PROCEEDINGS OF THE SAWTOOTH SOFTWARE CONFERENCE

April 2021

Copyright 2021

All rights reserved. No part of this volume may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from Sawtooth Software, Inc.

ENHANCE CONJOINT WITH A BEHAVIORAL FRAMEWORK

Peter Kurz Stefan Binner BMS - MARKETING RESEARCH + STRATEGY

BEHAVIORAL FRAMEWORK

As shoppers process information and act on it, they are not simple stimulus-response robots. Creating a behavioral framework prior to answering choice tasks therefore helps respondents select from choice tasks as if they were in a real purchase situation. If price and assortment changes are the focus of the research, it is particularly important to understand shopper perceptions of prices and values. Again, a behavioral framework is useful for interpreting consumer decisions, as simulated by the results of the choice model, in the appropriate context.

To create such a behavioral framework, prior to each conjoint exercise, we apply nine standardized, binary "Behavioral Calibration Questions" regarding each respondent's individual shopping behavior for the focal category. Based on principles from behavioral economics, these questions help consumers recall their usual buying habits. "Behavioral Calibration Questions" are also used to describe the context of consumer choices, including how purchase decisions are made within a specific category, as they reveal typical patterns of buying habits, purchase repertoires, and brand value perceptions, as well as price knowledge.

BEHAVIORAL CALIBRATION QUESTIONS

We use the derived contextual information about each respondent's individual disposition towards brand and price knowledge (or lack thereof), past behavior, and perceptions within the category in our analysis. Retrieving a prior shopping situation and their individual dispositions helps consumers to make decisions in the following choice experiment. Currently, the set contains nine "Behavioral Calibration Questions" (semantic differentials) with respect to buying habits along three dimensions: brand, price, and innovation.

"Behavioral Calibration Questions" are used in our research context for several purposes:

- to establish a behavioral framework before respondents answer the choice tasks,
- as covariates in the hierarchical Bayes estimation process, and
- as segmentation/filter variables in the choice simulator.

Furthermore, we store the responses to the "Behavioral Calibration Questions" in a benchmark database to anchor further conjoint studies in the different product categories.

OUR STANDARD BEHAVIORAL CALIBRATION QUESTIONS

In each of our conjoint questionnaires, we combine the binary questions with our nine semantic differentials and ask respondents which of two statements (left or right) is more related to their last shopping trip.

V F t V S	We would like to learn a few things about you and your general thoughts, feelings, and opinions when it comes to home upkeep, construction adhesives. Please read each pair of statements. For each pair, please indicate whether you agree with the statement on the left or the statement on the right more, and how much more. If both statements describe your opinion well, choose the one that best describes you. If neither seems to describe you well, choose the one that comes the closest. Select one response for each.									
		Agree Left	Agree Right							
	I think that brands differ a lot	\bigcirc	\bigcirc	I think that all brands are more or less the same						
	I always know exactly what brand I'm going to buy before I enter the shop	\bigcirc	\bigcirc	I decide what brand I'm going to buy when I'm standing in front of the shelf						
	I always buy the brand I bought last time	\bigcirc	\bigcirc	I switch between different brands						
	I compare prices very carefully before I make a choice	\bigcirc	\bigcirc	To be honest, I compare prices only superficially						
	I always search for special offers first	\bigcirc	\bigcirc	Special offers are not the first thing I look out for						
	I always know the price of the products I buy	\bigcirc	\bigcirc	I never really know what products cost						
	I'm always interested in new products	\bigcirc	\bigcirc	I prefer to stick to what I know						
	I think that products in this category need to be improved	\bigcirc	\bigcirc	I'm completely satisfied with the products as they are						
	I find it easy to make the right choice for me	\bigcirc	\bigcirc	I find it very difficult to make the right choice for me						

Example from R&D study in US (2020, context: construction adhesives)¹

Of the nine semantic differentials, three are related to the "Role of Price," three to the "Role of Brand," and three to the "Role of Innovation." This approach allows respondents to recall past behavior when buying a product in this category.

Over the years, we have adapted sets of nine semantic differentials to different product categories, as not all statements behave similarly in distinct shopping situations or categories. For example, when buying a new car, virtually no respondents would answer, "I never really know what the car I buy would cost." In this situation, one needs a question such as "I never really know what the competitive brands cost; I more or less compare only within my preferred brand." Such adaptations for each category are necessary to produce a valid framing of respondents from different target groups.

¹ We are uncertain of the origin of these questions; we first encountered them in a segmentation approach from Research International in 2008. In this approach, the questions were asked as scale questions and used to derive consumer segments.

	3-Scherheit der Wahl vs. Ambigalerz		1- Minedemiet mit Status Que		1-Interesse an neuen Innondiomen		f-Resilentitis		5-Sonderangebooe		4-heisegede)-Nederlageliter, (Acception		1- wichtgliet der Warte bei der Gaufertscheidung)- Naterofferenseug	
Denotige Wahlfurmon zu helfen fältnir softwar	Denotinge Nati für mots zu Intellen, fäll mir leicht	ich tie vol scheden mit der Problem, wie se sind	ich glaube, die Produite müssen verbessert werden	ich bieber beitern bei ben, was ich schur Hanne	lot bir inner al neuen Produkter interessent	ich well ne genau, was die Produke Lusser	ich lanne ganau die Franse dar Frankille, die chi laufe	Sandwargiptute said noti dae Entek, worsch ich suche	ich suche immer zuerschach Sanderzeigkoben	Enito paag vergleche ch Prese eher uherfachlich	loh vergilecte Praise sehr genau, tenorich laufe	ich laufe inner die gebie Natie Vorscheiteren Barten	kh enborede nuch eschen Nauf für eine Nach	ich veräligistes kärigeraas, versite läiste ich laaviter versite	De liaden urterscheiden sich staht nommander Alle Tanter sich nicht nicht werüger gleich	PMC5
Derichtige Natif für mich zu Innflein Stilter schwar	De richtige Natif für mich zu herften, fält mir eicht	ich im voll zuhleter mitten Angeboten, von sie sind	ich sterke, das Angebot mas noch verbasset werden	ich seine leiter berühmt, was ich achun leime	ld tin immer an neuen Angeborten interessiert	ich meis ze genau, was die einzelnen Angebote loote	ld lene de Reie de far nich interesanten Argebote	Sandarangahata sind noti das Estat, women bit suche	ld sole mer zwis net Sinderingsholer	Enici yasığı veryeche ch Preve drer sterilichich	ich verpleche Prese sehr genau, bevor ch ubechliebe	ich wähle nimer ben gleicher Anbeiter Ich nichtste zwischen Verschiedenen Anbeitern	ich entscheite mich entstate Abschluss für den Anbeiter	kt med jetzt schon, vetorer Antieter ich wilfte	De Anbeter unterscheiden schlisten nuterscheider verligte glech verligte glech	Media (Pay TV, VoD)
De notige Nati for mot to befor, fait mi schwei	De nortige Natif für mot zu terften, falt mir leicht	Andreetweet, we se and	th danke, das Angebot music rich vertreuert merzer	ch schun lenne	tt bn imer an selen Mobileletoren intersset	th well ne yerau, was de einzelnen Modelle losten	ich leine die Preise der Ein mich mensionen Wodelle	Socialization and not das Erals, virradi til suche	th suche miner zuerst mach Sunderangeboter	Britist gesagt versjecte de Frese etter oberfectiot	lch vergielche Prose sehr genau, bevor ch laufe	tit karle inner de gløtter Narte Itt methek zwischen Herschiederen Narten	th entscheide mich erst bein Kauf für eine Table	kti velijetiti schan, vetice liete chiaufen verite	Se læver unterscheder sich statt vonetander Alle Narter sind mitr oder værliger plack	Mobilteletone
De schlige Walk for mich zu Teeffen: Salt mit schwer	De nortige Nadi Grancet zu techen, Silt ner kont	ich ber voll sufreiden mit der Computerin, wie sie sind	ici tanic, das Angebot mass rach verbenant menter	ch bieber leit bert, was ch schon lærne	chte nner är naven Computern riterssief	ch veid në geran, mas dë einzelner Wodelle inster	ch levre de Pese der fø nich niteressaten Nodelle	Southerangenite and notifi data State, wanted ich suche	ch suche immer zuerst nach Sonderangeboten	Brich gesagt vergleiche ich Preise einer icher fachlich	ich vergleiche Freise sehn genau, bevor ich kaufe	chade mer te jede llate ch vedse zwoden vescheben llaten	Chertscheite mich ans bem Kauf für eine Nache	ich weiß jetzt schon, wetche Namie die Vaurben werche	Die Namer unterscheiden sich staht undersander Alle Namer sich mehr oder weniger gleich	PC Notebook
Dentifige Real for mich zu- herfinen fälterne schwer	De richtige Walkformón zu henfen, Natione Hacht	forter virt schieden mit der Fermæltern, vie se stof	lo texis, das Angebot mus voci vectassiet verdiet	ich blebe leber bei Jon, was ich schur lenne	lo bi inner ænslett Fernsellern nieressert	lch weid tie genaa, was die einzelnen Modelle tuster	ich lanne die Preise der Tit nich nichtesaster Mindelle	Sonderængebote sold nicht dae Erste, wonach ich suche	lch suche inner zwest nach Sonderzegeboten	Brido gesagt vergleche ch Prese eter ster fiscolich	ich vergieche Ansee sahr genau, bevor ich kaufe	loi laute mer ste gebte Natie Ion wecke zwischen verscheiteren Barten	ko enaceste not encleen. Kaaf far ene Narte	lo vel jett schor, veche llarte di kurist veche	Die Hanken unterscheiden sich staht novemander Alle Tierten sich mehr sofer versiger giech	Femseher
Derichtige Nati für nich zu herten, fäll mit schwät	De notige Nali for nich zi hoften, fall nir lecht	in in vitatheler state Autos, vie se and	ict daria; das Angebot musi roch ierbezeitt werden	ich bieder leich fann, waar Ich schwin leinne	let bin inner an reuer Autos interessiert	ist weit ne genau was de einzelsen Modelle kosten	lch lenne fie Prese der für mich nierenaarker Madelle	Sonderzegelacht sind nicht Bast State, worsch ich sinche	ld aute mer zweit nich Sanderzegisteter	Brid yeag regiside on Prese eller skenischich	lah vergleche Pesse sehr genau, bevor ah kavie	ich nähe mer de gebte Automarke Ich nachsie zunschen erschlecherel Automarken	ld etstedt nich est umheber vor den Kad	ich mell jetti schen, welche läche ch laufen werde	unterscheiden sich statt unterscheiden sontersteter Alle Automachen sind mehr oder weniger gietzt	Auto
Die notige Wald für mich zu hefter. Fall mit schwer	De notige Nall for not za terber, falt ne lecti	bit the notice find the Tarfeet, with the set	bi dania, disi Angebot masi radi vetasari werter	ich beder leicher bei dem, was ich schon lenne	ti bin inner an reven Tarrien interessert	bt weiß ne genau, was die einzeleen Tarife tosten	th leave de Prese der Sv noti recreaanter Taulte	Sondersongelande sond motte date Erster, verrach och sucche	bh suche miner zuerst nach Sonderzugeboter	Bridt gesagt vergleiche ich. Prese einer oberfächlich	bh vergleiche Preise sein gerau, bevor ich absolhielle	bt wähe inner der jeicher Anbeiter bt websie zwischer reschiedenen Anbeitern	th entschede mich erst bei Abschlass für den Anbieter	bit veli ježit schon, velcen Anbieter ich wähe	De Maketer unterscheider sich stahl vonstanster Alle Maketer sich rahr oder weniger giech	Mobiliarite
Derschige Wahl formich zu befreit. Bilting schwer	Die indrige Vasif für nicht zu herfen, Säthar leicht	tation via suffection of the Tanlera, vie se and	ct denie, das Angebot muss wich verbestaert wenden	ch bieder leiden hei dem, was ch schum leinne	ch be inner al sous Tanlee interessert	ch well ne gerau, mas de einzelnes Tarife kosten	chierne die Prese ber St nich interssanden Tanlie	Sorderungstate sind richt das Erste, wonsch ich suche	ch suche immer zuerst nach Sonderwinjebatien	Entot yesyt verpleche ich Preise eller ohenflichlich	ch vergleiche Ansie seiter genau, tervor ich abschlieble	ch wilke immer den glechen Anthelter dri verchelte zu hochen verscheiteren Aufheiterin	ch erischeite nich ersitte Abschluss für den Anbieter	ich weiß jetzt schon, weichen Anbeiter ich wähe	be Anheter uterstecter act start overander Ale Anheter ant nehr ofer wenige gieth	Telekontarife
Die richtige Riefi für mich zu breften, fällt mir schwer	Denotige Nati formot za Inofes, fall nor exist	to bir vil scheden milder Tanlen, weise sind	lc) devia, das Angebot musi rucci vetvesent werden	ich bleder letter bei dem, was ich schwil lettere	lot bis inner ærnsvet. Tænten riferssæt	lch well ne genau, was die einzelnen Tantle luster	lch lærne de Prese der Tør mich alexesanden Tarlfe	Sondarangelote sed noti des Electe, worsch ich suche	ldt suche mmer zwesch nach Sonderze geboten	Enio yeag eryecte ch Prese eter sterfachich	loh vergleche Prese sehr genau, bevor ich abschliebe	lch withle inner den glecker Andreter Ich wechste zuwischen verschiederter Aubertern	lch entscheide nich erst bei Absohless für dem Anbieter	lch veräljettil schon, vector Anheter ich veläke	Die Anbieter unterscheider sich sich normander Alle Anbieter sich nehr sider weruger glech	Stromtanle
Die notifige Natil für mich zu herften. Silt mit schwer	De richtige Natif für nich zu herten, falt nir eicht	de per vol zu fieder mit den Gerokansten, wie sie sind	ich starter, das Angebot muss noch netweset versten	id beite letter beiden, was id schur letter	b in mer an neven Grokunts-Angeboten Interessert	ici nei, regeau de Kandtionen meires Girokuntas	id ione peau de Konditionen meines Gerokuntus	Senderkonditionen sind nort bas Erster wonach ich suche	b) such imer such such Sanderkundhörren	ich vergleiche Konditionen eher aberflachich	bb vergleiche Konditionen schrigerau vor Abschlass	bt wähe mer är gleche Bank bb methek zurschen verscheidener Banken	ist estabeds not erst kurz vor Abschlass für d. Bank	lch mell jetzt schen, welche Banik ich wähle	De Banken unterscheite sich sich nitretauster Alle Banken sich nicht sicht veruger gleich	Geokanto
Die nothige Vladi für mich zu befüre. Fält mit schwer	De rottige Valif für nich za berfan, fält nir lecht	bit bin vol zufreder mit der Andegren, mit der sind	bt danke, das Angebot hans noch vertessert werden	bi bieder leicher ter dem, wass cht schon lenner	bi bin miner an neuere Anlage-Angeboten Internaert	ch velute grau de Kondinnen meiner Anlagen	ttiene jenu že Kontikoven melser Anlagen	Sonderkonditionen sind nicht das Erste vinsich ich suche	bt suche immer zuerst nach Sonderkondfillonen	th vergleiche Kandilionen eher oberflichtot	ch vergleiche Kandhinnen sehr genas vor Abschluss	th wille mer är sjaden Aubeter I hvecke zvischer I orschelene Aubetern	tt eischeit ers karz vor Abschuß über den Anbeter	ich versigend schon, versichen Antoleter ich wähle	Se Andreier unterscheiter sich stati vonerander Alle Andreier sich nahr oder weniger gleich	Span-Anlagen
Dendrige Wall formch zu heffet. Bit mit schwer	Dendrige Wall formoto zu terfen, 188 nar lecti	stationaliset, ne se and	ch denie, das Jugebot muss nuch verbesauf werden	ch bleder leider har ben, was ch schun lærne	ci bir mer ar rever Angeboten interesset	ci vel, re jera, nes nene Vrakenkasse lostet	chierne gena die Kasten meteor Konkenkusse	Sonderkanditionen sind holit das Erste winach ich suche	ch suche misse zuerst nach Sonderkundbünnen	Ento yang veglebe bi Konten eter nerfisitio	ich vergleiche Kustern sehr genau von Abschluss	ch wähe imer de jeiche Kraitenbasse ch vechet zu schett verschetten Kraitenbassee	ct erischelle nich ans kann wor Abachluss	ich veräl jettil schom, vechte Kranivenkasse ich wähe	artenskensen antensken sit sitt Alt kontentaant sit Alt kontentaant sit	Krankenkasse
Denotisje Natiliär möt zv heften fält nir softwar	Derottige Natificiation an Institut, Sation Bedit	ich im vol acheden mit WEZ- Taniten me se sold	lch sterier, das Angebot nuus roch verbeseet verben	ich Steiter Leistern, was ich schar Leiner	ld bir inner annaven. Angeboten istemasert	ldt velå te gana, was meine Versiderung loste	ich lærne gensu fre Kustern melmer Versicherung	Soderkandhonen sid noti da: Eisik wonach idt suche	ld sude mer sversligst Sonderkandfilanen	Enici yeap enjecte ch Kostee eter oberfactlich	ldi vergleche Kosten sehr genau vor Abschluss	lo wähe mer le jische Versicherung Io wechst zuscher versicherungen	tin enscheite nich ess kanz vor Abschlass	lch verså jettel schove, versocher Versächer nang lich wähle	ofersonenige ofersonenis operatie Ne Vesideringer sid net nie weige ged	VF2-Versiderung
Denotige Nati fir mich zu heften fäll mit schwer	De notige Nati far nati za traften, falk mi lecti	ich bin vol zuhlechen mit dem Fluggangebot, wie eis ist	bt darie, das Angebot musi noch instesset verden	ich beside leiter besiden, was ich schur leinee	ld bi inne ar reaer Rug- Angeboten nteressert	bt well ne genau, was de Filige loster	le leve no te Rugreset plas	Sinderangebole sind nicht das Erste, verrach ich suche	ld such mer perst such Sindersrjøtner	Bridt gesegt vergleiche ich Preise eher sterflichlich	ld vergleiche Preue sehr genuu, tevror dit buche	ich turche faut immer die gesche Anthee Ich wechsie zurschen wescheidenen Anthees	tri estatede not est bei Buchung für eine Airline	lch versk immer genau. verske Avrike of buckern verske	De Antines untersteden sch stati voeinander Alle Antines sich nehr ster veriger glech	Airines
De notige Hall for mot zu Telfen. Bit mit schwer	De notige Natifiarmot za hafter, falt milecti	tit bin vol zufinder nit dem Reiseangebot, wie is st	tit danke, das Angebot muss och vertessert verder	ch schon lietme ch schon lietme	ti to inner ar reuer Fraschalergeboten tieresert	b) veli në garai, van de Reisen luster	thiere not te Telepineer pit au	Sondersingedinte sond nucht lass Erste, without ich suche	th such a timer cuerch rect. Somerenighter	Britch gesugt vergleiche ich Freuer einer oberfächlich	ch vergleiche Ansee sehr perau, liever ich buche	pecter pecter d) vectete 2000/80 escheterer escheterer	bt etscheite nich erst bei Buchung für einer Anbieter	bh weiß genau, welchen Venanstalter ich wählen werde	unerscheider sch stahl noreinander Hie Reisenermatalter and Hie noter werige glech	Pauschalreisen

BEHAVIORAL CALIBRATION QUESTIONS

The first insight we derive from these nine questions is the identification of four respondent segments. Two semantic differentials, "brands differ a lot" and "always buy the same brand," can be used to classify consumers according to "Brand Loyalty" and "Category Involvement," thereby providing useful insights about the product category in general. Quantifying these different buyer segments is useful for identifying the best-performing strategies for products under investigation in the choice model.



This classification mostly refers to consumers' attitudes towards brands. A consumer classified as "Indifferent" is not necessarily indifferent to other attributes. Segment names should not be taken too literally, as classifications represent only a rough outline of consumers' personalities. For instance, a "Loyal" consumer may actually have a relevant set of two or three brands. What makes her a "Loyal" consumer is her self-perception as someone who sticks to her brand(s) (as opposed to consumers who are indifferent to brand), and her belief that the difference between her brand(s) and others really matters.

Loyal

highly involved, and committed to one favourite brand

Critical

highly involved, but not committed to one favourite brand

Routine

uninvolved, habitually buying the same brand(s)

Indifferent

uninvolved and uncommitted



The figure above shows an example of the distribution of the four consumer types within the "laundry detergent" category. We see this as an initial blueprint for each category to begin interpreting the results of our choice models. Based on the benchmark from past studies, the client can easily determine how target consumers think about this category.

EXPERIENCE WITH BEHAVIORAL CALIBRATION QUESTIONS

Asking the nine "Behavioral Calibration Questions" before our choice exercise helps respondents to recall their behavior during their last shopping trip in a specific category. Therefore, we assume that the nine questions improve their decision-making process in the subsequent choice exercise, supporting a realistic answering behavior comparable to real shopping situations. Therefore, this approach helps generate more realistic data. Using the derived shopper classifications as segmentation variables in the choice simulator provides deeper insights into respondents' preference structure. Based on our findings from numerous conjoint exercises, we learned that answering the nine questions results in better "Share of Choice" estimates as compared with conjoint exercises performed without the calibration questions. Furthermore, part-worth estimates, which include the "Behavioral Calibration Questions" as covariates, further improve share predictions against holdout samples (ensembles with the questions and other covariates offer marginal improvement in results).

EMPIRICAL VALIDATION OF BEHAVIORAL CALIBRATION QUESTIONS

For validation purposes, we conducted nine empirical R&D studies over the last two years, in which we asked 50% of respondents the nine "Behavioral Calibration Questions" prior to answering the choice model, whereas the other 50% answered the choice model without being exposed to the semantic differentials prior to the choice tasks.

We addressed the following hypotheses in this paper:

- The framing offered by the "Behavioral Calibration Questions" results in improved answering behavior among our respondents, leading to part-worth estimates that are more stable and valid.
- Adding the answers to the "Behavioral Calibration Questions" as covariates in the HB estimation further improves the part-worth estimates.
- Using the questions as filter/segment variables in the choice simulator provides additional insights in the data as "Shares of Choice"; elasticities differ according to the derived segments based on roles of brand, price, and innovation.

All studies were conducted with respondents recruited from online access panels in 2019 and 2020 and the samples were split as outlined above (i.e., Behavioral Calibration Questions shown or not). The studies varied in terms of categories, number of attributes, number of levels, number of concepts, and number of tasks. Sample sizes depended on the number of parameters to be estimated and varied between 250 and 1,000 respondents.

Only one study ("super glue") differed slightly from the others, as we conducted 4 sample splits to create an opportunity to validate the estimation samples with separate validation samples. (For the two estimation samples, n=500 interviews, and n=250 interviews for the two validation samples.) These four split cells enable cross-validation of the part-worth estimates derived from asking or not asking the "Behavior Calibration Questions" and including or excluding them from the hierarchical Bayes estimation.

Project	N=	Attributes	Levels	Tasks/Concept per Task	Model Specifics	Covariates
Detergent ADW	1006	6	10+2*2+2*3+6	12/8+None	502/504	Socio-demographic, Purchase Behavior
Construction adhesives	510	30	6	15/4+None	250/260	Socio-demographic Purchase Behavior
Drops	1030	3	47+3+47*3	15/12+None	530/500	Socio-demographic, Purchase Behavior
Edible Fat	2030	12	12+6*2+3*2+2*5	15/6+None	1030/1000	Socio-demographic, Purchase Behavior
None Electric Air freshener	500	5	7+2*9+2+7	15/5+None	250/250	Socio-demographic, Purchase Behavior
Hair Shampoo	1016	46	96+ 40*2 +3*5	15/12+None	509/507	Socio-demographic, Purchase Behavior
Potato Chips	800	38	45+2+2+30*5	15/5+None	400/400	Socio-demographic, Purchase Behavior
Laundry Detergent	980	16	96+15*5	15/12+None	580/400	Socio-demographic, Purchase Behavior
Super Glue	1500	23	22+ 22*5	15/12+None	500/500/250/250	Socio-demographic, Purchase Behavior

BUYING HABITS AND INVOLVEMENT

The four segments derived from the "Behavioral Calibration Questions" have real potential to differentiate between categories and to identify promising strategies. For example, a significant proportion of "indifferent" consumers may have a larger effect on strategies for new product development, compared with a large share of "critical" or "loyal" consumers.



A comparison of the nine empirical studies shows that the different product categories have different compositions in terms of the four consumer segments. For example, in the "super glue" category, the study identifies an equal number of "Indifferent" and "Routine" consumers, whereas only a small proportion are "Loyal." In contrast, in the category "laundry," "Loyal" customers are by far the largest group, followed by the "Critical," "Indifferent," and "Routine" consumers. Furthermore, the "non-electric air freshener" (NECA) category has by far the highest share of "Critical" consumers. In the context of introducing new, innovative products, this category seems to offer many more opportunities as compared with the "super glue" category.



BEHAVIORAL ROLES

The three behavioral roles represent a second possible usage of the nine calibration questions to understand the behavior of the respondents during shopping trips in a

particular category. We found these roles helpful for interpreting the "Share of Choice" from simulations. They allow deeper insights into the reasons respondents behave differently in their choices.

The three roles derived from the "Behavioral Calibration Questions" are:

- Role of Price
- Role of Brand
- Role of Innovation

Each is represented by three semantic differentials that ask (in a binary manner) whether the left or the right statement better corresponds to the respondent's last shopping trip in that category.

The following results, based on nine empirical studies, demonstrate the diversity of behavior within the different categories.

ROLE OF PRICE

The "Role of Price" (RoP) is based on the following three semantic differentials:

"I compare prices very carefully before I make a choice"

"To be honest, I compare prices only superficially"

"I always search for special offers first" vs "Special offers are not the first thing I look out for"

"I always know the price of the products I buy"

VS

"I never really know what products cost"

The following table shows how differently consumers behave when buying within these nine categories:



The number of consumers who always compare prices carefully varies between 41.9% ("cough drops") and 86.4% ("NECA"). In the "edible oil" category, 40.8% are looking for special offers first, compared with 75% in the "automatic dish washer detergent" ("ADW") segment. Price knowledge varies between 42% ("edible oil") and 78% ("laundry detergent").

Such differences in consumer behavior are useful for interpreting results from choice models. For example, in the "cough drops" category, a price increase is more likely to be accepted, given that 58% of the consumers do not compare prices. In contrast, only 15% do not compare prices in the "NECA" category, so price increases could have a much higher impact on preference shares.

ROLE OF BRAND

The "Role of Brand" (RoB) is represented by following differentials:

"I always buy the brand I bought last time" vs "I switch between different brands"

"I think brands differ a lot" vs "I think that brands are more or less the same"

"I always buy the brand I bought last time" vs "I switch between different brands"

These three differentials provide insights into the RoB, thus deepening understanding of consumers' behavior in this regard. This approach provides further insight when interpreting simulations based on choice models.



Again, there are significant differences between the segments: The brand-switching attitude varies between 21% ("NECA") and 56.8% ("laundry detergents"), representing a significant difference when a company aims to "introduce a new brand" into a category. Because 72% of "super glue" customers think that all brands are more or less the same, compared with only 14% of "NECA" customers, it seems that having a strong brand has more equity in the "NECA" category as compared with "super glue."

ROLE OF INNOVATION

For "Role of Innovation" (RoI) the following three differentials are used:

"I'm always interested in new products" vs "I prefer to stick with what I know"

"I think products in this category need to be improved" vs "I'm completely satisfied with the products as they are"

"I find it easy to make the right choice for me" vs "I find it difficult to make the right choice for me"

With these differentials, we can derive insights about the opportunities for new products in the different categories.

	Construction A	(Rettinery			Calible				have been been been been been been been be	gent .	
	w	10			44.4	81.0			-	8.0	-
	-	86.1			m	Ra	manufactoria		N/I	- 140	
And a survey of the later terms	- 111	14	the state of the second st		4	*	fearing the second second	- the second strength of the spectrum of the s		144	THE R. LEWIS CO., No. of Concession, Name
	Ratematic Obtained	of Classorgiantia			Putato	(hips					
	- 944	447			194	948				80	
the foregroup of a series			Transmission and the set of the	interference of the same of th	e .	ALE .				- Mark	
	-		manager and the second	instantonicapitan Inte		W	Charles and the state of the spin			84.8	The second second second
	Cough Der				Nan Gartris	lår fingsflærner					
	21	1998/				18,4	-				
I	-	762	In constraint of the sector		ni	-	il and the second		-	ni	
*****	94.0	-961	(installing to open a set	international the special set	6	3	B Joseffanger			144	

"Satisfaction with current products" ranges from 16.6% ("NECA") to 82.2% ("Potato Chips"), representing a large difference in suppliers' opportunity to develop new products. Another example: 39% find it "easy to make the right choice" in the dishwasher detergent category, compared with 92% for "NECA." This may indicate the need for differentiation, such as by developing and clearly communicating specific USPs for different products.

INITIAL CONCLUSIONS

Results from the nine "Behavioral Calibration Questions" indicate that they have a potential to differentiate between respondents' buying habits. Regarding our hypothesis, behavior during the most recent shopping trip (within a category) influences the answering behavior in the choice exercise. Considering this, these questions should help respondents to recall their decisions during their last shopping trip more effectively; therefore, responses to the following conjoint tasks should be much easier and clearer to them. Bivariate analysis of the "Behavioral Calibration Questions" suggests that our hypothesis may be correct and that it is worthwhile to invest the additional interview time to improve the answering behavior of respondents on the choice task.

ENHANCE CONJOINT

To explore how the calibration questions enhance the conjoint exercise that follow, we consider three different mechanisms:

- Simply asking the questions helps respondents to recall their most recent shopping trip, which results in more reliable answers.
- Using these questions as covariates improves the Bayesian estimation of the part-worth utilities and results in better "Share of Choice" estimations, better hit rates, and less error.
- The three roles may provide further insight when using them as segmentation variables in the choice simulator.

All nine empirical studies were analyzed using the same settings to avoid methodological bias. We used Sawtooth Software CBC/HB (190,000 burn-in-draws, write out 1,000 draws by using every tenth draw). For SoC simulation, we used the average over these 1,000 draws, as well as the Sawtooth Software default settings for prior variance and degrees of freedom (1.0/5), with an acceptance rate of 30%. For the comparisons, we used three different estimations for the sample split cells with "Behavioral Calibration Questions":

- Standard HB estimation,
- HB with the nine binary questions as covariates, and
- Ensemble of nine estimation runs with one of the questions used as a covariate in each run.

One of the great achievements of machine learning is certainly the use of ensembles. An ensemble approach generates multiple diverse models, include HB estimations with different covariates as in this study. First, we can make predictions with each of the specific HB models individually. Due to the different covariates, these models are diverse in the sense that each provides different predictions and has its own unique strengths and weaknesses. For the ensemble approach, we take the nine different models and blend the SoC predictions to reduce bias from the individual models, thereby generating more robust and accurate predictions.

SHORT REMINDER: HOW WE MEASURE THE VALIDITY OF CONJOINT STUDIES

Before describing the results of the different approaches, we would like to review how the measures of validity are computed.

The standard approach is to use one predefined choice task not used for the estimation process as a "holdout task." This task is then simulated and the MAE (mean absolute error) or the MSE (mean square error) for the whole sample is calculated taking into account the number of concepts in the task.



The alternative approach is to individually simulate the "holdout task" for each respondent and match it with his or her actual answer to this task during the interview.

Alternative Solution:	Indiv Hit R	idual ates			
Choice Task	Estim part	ation of worths	 ;	Simulation o	of k
Choice Task	(Ut	ilities)		Ţ	
Choice Task		Ideal world	Concept 1	Concept 2	Concept 3
Choice Task		Concept 1	100	0	0
Choice Task		Concept 2	0	100	0
Haldard Trak	-	Concept 3	0	0	100
Holdout task			100	100	100

If real market data (e.g., market shares) are available, the root mean squared error (RMSE) is used.

EMPIRICAL RESULTS

The nine studies analyzed confirm our hypotheses: the "Behavioral Calibration Questions" are effective, and hit rates can be significantly increased by asking these questions up front, even if they are not used in the HB estimation. Using the "Behavioral Calibration Questions" as covariates in the HB estimation further improves hit rates. Finally, an ensemble of part-worth utilities from nine estimations based on one calibration question as covariate in each run results in slightly higher hit rates compared with a single HB run that use all nine questions as covariates. Within a category, the more specific the three different roles of our behavioral calibration questions are, the more the hit rates can be improved by using this additional information in the estimation. For example, in the "NECA" category, where numerous consumers compare prices and search for new innovative products, the hit rate could be improved from 43.5% to 53.5%.

Behavioral Calibration Questions									
Hitrate in %	Chance-Rate	not shown	shown	used as covariate	Ensemble				
ADW	11,11	36,50	41,60	41,90	43,20				
Construction adhesives	20,00	53,90	55,30	55,60	57,10				
Cough Drops	7,69	32,40	39,10	40,20	41,90				
Edible oil	14,29	41,20	49,30	51,10	53,20				
NECA	16,67	43,50	52,40	52,80	53,50				
Hair Shampoo	7,69	30,90	32,10	33,00	33,40				
Potato Chips	16,67	47,10	52,40	52,60	52,90				
Laundry Detergent	7,69	31,80	36,20	37,20	37,80				
Super Glue	7,69	34,20	38,70	39,10	39,60				

Out-of-sample calculations were done by splitting the samples into estimation and validation samples (80% and 20%, respectively).

Only the "super glue" study design consisted of four sample splits, such that the possibility of using validation samples and estimation samples was built into the design. In this study, the separate validation samples each had 250 respondents answering or not answering the Behavioral Calibration Questions, whereas the estimation samples had 500 respondents each.

As we could not calculate part-worth estimates for out-of-sample tests, and therefore could not simulate preference shares in the traditional way, we used logCounts (described in Johnson, Orme, Pinnell 2006). The conclusion is roughly the same as that for the above-mentioned hit rates: almost all studies have better RMSE values when "Behavioral Calibration Questions" were implemented. Only the "Shampoo" study seemed to not benefit from use of the "Behavioral Calibration Questions," but the framing did not harm the results.

		wit	hin-sample_	out-of-sample				
RMSE	not shown	shown	nown used as covariate Er		not shown	shown	used as cova Ensemble	
ADW	2.12	2,01	1,97	1.06	2,67	2,48	2,31	2,26
Construction adhesives	1,74	1,69	1,66	1,61	2,19	1.98	1,89	1,14
Cough Drops	2,51	2,43	2,41	2,35	3.21	3,17	2,94	2,89
edible oil	2,45	2,42	2,40	2,39	3,39	3,25	3.11	\$,06
NECA	2,32	2,62	2,58	2,57	3,94	3.37	2,89	2,93
Hair Skampoo	1.21	3,23	0.22	1,20	4,63	4.65	4.71	4,11
Potate Chips	2.16	2,05	2,01	1,95	3.12	2,95	2,73	2,67
Laundry Detergent	2,38	2,19	2,34	1,99	2,99	2,74	2.54	2,44
Super Glue	1.84	1.79	1,67	L66	3.87	2.56	2.17	7.05

Only two of our nine studies have reliable, "real" market shares information and can therefore be compared against them. In the "super glue" case, we estimated separate models for the validation and estimation splits and compared them with the RMSE measure. The results support the same deductions as the above comparisons: simply asking the "Behavioral Calibration Questions" improves the predictions. The inclusion of these questions as covariates or in an ensemble approach further improves the "Share of Choice" simulations.

		Share of Choice - Market Shares								
	RMSE	not shown shown		used as covariate Ensemble						
Ca	onstruction adhesives	5,68	5,21	5,19	4,96					
	NECA	10,23	9,63	9,60	9,38					
	Super Glue	8,36	7,56	7,47	7,18					

USE BEHAVIORAL CALIBRATION AS SEGMENTATION

Our third approach in using the "Behavioral Calibration Questions" is based on the three "Roles." For each, we calculated a filter variable based on the three questions to derive specific "Share of Choice" values for the splits.

The following example shows different price elasticities for one SKU in the "edible oil" study. Again, it is clear that asking the "Behavioral Calibration Questions" results in different elasticities:



The different elasticities correspond with our expectations regarding the role of price, in that price-sensitive buyers with brand-switching behavior (i.e., respondents who switch to a different brand when price increases) have higher elasticities:

Role of Price									
R -		-0-	-	-					
6.70-	-5,10-	5.10	4.50	-8					
	122.00	63.49	£3.99	64.49					



Innovation seekers are less price sensitive. Simply exposing respondents to the "Behavioral Calibration Questions" results in different elasticities ("edible oil" study):



For a more detailed inspection of these effects, we calculated the arc-elasticities of demand for the different segments. The differences between the segments provide detailed insights into the influence of consumer behavior on price and can be leveraged for more insightful recommendations to clients.

	1	ARC - Elastic	ities of Dema	nd	
	€	2,49-2,99	2,99-3,49	3,49-3,99	3,99-4,49
	yes	-1,00	-1,44	-2,14	-1,00
Role of Brand	no	-0,94	-1,19	-0,61	-0,53
	all	-0,94	-1,22	-0,78	-0,58
Role of Price	yes	-1,57	0,00	-0,88	-1,16
	no	-0,89	-1,33	-0,77	-0,53
	all	-0,94	-1,22	-0,78	-0,58
Polo of Innovation	yes	-1,29	0,00	-0,52	0,00
KOLE OF ITTOVALION	no	-0,89	-1,41	-0,83	-0,69
	all	-0,94	-1,22	-0,78	-0,58
Behavioral	shown	-0,94	-1,22	-0,78	-0,58
Calibration	not <u>shown</u>	-0,26	-0,61	-0,49	-0,42

FINDINGS

Based on our nine empirical studies, we can conclude that the "Behavioral Calibration Questions" represent a useful extension to DCM exercises. Our findings suggest that all three hypotheses may be verified. The "Behavioral Calibration Questions" help the respondents to recall their most recent shopping trip in a particular category and thereby positively influence answering behavior in the ensuing conjoint model. The data-generation process comes closer to representing a real shopping trip. Using the questions as covariates can also help improve the estimation results, rendering them more meaningful for simulations. The use of nine different estimations based on the "Behavioral Calibration Questions" in an ensemble approach slightly improves the results and always performs slightly better than a single estimation with nine covariates. Due to the modest improvement, one should decide if the additional effort required by this approach is justified. Using the "Behavioral Calibration Questions" as filter variables provides more detailed insight into the data structure and helps to improve recommendations for clients.

Consequently, it seems that further investing in these additional questions is worthwhile to improve our conjoint models.

FUTURE RESEARCH

The nine "Behavioral Calibration Questions" are a good starting point for further developments. To take advantage of such a framing exercise, the "Behavioral Calibration Questions" could be extended to include more than the three roles. For instance, three additional semantic differentials about the importance of features could be added to generate a fourth role. More specific wording for different categories should also be developed and validated.

From a more methodological point of view, a next step could be the use of the questions as an input for a Bayesian variable selection model to improve part-worth estimates.

Benchmarks should be established by building a database of Category Behavioral Calibration results to position tested concepts in commercial studies.



Peter Kurz



Stefan Binner

REFERENCES

- Allenby, G.M.; Rossi, P.E. (2006): Hierarchical Bayes Models, in: Grover, R.; Vriens, M. (Eds.): *The Handbook of Marketing Research: Uses, Misuses, and Future Advances*, 418–440, SAGE Publications Inc., Thousand Oaks.
- **Binner, S. (2006):** Do Individual Hit Rates Matter at All? (Design & Innovations Conference München 2006).
- Hein, M.; Kurz, P.; Steiner, W. (2013): Limits for Parameter Estimation in Choice-Based Conjoint Analysis: A Simulation Study, European Conference on Data Analysis 2013.
- Hein, M., Kurz, P., Steiner, W. (2019): Analyzing the Capabilities of the HB Logit Model for Choice-Based Conjoint Analysis: A Simulation Study. In: Journal of Business Economics (forthcoming).
- Johnson, R., Orme, B., Pinnell, J. (2006): Simulating Market Preference with Build Your Own Data. Proceedings of the 2006 Sawtooth Software Conference.
- Kurz, P; Binner, S. (2011): Added Value through Covariates in HB Modeling?, Proceedings of the 2011 Sawtooth Software Conference.
- Liakhovitski, D.; Shmulyian, F. (2011): Covariates in Discrete Choice Models: Are They Worth the Trouble? ART Form Presentation.
- **Research International. (2010):** Using the landscape questions in pricing research (promotional material) RI Hamburg.
- Sentis, K. and Li, L. (2001): One Size Fits All or Custom Tailored: Which HB Fits Better? Proceedings of the 2001 Sawtooth Software Conference.
- Sentis, K.; Geller, V. (2011): The Impact of Covariates on HB Estimates, Proceedings of the 2011 Sawtooth Software Conference.