.236	0.221	0.225
	Honest	1.050	1.511	1.318	1.290	0.143	0.177	0.172
	Wholesome	1.760	1.901	1.960	1.777	0.214	0.208	0.230
	Cheerful	0.992	0.924	0.754	0.821	0.148	0.118	0.118
Exciting	Daring	1.544	1.662	1.637	1.570	0.185	0.189	0.184
	Spirited	1.393	1.531	1.661	1.449	0.171	0.177	0.184
	Imaginative	1.491	1.854	1.869	1.679	0.177	0.208	0.209
	Up-to-Date	1.182	1.675	1.022	1.323	0.159	0.199	0.144
Competence	Reliable	1.603	1.318	1.887	1.642	0.190	0.166	0.212
	Intelligent	1.521	1.179	2.270	1.666	0.185	0.176	0.268
	Successful	2.208	1.421	1.930	1.879	0.246	0.182	0.232
Sophistication	Upper-Class	1.522	1.011	0.258	1.930	0.241	0.235	0.173
	Charming	0.395	0.969	1.049	0.421	0.087	0.183	0.240
Ruggedness	Outdoorsy	2.582	2.596	2.374	2.239	0.291	0.262	0.281
	Tough	1.142	1.763	0.865	1.276	0.152	0.184	0.120

TABLE 2
ESTIMATED BRAND-SPECIFIC PERSONALITY FACTOR-SCORES

	Posterior Mean					Posterior Standard Deviations				
Brand	Sincere	Excitg.	Comp.	Soph.	Rugged	Sincere	Excitg.	Comp.	Soph.	Rugged
Car average	-0.024	0.169	0.140	0.552	-0.028	0.019	0.021	0.026	0.040	0.023
Honda	0.262	-0.015	0.187	-0.151	0.045	0.049	0.043	0.044	0.049	0.036
Porsche	-0.534	1.036	0.300	1.651	-0.171	0.073	0.091	0.072	0.175	0.055
Chevy	0.475	0.005	0.082	-0.076	0.651	0.074	0.070	0.066	0.081	0.081
Lexus	-0.229	0.206	0.305	1.226	-0.282	0.060	0.066	0.064	0.161	0.045
Saturn	0.187	-0.172	-0.143	-0.126	-0.006	0.067	0.075	0.066	0.084	0.049
VW	0.421	0.047	0.040	-0.123	0.005	0.075	0.071	0.066	0.085	0.049
Pontiac	-0.096	-0.327	-0.416	-0.246	0.117	0.058	0.083	0.078	0.088	0.059
Mercedes	-0.241	0.393	0.682	1.575	-0.366	0.068	0.074	0.097	0.171	0.062
Volvo	0.109	-0.162	0.192	0.417	0.058	0.063	0.067	0.063	0.153	0.059
Jaguar	-0.597	0.674	0.169	1.375	-0.329	0.080	0.082	0.067	0.157	0.063
Jeans avg.	-0.247	-0.294	-0.520	-0.174	0.146	0.018	0.026	0.029	0.034	0.017
Levi	0.916	-0.029	0.418	-0.115	1.491	0.087	0.035	0.099	0.046	0.092
Lee	0.325	-0.659	-0.440	-0.358	0.951	0.070	0.076	0.088	0.083	0.077
Guess	-0.508	-0.373	-0.664	-0.123	-0.184	0.065	0.062	0.089	0.088	0.054
Fubu	-0.841	-0.654	-1.342	-0.748	-0.485	0.073	0.072	0.097	0.133	0.052
Polo	-0.279	-0.182	-0.401	0.058	-0.047	0.054	0.060	0.077	0.086	0.045
Tommy Hilf.	-0.515	-0.099	-0.360	0.052	-0.053	0.065	0.061	0.074	0.093	0.049
Calvin Klein	-0.383	-0.131	-0.235	0.057	-0.164	0.057	0.057	0.071	0.097	0.052
Gap	-0.028	-0.130	-0.403	-0.206	0.063	0.057	0.060	0.072	0.071	0.054
Diesel	-0.623	-0.596	-1.186	-0.277	0.011	0.070	0.074	0.094	0.108	0.053
A & F	-0.535	-0.088	-0.588	-0.074	-0.129	0.067	0.059	0.091	0.078	0.049
Mags. Avg.	0.031	0.110	0.149	-0.267	-0.334	0.020	0.020	0.024	0.035	0.017
Time	0.044	0.126	0.550	-0.287	-0.439	0.037	0.035	0.066	0.047	0.031
Readers Dig.	0.760	-0.159	0.222	-0.212	-0.245	0.092	0.067	0.064	0.078	0.060
Cosmopol.	-0.539	0.338	-0.152	-0.168	-0.861	0.077	0.066	0.069	0.093	0.080
Nat-Geo.	0.809	0.825	0.583	-0.193	1.602	0.093	0.083	0.080	0.078	0.129
Money	-0.341	-0.471	0.154	-0.434	-0.772	0.057	0.080	0.062	0.119	0.067
People	-0.462	0.089	-0.270	-0.307	-0.763	0.064	0.065	0.069	0.086	0.072
Parent	0.793	-0.028	0.116	-0.121	-0.638	0.103	0.064	0.064	0.107	0.058
GQ	-0.566	0.007	-0.173	-0.181	-0.305	0.075	0.064	0.067	0.089	0.056
Rolling Stone	-0.411	0.687	-0.090	-0.339	-0.482	0.069	0.077	0.061	0.099	0.056
Cnsm. Report	0.223	-0.315	0.549	-0.426	-0.433	0.064	0.073	0.083	0.084	0.052

TABLE 3

ESTIMATES OF BRAND-SPECIFIC ATYPICALITY FROM OWN CATEGORY, AND FIT WITH POTENTIAL EXTENSION CATEGORIES*

	Posterior Means				Posterior Standard Deviations			
		Fit				Fit		
	Atypicality	Cars	Jeans	Mags.	Atypicality	Cars	Jeans	Mags.
Honda	0.636	0.791	0.928	0.490	0.030	0.064	0.057	0.049
Porsche	0.864	1.515	2.447	2.221	0.022	0.131	0.151	0.167
Chevy	0.763	1.075	1.125	1.115	0.029	0.078	0.082	0.084
Lexus	0.622	0.779	1.761	1.533	0.078	0.129	0.140	0.168
Saturn	0.665	0.846	0.629	0.583	0.044	0.083	0.071	0.070
VW	0.658	0.834	0.961	0.573	0.046	0.085	0.076	0.073
Pontiac	0.775	1.112	0.265	0.869	0.030	0.086	0.070	0.077
Mercedes	0.812	1.256	2.296	1.964	0.034	0.135	0.154	0.172
Volvo	0.340	0.429	1.021	0.853	0.082	0.076	0.112	0.135
Jaguar	0.792	1.174	2.048	1.854	0.034	0.115	0.139	0.153
Levi	0.897	1.944	2.034	2.063	0.011	0.111	0.102	0.105
Lee	0.711	1.721	1.082	1.648	0.031	0.091	0.073	0.082
Guess	0.322	1.295	0.475	1.119	0.057	0.090	0.060	0.084
Fubu	0.801	2.339	1.383	1.964	0.023	0.109	0.089	0.098
Polo	0.221	0.864	0.365	0.833	0.059	0.087	0.063	0.079
Tommy Hilfiger	0.338	0.915	0.492	0.898	0.060	0.088	0.066	0.077
Calvin-Klein	0.380	0.804	0.540	0.731	0.058	0.088	0.065	0.078
Gap	0.193	0.994	0.334	0.740	0.052	0.086	0.057	0.071
Diesel	0.604	1.850	0.852	1.691	0.046	0.100	0.080	0.091
A&F	0.330	1.132	0.483	1.001	0.055	0.092	0.059	0.087
Time	0.445	1.029	1.331	0.423	0.076	0.069	0.063	0.061
Readers Digest	0.739	1.176	1.328	0.798	0.045	0.092	0.083	0.081
Cosmopolitan	0.777	1.273	1.290	0.882	0.036	0.095	0.081	0.078
National Geo.	0.958	2.138	2.396	2.254	0.006	0.112	0.110	0.118
Money	0.763	1.436	1.199	0.848	0.038	0.107	0.078	0.074
People	0.736	1.292	1.062	0.790	0.040	0.088	0.075	0.071
Parent	0.766	1.250	1.486	0.858	0.043	0.099	0.086	0.091
GQ	0.689	1.027	0.737	0.705	0.052	0.090	0.070	0.077
Rolling Stone	0.739	1.222	1.274	0.795	0.042	0.097	0.076	0.074

Consumer Reports	0.658	1.268	1.337	0.656	0.055	0.089	0.082	0.072

*A higher atypicality index means the brand possesses more brand-specific personality imagery, not attributable to its parent category. A lower fit index between a brand and a new category indicates a smaller brand-to-category distance and thus higher fit.

Journal of Marketing Research, Article Postprint, Volume XLVI

TABLE 4
VALIDATION TEST #2: REGRESSION RESULTS CONFIRM THAT FIT-ESTIMATES PREDICT JUDGMENTS OF APPROPRIATENESS AND NOVELTY OF EXTENSION [a]

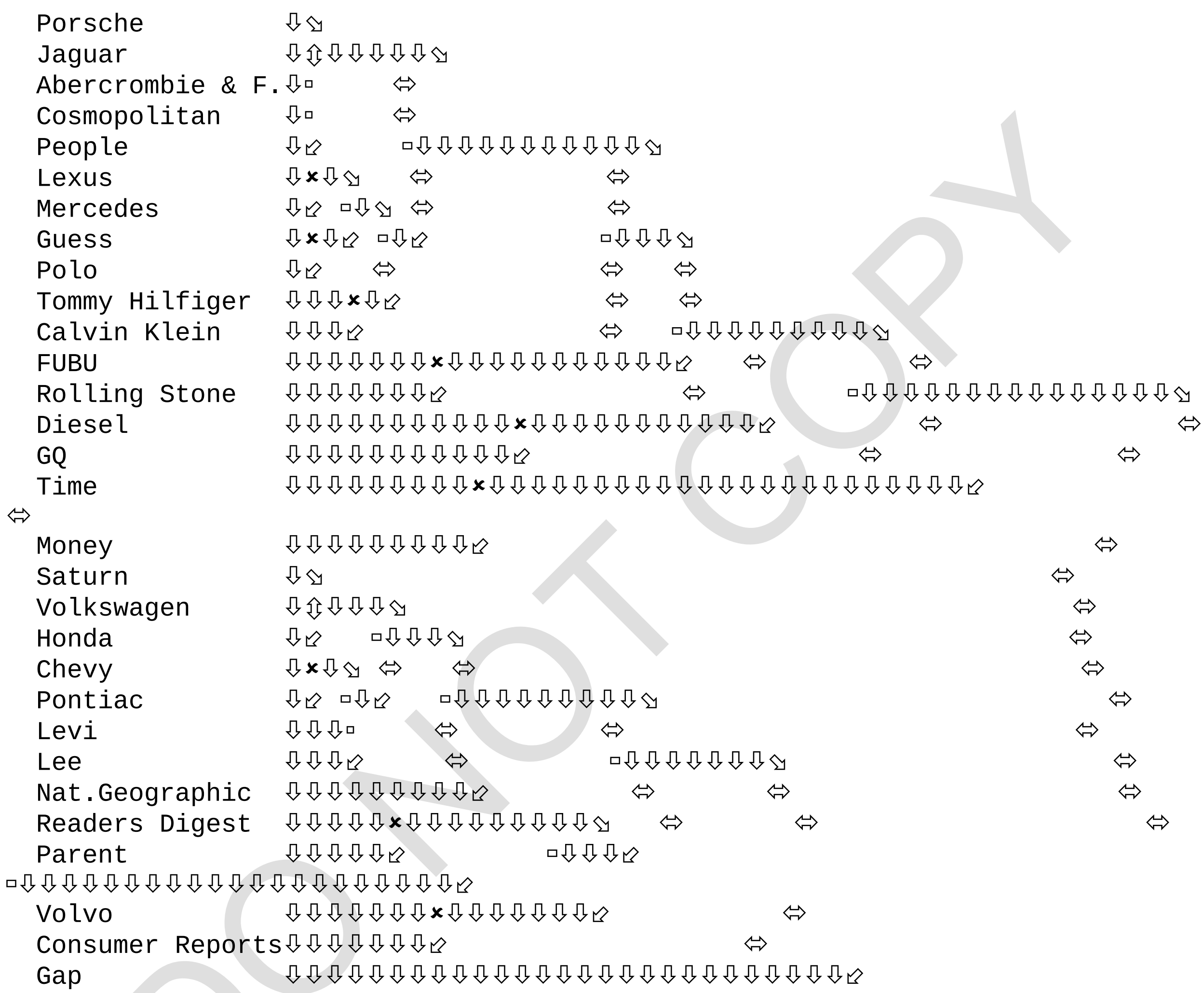
	Appropriateness	Novelty
Eigen value	6.836	1.356
% Variance	68.36	13.56
Factor Loadings		
LOGICAL	**0.887**	0.213
MAKES SENSE	**0.881**	0.224
APPROPRIATE	**0.874**	0.243
FIT	**0.863**	0.256
FAVORABLE	**0.845**	0.338
GOOD	**0.838**	0.343
POSITIVE	**0.830**	0.342
ORIGINAL	0.216	**0.887**
NEW	0.234	**0.881**
SPECIAL	0.398	**0.789**
Random Effects Multivariate Regression (Standardized Coefficients)		
Atypicality[b]	-0.129***	-0.010
Fit	-0.122***	0.053***
Prior Attitudes: from Brand	0.129***	0.118***
Knowledge: from Brand	0.001	0.061***
Prior Attitudes: Target Category	-0.015	-0.062***
Knowledge: Target Category	0.045*	0.024

* p-value < .10
** p-value < .05
*** p-value < .01
a. Concept evaluations and brand impression and knowledge were separately standardized within subject. Four subjects were deleted because their evaluations were constant across concepts or brands.
b. Atyicality and Fit have a correlation of 0.609. The residual from regressing Atypicality onto Fit is used in the analysis. The residual and Fit are uncorrelated. The model was Atypicality = 0.4763 + 0.1639*Fit

Journal of Marketing Research, Article Postprint, Volume XLVI

FIGURE 1

BRAND PROXIMITY IN CLUSTER ANALYSIS[a] SHOWS OPPORTUNITIES FOR CROSS-CATEGORY BRAND LICENSING AND CO-BRANDING DEALS



[a] Clustering using Brand - Category Means, standardization within brands

Journal of Marketing Research, Article Postprint, Volume XLVI

Brand Extension Strategy Planning:

Empirical Estimation of Brand-Category Personality Fit and Atypicality

Rajeev Batra, Peter Lenk, and Michel Wedel

Web Appendix A: Brand and Category Personality Model

The model specification for Equations (1) to (3) can be summarized as:

$$\begin{aligned} & U_{i,j,b|c} = \varphi_i + \mu_j + L'_{j,c}\left(\alpha_i + \alpha_{i,c} + \alpha_{i,b|c}\right) + \varepsilon_{i,j,b|c} \\ \text{(A.1)} \quad & \text{Var}(\varphi_i) = \tau^2 \text{ and } \text{Var}\left(\varepsilon_{i,j,b|c}\right) = \sigma^2_{j,b|c} \\ & \text{Var}(\alpha_i) = \text{diag}\left(\lambda^2_f\right); \text{Var}\left(\alpha_{i,c}\right) = \text{diag}\left(\lambda^2_{f,c}\right); \text{ and } \text{Var}\left(\alpha_{i,b|c}\right) = \text{diag}\left(\lambda^2_{f,b|c}\right) \end{aligned}$$

for subject i, personality item j, and brand b nested within category c. All of the random components are normally distributed with mean zero. These random components induce a covariance that reflects the structure of the data. Table A.1 details the covariances. In our application, there are J=15 brand personality items, B=10 brands per category, C=3 categories, and P=5 brand personality dimensions. The total number of personality items are J×B×C = 450. An unconstrained covariance matrix has 101,475 unique parameters. The factor model has 666 parameters to model the covariance, excluding the J=15 mean parameters. There is one variance for the subject, random effect; 450 (J×B×C) error variances, 5 (P) variances for subject personality factors, 15 (C×P) variances for category personality factors, and 150 (B×C×P) variances for brand personality factors. The total number of loadings is 45. To identify the CFA model, we set the first, subject-level factor variance λ_1^2 to a constant. In special cases, such as C=1 or B=1, other constraints would have to be imposed on the model.

Journal of Marketing Research, Article Postprint, Volume XLVI

Table A.1. Implied Covariance Structure from Hierarchical Factor Model

Items	Brands	Categories	Variance or Covariance
j	b	c	$Var\left(U_{i,j,b\|c}\right) = \tau^2 + \sum_{f=1}^{P}\left(\lambda_f^2 + \lambda_{f,c}^2 + \lambda_{f,b\|c}^2\right) l_{f,j,c}^2 + \sigma_{j,b\|c}^2$
j ≠ j*	b	c	$Cov\left(U_{i,j,b\|c}, U_{i,j^*,b\|c}\right) = \tau^2 + \sum_{f=1}^{P}\left(\lambda_f^2 + \lambda_{f,c}^2 + \lambda_{f,b\|c}^2\right) l_{f,j,c}\, l_{f,j^*,c}$
j & j*	b ≠ b*	c	$Cov\left(U_{i,j,b\|c}, U_{i,j^*,b^*\|c}\right) = \tau^2 + \sum_{f=1}^{P}\left(\lambda_f^2 + \lambda_{f,c}^2\right) l_{f,j,c}\, l_{f,j^*,c}$
j & j*	b ≠ b*	c ≠ c*	$Cov\left(U_{i,j,b\|c}, U_{i,j^*,b^*\|c^*}\right) = \tau^2 + \sum_{f=1}^{P}\lambda_f^2\, l_{f,j,c}\, l_{f,j^*,c^*}$

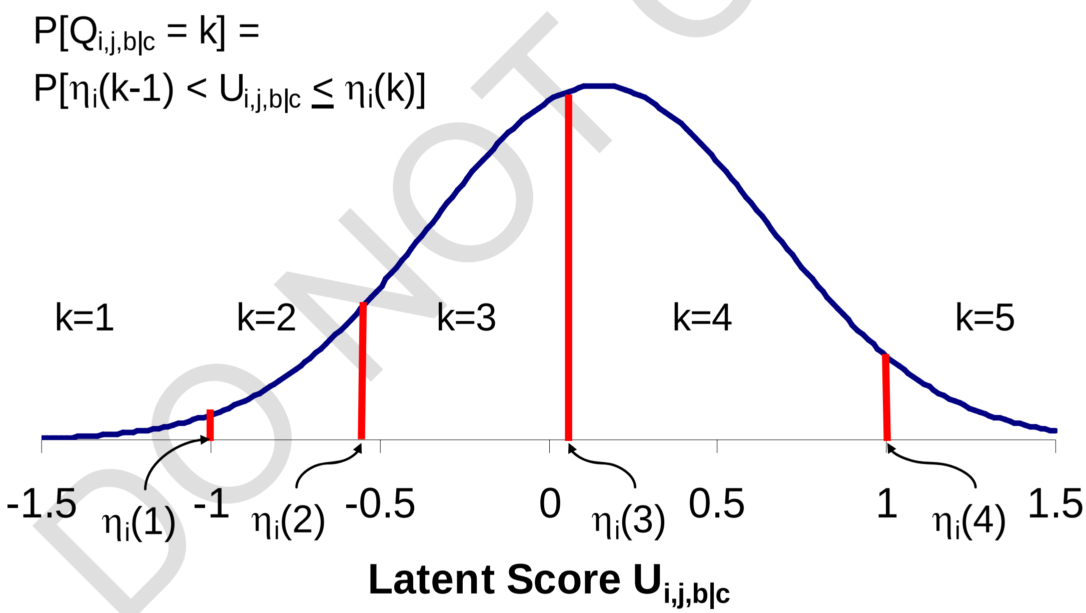
A cut-point model relates the latent variables $U_{i,j,b|c}$ to the observed ratings $Q_{i,j,b|c}$. All personality items were assessed on a nine-point scale. The cut-point model relates the ordinal personality scale responses to an underlying, normally distributed random variable. These models have a long tradition in statistics. Aitchison and Silvey (1957) proposed an ordinal probit model, and McCullagh (1980) used threshold models for ordinal regression. These models effectively relax the assumption that the observed variables have a normal distribution. The ordinal data can be very non-normal – e.g. skewed or bimodal – and the latent variable, which is used in the factor models, is still normally distributed. Gelfand, Smith, and Lee (1992) developed MCMC methods for ordered data. Rossi, Gilula, and Allenby (2001) and Johnson (2003) used heterogeneous threshold models to correct for scale-usage effects. The cut-point model assumes that respondent *i* selects scale point *k* for item *j* if the latent variable $U_{i,j,b|c}$ falls between two consecutive cut-points:

(A.2) $$Q_{i,j,b|c} = k \quad if \;\; and \;\; only \;\; if \;\; \eta_i(k-1) < U_{i,j,b|c} \le \eta_i(k),$$

where there are K scale categories. Figure A.1 shows the cut-points for the situation where $U_{i,j,b|c}$ has a normal distribution, the standard case, and there are 5 categories. In this paper, the cut-points are specific to the respondent and are ordered as follows:

$$\begin{aligned} &\eta_i(0) = -\infty \textit{ and } \eta_i(1) = -1 \\ &\eta_i(k-1) < \eta_i(k) \textit{ for } k = 2,\ldots,K-1. \\ &\eta_i(K-1) = 1 \textit{ and } \eta_i(K) = \infty \end{aligned} \tag{A.3}$$

To identify the model, we set $\eta_i(1)$ and $\eta_i(K-1)$ to –1 and 1, respectively. There are K-3 unknown cut-points per respondent, and these are estimated from the data. The model accounts for individual differences in the cut-points to accommodate idiosyncratic scale use, but we are not interested in the values of the cut-points for individual respondents per se.



The distribution of the observed ordinal personality data is derived from the latent variable and the cut-points are as follows:

Journal of Marketing Research, Article Postprint, Volume XLVI

(A.4)
$$P\lfloor Q_{i,j} = k \rfloor = P\lfloor \eta_i(k-1) < U_{i,j} < \eta_i(k) \rfloor \; for \; k = 1,...,K$$
$$= \int_{\eta_i(k-1)}^{\eta_i(k)} dF_{i,j}(u) \quad ,$$

where $F_{i,j}$ is the cumulative distribution of $U_{i,j,b|c}$, which is taken to be the normal distribution in this paper.

We use standard prior distributions. The prior distributions for all means $\{\mu_j\}$ are normal with mean zero and large variances. In CFA, the free loadings are assumed to be positive, and their prior distribution is a normal distribution with mean zero that is truncated below at zero, also known as "half-normal." All variance parameters come from an inverse gamma distribution. In the cut-point model, the free parameters are uniformly distributed on $\eta_i(2) < \ldots < \eta_i(K\text{-}2)$.

A final feature of our approach is that it enables imputation of missing data (if any), arising due to the split-structure of the questionnaire. Splitting the questionnaire, by randomly assigning blocks of questions to subjects as we did in the present study, creates missing data, hindering the application of traditional factor models. We impute the missing data at the same time as we estimate the model, by drawing from the "predictive distribution" of the data, an approach that is made possible by the application of the Gibbs sampler (Gelfand and Smith 1990). The estimation of the model adapts the MCMC procedures described in Ansari and Jedidi (2000), Ansari Jedidi and Dube (2002), and Song and Lee (2001 and 2002).

Web Appendix B: Summary of MCMC Sampling Scheme

We assume that the reader is familiar with the updating the full conditional distribution for standard models using MCMC.

Update cut-points $\{\eta_i(k)\}$ for k = 2, …, K-2.

The ordinal response $Q_{i,j,b|c} = k$ if $\eta_i(k-1) < U_{i,j,b|c} \le \eta_i(k)$. Thus, the full conditional distribution for $\eta_i(k)$ is uniform on the interval with lower endpoint $\max(U_{i,j,b|c} = k, \eta_i(k-1))$ and upper endpoint $\min(U_{i,j,b|c} = k+1, \eta_i(k+1))$. We need to include $\eta_i(k-1)$ and $\eta_i(k+1)$ in the updating if none of the observations for subject i take the values k or k+1, respectively. We exclude missing observations.

Update the latent variables $\{U_{i,j,b|c}\}$.

If the ordinal response $Q_{i,j,b|c} = k$, then $\eta_i(k-1) < U_{i,j,b|c} \le \eta_i(k)$ where the latent variables $\{U_{i,j,b|c}\}$ are independently distributed normal distributions with means and variances:

$$E\left[U_{i,j,b|c}\right] = \phi_i + \mu_j + L'_{j,c}\left(\alpha_i + \alpha_{i,c} + \alpha_{i,b|c}\right) \text{and } \mathrm{Var}\left(U_{i,j,b|c}\right) = \sigma^2_{j,b|c}.$$

If $Q_{i,j,b|c} = k$ is observed, then generate $U_{i,j,b|c}$ from a truncated normal distribution on the interval $\eta_i(k-1)$ to $\eta_i(k)$ with the above mean and variance. We use the inverse cdf normal method for generating truncated normal distributions. If g and G are the unconstrained normal density and cumulative distribution functions for U, then generate a < U < b by: $U = G^{-1}\{ V[G(b) - G(a)] + G(a) \}$ where V is a uniform random deviate on (0,1). If $Q_{i,j,b|c}$ is unobserved, then generate $U_{i,j,b|c}$ from a unconstrained normal distribution with the above mean and variance.

We then analyzed the rest of the model with one of two methods: using the generated $\{U_{i,j,b|c}\}$ for both the observed and missing data, and using just the observed data. The first method corresponds to Bayesian data imputation methods. The results of the two methods are theoretically identical and the numerical results were similar. The paper reports the results using just the observed $\{U_{i,j,b|c}\}$ because they are missing by the design of the survey: subsets of subjects evaluated different subsets of the 30 brands. If there were non-responses to items on the questionnaire, we would impute the missing values. However, our dataset had compete responses to all items on each questionnaire.

Update the subject-level random effect $\{\varphi_i\}$.

Because the random effects have $N(0,\tau^2)$ distribution, and

$$\phi_i = U_{i,j,b|c} - \left[\mu_j + L'_{j,c}\left(\alpha_i + \alpha_{i,c} + \alpha_{i,b|c}\right) + \varepsilon_{i,j,b|c}\right]$$

has a normal likelihood, the full conditional is also normally distributed with the standard, updated mean and variance.

Update the variance τ^2 of the random effects.

Because the prior distribution for τ^2 is an inverse Gamma distribution, and the $\{\varphi_i\}$are iid $N(0,\tau^2)$, the full conditional is also inverse Gamma with the standard updating of the prior parameters.

Update the population mean μ_j for item j.

Because the prior distribution for μ_j is a normal distribution and

$$\mu_j = U_{i,j,b|c} - \left[\phi_i + L'_{j,c}\left(\alpha_i + \alpha_{i,c} + \alpha_{i,b|c}\right) + \varepsilon_{i,j,b|c}\right]$$

has a normal likelihood, its full conditional is also normally distributed with the standard, updated mean and variance.

Update the error variance terms $\{\sigma_{j,b|c}^2\}$.

Because the prior distribution of $\sigma_{j,b|c}^2$ is inverse Gamma distribution and the $\{U_{i,j,b|c}\}$ has a normal distribution with variance $\sigma_{j,b|c}^2$, the full conditional of $\sigma_{j,b|c}^2$ has an inverse Gamma distribution with the standard updating of the prior parameters.

Update the factor loadings $\{L_c\}$.

In CFA, the priors of the free loadings have truncated normal distributions that are greater than 0. Because the loadings in

$$L'_{j,c}\left(\alpha_i + \alpha_{i,c} + \alpha_{i,b|c}\right) = U_{i,j,b|c} - \left[\phi_i + \mu_j + \varepsilon_{i,j,b|c}\right]$$

given the factor scores have a normal likelihood where the variance $\sigma_{j,b|c}^2$ are scaled by the factor scores, the full conditional also has a positive, truncated normal distribution with the standard updating of the mean and variances. We use the inverse normal cdf method for generating truncated normal deviates. In EFA, the priors for the free loadings are normally distributed, and the full conditionals also have normal distributions with the standard updating of parameters.

Update the factor scores $\{\alpha_i, \alpha_{i,c}, \alpha_{i,b|c}\}$.

The model assumes that the factor scores are iid normals with mean 0 and variances $\{\mathrm{diag}(\lambda_f^2), \mathrm{diag}(\lambda_{f,c}^2), \mathrm{diag}(\lambda_{f,b|c}^2)\}$. Because the factors scores in

$$L'_{j,c}\left(\alpha_i + \alpha_{i,c} + \alpha_{i,b|c}\right) = U_{i,j,b|c} - \left[\phi_i + \mu_j + \varepsilon_{i,j,b|c}\right]$$

given the loadings have a normal likelihood with scaled variances, their full conditional distribution also is a normal distribution with the standard updating of the means and variances. In the updating one needs to keep track of which observations are used for which factors. All of the observations for subject i are used in updating α_i. All of the observations for subject i and category c are used in updating $\alpha_{i,c}$. Only the observations for subject i and brand b in category c is used in updating $\alpha_{i,b|c}$.

Update the factor variances $\{\lambda_f^2, \lambda_{f,c}^2, \lambda_{f,b|c}^2\}$.

Because the prior distributions for these variances are inverse Gamma and the factor scores are normally distributed with mean 0 and these variances, the full conditionals are also inverse Gamma distributions with the standard updating. To identify the model in CFA, the first component of λ_f^2 is set to a constant and is not updated.

We fitted the model with Markov chain Monte Carlo (MCMC) methods (Gelfand and Smith 1990), using a program we wrote in GAUSS. The MCMC algorithm ran for 100,000 iterations. The initial transition period consisted of 50,000 iterations, which were not used in estimation. Of the next 50,000 iterations, every tenth iterate was used in the analysis for a total of 10,000. Traces of the iterations indicate that the chain reached the stable distribution well before the 50,000 iteration, and simulation studies using artificial data indicated that 30,000 iterations were more than sufficient: the chains using simulated data frequently converged within 3000 iterations and remained stable at the true parameters thereafter. The algorithm simultaneously imputed the data for brands that a respondent did not evaluate, thus simplifying the analysis.