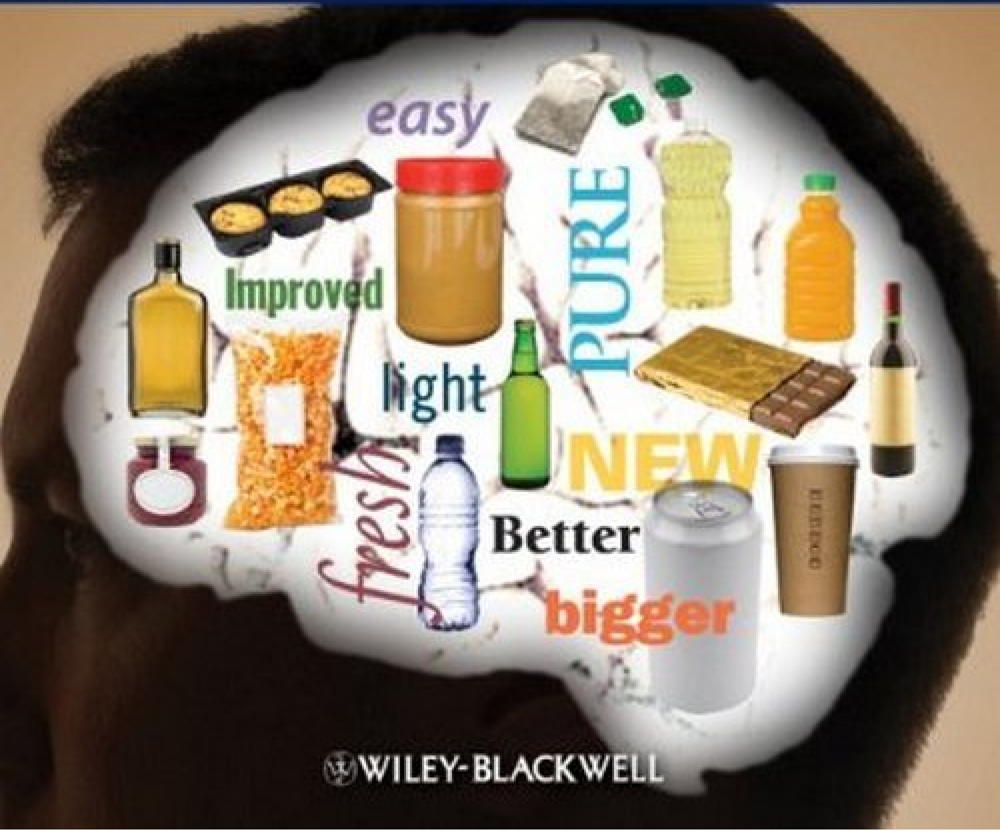
Packaging Research in Food Product Design and Development

Howard Moskowitz, Michele Reisner,
John Ben Lawlor and Rosires Deliza



WILEY-BLACKWELL

Packaging Research in Food Product Design and Development

Packaging Research in Food Product Design and Development Howard R. Moskowitz, Michele Reisner, John Ben Lawlor and Rosires Deliza
 ISBN: 978-0-813-81222-9

Packaging Research in Food Product Design and Development

Howard R. Moskowitz, Ph.D.
President and Founder
Moskowitz Jacobs Inc.
1025 Westchester Avenue, 4th Floor
White Plains, New York 10604 USA
Phone: 914-421-7400
E-mail: mjihrm@sprynet.com

Michele Reisner
Vice President of Research
Moskowitz Jacobs Inc.
1025 Westchester Avenue, 4th Floor
White Plains, New York 10604 USA

John Ben Lawlor, Ph.D.
GlaxoSmithKline
Nutritional Healthcare
Coleford, GL16 8JB, UK

Rosires Deliza, Ph.D.
Senior Research Scientist
Embrapa Food Technology
Av. Das Americas, 29.501-Guardaratiba
23022–470 Rio de Janeiro, RJ-Brazil

WILEY-BLACKWELL
A John Wiley & Sons, Ltd., Publication

Edition first published 2009

Blackwell Publishing was acquired by John Wiley & Sons in February 2007. Blackwell's publishing program has been merged with Wiley's global Scientific, Technical, and Medical business to form Wiley-Blackwell.

Editorial Office
2121 State Avenue, Ames, Iowa 50014-8300, USA

For details of our global editorial offices, for customer services, and for information about how to apply for permission to reuse the copyright material in this book, please see our website at www.wiley.com/wiley-blackwell.

Library of Congress Cataloging-in-Publication Data
Moskowitz, Howard R.
Packaging research in food production design and development / Howard R. Moskowitz, Michele Reisner, John Ben Lawlor.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-8138-1222-9 (hardback : alk. paper)
1. Food–Packaging. 2. Packaging–Research. I. Reisner, Michele. II. Lawlor, John Ben. III. Deliza, Rosires. IV. Title.
TP374.M67 2009
664′.09–dc22
2009009724

A catalog record for this book is available from the U.S. Library of Congress.

Set in 10 on 12 pt Times by SNP Best-set Typesetter Ltd., Hong Kong
Printed in Singapore

1 2009

Dedications

Writing a dedication to a book means pouring out one's heart to those who have inspired, helped, and even on the unwelcome occasion, critiqued the work. It is with this in mind that I would like to dedicate the book to a number of people who have given me the inspiration to write it. First, of course, is my late, beloved, occasionally feared, professor of psychophysics, S.S. (Smitty) Stevens, The Professor of Psychophysics at Harvard University. Smitty couldn't have known it, but it was the years of graduate studenthood in Harvard's Laboratory of Psychophysics 40+ years ago that would prepare me. Second, and perhaps to his surprise, is Emeritus Professor of Psychophysics, Eugene H. Galanter of Columbia University, whose own work in psychophysics gave me the courage to do what you see and read here—apply psychophysical thinking to the new world of packages. And finally, to my wife Arlene, and to my sons, Daniel and David. You, my family, are the reason why I still go on creating.

Howard Moskowitz

I want to thank four people that have blessed and shaped my life. Between these lines and words are my immeasurable gratitude and love.

To my beautiful daughter, Casey Reisner. The joy of my life. When did the hourglass turn and she become such a strong, wise young woman? I rely on her advice as she inspires me to challenge myself and to exceed my expectations. She motivated and kept me focused while working on this book. I am proud to call her my daughter and blessed to have her as my friend.

To my mom, Kay Mayville, my role model. She taught me through example that I could attain anything I set my mind on through hard work and perseverance. Her unconditional love, praise, and strength throughout my life are priceless gifts to me. I wouldn't be standing where I am today without her.

To my dad, Paul Mayville. With a guiding and loving touch he taught me at an early age there was a big world out there to explore and that I should leave my footprint. This realization took me on many paths that I may not have walked. Always a king in my eyes.

And in loving memory of you, my beloved husband, Bob Reisner. You believed in me, encouraged me and your kind and gentle love changed my life. I carry you always in my heart.

Michele Reisner

I would like to dedicate this book to my parents, Pat and Benny, and to the rest of my family—thank you for always believing in me. I would also like to thank Rachel Katz and Barbara Itty. Thank you, Howard—I would never have made it this far without you, that's for sure. Finally to Mimi Moustey for being there for me. Merci, Mimi.

John Ben Lawlor

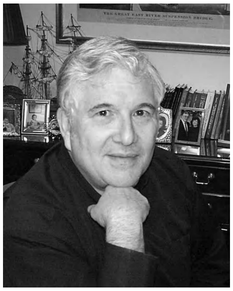
My thanks to the authors, especially to Dr. Howard Moskowitz, for giving me the opportunity of sharing this journey. I enjoyed working with him and learned a lot during our work.

To Amauri and Felipe—my husband and my son—the loves of my life!

Rosires Deliza

Contents

Authors



Howard R. Moskowitz, Ph.D., Moskowitz Jacobs Inc.

Dr. Howard R. Moskowitz is President and CEO of Moskowitz Jacobs Inc., a firm he founded in 1981. He is both a well-known experimental psychologist in the field of psychophysics and an inventor of world-class market research technology. Widely published in the scientific press, Dr. Moskowitz is known worldwide as the leading thinker and creator of advanced research technology in the area of new product and concept development. His extensive speaking engagements span both scientific and market research conferences, as well as guest lectures at leading business schools. He lectures in the U.S. and abroad and serves on the editorial board of major journals.

Dr. Moskowitz has been the recipient of numerous awards, including the David R. Peryam lifetime achievement award by the American Society for Testing and Materials, the Charles Coolidge Parlin Award from the American Marketing Association, considered to be the "Nobel Prize" of market research, for his lifetime contributions to the field, ranging from product work to the optimization of consumer concept and package designs and the 2006 ARF Innovation Award for the development of the most innovative research idea. From November 2004 to November 2006, Dr. Moskowitz appeared weekly as the "Food Doctor" on ABC News NOW, where he anchored a 10-minute spot, featuring food and beverage entrepreneurs. He is the co-author of the recently published book *Selling Blue Elephants: How to Make Great Products That People Want Even Before They Know They Want Them* (Wharton School Publishing). This book summarizes new approaches and applications that are implemented in MJI's commercially available software (IdeaMap®). A worldwide best seller, the book is currently printed in 15 editions.

Michele Reisner, Moskowitz Jacobs Inc.

Michele Reisner is Vice President of Research at MJI and has more than 18 years of sensory experience concentrating heavily on food and beverage optimization.

A member of the Institute of Food Science and Technology, Ms. Reisner is a contributing writer for several food magazines such as *Food Product Design* and *Bakers Journal*. She has been published in *Food Quality and Preference* and the *Journal of Food Technology*. She is a member of Who's Who of Executives Across America.

Her degree from the Culinary Institute of America with 15 years of experience in the field of culinary arts broadens her appreciation of the issues and challenges facing the food and beverage industry.

John Ben Lawlor, GlaxoSmithKline, UK

John Ben Lawlor is a Senior Scientist at GSK Nutritional Healthcare working on sensory and consumer science research and development activities. Previously he worked at University College Cork, Ireland, as a post-doctorate researcher, at PepsiCo Ireland, at CESG France, and at the University of Manitoba, Canada. He trained as a food scientist at University College Dublin, Ireland, and graduated in 1995. He received his M.Sc. in nutrition in 1997, and his Ph.D. in Sensory and Consumer Science in 2002, from University College Cork, Ireland. His research interest is consumer psychology. He is a member of the Institute of Food Science and Technology.

Rosires Deliza, Embrapa Food Technology, Brazil

Rosires Deliza is a Senior Scientist at the Brazilian Agricultural Research Corporation (Embrapa) since 1990, working on sensory evaluation and consumer science. She was trained as a Food Engineer and has a Food Science M.Sc. degree both from the State University of Campinas—SP. She received her Ph.D. in Food Science from the University of Reading, UK, in 1996, and the experimental work was carried out at the Institute of Food Research, IFR—Reading. Her research interest is on consumer expectations of product performance and how these expectations influence attitude toward products, including their sensory properties.

She teaches graduate consumer science courses at the Federal Rural University of Rio de Janeiro, and sensory evaluation at the Chemistry Institute of the Federal University of Rio de Janeiro. During her research career, she has authored or co-authored more than 100 original research papers, book chapters, review articles, and technical publications. She was a Member of the Board of Directors of the Brazilian Society of Food Science and Technology from 2005–2008. She received the Scientist of Rio de Janeiro 2007 Award, which is offered to the state's outstanding scientists.

Preface

Many of us, such as the four writers whose words you will read, have worked in the field of product, concept, and package research for many years. But how does someone in business go about the task of assessing, optimizing, or perhaps even commenting on a package? For product work, there is no doubt that science and scientific methods play a part. You need only look at the large business and scientific literature that deals with making better products, and you will quickly realize just how serious research can be in the quest for better foods and drinks. Try doing the same search, but this time for "concepts" or blueprints about how to make a food (What should it contain? How should the advertising be phrased?), and you will find far less. Now do the same search, but this time search for package design. Certainly you will find articles, not many books, and even some science. Yet, to a great degree, what you will find are methods to evaluate packages but really not very much of a well-developed science.

As scientists, business people, writers, and researchers working with people, we noticed that when it came to package design, much of what we were told was in the form of "best-practices," albeit with little justification. That is, there were no really solid papers in the scientific literature, and also no solid sources of data itself to guide us as to what to do. We were on somewhat shaky grounds. We looked around, but the only guidance that seemed to come straight to us was informal, fairly discursive treatments of packages in the so-called "trade magazine" world. There was no science and certainly no good body of knowledge to speak of.

Armed with that insight, we began to formulate what would be the best way to create the rules for package design. We decided that we would be best off developing databases about how people respond to package design. We didn't want to create a single study, because frankly single studies don't go anywhere. They get buried in the literature, generally forgotten unless they open up a "hot area" and are otherwise quite unsatisfactory. Thus was born the idea of a series of experiments over the past few years, interesting to us and hopefully to you, dealing with different foods, yet commonly known to the consumer, and answering different fundamental questions. What you will read is the results of those efforts, which happily and admittedly with some trepidation, we call our exploration in this world of package design.

Design—The Smile in the Mind

We begin with a catchy phrase, courtesy of Johannes Hartmann of Unilever, in Rotterdam. When we discussed the nature of the book with Johannes, he brought up the interesting conundrum that we recognized we were going to face. Design is a big world, a world filled with art, a world filled with points of view (some very strongly held), and mostly a world that has not welcomed science, at least not historically.

Johannes' notion of "the smile in the mind" gets at the very elusive nature of design research, and gives us a way to ground our efforts. Rather than trying to understand the soul of the designer in the food industry, we decided to limit our efforts to scientific studies of how people responded to systematically varied stimuli. To understand and formalize "art" may be impossible and is certainly not something that we wanted to try. Yet, to formalize principles, to identify patterns in data, to provide some glimmer of a reality in people's responses to design factors, ah, well, that was reasonable and safe territory.

We recognized almost from the start that we were about to create a new discipline, akin to the creation of a discipline for "writing concepts" (see Moskowitz, Porretta, and Silcher, 2005). In that earlier book, companion to this book on package design, we faced some of the same issues. Concept development for the food industry is often considered an "art," something that cannot be easily quantified. Of course there are well accepted "concept tests" done by market researchers, and there is little in the way of overt hostility to the testing process. Yet, scratch any advertising agency writer responsible for creating concepts and you will find that same feeling, of "art" deep inside, of the same "smile in the mind." Nonetheless, we managed to create a science or at least a systematized body of data for concepts. There is every reason to believe that by following the systematic approach we may be able to quantify some of that "smile in the mind."

With these words of background, and with our intellectual souls bared, we invite you to join us on this journey into the world of graphics design for packages. You will see a number of different topics addressed for the world of food and drink. We hope you enjoy the reading as much as we enjoyed the writing.

Acknowledgments

Like any type of production, there are people behind the scenes that you don't hear about, however, their talents are instrumental and without them, the curtain, so to speak, wouldn't go up.

We would like to thank Barbara Itty, who helped to design many of the experiments in the chapters you are about to read. We would present Barbara with an idea and she would tirelessly search the Internet for photos and research on the topic at hand.

To David Moskowitz, thank you for your creative eye and style. You were given the nearly impossible task of artistically creating designs from photos and snippets of information we gathered and making them into a reality for us to use in our experiments.

To Paolo Gentile, thank you for your expertise in creating complex graphs and tables in each of the chapters.

Our special thanks to Linda Lieberman, our editorial assistant. It wasn't always easy keeping four authors with their unique writing personalities organized and on track meeting deadlines. Linda, you have been our "special sauce" who held the project together as we moved from data to manuscript, from manuscript to publication.

Our thanks also to Alexis Fried of Johnson & Johnson for her memories of the FunFeast project and to Alex Gofman for his incisive chapter on eye tracking.

We also wish to thank our colleagues at Tobii, Nick Harmsen and Johanna Fyrebörk, for their inspiration, guidance, and collegial help as we got to experiment with and learn about the Tobii system.

And finally to our associates at Wiley-Blackwell, Mark Barrett and Susan Engelken: You both kept us on our toes, gave us the encouragement when we needed it most, and of course kept us to our word, our vision, and our contract! Thank you.

Packaging Research in Food Product Design and Development

Packaging Research in Food Product Design and Development Howard R. Moskowitz, Michele Reisner, John Ben Lawlor and Rosires Deliza
 ISBN: 978-0-813-81222-9

Part I

Methods, Materials, and Mind-Sets

Packaging Research in Food Product Design and Development Howard R. Moskowitz, Michele Reisner, John Ben Lawlor and Rosires Deliza
 ISBN: 978-0-813-81222-9

Chapter 1

A Practitioner's Guide to Research, or What You Should Know

Introduction

When we began this book on consumer response to package design we had hoped to "dive right in" to case histories, to illustrate some of the approaches and, of course, the interesting byways that one could follow. After our initial draft was about half completed, however, we realized that there was a lot of "research savvy" that we wanted to impart to our readers. It wasn't sufficient just to provide you with topics, case histories, insights, and a global view of the problems in package design. It became increasingly clearer that we needed to share with you research approaches, ideas, as well as ways of thinking. By the term "research approach," we don't mean the dry, boring, often brain-twisting minutiae that research could become, but rather the excitement that comes with studying consumer responses.

With that caveat in mind, we now invite you to join us for our trip into some of the interesting worlds of research, provided from the perspective of active researchers who are involved in the practical applications of design, as well as in the creation of a new science.

Time Travel—Let's See Where Some of This Thinking Started

Let's travel back in time to the 1950s, to a laboratory at Harvard University, and visit with some researchers who worked about 60 years ago. As we walk in the door, we enter a large, cavernous building called Memorial Hall, situated a little north of Harvard Yard, at the intersection of Oxford Street and Kirkland Street. We pause for a moment, look at this monument by Harvard graduates to the dead of the Civil War, and then go into a connected warren of nooks and crannies. Along one wall we see a sign: "Laboratory of Psychophysics." We walk in and watch an animated discussion taking place between a rather dour, yet impressive man, whom we later meet, and a group of excited students.

During the course of this animated conversation, we manage to see this professor, who we discover to be a man named Smitty (S.S.) Stevens, and a group of graduate students. The animated conversation is about the relation between the size of an object and how big it seems to the viewer. Another student talks, not about squares and circles, but about finding the relation between the heaviness of some blocks of metal and the grams.

Here, we're in the world of psychophysics. As we mentioned before, psychophysics is that arcane branch of experimental psychology devoted to understanding how the mind transforms the physical stimuli of nature, squares, circles, weights, tones, sugar solutions, the smell of alcohol, and the like into private sensory experience. The discussion is a bit amazing, as ideas fly back and forth.

Perhaps the most important thing to stand out for us, unexpected visitors, is the sense (no pun intended) that perhaps these people are "onto something." Maybe there *is* a lawful relation between physical magnitude as we measure it in the laboratory and the sensory experience that comes about from those stimuli.

Before we continue the story, we should say that we didn't happen in on this meeting by accident. No. We're packaging engineers now, not dilettantes. And our quest? That's simple. We have heard from our industry meetings, or perhaps have read in one of the trade magazines, that some interesting work is going on at Harvard. We've heard mention of a psychophysics laboratory, a place where researchers are trying to relate the features of stimuli to perceptions. We were wondering whether this new area of psychology could help us understand how our customers respond to our packaging changes. Specifically, we know that there are issues about size and weights. Can psychophysics help us here? And the questions we could ask go on and on and on.

 ISBN: 978-0-813-81222-9

Well, the story could go on. As we listen in, not exactly comprehending everything that's being said, we get the gist of what's going on, and that satisfies us. We learn that there's a new movement afoot, whose aim is to bring real, useful "numbers to perception." The group inside Memorial Hall is animated while they talk about applying this new branch of psychology to help researchers understand perception, such as how sweet the taste of Coca Cola is. Another person is talking about the perception of heaviness. That draws our attention, and we drift over to where people are talking about the perception of color, and even such perception as "numerousness" or "density," as well as the brightness of a color.

All in all, these psychophysicists are speaking our language. As we emerge from this reverie, this "dream trip to the past," we are struck by the excitement around us. We see, or rather we feel, a sense that here is a branch of psychology that someday we will use, that at some future date will change the way we think about the stimuli with which we work.

And then, as quickly as we came, we move from the past of Memorial Hall and its now departed psychophysicists to the reality of today's packaging laboratory. We know now that there could be a science of subjective perception, and are certain that science could have a great effect on us, if only we would follow the systematic approach.

What's Important? Answering the "Major" (First-Order) Questions

We begin with a discussion of the "major" questions. If the term "major" is a bit strange, then let us suggest the term "first-order" questions. By this, we mean the questions that are really worth answering. Of course, in business there are lots of tactical issues that arise daily that must be answered and whose answers, once provided, disappear. Both the question and the answer disappear because other than the momentary need to know a specific "factoid," there wasn't much to the research at all. The researcher did not become truly better at his craft, and the client wasn't served beyond his simply finding the answer to the momentary question.

We authors ought to approach the issue of "major" questions by first defining what's valuable to know. As you will see again and again, we believe very strongly in the value of learning the "relations between variables." In simpler terms, we want to learn about "how the world works." For example, we can easily measure the difference between two packages. It's a lot more interesting to know how to engineer the packages so that they look similar, or if we wish, look different. We want to discover the rules of perception as nature lays them out.

We can trace this desire to learn the rules from our history as experimenters, with a specialty in psychophysics. Psychophysics is the branch of experimental psychology that deals with the relation between physical stimuli and subjective responses. Originally a branch of physiology begun by the German physiologist E.H. Weber in the 1830s (Fancher, 1996), psychophysics grew to a science under a host of experimental psychologists during the twentieth century. Most notable was S.S. Stevens of Harvard University, that gentleman we just described above in our mind's journey. The important thing to keep in mind is that the psychophysicist looks for quantitative relations between the variables that they can control and the subjective responses. That is, the psychophysicist tries to create an orderly representation of the world.

In the world of package design, there are many first-order questions including the types of package features that drive acceptability or the existence of groups of people with different mind-sets. We will pose a number of different first-order questions. During the journey, you will see traces of and echoes from psychophysics as we wrestle with "rules" that the package engineer can modify, and in turn what the consumer respondent perceives.

As you, the reader, delve into what we present, you might want to ask yourself about the topics you read and the ideas that these topics spark. The goal of such an exercise is to begin to recognize these major ideas—these first-order problems—and to distinguish them from second order, less important topics. If the idea opens up new possibilities to you, if you feel that you can answer questions that you could not answer before, or even if you feel that intuitive sense that you're onto something here, then chances are you are facing a first-order question or a major problem. On the other hand, if after reading something you feel that what you read is just a rehash of old ideas, that you sort of "knew it" already, that you are not particularly enriched by what you read, then odds are you have encountered a second order issue or a minor problem, and it's time to move on to something that teaches you.

What a Scientist Looks For

As we go through the different chapters in this book, you will see us following a path. We begin with methods, but quickly veer off into the types of interesting topics that one might ask in a business environment. Our approach begs the question: Just exactly what does a scientist look for when taking such a journey?

It might surprise you to realize that the scientist generally has no master plan or preconceived way that he believes the world should be organized. Many of us are educated by reading about the history of science and culture, where we learn about great organizing principles, or single breakthrough ideas. Newton, of course, comes to mind, but there are many others who had brilliant insights from which modern science grew.

That's the good news. The bad news is that few scientists in their lifetime really know precisely what they are looking for. Few scientists work with a worldview that dictates the next logical experiment, the next key hypothesis to be proved or disproved by the just-right experiment, the so-called "*experimentum crucis.*" The reality is much different, much more ordinary. Scientists, like nonscientists, are on a journey. They are fascinated by what they are doing, just as we are. Look at the life career of a successful scientist or equally the short history of a fascinating project that started and finished in six months. Rarely will you find strict adherence to the path set up at the start. Rather, you will find interesting byways, time taken to digest new observations and new ideas, and time taken to rethink hypotheses.

In the end, scientists are normal, ordinary people leading their lives, looking at the phenomena of nature in a bit more orderly fashion, and trying to make sense out of the data by finding patterns. That's all. Rather than having a grand plan to understand the world, it's rather like the evolving play of a child charmed with a toy, seeing things that were not seen before, cherishing the minutes before he has to come in from play to eat dinner. So, in a sense, the scientists look for reality, just like we all do. The scientist's big difference is perhaps a bit more discipline in one particular area—designing-measuring-recording.

The Role of a Worldview

Quite often, researchers begin by answering questions. Then they get so fascinated with the questions that they slowly transform from novices to serious investigators. This is wonderful and should be welcomed. The passion to learn, to discover, to create, and to answer real problems converts the novice to a professional. At the same time, however, something is missing—a worldview, an organizing principle, a direction. Without that organizing principle, this newly minted professional wanders around like a dilettante, investigating interesting properties of products, publishing here and there, and moving on like a butterfly.

For the researcher to be truly successful and to have an impact, it's important to operate within the limits of a worldview. Worldview, or the more romantic German word *weltanschauung*, refers to a coherent way to organize one's view of reality. Originally a term taken from philosophy to describe the way a philosopher looks at questions, worldviews are important to scientists and even more important to budding researchers. They dictate the way a person goes about asking questions, collecting data, and determining what is a valid answer and what isn't. Armed with a worldview, the researcher has a weapon to fight the unknown and a light by which to see that which is discovered.

A worldview is as important in consumer research as in any other branch of science, and may be even more important when it comes to the science of design. Why? The answer is straightforward. There are no principles of design research. For many years, design has been enshrined as art, perhaps not pure art, but art, nonetheless. Design is a form of industrial or commercial art, but the artistic production values still shine through. It's hard to be a scientist in a world of artists. The worldview gives the hopeful scientist an equal home, not in art, not as a secondary citizen, but as a citizen of an equally valid world.

Testing Isn't Science

Throughout this book we refer to different types of "tests" that we run with packages. It is important to understand the difference between "testing" as a discipline and "science" as a body of knowledge. Often the differences are subtle, especially when we are in the business environment and the testing is both to understand (aka "insights") and to answer specific problems of the very moment that need empirical data from consumers.

The basis of empirical science is, of course, a collection of data or observations that have been categorized. They have been tapped at various levels to identify recurrent patterns or "rules." Science is not simply "good practices" in executing tests, although without such good practices it would be hard for a set of data to generate a science. The data and the uncovered relations in the data must be reliable (reproducible) and valid (must mean something). Reliability and validity are topics unto themselves and lie at the foundations of empirical science. We don't need to go over them in particular detail because we are focusing on other issues.

Good testing is often confused with good science, especially by market researchers and sensory professionals, who are in the business of "measuring and testing." Good testing protocols mean that the tests are laid out, the controls are proper, the statistical analysis is appropriate, and the inferences have real meaning, rather than being just a rehash of the observation. ***Good testing provides the data that a science needs, but good testing is not the science.*** Thus, if this book were to concentrate primarily on methods for testing designs among consumers, then we could not legitimately use this book as a possible foundation of science. Rather, we should call the book "methods for testing." This is not our aim. We are interested here in creating a science, using good tools to understand the relation between features of designs and consumer responses. We are not interested in a "cookbook" of good test methods. And, besides, in the world of design there doesn't seem to be such an accepted cookbook, perhaps because design research has a relatively short history.

On Thinking about Problems, or "What Should I Investigate?"

Finding and Recognizing That First-Order Problem

It may seem a bit strange in a book like this to talk about how to find problems to work on. After all, most people at scientific conferences come to these meetings flushed with their latest scientific triumph, certain that the work they are doing constitutes "the breakthrough" in the field. They are hopeful that others will become as enthusiastic as they appear to be when they present their papers to the crowd of other scientists.

The same scene happens again and again among researchers. It doesn't have to be a convention of scientists. It could be a meeting of researchers in some applied discipline, or a meeting of chefs, and so on. The pattern inevitably repeats, almost predictably, at times all too predictably. The professional about to meet his or her colleagues waxes enthusiastic about a stream of research or whatever he is doing, and from there, sallies forth to meet colleagues.

All that said, just how does the young researcher, the young designer, discover this wellspring of excitement? How does one discover a first-order problem on which to work, for months, for years, perhaps decades? Is there a magic, or does knowledge of the efforts done before by one's professional predecessors some how gel into the perception that "*Aha, here is an important problem on which I should work?*"

In the world of science and research (they're not the same, you now realize), people choose that with which they feel comfortable. What does this mean when we talk about finding a first-order problem to work on? It simply means that not everyone will find the same problem interesting or worth working on. With the realization that scientists and researchers are different, we offer the following suggestions, more as guides than as prescriptions:

1. *True meaning is important.* Think about what the problem really "means." Are you attacking an issue that comes up once in while, or is what you are attacking something that has lots of "legs," which can stand the test of time? You probably cannot answer this question, but you should try very hard to do so. You are going to spend time and money working on the issue you have chosen. It might as well be worthwhile.
2. *Do the "mother test."* Try to describe the problem in as simple terms as possible. Explain the problem as if your mother asked you to describe what you are doing. Avoid jargon. Avoid the tendency to inflate the importance of the problem. Just try to explain what you are doing and why you are doing it. The simpler you make the explanation, the more easily you will "sense" the value of what you are doing. Cut away the jargon, the nonsense, the self-importance that comes from thoroughly confusing your audience. Dig into the heart of the matter in as simple a way as you can. Make nature—the nature that you are studying—as simple as you can without losing the essence of what you are doing.
3. *Love what you do, and do what you love.* Perhaps the most important thing about a first-order problem is whether you have fallen in love with it. Does it excite you? Do you think of all the areas that it can lead you? Does it have "legs" on which you can create other projects or other pieces of research? If you answer "yes" to these questions then the chances are high that you have encountered a first-order problem. You are thinking of where it can take you.

Recognize what you do and come to grips with it If, however, you answered no to the foregoing questions, if the problem was limited and, in fact, a bit pedestrian and boring, then chances are that you have not encountered a first-order problem. There is so much to do in today's business and academic worlds that we often settle for these rather boring, second-order problems, which occupy our time if not our mind. They're safe, limited in time, and require no additional energy to face.

Design and Packaging Are Filled with First-Order Problems

Now that you have heard some of the bad news, let's focus on the good stuff. Over the past decades, design and packaging have been left to two groups of individuals, neither of whom can particularly be said to espouse consumer research. On the design side, we have the package designer, a commercial artist whose job it is to create attractive, "pulling" designs that get the consumer to take the product off the shelf and put it in their "figurative" shopping cart. On the packaging side, we have the packaging designer and engineer, both of whom are trying to figure out how to solve the problem—store the product contained therein, in an efficient, safe, cost-effective manner.

Neither packaging design nor packaging itself has received much attention from consumer research. There are very cogent reasons for this inattention. Some comes from the nature of the designer, others from the nature of the stimulus. We end this section with a discussion and a suggestion about how to find your own first-order problems on which to spend time.

We begin with designers. Designers are artists, who by their very nature, do not like to be judged. "Like" is a bit weak, but an acceptable word that is better than the more accurate word—"hate." Research, in turn, is perceived as judging the artist. Indeed, for many years, research did in fact judge the work of such designer artists. When a package design was tested, either by a focus group or in a larger-scale quantitative project, the underlying goal was to measure "performance," either good or bad. Certainly, no one really emphasized the fact that the research was to assess the performance of the design, using the type of harsh language we just used. Rather, the research was "positioned" to the designer as providing guidance, feedback, and even that most wonderful of all words, "insight." However, there were, at most, a few designs being evaluated and no systematic way of taking the data from the consumer respondents to identify a pattern, a trend in the data that could make the designer smarter. If any patterns were to emerge from the study, then they would emerge from the intelligent observations of the researcher, possibly unplanned, but happy and fortuitous all the same.

The approaches that we espouse look for patterns rather than for right versus wrong. In the long run, designers love this type of information, which shows them "how the world works."

We end this section by discussing package engineers, who work with the other side of their brain, the left or rational side. They want to know what features to put on the package and whether the package will do its job. Of course, there is an aesthetic side to package engineers, but that is not why they are hired. The engineer's job, first and foremost, is to ensure that the package "does the job," maintains the quality and safety of the food and beverage, and by the way, provides the customer with a pleasant experience interacting with the food.

The news is good here as well. We espouse a scientific approach that appeals to the package engineer, who is afforded a new type of consumer data, but in the same general format, namely relations between variables, rather than simply good versus bad.

Addressing Practical Questions

What Do I Do and What Do I Get? The Designer's Question

It may sound a bit silly to ask a question about the number of stimuli, without actually specifying a problem to be answered, or an approach to be followed. However, in the world of applied testing of packages, as well as almost anything else, one of the first questions that is asked has something to do with "what do I have to do to ensure that this test works?" When you talk to a researcher, he (we use he instead of they), in turn, will have to talk with someone who must produce the test stimulus. The immediate question that comes from the developer is "How much am I being committed to do?"

In the original days of design research as well as package research, a lot of the focus was on evaluating a few fixed concepts, whether these are package designs of a simple graphic nature, or actual prototypes with specific physical dimensions and features. The labor involved in such activities could be considerable. Furthermore, since the efforts were focused on evaluating what would turn out to be "finished stimulus," the question about "how many" was understandable and, in those situations, quite reasonable. Often, the designer would sit in a focus group, perhaps behind the glass mirror and wait casually or perhaps not so casually, for his creation to be praised or, more often than not, torn apart with less than kind words. Such discussions often turn into toxic situations. Research, commissioned to generate insights, actually and inadvertently becomes a "star chamber" where one's creations, and by extension oneself, are put on trial and judged.

Happily, in recent years, some of the movement among the more advanced marketers, designers, and researchers has been toward using research to understand and discover patterns. In these more enlightened situations, the designer is asked to come up with an array of different stimuli, often with some type of underlying structure. The work need not be polished, because the objective is to discern patterns. One gets a lot fewer issues about effort in such collaborative situations, where the outcome is not good/bad but rather "What is nature trying to say?"

The happiest of all situations comes from those newly emerging technologies that put stimuli together on the computer, in two dimensions, or simulated in three dimensions. These technologies take away some of the onerous work that the designer had to do. The enlightened designer who recognizes that these many systematically varied products help uncover patterns doesn't need to ask "How many stimuli are enough?" There is no top number for the designer, for he only does the initial design portion. The computer does the heavy lifting of combining the elements, presenting the results, acquiring data, and then doing the analysis.

What Do I Do and What Do I Get? The Researcher's Question

For researchers, typically, the greater the number of test stimuli, the more solid the answers will be. That is, it is quite difficult to answer any but the simplest problem with one or two stimuli. Certainly, you could present a respondent with one test stimulus and ask "Do you like this or dislike this?" or, perhaps, give the respondent a scale to use. Then, you would average the responses from many individuals and present the results. However, notice that the results themselves are not couched simply in terms of the responses to that one product that you tested. No. When you present that one average, you or the researcher will inevitably be asked "What are the norms behind this number?" or "How shall I interpret what I just got?" At that point, you have to invoke norms, or other numbers, to put the data into context. We know that the foregoing explanation was pretty long winded, but it needed to be said. There is no free lunch in research, when you are testing stimuli to make a decision. If you test one stimulus for a go/no/go decision, you must invoke norms.

Let's move to the world of multiple stimulus testing. Our goal, stated again and again, is to develop or uncover patterns relating to variables. Most of the time, we will deal with systematically varied stimuli, which means that we need to understand how much variation among stimuli we need and, in fact, what the minimal amount of variations or test stimuli is, which will provide the answer we need.

We begin with the fewest number of stimuli—one. There is very little that you can do with one stimulus. Of course, that sad state of affairs does not stop people from feeling that they can learn a lot about patterns by asking many, many questions of this one stimulus, from many people, of course. They then correlate the different attribute ratings. That is, with 100 people rating the one test stimulus on purchase intent and uniqueness, they savor the opportunity to run a statistical correlation between ratings of purchase intent and ratings of uniqueness. The problem is that there is only one stimulus. The researchers are basing their correlations on error, variation among people, for the one stimulus. *There is no pattern to be discovered, at least no pattern pertaining to the test stimulus.*

Quite often, the research involves two stimuli compared directly to each other. There is a sense in the research community that it's okay to test a single stimulus, but that you get a lot more information if you test the stimulus against some type of "benchmark," perhaps a product that performs well in a category. With two stimuli the typical analysis compares one to the other. In this case, the respondent has to choose either a set of attributes (paired comparison method) or he chooses to scales each product and then the researcher does the comparison of the two stimuli afterwards. Again, this paired comparison method does not lead to the discovery of patterns. Of course, the researcher knows how well the stimuli perform and which one is preferred to the other. On the other hand, there is no sense of any pattern emerging in the data, other than the fact that the preferred stimulus overall was preferred on a series of attributes, and not preferred on some other attributes. This is an eminently dissatisfying approach if we are to build a real science, but a popular approach, nonetheless.

When we come to the third approach, multiple stimuli, things become interesting. By the time the researcher works with three or four or, hopefully, many more stimuli, a lot can be learned from the data. Indeed, far more can be learned from one study of eight products than from four studies of two products, or eight studies of one product. When we deal with multiple stimuli, we have a chance to look at relations among the different dependent variables. For example, consider the data in Table 1.1. We have the results for eight test stimuli, rated on interest, uniqueness, fit to a specific end use, and relative frequency of purchase. Each of these attributes is presumed to be independent of each other. What can we learn from the data in Table 1.1, and why could we not learn this from studies with, say, one or two stimuli?

If we look closely at Table 1.1, we see a few things that tell us about the world of packages, even if we don't know what the stimuli are. As we go through these observations, think about what we might say if we were limited to only one or two of these eight stimuli.

Question 1: Do All of the Package Stimuli Score the Same, or Do We See a Range?

If you look at Table 1.1, you will see that the stimuli are ranked in descending order in terms of purchase intent (top-3 box percent, i.e., 7–9 on a 9-point scale). We see a 13-point difference, with stimulus A scoring 60% top-3 box, and stimulus H scoring 47% top-3 box. Rather than discussing whether these percents are significantly different from each other, let's look at the problem from the point of view of what we learn in the "here and now." We see that there is a large difference of 13%. Intuitively, we conclude that whatever we did to the stimulus made a difference in terms of driving people to say that they would buy the product.

Table 1.1 Ratings of eight stimuli on four attributes (interest, uniqueness, fits end use, and "relative" purchase frequency)

Stimulus	Purchase interest % top-3 box (7–9)	Uniqueness	Fits end use	Purchase frequency
A	60	65	55	5.1
B	57	53	49	4.8
C	56	54	50	4.7
D	53	61	47	4.7
E	52	50	48	4.8
F	52	60	47	4.6
G	50	45	43	4.4
H	47	50	43	4.4

Question 2: If There Is a Range of Purchase Intent, Then Where Do the Different Packages Lie on This Range?

By this question, we try to address the issue of whether the stimuli cluster together at the top (i.e., most of them are highly acceptable, with one or two less acceptable), cluster at the bottom (i.e., most of them are less acceptable, with one or two more acceptable), or scatter across the range (i.e., what we do generates a range of effects). Looking at Table 1.1, we see that the eight-package stimuli distribute across the range of acceptance. So, we don't have a situation where we deal with a lot of highly acceptable or a lot of only modestly acceptable stimuli. We also infer that the different variables involved, which we don't know as yet, create a relatively broad range of acceptance. Thus, we know intuitively that probably we could "drive acceptance" by choosing the correct variables to change.

Question 3: Now That We Have Sorted the Eight Stimuli by Purchase Intent, What about the Other Attributes? How Do They Behave?

We know that these eight stimuli are different from each other, because we created them to be different. We also know that we could look at average data from all the respondents who evaluated these stimuli. In Table 1.1, we are looking at average data. So, the variability across the stimuli that we see is due to the stimuli themselves and not due to the differences between people. Researchers often forget this simple, but crucial, fact. It's appropriate to look at relations between attributes if we are dealing with different stimuli, rather than one stimulus rated by different people. We look at average data, so we are really looking at best guesses for the attribute level for each of our eight stimuli. On the other hand, when we look at one stimulus, we really deal with one stimulus only. Our analysis to find patterns is founded on the variability across people, not across stimuli. What would happen if all respondents were identical? In that case, then, we would have no variability at all and nothing to analyze!

Question 4: As Other Attributes Change, How Does Purchase Intent Change?

We are not dealing with causality here. We really don't know what external forces drive purchase intent. We are simply interested in the relation between changes in one attribute (i.e., uniqueness) and changes in another attribute (i.e., purchase intent). The relation can be a straight line, a curve, or no relation at all. Scientists call this analysis R-R or response-response (Moskowitz, 1994). We look at the relation between two variables and try to guess what nature is telling us. Figure 1.1 shows this type of analysis. Keep in mind that Figure 1.1 is idealized.

What Scales Should People Use? The Researcher's Favorite?

There is probably nothing that brings out arguments as much as scales. This behavior itself is fascinating and worthy of study. What is it about the act of measurement that raises people's ire so quickly and so frequently? Are scales really so important in science that they are the *causa belli* of wars between researchers? Indeed, psychophysicists who use the method of magnitude estimation often fight doggedly with other scientists who use so-called category scales, which have a limited number of scale points. Magnitude estimation, in contrast, allows the respondent to assign numbers so that the ratios of the numbers reflect the ratios of perceptions. Whether these numbers are true ratio-scale values or not remains a continuing topic of discussion and argument.

To talk about scales in the foregoing way is to engage in a discussion not necessarily appropriate to this book. What is appropriate is an understanding of the scales, what they ought to do, and what some of the problems are with them.

Paraphrasing the late Gertrude Stein, "Scales are scales are scales." For the most part, the scales that researchers use try to measure the intensity of perception, from low to high. Most scales will do the job adequately. Perhaps one might want to avoid unnecessarily short scales (i.e., 3 points or so). These short scales won't be able to differentiate a group of say five to seven products, all which are noticeably different from each other. There's just not enough room on the scale to show that two products differ especially if every respondent uses the scale identically. In such a case, and with a 3-point scale, if we work with, say, five products, then at least two products that are discernibly different must lie on the same scale point.

Who Should Participate? Does It Really Make a Difference to the Data?

In the 1960s to the 1990s, market researchers spent a great deal of time "validating the interview." When the senior author was introduced to this notion of "valida-

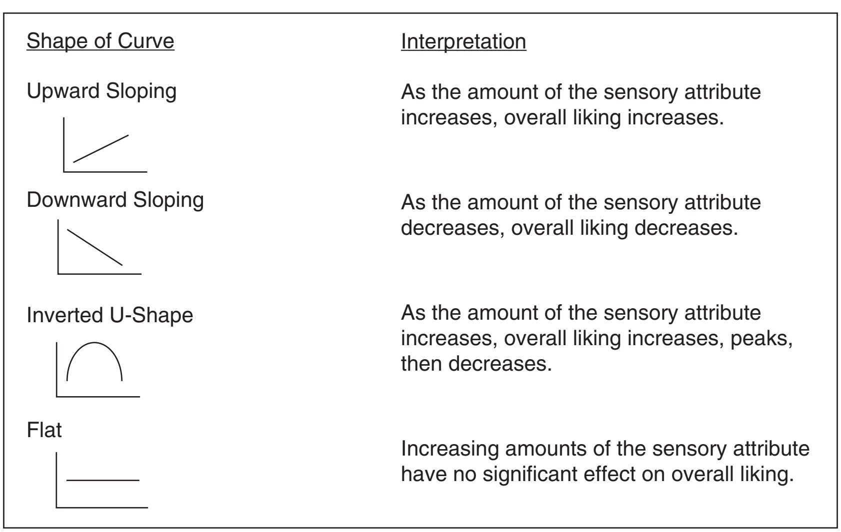


Figure 1.1 R-R Analysis.

tion" after a stint working as a government scientist, the idea seemed a bit far-fetched. Very simply, in the market research community, validating an interview meant a third party calling a respondent in an interview to ensure that the respondent actually participated!

It seemed to the then-young researcher that validation ought to mean something more profound, such as: "Did the market research study lead to significantly more business for the client?" Of course, this improvement in company performance is natural to wish for; otherwise, why bother doing research in the first place?

The real importance of the "validation step" was to ensure that validity of the data itself. In the world of research, validity does mean that the interview is "valid," and with such validity the right person was interviewed, for the appropriate period of time.

This notion of "validity of the interview" did not go away in the 1990s, but got much worse as the technology for interviewing became more sophisticated. In the 1990s, a lot of face-to-face interviewing began to migrate from the central location in a mall out to the Internet. The migration was caused by the "usual suspects"—namely, speed and cost. It was faster to do an Internet interview and, of course, the results were already in tabulated form since the interview was computer-based, and the programs for doing it compiled the data in an analysis-friendly format. What formerly took a week could be done in parallel overnight as the invitation to participate was streamed out by e-mail invitations to thousands of respondents simultaneously. It was also far cheaper to run the interview by machine than by trained interviewer. Machines cost a lot less and could work non-stop, 24 hours a day, around the world.

Over time, the Internet interview itself became the subject of abuse (Couper, 2000; Dillman, 2000; Konstan et al., 2005; Wright, 2006). At first considered to be a novel experience, the Internet interview soon became boring. Hundreds of companies jumped on this cheap, quick bandwagon, offering panels, "do-it-yourself" interviews, and the promise of a nirvana made up equally of inexpensive data and speedy results. This proved to be a fatal combination, as consumers were "oversampled," getting a half dozen or more daily invitations to participate in a panel, give their opinion, even make a living doing such interviews.

In the end, the quality of the Internet data dropped, as more people refused to participate. The issue of interview "validation" again emerged. It was no longer the old question of "Did you really participate in this study?" but rather "Are you a real respondent, or are you one of those hundreds of thousands of people who participate in hundreds of studies a year, in order to make a living, posturing yourself as fulfilling the criteria for the recruitment? Are you one day an older female, and the next day a younger male, just to enter the study, for a chance to win the sweepstakes?"

At the time of this writing (late 2008), various consumer research business professional organizations have focused on the severity of the problem. Yes, it is important to have the correct respondents, because otherwise we don't know whether the data that we get is truly representative of what the "target consumer" actually feels about the test stimulus, whether package, product, concept, etc.

The worldwide consumer research business is moving toward a solution. Although many researchers still use large samples of consumers who volunteer to participate (so-called "opt-in panels"), an increasing number of companies are now offering so-called panels. These panels comprise individuals about whom a lot is known. When a researcher wants to work with a specific type of individual (i.e., a female who buys a certain type of juice), the researcher can work with one of these "sample-providing, panel companies" to identify individuals who are known to fit the criteria. Since the company keeps a detailed record on the panelists based on a questionnaire that the panelists themselves complete, it is straightforward to recruit the appropriate individuals. Of course, these targeted panelists cost more money to recruit, but the researcher can feel assured that the respondent is the appropriate person for the study.

Making Sense of It All—The Art of Analysis

Analyzing Data—Relations between Variables Versus Difference and Error

If scales tend to be a tempest in a teapot, then data analysis tends to be a hurricane in the same teapot. Although most people in design research are not statistically oriented, but rather have some type of visual and artistic orientation, those who work in statistical analysis constitute two radically different groups. It is worthwhile describing them partly because the description is important as an aid to understanding, and partly because describing them is just good fun.

Unlike Julius Caesar, who in his history of the Gallic Wars said, "All of Gaul is divided into three parts"

(*Omnis Gallia est divisa in tres partes*), we might say that statistics data analysis is divided into two parts, often at war with each other. One part looks for differences, while the other part looks for patterns. These two mind-set segments walk around with different "maps of reality" in their heads.

Those who look for differences conceive of nature as a series of centroids, center points, around which "stuff" floats, at closer and farther points, relative to the centroid, which is the mean or the "real essence" of what is being measured. Around this mean float real cases, real instantiations. Thus, we might have a container of a different shape that is perceived to have an average level of "liking." This average level is the mean. Around this mean float the different cases that make up the mean. To this first segment, the world comprises averages and deviations from the average. To these people, science looks for the "average" and measures the variability or deviation around the average as a matter of course. This segment loves to report means, measures of central tendency. The segment is a real stickler when it comes to the numbers. The tightness of measurement, the accuracy is all-important. This segment does not look for patterns. Rather, it looks for consistency, reliability, and the "true value" of what it is measuring.

Let's now move to the other segment, to those who look for patterns in data. These individuals look at the stimuli and try to connect the dots. They don't seek statistical significance as much as they look for some clue about how the world works. These individuals are scientific detectives. They're not necessarily better than the first segment, which was a stickler on significance, representation, and solidity of data. Rather, this second segment is involved with nature in a more profound way. They want to know what will happen, how nature is constructed, etc. We four authors of this book count ourselves as members of this second segment. We don't disparage the efforts for precision. Rather, we are far more excited by patterns in the data that show us the elegance of design and packaging as revealed in the response of real consumers, in structured, interpretable, insight-creating experiments.

Summing Up

In this chapter, we have tried to describe the mind of a scientist approaching the world, with the aim of helping you to understand what he thinks, or at least ought to think. In some cases, we have abandoned the natural gravitas of authors and just "said it like it is." We are excited about design and package research. Otherwise, of course, we wouldn't have written this book. At the same time, we wanted to put our imprint on the science that is being created now. It's a wonderful time to do research, to get involved in a field that has its origins, on the one hand, in psychophysics and, on the other, in business, with findings and insights just waiting for the right investigator to happen over them and reveal them to the world. So, in other words, let's move on to our exciting journey into substantive topics, now that we have shared with you our beliefs, biases, and vision.

References

Couper, M.P. (2000) Review: Web surveys: A review of issues and approaches. Public Opinion Quarterly, 64, 464–494.

Dillman, D.A. (2000) Mail and Internet Surveys: The Tailored Design Method. John Wiley & Sons, NY.

Fancher, R.E. (1996) Pioneers of Psychology. W.W. Norton & Company, London and New York.

Konstan, J.A., Rosser, B.R.S., Ross, M.W, Stanton, J., and Edwards, W.M. (2005) The story of subject naught: A cautionary but optimistic tale of internet survey research. Journal of Computer-Mediated Communication, 10. Retrieved December 5, 2008 from http://www3.interscience.wiley.com/cgi-bin/fulltext/120837940/HTMLSTART

Moskowitz, H.R. (1994) Product testing 1: The distinction between S-R (stimulus-response) and R-R (response-response) models. Journal of Sensory Studies, 9, 313–321.

Wright, K.B. (2006) Researching internet-based populations: advantages and disadvantages of online survey research, online questionnaire authoring software packages and web survey services. Journal of Computer-Mediated Communication, 10. Retrieved December 5, 2008 from http://www3.interscience.wiley.com/cgi-bin/fulltext/120837952/HTMLSTART

Further Reading

Dillman, D.A., Smyth, J.D., and Christian, L.M. (2008) Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method. John Wiley & Sons, Inc., Hoboken, NJ.

Gescheider, G.A. (1997) Psychophysics: The Fundamentals. Lawrence Erlbaum.

MacElroy, B. (2000) Variables influencing dropout rates in Web-based surveys. Quirks Marketing.

Moskowitz, H.R. (1983) Product Testing and Sensory Evaluation of Food. Food and Nutrition Press, Westport, CT.

Moskowitz, H.R., Porretta, S., and Silcher, M., Eds. (2005) Concept Research in Food Product Design and Development. Blackwell Publishing, Ames, IA.

Stevens, S.S. (1946) On the theory of scales of measurement. Science, 677–680.

Stevens, S.S. (1975) Psychophysics: An Introduction to its Perceptual, Neural, and Social Prospects. John Wiley, NY.

Chapter 2

Consumer Packaging: Important Functionally, but Not Attitudinally

As the marketplace for food and beverage products becomes increasingly competitive, we find more manufacturers searching for the Holy Grail—a new product for a new occasion. Some might call this the Blue Ocean strategy for product design, using the term from marketing strategy introduced by W. Chan Kim and Renée Mauborgne, professors of strategy and international management at INSEAD, the French business school (Kim and Mauborgne, 2005).

Does packaging provide this Blue Ocean opportunity? This chapter looks at the role packaging plays in the grand scheme of things. We will treat packaging in two ways:

1. How important is packaging, in general, when people have to identify what is important to them about selecting a food or beverage? Packaging, as we will see in the grand scheme of things, comes out to be at best a modest performer. However, for some products packaging is more important than for others.
2. When people respond to concepts, do the elements dealing with packaging come to the fore, or is packaging just "there"? We will look at the results of a set of studies that show the importance of different packaging elements for beverages, and how the subjective importance of elements changes when the nature of the end use changes. We will see that specifying the end-use of the beverage makes a difference in terms of how strongly the package elements perform.

The important thing to remember for this chapter is that we deal here with attitudes toward packaging and labels, not with the actual packages and labels themselves.

By Way of Background

Many of the results you will read in this book come from single-minded studies, whose objective was to learn about packaging, either directly or as part of the larger product issue.

We want to move away from these later one-off studies, at least for this chapter, and into a world that can be called the "consumer mind on the shelf" (Moskowitz and Gofman, 2007). During the latter part of 2001, a group of investigators including the senior author and Jacqueline Beckley of the Understanding & Insight Group, developed the organizing principle that much could be learned about food and drink by experimental design of ideas done, however, on a much larger scale. The vision that led to this notion of encapsulating the consumer mind came from the realization that you learn a lot both within a study, and across comparable studies.

The cross-study learning can even be magnified if the researcher runs several studies simultaneously, using the same scale, the same approach, and even the same basic configuration of test stimuli. In this way it's possible to compare the results of "similar" or, perhaps, even identical stimuli across studies. The studies might deal with different foods or drinks, in which case we would discover just how well or poorly an idea performs when paired with a specific product.

The first of these multi-product studies was Crave It!™, executed several times, but beginning just around September 2001, right before September 11 (Beckley and Moskowitz, 2002). That project dealt with 30 different foods and beverages, ranging from coffee and cola to barbecue ribs, tacos, etc. The objective was to discover what messages about products, emotions, brands, safety, etc., resonated with the consumer. The underlying design, really the skeletal structure, for the Crave It!™ projects was carried through all the 30 different foods/beverages, each evaluated through systematically varied concepts (see Moskowitz, German, and Saguy, 2005).

After evaluating some 60 unique test concepts comprising 36 elements in different combinations, thus being

 ISBN: 978-0-813-81222-9

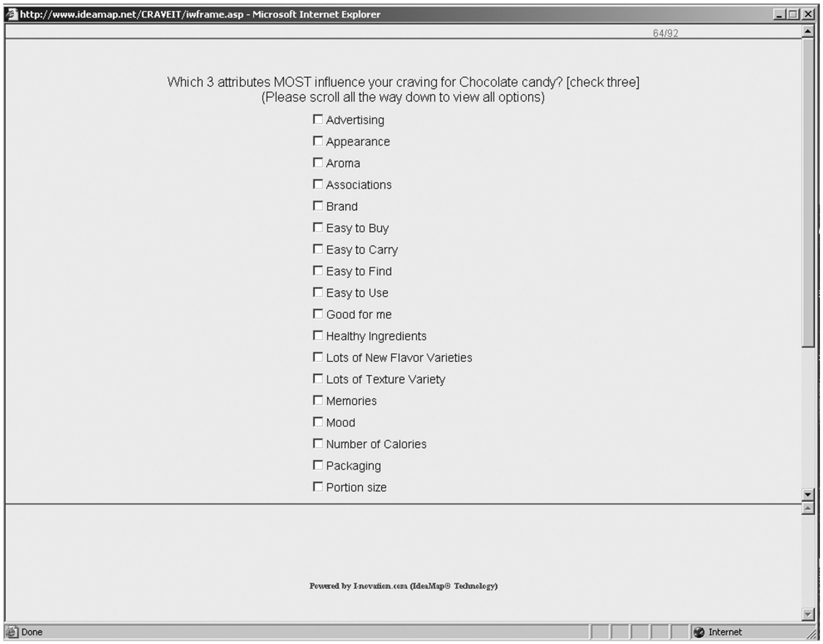


Figure 2.1 Part of the classification question from Crave It!™ dealing with the selection of what drives craving. This particular question came from the chocolate candy study. Each individual food had the same question asked, but particularized to that food.

thoroughly steeped in the product, the respondent completed an extensive classification questionnaire. One of the classification questions required the respondent to select the three most important factors in a food or beverage that drove selection. Look at Figure 2.1 to get a sense of how the question was asked, keeping in mind that the screen with the appropriate food/beverage name was presented to the respondent. Only the name of the food or beverage changed; the questions did not.

When we tallied the results from the different studies, we learned that packaging was relatively unimportant when compared to the sensory experiences offered by the food or beverage. From these initially discouraging results, we soon realized that the notion of "packaging" per se was simply too general. That is, asking a person about packaging simply fails to paint a word picture to that individual, whereas asking the same person about aroma, texture, taste, etc., seems to conjure up a picture. It is no wonder, therefore, that we ended up with the results we did, shown in part in Table 2.1. The role of packaging seems to be very small when we use the word "packaging" alone, without painting a concrete word picture.

Table 2.1 Percent of respondents who selected "packaging" as one of the three main drivers of "craving a food." Results from the 2001–2002 Crave It!™ database with permission from It! Ventures, Inc.

Food	Packaging selected among the top three (%)	Packaging not selected among the top three (%)	Base of respondents
Snack Mix	7.5	92.5	240
Pretzels	5.0	95.0	240
Tortilla Chips	4.2	95.8	238
Salad	4.2	95.8	239
Nuts	3.3	96.7	240
Popcorn	3.3	96.7	242
Cola	2.9	97.1	272
Hot Dogs	2.1	97.9	239
Potato Chips	2.1	97.9	242

Where Packaging Fits—Results from the Drink It!™ Database

We now delve into detailed data derived from 6,000+ respondents in the Drink It!™ database. This study, like the aforementioned Crave It!™ project, is part of our ongoing effort to create systematic databases of different products in a category such as beverages, in order to understand what drives consumers to be interested in, and to select a beverage.

We set up the Drink It!™ database in the following straightforward way:

1. The respondent received an email, and if interested, the respondent clicked on the embedded link. Those who responded were guided to a "wall" showing the available studies. The respondent chose one of the available studies listed. When a study was reasonably "complete" (i.e., had more than 250 respondents), the study was no longer available on the "wall." This strategy ensures that the studies all complete with approximately the same number of respondents. Figure 2.2 shows the wall.
2. As in the Crave It!™ project, each of the 30 studies used experimental design of ideas, with four silos, each comprising 9 elements (36 elements total). Every respondent evaluated a unique set of 60 combinations. This unique set is important because it prevents any hidden biases from affecting the data, such as a limited selection of combinations. These biases are eliminated by being simply "randomized out," swamped by the natural variation of the many different concepts. (See Moskowitz and Gofman, 2007.)
3. At the end of the interview, the respondent completed a classification question about who he was, as well as factors that he considered to be important in the selection of the particular beverage being evaluated. It is that "importance" question that will hold our focus now.
4. For the importance question, the respondent selected 3 factors from a total of 22 that he considered to be the top ones.

Let's see what respondents selected. Looking at Table 2.2, comprising almost 6,400 respondents in the total panel, we see taste coming up as the key factor. This is to be expected. After all, taste is the most critical factor in a beverage, or at least people talk about the taste of the beverage. Almost all of the respondents put taste among the top three factors of key importance (89%).

What surprises us, however, is that respondents don't really pay attention to factors other than either the sensory characteristics of the beverage (taste,

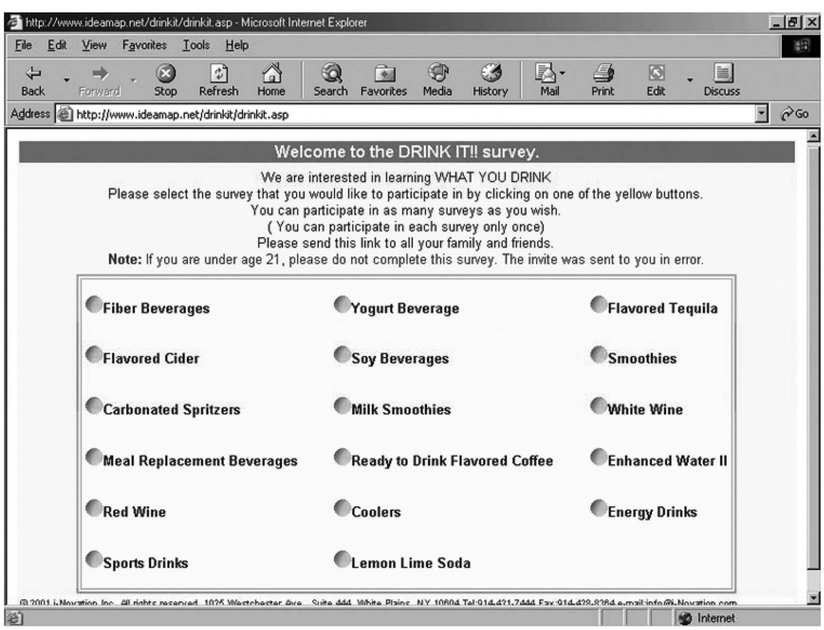


Figure 2.2 Example of the "wall" from the Drink It!™ database. The screen shot was taken late in the course of the project, when many studies had their requisite number of respondents. The less popular beverages remained available for the respondent to choose, whereas the more popular beverages were no longer available.

Table 2.2 Percent of respondents choosing each factor as being one of the top three most important factors for drinking a beverage

	All 6,346 respondents (% selecting)
Taste	89
Appearance	52
Aroma	32
Temperature of beverage	31
Variety	24
Mood I'm in	19
Brand	17
Healthful ingredients	15
Texture	15
Not mixed, pure beverage	11
Health considerations	9
Convenient to use	9
Fits with my meal	8
Portion size	7
Social situation	5
Convenient to buy	4
Convenient to carry	3
Memories	3
Package	2
Associations	2
Advertising	1
Package	3
Convenient to carry	3
Portion size	7

appearance, aroma, temperature, variety), or brand and mood, respectively. The option packaging per se is rarely selected (2%). Again we learn from this that packaging itself is a general word, like the word advertising. We know that taste and appearance are general words as well. However, taste and appearance are subjectively more real, and associated with specific, reinforcing stimulation. We often talk about the taste of a product, or the appearance of package or product. We don't think about the specifics of a the package when we talk about appearance. We just think that the package looks good on the shelf, or that there is something noteworthy about the package. So, we learn that when "taste" and "appearance" are cast in general terms, they seem able to maintain their potency. In contrast, when "packaging" changes from a specific feature to a general idea, it loses its potency.

The same thing happens with the word "advertising." We think of funny advertisements; they are important and often memorable. However, the general notion of "advertising" is not important to people. Neither the general words "advertising" nor "packaging" paint strong, reinforcing, positive pictures in the mind.

When we replace the word "packaging" with the more concrete phrase "convenient to use," the selection jumps from 2% to 9%. Convenience is the more important, effective expression of packaging. The word "packaging" is too general. The phrase "convenient to use" is more dramatic and paints the necessary word picture in a person's mind.

What Beverages Are Most Sensitive to Packaging?

In our explorations, we saw that only 2.56% of the respondents selected packaging, in general, to be important. We obtained this low percent by aggregating the data across all of the beverages. Perhaps we are missing something important here. We know that packaging is key to a product's success, and that the five Ps of marketing (price, product, position, promotion, and package) always feature the package. Could it be that we have different types of products, some of which involve packaging more than others?

To answer this question, look at the percent of respondents who selected one or more packaging-related phrases for each of the beverages (Table 2.3). We saw that packaging as a general term plays little role. Respondents don't select packaging as a key factor for any of the beverages, except Kids Beverage and Sports Drink. When, however, the end use is made more concrete by the terms buying, carrying, and especially using, packaging becomes much more important. The notable finding here is that the term is not packaging, but rather the concrete benefit that comes from packaging (easy to buy, easy to carry, easy to use). The benefit paints the necessary word picture.

Summing Up

One of the most valuable real estate locations in the food-marketing environment is the food package (Aaron et al., 1994; Coulston, 1998). Packaging is important (Deliza and MacFie, 1996; Bredahl, 2004; Caporale and Monteleone, 2004). Yet, the overall general phrase packaging does not particularly spark the respondent, perhaps because the word "packaging" does not paint a suffi-

Table 2.3 Drink It!™—Percent of respondents who selected one or more packaging elements

	% selecting the term ***packaging*** itself
Kids beverage	7.3
Sports drink	6.5
Enhanced water	5.1
Flavored cider	4.6

	% selecting ***convenient to buy***
Shakes	7.6
Smoothies	6.5
Sports drink	5.6
Enhanced water	5.6
Flavored tequila	5.2
Yogurt beverage	4.8
Flavored cider	4.6
Lemon lime soda	4.2
Cola	4.2
Soy beverage	4.1
Kids beverage	4.1
Coolers	4.1

	% selecting ***convenient to carry***
Enhanced water	10.2
Kids beverage	8.2
Sports drink	6.5
Smoothies	6.0
Iced tea	4.5
Yogurt beverage	4.3
Meal replacement Beverage	4.2

	% selecting ***easy to use***
Meal replacement Beverage	18.5
Kids beverage	18.2
Enhanced water	15.3
Fiber beverage	14.5
Hot chocolate	12.7
Yogurt beverage	12.2
Sports drink	11.6
Smoothies	11.5
Juice	9.4
Soup	9.3
Flavored coffee	9.1
Milk smoothie	8.9
Hot tea	8.1
Cola	8.0
Iced tea	7.7
Flavored cider	7.7
Carbonated spritzer	7.5
Lemon-lime soda	7.2
Soy beverage	6.2
Coolers	5.9
Flavored tequila	5.7
Milk	4.4

ciently specific word picture in the respondent's mind. Rather, the key appears to be packaging in terms of an end use. Respondents react to specifics, just as they do in concepts. It's not the particular package, or even the notion of packaging, but rather how the package and the person interact. If the package simplifies the life of the consumer respondents, then it's likely that packaging will be chosen as an important aspect. The rule here is not "packaging" per se, but rather something that is relevant to the respondent in daily life, something specific, something familiar.

Do Consumers Respond to Packaging Features in the Body of a Concept?

This second section of the chapter focuses on specific packaging features for one product, a carbonated soft drink. Look at any supermarket or the local convenience or C Store. The beverage aisle is quite abuzz with the comings and goings of different brands, the fickle tastes of consumers and, most of all, the ongoing fight between powerhouse brands like Coca-Cola® with the emerging interest in good-for-you products such as bottled waters and energy/health drinks.

Doing the Experiment—A New Twist on a Familiar Theme

The approach we used is similar to the experimental designs that we describe throughout this book. There is only one major change, however, and it is that change which produces the information. It is "fit to and end use" that tells the story.

The process followed these steps:

1. Create the experimental design as we had done for the Crave It!™ and Drink It!™ databases. The design comprises four silos of nine elements each. These elements were mixed and matched by the experimental design. Every respondent evaluated a different, unique set of 60 combinations. These 36 elements were all text elements rather than pictures. Most of the elements were descriptions of flavors or statements about emotion. Three elements, however, dealt with packaging, which were always associated with a benefit for the respondent:

   With a thermal barrier your drink will stay colder longer.

   The mini-drink 6-pack … the perfect size for children and people on the go.

   Available in gallons to quench that giant thirst.
2. The respondents were led to a "wall" that showed the available studies from the set of 12 studies. The 12 studies were actually the same, except for "end use" (e.g., appropriate for breakfast, appropriate for Coca-Cola®, etc.). We saw a similar approach in Figure 2.2, where we dealt with the "wall" for Drink It!™ The only thing that changed was the rating scale. For example, for the Coca-Cola® study, the respondent read a concept and rated how well it would fit Coca-Cola®. For the breakfast study, for example, the respondent rated the concept on how well it would fit a breakfast beverage. So, with this type of scheme, all that really changes is the respondent's mind-set. Everything else is pretty much the same.
3. When we analyze the results, we look for the effect of the "end" use on the impacts or utilities of the individual elements. Does the impact change by end use? If so, then tuning the respondent's mind to the end use affects how he reacts to the concept.
4. Most important for this analysis is the researcher's ability to look at the impact of each of three packaging elements as the end use changes. The question that we can answer is simply "Does end use or need state change the impact of a package element?"

So, Where Is Packaging in This World of the New Carbonated Beverage?

Let's look at the impact values for the 36 elements in Table 2.4. As we typically do, we sort these elements from highest to lowest, in order to get a sense of "what's hot and what's not." When we look at the 36 impact values, we cannot help but be struck by two things. First, the winning elements are taste/flavor, with maybe a little bit of "health" and "good for you" thrown in. Second, packaging is not good, not bad, for the elements that we chose. Packaging is, in a phrase, "just there." However, when we get to the statement "gallons of beverage," we discover that the utilities of this packaging element are negative.

So what are we to make of this? By itself, packaging is functional, and not a "destination element" that drives acceptance. This is very important. People drink with their eyes and, of course, with their mouths. Taste, flavor, even ingredients are important. Packaging is important, but in a different way. The package stores the product and makes it available. But packaging is not the major driver of acceptance that flavor is. At least packaging is not the major driver for the total panel, for a carbonated beverage.

Does Positioning or End Use Change Response to Packaging?

We know now that packaging is only modestly important, that the specific package feature affects the utility, and, finally, that taste/flavor. Good-for-you messages do far better or at least the correct messages do far better. We also know that it is possible to have packaging messages that don't do well. (For example, "*Available in gallons to quench that giant thirst*" didn't do well, with an impact of −5, for the unpositioned, that is, otherwise unspecified carbonated beverage.)

What about the remaining 11 studies in this database, where the product is positioned, both in the introduction and in the rating question? Does positioning the carbonated beverage in terms of end use or for a specific brand or age group make any difference in the performance of the packaging elements? This is an interesting question because one of the elements, "*The mini-drink 6-pack … the perfect size for children and people on the go,*" does talk about situation.

Let's answer this question by looking at the performance of the three packaging elements in the 12 different studies listed for us in Table 2.5. We will not look at the other 33 elements, despite the richness of the results and the temptation to "get lost in the data and look for insights." Rather, we will focus our attention on the packaging elements. Furthermore, we will look at the performance of these elements by total panel, as well as by subgroups, in the population (genders, ages). To make

Table 2.4 Impact or utility value of the 36 elements for concepts about an otherwise unnamed, unpositioned carbonated beverage. The three packaging elements in the set of 36 elements are shaded.

B8	A thrilling burst of unique cherry flavor and a sweet, crisp taste that gives you "more to go wild for"	8
C1	Delivers at least 100% of the recommended daily intake of Vitamin C, 15% of folate, and 14% of potassium per 8-oz. serving	7
D7	An energizer that keeps you going … without the caffeine	6
D5	Rich and creamy with no caffeine … the perfect drink to satisfy the whole family	5
B5	Enjoy a smooth slightly translucent drink that's intriguing from the very first sip	5
B2	With a little splash of vanilla flavor … sure to delight	4
C4	Enjoy a delicious taste, but without the calories	4
B9	Introducing new clear natural refreshments with a light hint of flavor	4
A2	100% organic … healthy for you and the planet	3
C8	Created for today's naturally healthy lifestyle	3
D4	Helps you to achieve peak performance when you need it most	3
C6	Quenches your thirst and stimulates your mind	2
D1	Enjoy a daring, high-energy, high-intensity, active drink	2
C7	For the health conscious … a sweet drink with no sugar or aspartame	2
C2	A healthful source of calcium	2
D9	So light, so crisp, so refreshing	2
C3	Provides you with the balanced nutrition you need to live a healthier life	1
D6	A drink that kids thirst for and moms will love	1
B4	An eclectic mix of fruit and other intriguing flavors	1
A5	With a thermal barrier, your drink will stay colder longer	1
C9	Keep trim with a reduced-calorie drink	1
A1	The mini-drink 6-pack … The perfect size for children and people on the go	1
B7	Satisfy your thirst … with real plum juice, ginseng and honey	1
A6	A non-carbonated drink … that won't weigh you down	−1
B1	Invigorate your senses with shocking lemon-lime flavor	−1
C5	A drink which eliminates stress	−1
D2	A bold, energetic, unstoppable drink in a glow-in-the-dark container	−2
B6	Introducing new and exciting flavors such as blueberry twist and wacky pink watermelon	−2
A3	A drink that appeals to your senses … with a unique aroma	−2
A8	Slightly carbonated thirst quenching drink	−3
A7	Enter a whole new universe with a blend of enticing aromas	−3
D3	A refreshing alternative to coffee … with a burst of caffeine	−4
A4	Available in gallons to quench that giant thirst	−5
D8	A jolt of caffeine to awaken your senses	−5
A9	Kick it up with a new highly carbonated drink	−5
B3	Comes in a variety of flavors and crazy opaque colors like punky purple, brilliant blue, and goofy green … you got to try them all	−6

the analysis easy, we sorted the studies for each element by the impact for the total panel.

Take a look at the first element: "*The mini-drink 6-pack … the perfect size for children and people on the go*." We see a very clear pattern:

1. When the carbonated beverage is positioned for kids, ages 7–11, the mini-drink 6-pack is perfectly appropriate, with a utility of +16. This finding makes sense because the product has been positioned for kids. On the other hand, it is definitely inappropriate for a supper or after-supper beverage. The respondents are saying that this mini-drink 6-pack idea simply doesn't work if the beverage is positioned to be consumed at home. For the other situations, this mini-drink 6-pack idea is irrelevant.
2. Just because the total panel reacts strongly to an element doesn't mean that everyone will react as strongly. The younger respondents (ages 21–30 and 31–40) find this packaging idea appropriate. On the other hand, when we get to the older respondents, 41–50 years old, the mini-drink 6-pack just doesn't score as well. So we learn here that the performance of the packaging element emerges from the three-way interaction of the particular packaging element, the end use to which the element is being positioned, and the nature of the respondent who is doing the evaluating.
3. We shouldn't look for general rules. It's sufficient with these types of data to find out what works "in general" (total panel). It's with the total panel where we expect to find rules of what works and what doesn't. When we get to the fine-grained detail, the granular data, such as the element-by-end-use-by-age, more than likely the general patterns will just not be clear. Often interactions are more subtle and more prone to random, confusing error, than are simple "main effects" of one element at a time.

Table 2.5 Impact values for three packaging elements, in 12 studies, each study position for a different end use, brand, or user group, respectively

The specific packaging element	Who is rating the test concepts Total	Male	Female	21–30	31–40	41–50
The mini-drink 6-pack … the perfect size for children and people on the go						
For Kids (7–11)	16	11	17	19	23	4
For Coke	3	–3	4	18	5	–3
For Lunch Time	2	9	0	16	7	–3
For Younger Teens (12–15)	2	–5	4	6	19	–14
For A New Drink	1	–10	3	–6	–3	3
For Afternoon	1	–5	4	–8	15	1
For Older Teens (16–19)	1	–13	7	16	13	–3
For Breakfast	–1	1	–2	1	–2	–1
For Mid-Morning	–1	4	–2	–20	–1	–1
For After Sport	–3	–6	–1	–7	–6	–1
For Supper	–6	–8	–5	1	5	–8
For After Supper	–6	–9	–4	–4	–6	–3
Available in gallons to quench that giant thirst						
For Younger Teens (12–15)	5	6	4	8	6	–14
For After Sport	4	–1	8	7	6	–2
For Kids (7–11)	2	–5	4	6	5	4
For Older Teens (16–19)	2	2	2	–11	3	–1
For Coke	0	–11	3	5	1	2
For Supper	–1	2	–1	0	8	–1
For Afternoon	–2	–3	–2	5	2	–5
For After Supper	–2	–4	0	–2	3	–3
For Breakfast	–3	–3	–3	0	–3	0
For Mid-Morning	–4	–4	–4	–15	–10	0
For Lunch Time	–4	–8	–4	–7	2	–4
For A New Drink	–5	–2	–5	–2	–6	–5
With a thermal barrier your drink will stay colder longer						
For Coke	7	7	7	11	5	4
For Older Teens (16–19)	3	–2	4	2	3	1
For Afternoon	2	0	3	–11	5	3
For After Sport	2	–1	4	–3	6	–2
For A New Drink	1	1	1	2	–8	1
For Kids (7–11)	0	–24	4	6	1	–7
For Supper	–2	–1	–2	0	–4	–3
For After Supper	–2	–2	–2	–3	1	–1
For Older Teens (12–15)	–3	–8	0	7	1	–6
For Breakfast	–4	–3	–5	–4	–3	–6
For Mid-Morning	–4	–8	–3	–7	–10	–3

Let's now move to the second element: "*Available in gallons to quench that giant thirst.*" It is clear that this is a far different proposition for the package.

1. This element is appropriate for teens and following sports events.
2. The element is not at all appropriate for an unpositioned carbonated beverage (a new drink). Telling the respondent that the beverage is available in gallons reduces the impact of this packaging element.

Finally, let's move to the third packaging element: "*With a thermal barrier your drink will stay colder longer.*" This element is most appropriate for Coca-Cola®, and least appropriate for breakfast and midmorning. Packaging that keeps the beverage cold is relevant for Coca-Cola®, consumed in different venues, and virtually irrelevant for a product that would be consumed in the morning and assumed to come right from the refrigerator.

Summing Up

We see here that packaging is a factor in the marketing mix, but not necessarily the key factor in the respondent's mind. Although at some level we might feel let down that packaging does not act as the primary motivator for a product, it is important to keep that fact in mind. For most people, and indeed for business, packaging is *functional*, doing the job of protecting the product, ensuring that it can be merchandised properly, attracting the customer through graphics, and presenting the appropriate information. These are jobs to be done, not necessarily attractions to a customer.

As we proceed in our study of packaging, we see that by itself, packaging as an "idea" is not particularly strong in the mind of respondents. That is, people don't think of packaging in the same way that they think of aroma, taste, calories, etc., as a driver of the desire to eat a food or to drink a beverage. Thus, for most people, we conclude that packaging is a "general idea." There is only some, not great, sensory pleasure that a person gets with packaging, and for the most part when the food or beverage is consumed, the package is merely a memory.

References

Aaron, J.L., Mela, D.J., and Evans, R.E. (1994) The influence of attitudes, beliefs and label information on perceptions of reduced-fat spread. Appetite, 22, 25–37.

Beckley, J. and Moskowitz, H.R. (2002) Databasing the consumer mind: The Crave It!™, Drink It!™, Buy It!™ & Healthy You!™ databases. Paper presented at the Institute of Food Technologists Annual Meeting, Anaheim, CA.

Caporale, G. and Monteleone, E. (2004) Influence of information about manufacturing process on beer acceptability. Food Quality & Preference, 15, 271–278.

Coulston, A.M. (1998) President's page: Nutrition messages on food packages—Location, location, location. Journal of the American Dietetic Association, 12, 1470.

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

Kim, W.C. and Mauborgne, R. (2005) Blue Ocean Strategy: How to Create Uncontested Market Space and Make the Competition Irrelevant. Harvard Business School Press, Boston, MA.

Moskowitz, H.R., German, J.B., and Saguy, I.S. (2005) Unveiling health attitudes and creating good-for-you foods: The genomics metaphor, consumer innovative web based technologies. CRC Critical Reviews in Food Science and Nutrition, 45, 165–191.

Moskowitz, H. and Gofman, A. (2007) Selling Blue Elephants: How to make great products that people want before they even know they want them. Wharton School Publishing, Upper Saddle River, NJ.

Further Reading

Bateson, J.E.G., Reibstein, D.J., and Boulding, W. (1987) Conjoint analysis reliability and validity: A framework for future research. Review of Marketing, 451–481.

Cattin, P. and Wittink, D.R. (1982) Commercial use of conjoint analysis: A survey. Journal of Marketing, 46, 44–53.

Green, P.E. and Krieger, A.M. (1991) Segmenting markets with conjoint analysis. Journal of Marketing, 55, 20–31.

Green, P.E. and Srinivasan, V. (1980) A general approach to product design optimization via conjoint measurement. Journal of Marketing, 45, 17–37.

Luce, R.D. and Tukey, J.W. (1964) Conjoint analysis: A new form of fundamental measurement, Journal of Mathematical Psychology, 1, 1–36.

Mizuno, M. (1997) Conjoint analysis of consumer preferences in cyber space. Advances in Human Factors Ergonomics, 21 B, 475–478.

Moskowitz, H.R., Gofman, A., Itty, B., Katz, R., Manchaiah, M. and Ma, Z. (2001) Rapid, inexpensive, actionable concept generation and optimization—the use and promise of self-authoring conjoint analysis for the foodservice industry. Food Service Technology, 1, 149–168.

Moskowitz, H.R., Silcher, M., Beckley, J., Minkus-McKenna, D. and Mascuch, D. (2005) Sensory benefits, emotions and usage patterns for oliver: Using Internet-based conjoint analysis and segmentation to understand patterns of response. Food Quality and Preference, 16, 369–382.

Wittink, D.R. and Cattin, P. (1989) Commercial use of conjoint analysis: An update. Journal of Marketing, 53, 91–96.

Wittink, D.R., Vriens, M., and Burhenne, W. (1994) Commercial use of conjoint analysis in Europe: Results and critical reflections. International Journal of Research in Marketing, 11, 41–52.

Chapter 3

Starting at the Beginning: Experimenting to Discover What Shape "Wins"

We begin our exploration of design and packaging with a very easy project. Our goal here is to see just how far we can get with a project that required respondents to rate their interest in eight different chocolate candies, when these candies were shown as pictures. This project sounds pretty simple. In fact it was simple, although part of another much larger and somewhat complicated project. However, when it comes to evaluating a simple set of stimuli on a single attribute, what could be easier?

So, let us dive into the data and see what we discover, because it will be through the data, through the particulars, that we will learn what we need. Our particular project involved consumer evaluation of the appearance on a screen of eight different chocolate candies. This is the type of project that people enjoy doing, at least for the first few times, before they become jaded with interviews. Of course, ask anyone who works in a candy company about how he feels about testing candy, and most people will answer that after a while the projects stop being fun. The fun ends because candy can be fun just so long, before evaluating it or its package turn, out to be work, just like any other type of work. However, here, we are lucky. We are the experimenters, and we are working with respondents who, by the specific requirements of the project, had not participated in a study for at least three months.

The Stimuli and the Task

In a scientific study it is important to provide details about how the study was run. Even more important, however, is to place the study in the proper context. Why was the study done? What was the rationale behind selecting the specific test stimuli? What would be done with the data?

In this particular study, the objective was to measure how adults and teens reacted to different chocolate candy bars, when these bars were presented as simple pictures in black and white. This is a typical issue, a very simple problem to address, and a good place to start. The products were identified by the brand name. With the widespread consumption of chocolate candy there was no problem getting respondents to participate. The task was fairly easy since people were familiar with most of the candies.

The Stimuli

We began by sketching out the different candies, just to get a sense of their comparative dimensions. You can see these eight different candies laid out graphically and schematically in Figure 3.1. To our pleasant surprise, the company's team members said that they had occasionally thought of doing this type of structured comparison across the different candies, but never got to actually do it. They thought, *"Someone must have done this analysis. I'm just not sure who, nor sure of when."* It turned out on further investigation that this exercise was truly new to the company.

The actual study was much simpler than people even imagined it would be, based on all of the up-front discussion at the candy manufacturer. The products were "shot" as pictures, shown for Mounds in Figure 3.2 and for KIT KAT in Figure 3.3.

A little digression is called for here. Even in these simple studies, it is important to instruct the respondents about what to do. Respondents don't do evaluations day after day, although researchers do and sometimes forget what respondents know and don't know. So what seems clear to the researcher may be ambiguous to the respondent. That's why we have the orientation page ahead of the interview, as we see in Figure 3.4. Furthermore, the orientation page might be considered a bit wordy. It is. The wordy explication gives the respondent a sense of comfort, leads the respondents through the task, and creates a quick, short relationship with the respondent.

Packaging Research in Food Product Design and Development Howard R. Moskowitz, Michele Reisner, John Ben Lawlor and Rosires Deliza
 ISBN: 978-0-813-81222-9

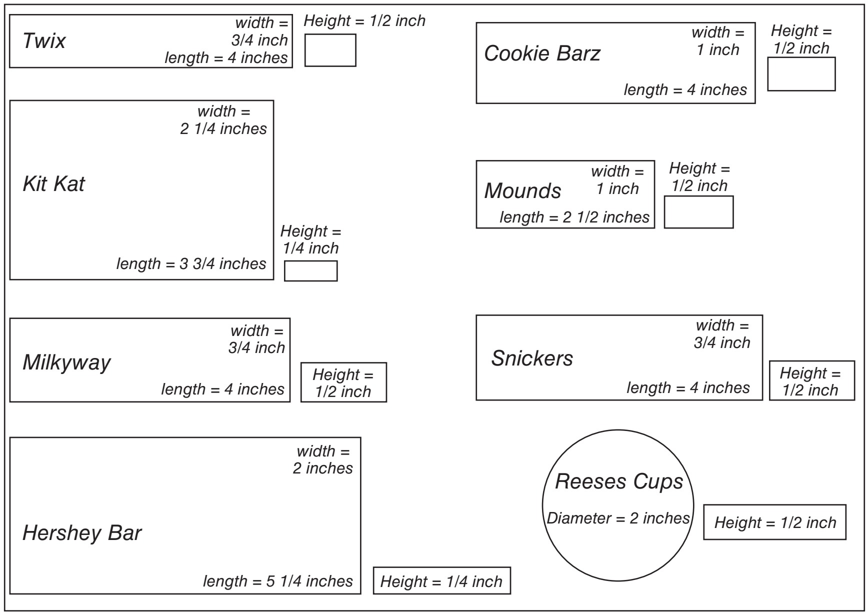


Figure 3.1 Schematic of the eight candies to be tested, showing their dimensions.

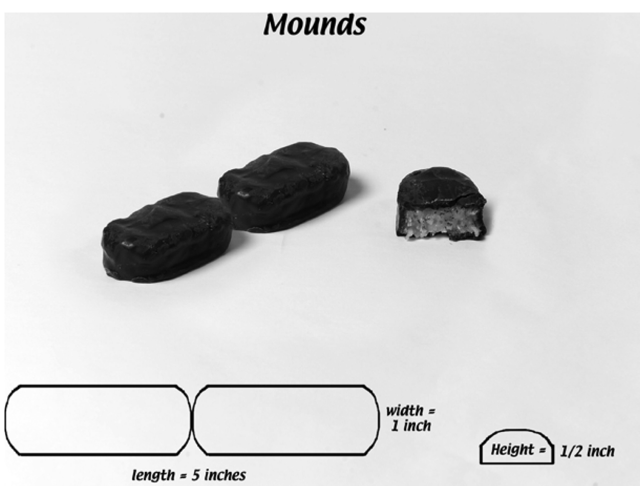


Figure 3.2 Example of the stimulus for Mounds.

All three outcomes make the task just a little less clinical, less impersonal. It's always good research practice to make the respondent feel comfortable at the start of the interview, and at that time sort out ambiguities about the stimuli and the scale.

The respondent evaluated each of the eight different "packages" of chocolate candies (really schematics of packages) in a unique random order. There is a reason behind this randomization. Most research with test stimuli presents them in a random order to avoid problems in the research, which often occur when the stimuli are tested in one single order by everyone. For example, respondents assign a higher rating to the stimulus tested first. This is the "tried-first" *response bias*. When you

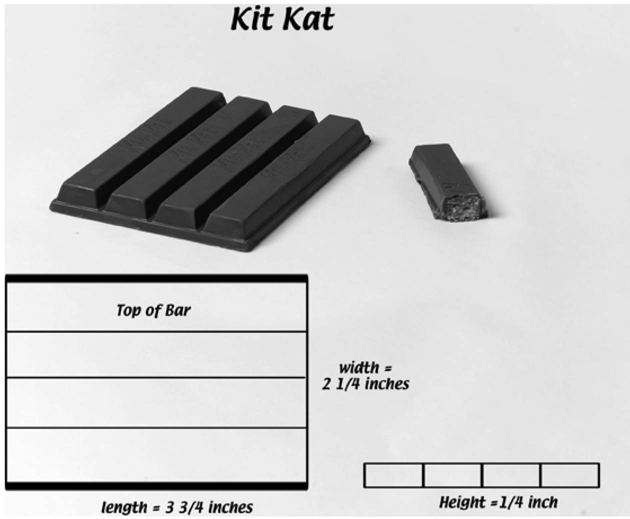


Figure 3.3 Example of the stimulus for KIT KAT.

All of the concepts you are about to see refer to a:

CANDY BAR

Please take your time and evaluate each concept (screen) thoroughtly. Once you have evaluated the concept, please enter your rating based on the following question. The entire conept should be rated as a whole. Please use the entire 1–9 scale to express your opinion.

How interested are you in purchasing this product?

← NOT AT ALL INTERESTED VERY INTERESTED →
1 2 3 4 5 6 7 8 9

PLEASE USE THE ENTIRE 1 TO 9 SCLE.
It is not necessary to press <enter> after your rating.

Figure 3.4 Instructions on the computer screen that tells the respondent what to do.

test the same stimuli in the same position, you will introduce this bias. The first stimulus will be "up-rated," which gives a false reading to that stimulus. If it is a poorly accepted product, then you may overestimate how good it is, just because of that boost from the first position. The most prudent thing is to either have a "dummy" first stimulus, whose data you discard (so-called training stimulus), or rotate the order of the stimulus. Many researchers do both, beginning with a test stimulus to "teach the respondents what to do," and then test the stimuli in a rotated or randomized order. The data from the first stimulus is irrelevant. Often this first test stimulus does not even come from the set of stimuli, but just something convenient to show in that first position.

Each candy appeared on the screen, one at a time, so-called *monadically*, or more correctly "sequentially monadic." The respondent looked at the candy and pressed the appropriate number to show how strongly he felt about the candy, based only on the picture, schematic of the package, brand name, and, of course, the particular rating question.

Some clarification is in order here. Often you will read about the results of experiments. Sometimes these experiments will be done for reasons that won't be totally clear from what you read. You might ask *"Why perform the studies in such a cumbersome method, when you can get the results more easily by simply doing the study in a different, more direct way?"* Chances are that when you read the results of the study and feel this way, the study was probably done in the way you are reading for reasons *other* than what is being immediately presented. And such is the case here. The study was done for a variety of other reasons. That is the reason for the unusual structure of the stimuli—brand name at the top, shapes and dimensions in the body of the stimulus. Here we discuss the study for didactic reasons—to illustrate how to gather and analyze.

Analyzing the Data—What Do We Look for?

Even before we begin to look at the substantive results (i.e., how the different candies performed), we might want to look at the way we measure acceptance. As you will see here and in other chapters throughout this book, we can take at least two different paths when we measure acceptance. We can look at the average level of feeling, the intensity of acceptance. Or we can look at a more "black and white" measure: accept or do not accept. The former, measuring intensity of feeling, comes from psychological science. The latter, accept or do not accept, comes from sociology. From sociology, this all-or-nothing membership in a group migrated into market research—a more applied discipline with a different intellectual history—a different perspective about what the numbers mean, and indeed a different worldview about the numbers that are really meaningful to look at.

Most scientists coming from either the "hard sciences" or "psychology" look at the intensity of feeling. That's what's captured in the rating scale we saw in Figure 3.4. The respondent who reads this question is thinking about the "intensity" of feeling, the degree of interest in this particular candy bar. The respondent doesn't think of "*groups of people*," nor "*Do I belong in this group of acceptors or would I rather be put into a group of rejecters?*" In fact when we think more deeply about the way the individual respondent must approach the scale, we realize that the only way that respondent can answer is on the basis of his own personal opinion. Anything else, average intensity or percent membership in a group must be a subsequent abstraction from the responses of many people.

One consequence of this intellectual heritage from the physical sciences and from psychology is that the appropriate measure of central tendency across many respondents is simply the mean or the median. The scale represents "amount of" and the average rating is the measure of central tendency (i.e., if nothing else were to be known about the product, then the average is the best guess). Now with that in mind, look at Figure 3.5. We see the arithmetic averages on the x-axis in this scatter plot graph. Each of the eight circles in the graph corresponds to one of the eight products. The abscissa or x-axis clearly shows that the eight products don't cluster in one region, but rather distribute across the scale. Certainly there aren't any poor candies (nothing averaging below 4.5), but on the other hand, that makes some intuitive sense. How can commercial chocolate candies be uninteresting?

Commercial research doesn't really care very much about the intensity of a single person's feeling, except when it comes to making changes to the product. The commercial researcher also doesn't necessarily care about the average degree of acceptance, at least when the

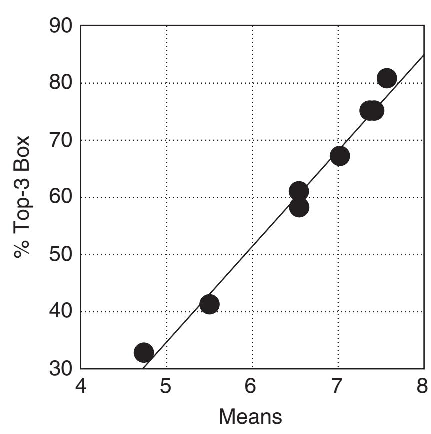


Figure 3.5 How the percent top-3 box (percent rating 7–9) covaries with the arithmetic average. Each circle corresponds to one of the eight candies. Although the two measures of acceptance correlate highly, they signify very different things about acceptance, and lead to different things that one can say about a product.

study is funded by marketing. A subtle change in focus takes place, but it is a change that is very important to highlight and always to keep in mind.

Although the individual respondent can only rate his or her feeling, the commercial researcher usually searches for the number or, more typically, the proportion of individual respondents who feel a certain way (i.e., those who exhibit a specific, predefined set of responses). In our case it is the percent of respondents who are interested in the product. In some cases it is the percent of respondents who are highly satisfied with the product, and in some other cases it is the percent of respondents who are dissatisfied with the product.

Operationally, there is a big difference between measuring intensity of feeling versus assigning a person to the group who is satisfied. S.S. Stevens, Professor of Psychophysics at Harvard University, would go out of his way to hammer home this difference, proclaiming that *"Nothing is quite as difficult in science as converting a continuous or reasonably continuous scale into a binary measure."* It sounds simple, but the thinking has to be clear. What is the rule by which we can convert this 9-point scale of a person's interest into that person's membership in the class of "*I like or accept the stimulus*" versus "*I dislike or reject the stimulus*?"

For our candy study, and as a matter of course, we *arbitrarily* chose the three high ratings—7–9 to represent a high degree of interest or acceptance. This is called the "top-3 box." A person who rates one of the test stimuli 7, 8, or 9 is assumed to "accept" that stimulus (i.e., to fall into the acceptor group). A person who rates that stimulus 1, 2, 3, 4, 5, or 6, respectively, is assumed not to "accept" that stimulus. Thus, for each stimulus a person evaluates, that person can either accept or not accept the stimulus, or perhaps more colloquially like or not like the stimulus. We will not analyze degree of liking, but instead simply tally up the number of people who accept versus not accept the stimulus.

Our specific choice, 7–9, is arbitrary. The top third of the scale is a fairly stringent measure. Of course we could make the criterion even more stringent, by looking only at the top two scale points (called top-2 box), or even the top scale point (top-1 box or top box). The reality is that the more stringent we make the criteria for "acceptance," the stronger the acceptance has to be (which is a good thing), (but the fewer the number of respondents in the pool (which is a bad thing). Therefore, looking at the top-3 box is a reasonable compromise. We have a reasonably strong level of interest that a person has to exhibit before being counted, but we don't make the level of acceptance so high that we make the data very sparse by having very few acceptors for any test stimulus.

Armed with this "newer" way of looking at acceptance, namely counting the number of respondents who rate a candy as 7—9, let's look at how these two measures covary. They should be reasonably correlated. The greater the number of respondents who like a candy (i.e., the more people who rate the candy as 7—9), the higher should be the average. We see this happy state of affairs in Figure 3.5. The abscissa or x-axis shows the mean rating, and the ordinate or y-axis shows the percent top-3 box. Each of the eight circles is one of the candies. The relation is almost perfect.

The approach of counting membership in the acceptor group rather than estimating average liking comes from sociology. Sociologists are interested in how many people exhibit a specific behavior, rather than being interested in the intensity of that behavior. So, when we deal with measures of acceptance throughout this book, for the most part we will deal with these percents, basing our approach on the sociological way of analyzing data. Market researchers have accepted that sociological approach and incorporated it into the way they think about problems. Parenthetically, when we deal with the consumer as a measuring instrument, to assess specific aspects or attributes of the package/product, we will revert back to the average or mean as the measure of central tendency.

What We Found

The most important thing in research is, of course, the results. That's why we do the study in the first place. The issue is, however, what do we look for? The question itself sounds a bit strange. After all, most people feel that when doing research one ought to begin the effort with a well-formulated question, some knowledge about the types of answers that one might get, and of course the ability to move from the data one observes toward either confirming the hypothesis or denying the hypothesis. We are taught this "linear," structured way of scientific thinking from the early grades when we learn about science. The same worldview pervades most of the work we do later on in our professional lives. Of course the reality is a bit different. We often have vague hypotheses, try to answer the questions, but always keep our eyes open for new, interesting side paths, results that intrigue and add valuable insights. And, often we don't even really need to have a simple hypothesis to prove or disprove. Rather, we enter the experiment looking to discover patterns. The search for meaning in nature, for patterns, systematics, rather than hypotheses, is what drives us. And that's what happened here.

When we began this specific research project on the shape of package designs for candy, we did not begin with a specific hypothesis to prove or disprove. Rather, we went in looking for patterns. The patterns that we seek tell us how nature works. We don't know the regularities, but we'll know them when we see them. For this specific study on chocolate, we are really looking for a simple pattern. That pattern can be described as "*The way the eight products line up. Are they the same, or do they score differently?*"

With that in mind, let us look at the summary data for the eight products, first for the total panel, and then look at the results from males versus females, and for adults versus teenagers. We are able to look at these "breaks" or "subgroups" because during the recruiting, we ensured that we had at least 40 teens and that the respondents were more or less divided evenly between male and female.

With that in mind, let's look now at the summarized data in Table 3.1. We see our eight candy packages, in the order of their "performance" (i.e., in descending order of top-3 box purchase). The results are quite remarkable, not so much for the performance of one or two products but because even within a popular product category we can see an enormous range of acceptance. Three of the candies perform exceptionally well (Reese's Peanut Butter Cups, KIT KAT, and Twix). Two of the candies perform poorly (Mounds, Cookie Barz).

Consumer researchers are accustomed to looking at subgroups of respondents in addition to the data from the total panel. These subgroups are typically defined either by geo-demographics (i. e., gender, age, income, market), or by purchase behavior (i. e., category usage, brand used most often, and the like). In our candy study, all of the respondents were recruited to be category purchasers, so there is no question that we are dealing with the correct target sample. The exceptionally high performance of three products and the poor performance of two other products cannot, therefore, be traced to an unusual group of respondents. All respondents were appropriate for the study, and furthermore, all respondents evaluated every single one of the candies.

By having each respondent evaluate all eight samples, we eliminate any bias that may be caused by the respondent. Each respondent serves as his own control. This powerful but relatively simple approach, having all respondents evaluate all samples, is called a "within-subjects design." You will encounter this strategy many times during the course of the book. It is a simple precaution that, at once, increases the strength of the data by reducing a host of biases.

Now let's look at what the data tell us. We will look at our statistic, the percent top-3 box. Keep in mind that this statistic tells us the proportion of individuals who we classify as acceptors for each candy, based on how they rate the picture of the product and the schematic of the package.

Table 3.1 Summary results from evaluation of pictures of eight candy bars, using a graphical display of the package. The numbers in the body of the table are percent top-3 box (ratings of 7, 8, 9) for purchase intent, rated on a 9-point scale. Each respondent rated all eight pictures.

	Total Sample	Males	Females	Teens 12–17	Adults 18–49
Reese's Peanut Butter Cups	81	73	89	89	73
KIT KAT	75	73	77	72	79
Twix	75	72	78	78	72
Snickers	67	64	71	63	72
Hershey's Chocolate Bar	61	55	67	57	65
Milky Way	58	60	57	67	49
Cookie Barz	41	37	46	44	38
Mounds	33	33	33	27	38

What do we see? Or course we see the data in a neat tabular form! But, again, what do we really see? Certainly we see some differences. Yet, being realistic, can we legitimately say that we see differences among groups, or are the differences that we see merely the result of random variability that always occurs in scientific studies, no matter how well controlled the studies may be? One way to convince ourselves that there are no differences by groups is by plotting the data in a scattergram. Parenthetically, "when in doubt, plot" is always a good idea in research to help clarify what the data might be saying.

Let's look at the plot of men versus women, and of teens versus adults in Figure 3.6. From the deviations we see that the complementary subgroups really respond quite similarly. Differences in the top-3 box are probably due to random error; there are no major discrepancies. It's this type of plot far more than statistical tests, that tells us what's going on. Indeed, if truth be known, the widespread use of easy-to-compute statistics without plotting often hides as much "truth" in the data as it reveals. Plotting the data is better than a lot of the inferential statistics, when we look for similarities and differences in populations.

Finding Individual Differences Through Segmentation

Throughout this book we will be talking about the value of segmentation as a way to understand designs, packages, and of course the people behind them, the consumers (Ennis et al., 1982; Jacobsen and Gunderson, 1986; Moskowitz et al., 2006; Meilgaard et al., 2007; Meullenet et al., 2007). We know that people don't agree with each other, that there is interpersonal variation, and that what one person likes another person may not like. The notion of percent top-3 box brings that idea to life. Since we focus on the percent of the respondents who rated a candy 7–9 on a 9-point scale, we instantly realize there are the others who did not rate the candy 7–9.

We saw no differences in general patterns when we looked at the summary statistics, by candy, across gender and across age (look at Table 3.1) So, how should we proceed? One way is to plot the distribution of the ratings for each of our eight candies, to see how these distributions appear.

Let's look at the distributions in Figure 3.7. We see our eight candy bars, with a so-called density distribution. You can't really see the individual circles (they're very small), but each column comprises a set of filled circles, one circle per respondent. The trick here is to see whether the people distribute across the scale, or whether they clump at one location. Distribution means that people differ; clumping means that the people are identical. Of course, if you have close to 80% top-3 box, you're not likely to have much distribution beyond 7–9.

What's quite interesting about these distributions is a sense that there may be different populations in our group of respondents. It's probably not the case for products like KIT KAT that score very highly across most of the respondents. It's more likely for those candies that score modestly well, or even poorly. A good example is Mounds. Mounds scored at the bottom of the group, at least based on the picture of the product, the structure of

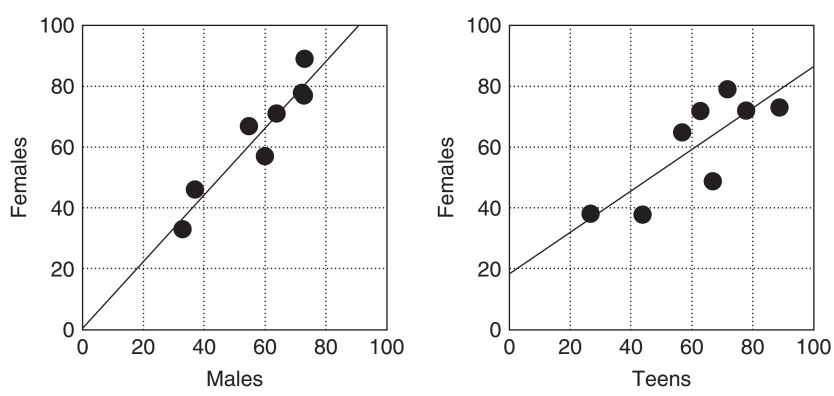


Figure 3.6 Scattergram of top-3 box values for the eight candy bars. Each circle corresponds to one of the candies. The respondents were divided into complementary subgroups and plotted against each other.

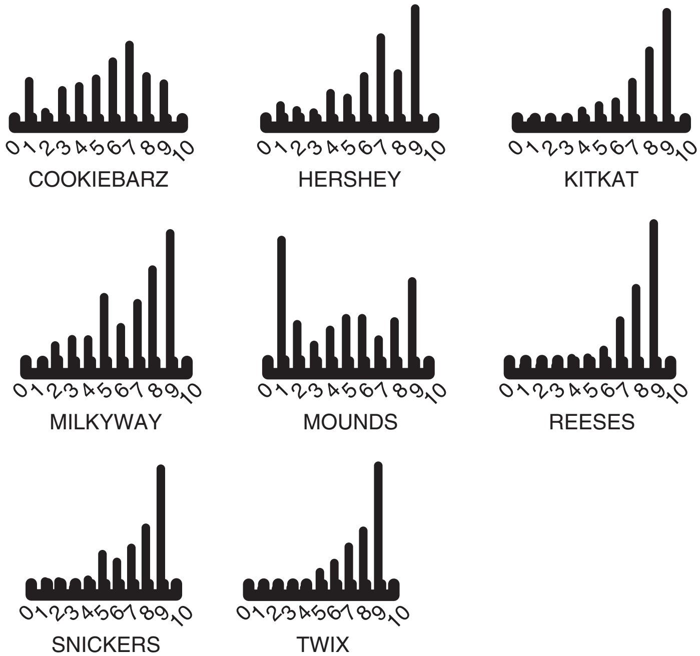


Figure 3.7 Distribution of 9-point ratings for the eight candies.

the product as depicted in the stimulus, and of course the brand name. Only 33% of our respondents assigned Mounds a top-3 box acceptance (i.e., rated the stimulus 7, 8, or 9, respectively). Yet, look at Figure 3.7, in the middle. We see the ratings for Mounds distributed in a way to suggest that there are some people who truly love the product. There is a bump of real Mounds "lovers," but also a whole group of other respondents, some of whom are indifferent to Mounds and others who really dislike the product.

Now that we know we may be dealing with different populations or different "mind-sets," or at least we think we may be, how can we isolate these groups? We'll talk a lot about segmentation during the course of our case histories. We might as well start here with the simplest data—liking ratings for a set of different products.

Fortunately, all of our respondents evaluated each of the products, so we're "good to go" on statistical grounds. We don't have any missing data, but if we were to have such data, we could use a variety of different methods to estimate the missing data, and then move on. We don't need to do so here, so let's move to dividing the people by a way that is both valid statistically and powerful enough to tell a story.

We begin with the recognition that all we have are the ratings of these eight products. Certainly we know about these products from other sources of information, but the only criteria that we will use for dividing the population will come from the eight ratings themselves. Let's now reason through what we ought to do, the why's behind the action, and the rationale for rejecting other actions at each step.

1. Lay out the products in an easy to use matrix. This first step is really a bookkeeping step, necessary for statistical analysis. The typical layout comprises rows as people, and columns as products, or columns as other measures on which we are going to divide the people. This layout is done to follow the requirements of statistical programs. By and large, the layout makes intuitive sense, but it's really following the rules of the statistical analysis package.
2. Reduce some of the redundancy in the variables on which you are going to segment. We have eight products. We would really like to make sure that we segment or divide people on variables that are truly different from each other. To do this we have to create a smaller set of truly independent "pseudo-products,"

and divide the people based on the pattern of responses to these pseudo-products. A good example of the logic behind this comes from dividing people on the pattern of their responses to questions, not products. Suppose we asked respondents to answer eight questions, but six of the eight questions had to do with taste, one had to do with appearance, and one had to do with texture. And, furthermore, most of the six questions on taste dealt with different aspects of the chocolate flavor. Dividing people on the pattern of their answers to these eight questions, six of which are flavor, and five of which deal with chocolate flavor, would give us a distorted division of people. We don't have to throw away our data. We just start again, reducing the redundancy of our eight questions or, in this case, of our eight products. We use the method of *principal components factor analysis* to reduce the redundancy of our eight products.

3. We divide the product set into three logical or pseudo-products that differ from each other. The number three is not fixed. The number of such logical or pseudo-products comes from statistical considerations in the analysis. It could be two, three, four, five, or even more pseudo-products. Each person now has three numbers, corresponding to their values of the three pseudo-products. We can plot the respondents, first as a total group, and then again after we have divided them into two, and then three groups. We divide or segment the people so that the patterns of the three pseudo-products for people in a segment are similar in that segment, and differ from the pattern of the three pseudo-products in the other segment(s).
4. Let's look at Figure 3.8. At the far left we see one undifferentiated mass of respondents. We don't have any segmentation. So, there's no reason to differentiate the respondents. Now let's move to the middle panel. We divide the respondents into two segments or clusters, based upon the pattern of their values on these pseudo-products. You can see that the circles move to the left front, and the Xs move to the right and to the back. We have two groups that seem somewhat different. Finally, on the right we have three segments. One group moves to the front, one to the right and back, and one to the left and back. We thus have three groups that make sense, at least by appearance.
5. We could divide the group even more finely into four clusters (segments), or into five, six, etc. However, that defeats the purpose. We're trying to divide the groups so that they make intuitive sense.
6. We can begin to understand how these clusters or segments behave by looking at the patterns of responses to the eight products. Once we know the segment to which a particular respondent is assigned by the "clustering program" (the program that does all of the "heavy lifting"), we can compute the percent top-3 box for each of our eight candies, for total, and for respondents in each segment. Let's examine Table 3.2 to see how the segments respond. We see immediately that the segments pull apart the products.
7. We don't exactly know the reason behind the segmentation. We would know more if we knew the characteristics of each. To some degree, we know a bit of that information already, since we have the physical dimensions and shape of each product in Figure 3.1. As yet we do not know anything else beyond the picture, dimensions, and brand. We need not go any

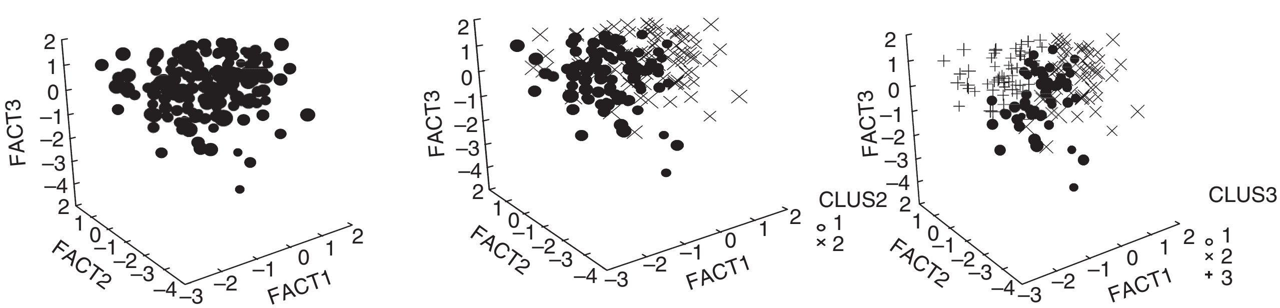


Figure 3.8 Location of the respondents on three "pseudo-products" (factors). The left panel shows the respondents before segmentation (filled circles). The middle panel shows the respondents divided into two segments (filled circles and X's). The right panel shows the respondents divided into three segments (filled circles, X's, and crosses).

Table 3.2 Top-3 box percents for the eight products, by segment. The products are showing a decreasing level of acceptance based on total panel (one segment). Products achieving more than 66% top-3 box acceptance are shaded.

Segment	Reese's Peanut Butter Cups	KIT KAT	Twix	Snickers	Hershey's Chocolate Bar	Milky Way	Cookie Barz	Mounds
1	81	75	75	67	61	58	41	33
1	78	81	91	46	58	55	51	14
2	84	70	61	87	64	61	33	50
1	66	79	87	36	66	68	60	23
2	81	69	54	88	76	76	43	58
3	95	79	89	72	39	28	21	12

further right now. We have what we need—a sense of how to do the experiment, what we measured, what we obtained, and the demonstration that segments may exist among our respondents.

8. Let's now proceed to the final discussion of what we have learned in the section Summing Up.

Summing Up

What has this chapter really taught us? If we stop for a moment to tally up what we have learned and what we're missing, we can come up with at least these six points:

1. *Test many stimuli, not just one*: It is possible to run a study with multiple stimuli and get meaningful data because the respondents act like measuring instruments. This sounds like a truism, but for many years people believed that the only really valid form of measurement was paired comparisons (prefer A or prefer B). You learn a lot by testing more products rather than testing fewer.
2. *Scaling works*: Respondents differentiate among products fairly easily when they rate or scale their feelings.
3. *Acceptance intensity versus acceptance "membership" is similar for group data:* Looking at degree of liking versus looking at belonging to the acceptor class gives similar results when we average the data from different respondents to get aggregate results. That is, if on the average a product scores high in acceptance, it's likely to have a lot of acceptors.
4. *Traditional geo-demographic subgroups show similar acceptance patterns*: Conventional ways of dividing the population (gender, age) may not generate large differences in product acceptance.
5. *Segment by the pattern of responses*: One can divide people by the pattern of their responses to products. It is best, in those cases, to remove some of the redundancy first, and then cluster or divide people on the basis of responses to "pseudo-products" whose redundancy has been eliminated.
6. *Segmentation is a statistical process and the insight underlying the segmentation comes from other information*: Segmentation on the basis of patterns of responses needs additional information about the product to make sense. Even though we can divide people by the patterns of responses, nonetheless, we cannot easily discover the underlying rationale for this segmentation. Later in this book we will look for patterns in the stimuli that provide a way to better understand what the segmentation means.

Postscript—Learning More about Segmentation

Most statistical packages on the personal computer, such as SAS or SPSS, have clustering programs that can analyze data of the type we have presented, and generate clusters or segments. These packages are easy to use, guide you through the different steps, and explain the output in user-friendly terms. There are no single "preferred" ways to cluster a data set. The packages offer various ways to measure "dissimilarity" between pairs of stimuli, and various criteria by which to create the different segments. For practical applications, it's best to start with a data set that you understand, and simply dive in, exploring what you get from different ways of clustering, and trying to understand the meaning of the segments that emerge. It's well worth the effort to explore a data set this way.

References

Ennis, D.M., Boelens, H., Haring H., and Bowman, P. (1982) Multivariate analysis in sensory evaluation. Food Technology 36, 83–90.

Jacobsen, T. and Gunderson, R.W. (1986) Applied cluster analysis. In: J.R. Piggott (ed.) Statistical Procedures in Food Research. Elsevier Applied Science, London and New York, pp. 361–415.

Meilgaard, M., Civille G.V., and Carr, B.T. (2007) Sensory Evaluation Techniques. CRC Press, New York.

Meullenet, J.F., Xiong R., and Findlay, C. (2007) Multivariate and Probabilistic Analyses of Sensory Science Problems. Blackwell Publishing Professional, Ames.

Moskowitz, H.R., Beckley, J.H., and Resurreccion, A.V.A. (2006) Sensory and Consumer Research in Food Product Design and Development. Blackwell Publishing and the Institute of Food Technologists, Ames.

Chapter 4

Patterns in Packages: Learning from Many Packages and Many Attributes

Introduction—By Way of History

As consumer researchers have plied their trade, they have been asked to do more with competitive analysis. The very existence and popularity of SCIP, the Society for Competitive Intelligence Professionals, attests to the art and science of understanding one's competitors. Sun Tzu, the famous Chinese strategist, said, almost 2,500 years ago in 400 BCE, "Keep your friends close, and your enemies even closer."

That being said, how can we learn from the competition? What should we do to become smarter? Or, can we? What are the tools? How should we look at data? Is what we do today even enough of the right thing? All these are good questions that need to be answered, and the principles underlying them put into practice. If we can understand the competition in a structured way, then perhaps we can use what they know already to our own advantage.

So, where do we begin? We might be happy doing the innumerable head-to-head comparisons that researchers are famous for. Perhaps we could report that "Package A is preferred 2/1 or 66/33 percent to Package B." The ever-enterprising and detail-oriented researcher might even go one step further, reporting such preference tests not only for the total package, but for each aspect of the package (e.g., graphics, closure, shape, size, density, etc., etc., etc.).

If you have ever attended a meeting where these so-called "paired comparisons" are the topic, you might well wonder about what you actually learned from this beauty contest. It's one thing to say that Package A is preferred to Package B 2/1 or even more realistically 54% to 46%. What do you do to the packages to make the losing package better? What do you learn that makes you smarter?

Now let's take this ritualistic paired comparison task and expand it. Instead of dealing with two packages analyzed in terms of preference (prefer A or prefer B, which shall it be?), think about preference in terms of different characteristics. Good examples are "Which package size do you prefer?" or "Which color do you prefer?" and so on. The list is endless, the consequences mind-numbing.

Now, to finish off this restrained critique of the hallowed method of paired comparison (like Noah's ark, two-by-two), think about the number of combinations you are going to have to test when you need to discover patterns among three, four, five, or even six, seven, and eight packages. The task seems impossible. Not only is the execution of the study complex in a two-by-two world, but after you have amassed the data it's not clear what to do with it all! How do you make sense of all these paired comparisons? You can certainly measure with accuracy, but a productive analysis to discover what's going on turns into an entirely different adventure.

Making Sense of the Competition—Profiles and Patterns

In the early 1980s, researchers at Moskowitz Jacobs Inc. began to develop research methods that would make sense of the competition. Some of the earliest work was funded by the Nestlé Corporation, in their U.S. offices located on Bloomingdale Road in downtown White Plains, New York, about one-half mile from where Moskowitz Jacobs Inc. has its headquarters. The story of the early work makes for a short and interesting digression, but most importantly, reveals to us how corporate needs turn into research approaches.

In the early 1980s, Nestlé became seriously interested in understanding the organizing principles that underlay perception of products that they produced. At that time, Ernst Schmid, a senior researcher in the marketing research department, began his quest to create better

 ISBN: 978-0-813-81222-9

research methods to understand Nestlé products. The first study revolved around coffee, specifically what characteristics of coffee from across the spectrum of commercially available coffees seemed to be the ones that consumers liked the most. Like most other companies at that time, Nestlé was wedded to the method of paired comparisons, so that it was unable to uncover these general rules of consumer preference. Through persistence, however, Schmid created a research project to evaluate 11 different coffees, with the goal of discovering a pattern.

We know that the basic relation between sensory attribute and liking is an inverted U-shaped curve. As a stimulus becomes stronger, it is first liked, then liked more, and then liked most. Any continuing increase in the stimulus intensity beyond this maximum degree of liking started to diminish acceptance. Schmid's question was simple: What did this particular curve look like for specific coffee attributes? And, another question arose: Could Nestlé discover groups of coffee consumers with different sensory-liking curves? That is, did there exist in the population groups of consumers who liked high intensities of sensory attributes (the so-called high impact people) and other groups who liked the low intensities of same sensory attributes (the so-called low impact people)? At the end of the day, the research program uncovered these groups, and made use of them in optimizing formulations. But, that is not the essence of our story. Now we want to learn how this approach of competitive analysis and pattern recognition translates into packaging.

What Do We Learn from Many Products?

When we test many products, instead of one product, we discover how they perform versus each other. That, in itself, is important especially when we use a scale of acceptance rather than the method of paired comparisons. If we live only in the world of pairs, then we certainly know that one package is preferred to another, and by how much. The task is harder, of course, when we deal with many products. Nonetheless, with enough data and enough theory, we probably could create some type of general scale so that the preferences that we observe are transformed into these scale values. Leon Louis Thurstone, the famous psychometrician, created just such an indirect method 80 years ago in 1927 (Thurstone, 1927).

We can do a lot better and be far more productive, however, if instead of paired comparisons we ask respondents to act as measuring instruments. Despite those critics who say that consumers cannot act as measuring instruments, the world of everyday experience shows that the consumer *can* and *does* act as measuring instrument in many everyday situations, whether shopping, driving, cooking, eating, or just about every other activity!

Now back to the problem of measurement. What should we do with ratings of acceptance, such as the sets of ratings for six hypothetical products that we see in Table 4.1? What can we learn from these results? Can the data make the package designer smarter? What new insights emerge to make the job easier? Certainly we can say that in the case of the first column of observations (A), the products all bunch together, that in the case of the second column (B) there are two good products and the rest poor products, and in the third column (C) the opposite occurs, so we see mostly good products, and a few poor products.

Analyzing Data for Shelf-Stable Milks

Let's look at an example of data from an actual study. The particular project dealt with shelf-stable chocolate milk drink where respondents evaluated nine different products, first by visual inspection (just looking at the product) and then, immediately afterward, by lifting and holding the product. For right now, we'll focus on the acceptance ratings of the nine products, to see how the products perform.

First look at the stimuli in Figure 4.1. Keep in mind that the respondents inspected the packages one at a time, in a randomized order. Randomizing eliminates any order bias, such as the often-observed bias of the product evaluated first to get a higher score in that first position than it would obtain in other positions.

Table 4.1 Three possible outcomes (results A–C) from evaluating six packages on liking, using a 0–100-point scale

	Result (A)	Result (B)	Result (C)
Package 1	68	68	68
Package 2	62	66	64
Package 3	56	45	63
Package 4	49	43	62
Package 5	42	49	49
Package 6	35	37	41

Figure 4.1 The nine shelf-stable milk products. Respondents first looked at the product, rated appearance attributes, then held the product and rated the remaining attributes.

Table 4.2 Performance of the nine products on acceptance (liking) and on purchase intent

	Nesquik	Hershey's	Grip'n Go	Tunes	Shamrock Farms 20 oz	Shamrock Farms 12 oz	Chug	Slam	DariGo
Appearance only									
Like the bottle overall	83	78	75	61	60	57	52	51	50
In hand									
Like the bottle overall	82	79	78	60	59	61	55	56	53
5-point purchase scale									
Top-1 box (definitely buy)	53	41	24	14	19	17	13	12	11
Top-2 box (definitely and probably buy)	88	73	62	38	44	47	29	29	30

We start the analysis with ratings of liking. What's special about these data is that they represent in-market products, so we know something about their performance. When we look at Table 4.2, we are struck by the fact that we are dealing with a "beauty contest" among the products. Certainly, we know which product wins and which loses. But, we don't know why and we're not particularly sure about what makes a good product.

Now What Do We Do?

If beauty contests only bring us part of the way, then we have to do something else. The question, of course, is what is this something else? Just knowing that the products performed a certain way might work if we could simply take that information and plug the acceptance scores into some type of predictive model. We might predict sales, and if the product were sufficiently acceptable, then we'd stop there, breathe a sigh of relief, and move on.

Life is hardly that simple. Testing is not science. Most tests that we might run are simply passive reports of an aspect of reality. We need the evaluation of what the products are, but we also need to discover specifically what to do once we have the data. Perhaps, in the hands of a skilled developer, these simplistic reports of package acceptance can be magically transformed into understanding why one package does well and another one does poorly. Unfortunately, that skilled developer is rare, more a matter of legend than reality. That skilled developer is akin to the wonderful senior doctors in novels such as Sinclair Lewis' Arrowsmith, who, from years of practice, can see a patient and instantly have a sense of what is causing the patient's distress (Lewis, 1925). Such expertise comes from years of clinical training, experience that other doctors (and hence, by analogy, package and product designers) simply do not possess.

With these caveats in mind, we move to the logical next step—adding more information to our study. Market researchers and sensory analysts call this information diagnostics, for the simple reason that the information diagnoses the characteristics of the product. Diagnostics come in the form of attribute ratings, usually assigned by the consumer respondent. However, diagnostic information about these products can be provided by marketers who deconstruct the products into the presence/absence of features, by expert panelists who give specific, structured sensory descriptions of the product, or by instruments that provide measures of package characteristics. Whatever the source of diagnostic information may be, the objective is to provide more data so that the data become useful. Those data allow the researcher to search for underlying relations among the variables in the data set.

The next question is what type of information is best? And, of course, who can provide that information? Consumer researchers are accustomed to having the respondent rate products (concepts, actual products, packages) on a myriad of questions, so there's no problem with the quantity of information. It's the quality of that information, however, that is important.

Let's move to an example of the type of information that one could ask about a product in a package. The focus of the evaluation is the package. Keep in mind that the respondent does not possess a particularly broad vocabulary for package features. Thus, looking at Table 4.3, we see a few descriptive questions that instruct the respondent to profile his response to the physical characteristics. Most of the other questions instruct the respondent to rate acceptance (liking, purchase intent), or image (e.g., appropriate for a specific end use). It is worth looking at Table 4.3 in detail because from it you will get a sense of the depth to which a questionnaire can probe the "package experience."

We see how large a data set might be, with many questions and many test stimuli. Now, the question is what do we do with these data? For our project on chocolate milk we worked with nine different products in the market, each rated separately on the full set of attributes that we see in Table 4.3. We can now generate a lot of data. The question comes down to "*How do we process the data from the nine milk products to give us direction*?"

It might seem that we are belaboring the point of "what to do" and, in fact, we are focusing and belaboring the point. It will be very important, both for this book and later on. If you are reading this book, then most likely you are interested in how to create better packages, better graphics designs, and better experiences. Most likely you are not particularly interested in the latest and greatest way to represent these data by a two- or three-dimensional picture.

Now, back to the question of what should we do to "make the data sing to us," or perhaps a bit less hyperbolic, how to extract insights from the data. Much of what you read in this book focuses on the relation between variables, in the spirit of psychophysics. We can do that investigation now, but we need to find the appropriate independent variable (x-axis) and the appropriate dependent variable (y-axis). We should keep in mind the following considerations, adapted from both development and science:

1. The independent variable should represent something that we can legitimately vary. We cannot really vary overall liking or purchase intent. Those are intrinsically dependent variables, results of our perception. We cannot order a package of a given level of acceptance. Nor can we really vary appropriateness for an end use. Like acceptability, the rating of appropriate is a judgment made after the respondent integrates the sensory information about the product with other criteria, such as price, brand, and the like.
2. With the criterion of actionability in mind (i.e., the developer can push the attribute in a direct and meaningful way, knowing what to do ahead of time), let us look at the different attributes in Table 4.3. For the most part, we see these attributes as dependent variables or results of different physical features.
3. Our observation about the preponderance of evaluative attributes is very important. Many consumer researchers focus on the description of a package (or product) in terms of how well it is accepted, and what the package is "good for." In the world of research this focus on evaluative and image attributes is under-

Table 4.3 Profiles of three commercially available, shelf-stable chocolate milk products on attributes, after both pure visual and hand evaluations of the packages, respectively

Brand	Nestlé Nesquik	Hershey Hershey's	Land O'Lakes Grip 'n Go
Bottle size	16 oz.	14 oz.	12 oz.
Overall ratings			
Like bottle overall	83	78	75
Overall purchase interest Top Box %	53	41	24
Top-2 box % purchase interest	88	73	62
Appearance attributes (before the bottle is held)			
Like overall appearance of the bottle	84	75	68
Like overall size of the bottle	81	77	67
Like overall shape of the bottle	79	75	74
Like label design on the bottle	82	72	68
Like color scheme used on label	79	74	73
Small versus large size of the bottle	67	64	51
Ease of reading overall text on bottle	85	77	74
Ease of reading flavor name on bottle	82	82	76
Ease of reading brand name on bottle	89	90	79
Overall size of the bottle (too little versus too much)	11	9	0
Bottle in-hand attributes (after the bottle is held)			
Like overall feel of bottle in hand	82	79	78
Like shape of bottle in hand	83	80	79
Like overall appearance of the bottle cap	79	73	65
Like color of the bottle cap	78	75	66
Like overall size of bottle cap	75	72	68
Comfortable to hold	84	82	82
Easy to grip	83	84	83
Ease of reading overall text on bottle	84	78	76
Ease of reading flavor name on bottle	82	81	77
Ease of reading brand name on bottle	90	88	78
Overall color of bottle cap, light versus dark	62	62	56
Overall size of bottle cap, small versus large	63	64	58
Amount of information on bottle (too little versus too much)	10	12	16

Brand	Nestlé Nesquik	Hershey Hershey's	Land O'Lakes Grip 'n Go
Imagery attributes			
Uniqueness of bottle	77	78	71
Cool looking	81	74	68
Fun looking	86	67	76
Quality of product	87	86	75
Good tasting	88	87	76
Miscellaneous attributes (percents)			
Consumer group product is appropriate for			
Children	85	62	90
Teenagers	72	73	55
Adults	48	73	29
Meals product is most appropriate for			
Breakfast	68	64	71
Lunch	55	58	51
Dinner	25	25	21
In between meals	75	74	64
Other	25	25	17
Occasions when the respondent expects to drink the milk product			
At home	65	66	67
At school	60	58	62
At work	46	54	32
For when you are on the go	64	62	51
When you want something good for you	30	32	25
After you play sports or play hard	18	11	14
When you are thirsty	35	28	27
When you want something fun to drink	56	43	45
When you are with your friends	40	33	29
When you want a snack	64	62	55
Other	15	20	14
None of the above	0	0	1
Types of food with which to drink the product			
Desserts such as cookies, cakes, pies, doughnuts	65	66	64
Muffins, bagels, or toast	57	54	51
Cereal	35	31	26
Sandwiches	43	45	41
Pizza	19	17	16
Dinner entrees	22	17	19
Ice cream	21	25	21
Chocolate	15	15	15
Other	32	33	26

standable. Researchers, especially consumer researchers, do not fancy themselves as product developers. Rather, they think of themselves as measuring the opinion of the consuming public. In turn, this consuming public is presumed to focus on "WIIFM" (what's in it for me) (i.e., what does the product do for ME?).

4. Let us identify some variables that we can change. These will be our independent variables. One of these is size. The package developer can change size in a straightforward way.
5. Now that we have identified size as an independent variable that we feel to be actionable, let us look for relations between size and the subjective response. As we stated above, basic and applied research both instruct us that as a sensory attribute increases, liking first increases, peaks, and then drops down. This relation, often described as an inverted U, appears to first have been discovered for liking versus sensory intensity, but may be a general principle.
6. Although we did not systematically vary the size of the bottle, let us see how size covaries with a few other attributes. To discover what is happening, we create a scatter plot. The x-axis is the perceived size of bottle, and the y-axis is the attribute rating. Typically, the data scatter, but there may an underlying curve or line. We're looking for the shape of the relation, even though we recognize it's not necessarily a good fitting curve. Look at Figure 4.2, showing how the perceived size of the bottle drives some other attributes. We fitted a scatter plot to the data and drew the best curve. We don't show the points but show the best fitting curvilinear relation. Remember, we have not done an experiment. Rather, we are trying to figure out what message nature is trying to tell us. There are four patterns that we can deduce from Figure 4.2, each of which emerges from the "fitted relation" between the perceived size of bottle as a consumer rating, and either rating of an attribute or appropriateness.

So, what can we conclude about these data, and thus what do we learn about the dynamics of our milk package?

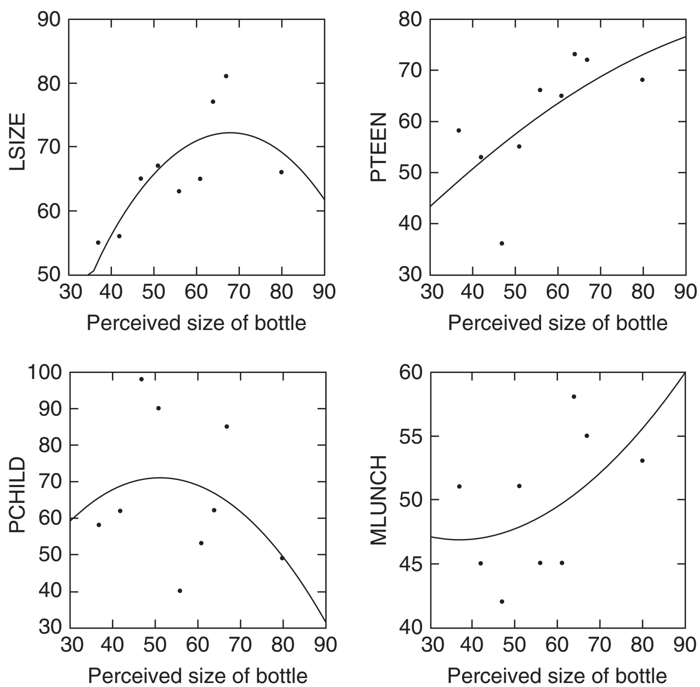


Figure 4.2 How the perceived size of the bottle "drives" other attributes (at least covaries with them)

1. The data do not precisely fit the curve. The reason for the scatter makes intuitive sense. The bottles varied on many features. We are simply plotting two subjective attributes and looking for a fit between them.
2. We can assume the relation is linear or quadratic. If we assume a quadratic, or nonlinear relation, then there is the possibility that the curve peaks somewhere in the middle sensory range. That intermediate optimum is the case for the attribute "liking of size." Bottles above the optimum size don't appear to be liked as much. We might never have uncovered this relation without testing the nine different bottles that we did here. In this first graph, the respondent rated both the perceived size of the bottle and the liking of the size.
3. As the bottle gets larger, more respondents select the bottle as appropriate for teens. Here the respondent rated perceived size, but voted "yes/no" about appropriateness for teenagers.
4. As the bottle gets larger, fewer respondents vote that the bottle is appropriate for a child.
5. As the bottle gets larger, more people vote that the product is appropriate for lunch.

Summing Up—What We Learned and What We Did Not Learn from This Exercise

It is clear that we can learn more from testing several products than we learn from testing one product. With one product we cannot uncover a pattern. Attempts to discover the pattern by looking at how different people respond to the same product are invalid, since there is only one stimulus. The different people are sources of variability. We cannot create knowledge by processing the variability, despite how attractive it may seem.

We also learned that in order to discover underlying patterns we need to have variables to generate these patterns. That is, we need to have at least two attributes, which we plot against each other. Just working with liking or another evaluative criterion does not tell us what we need to know. It helps to have many attributes with which to "play."

Third, the consumer respondent need not rate every attribute on a common scale in order to let the research uncover the patterns. The respondent can rate an attribute (e.g., size of the bottle), rate another attribute (e.g., liking of size), but simply select a third attribute (e.g., vote yes or no for appropriate for a specific situation or type of respondent). The analysis does not differentiate between scale values and percentages.

Fourth, in an ideal case we should have many independent variables to explore. To the degree that we can work with many more continuous variables (e.g., heaviness, length, width, etc.), we will be able to create more independent variables, and thus test for more relations.

Fifth and finally, the data need not be perfect. We can live with variability and imperfect fits. Certainly we would like the data points to lie close to the fitted curve, but the fit need not have to be high; it has to be reasonable. (The definition of reasonable is an entirely different topic, not relevant here.) When we work with many in-market products, our goal should be to discover patterns for future use.

References

Lewis, S. (1925) Arrowsmith. Harcourt, Brace & Company, New York.
Thurstone, L.L. (1927) A law of comparative judgment. Psychological Review, 34, 273–286.

Chapter 5

A Gentle Introduction to the World of Systematics

Introduction

By systematic research, we mean the creation of test concepts or stimuli, comprising systematically varied components. When we test these combinations (i.e., package fronts, concepts, or even actual foods), people inspect or actually eat the product and then rate the test stimulus on different attributes or scales. There is no deep introspection. Rather the action is simply sample and sense, then rate. Pure and simple—nothing more. It is an example of S-R, stimulus response.

With that in mind, let's now explore what we can learn from systematically varying the test stimuli, measuring reactions, and then discovering patterns where they may exist. This chapter lays the groundwork for the rest of the book. So, in that spirit, it might be a good idea to read the chapter twice over, just to get a sense of the tools that are available. You don't need mathematics. You simply need an appreciation of disciplined investigation, also known as science. We call this approach systematics, because that's what it is—systematic exploration.

The Simplest Case—Change One Variable and See How the Person Responds

When we systematically vary a stimulus and measure the reaction, we move beyond simply assessing how people respond, and into *rules* about their responses. Let's look at a typical example from the world of psychophysics, the branch of experimental psychology that deals with the relation between physical magnitude and sensory magnitude. For example, take a simple "set of dots," and vary its density (i.e., its numerosity). Is there a relation between the actual density of the dots and the perceived numerosity of the figure?

We can do that exercise fairly easily today using computer technology. Now, following the notion of systematics, let us create six different levels of numerosity or density, from levels that we know are very sparse to levels that are very dense. Look at Figure 5.1 as an example.

Now we present these six different pictures of densities in randomized orders to a person. The randomized order is important to ensure that the person actually pays attention to the task and doesn't simply "see the progression" and give a mindless pattern of increasing numbers. But we get ahead of ourselves here. Let's return to the task.

The person rates the perceived numerosity or denseness of the package. The more dense the picture (i.e., the greater the number of dots per unit area that our respondent perceives), the higher the number that will be assigned. The real question is what is the relation between the density of dots that we can control as experimenters and the subjective rating of "density or numerosity"?

If you understand this principle of "relations between variables," then you will understand the entire book. We are exploring the relations between the variables of package design and the subjective response of consumers who examine these systematically varied designs and rate them on a scale. Depending on what we vary (the stimulus), and what the consumer rates (the response), we can learn a great deal about how the consumer's mind transforms the package information into a judgment.

Of course, we need not have the respondent rate density or numerosity at all. Rather, we could embed a text or a number in the display of dots and ask the respondent to rate how hard it is to discover and then read the text or number. This approach is a different task. The respondent is performing an action in the presence of a systematic stimulus variation. The person is not rating the perception, but rather attempting to do an assigned task in the presence of different levels of a stimulus.

Packaging Research in Food Product Design and Development Howard R. Moskowitz, Michele Reisner, John Ben Lawlor and Rosires Deliza
 ISBN: 978-0-813-81222-9

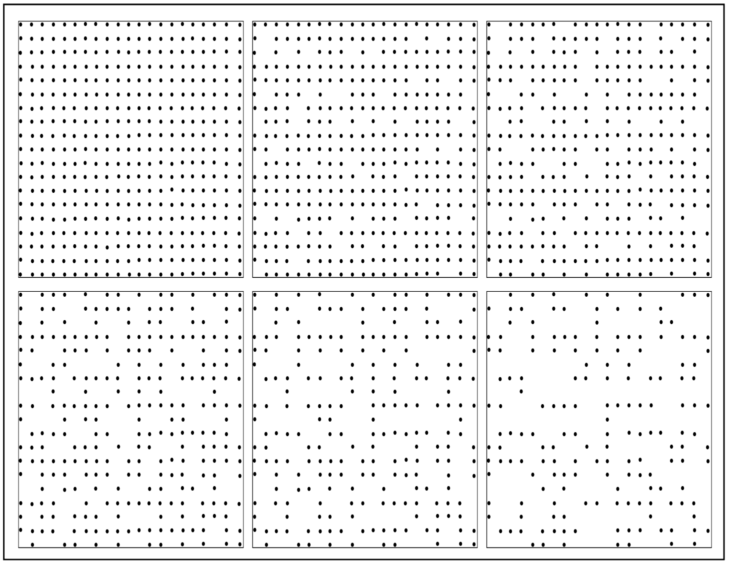


Figure 5.1 Six visual stimuli, showing different levels of "density" of dots.

At this point, you're probably thinking to yourself "Okay, this is nice to know. I'd probably read it in a book somewhere. It's a nice factoid that I'll use at the next party as a conversation opener (or closer). Yet, so what? Why is this information important? What can I possibly do with this piece of information? What can knowledge of these relations do for my practical work?"

Good questions. Perhaps we are working with an e-commerce site and want to put some mechanism into place that prevents a "bot" from reading the information, but yet allows the person to read the information and then type what is read. Now let's imagine that we want to make a set of dots more dense, but not too dense. If we change the physical density, then how dense does it look, and more importantly, how comfortable is it to read a number embedded in those dots? By doing the experiment, we can discover how dense the rectangle should be, to ensure that it is still readable, but that it defeats the bot.

We have just been talking about the world of the experimental psychologist, and particularly the psychophysicist. We have looked at a private sensory experience, and asked ourselves how to change that experience. We know we cannot just add or subtract sound pressures in a willy-nilly way. Rather, there is lawfulness in nature that we must appreciate. Changes in what we present to the test subject result in changes of perception.

Reality Is Complex, and Variables Can't Always Be Dialed

We just finished dealing with the simple case of the experimenter who presents different densities of dots, and instructs the respondent to rate the perception or do a task. What happens if we cannot "dial" the stimulus? Let's take what we have learned and proceed on our journey.

Nature doesn't always present us with this type of wonderful situation, where we can dial or titrate (i.e., systematically change the test stimulus). What happens in a case where we have two alternatives, either present or absent? Let's move out of the world of rectangles, dots, and numbers and move now into the world of package design, where the variables may not be continuous, but rather discrete, and where there may be several options of the variable, not just one or two, not just off or on.

Beyond One at a Time—Looking at Several Variables at Once

Most scientists are educated to look at one variable at a time. In this way, they feel that they better or more clearly understand "nature." That is, they believe that by looking at how a single variable "drives" a response they then "understand" how nature works. This heritage is admirable and pervades a lot of the way people think about the world. In fact, it would be fair to say that much of today's intellectual growth in science comes from this one-at-a-time analysis of variables in the world. The truth of the matter is that most psychophysicists spend their lives understanding the world, one variable at a time.

In the commercial world of design, things are not quite as simple nor are they orderly. Yes, one-at-a-time variation is satisfying, but it doesn't necessarily answer business problems about what to put on labels, what factors influence perception, and what drives the occasionally momentary impulse to buy the food when one is shopping in a store. Although the one-at-a-time method eventually uncovers the key drivers of responses to packages, the strategy is inefficient, and the timelines are just very long. It takes time to do things one at a time.

Most of you who read this book work in the world of business, where the research efforts have financial consequences. Business questions have to be answered quickly. For the most part, these business questions involve a specific goal, such as increased purchase frequency or better communication of nutrition. One variable at a time simply does not do the job, or if it does, then the problem is unusually simple.

When it comes to several independent variables at one time, matters can become complicated. When we deal with only two options for each variable, we might be able to keep things to a reasonable number. The math is pretty easy to do. If each variable has two options, then one variable requires two combinations, two variables require four combinations, three variables requires eight combinations, etc. The numbers don't really start mounting until we reach five or six variables, at which time we have 32 or 64 combinations, respectively. The rule is simple—with two options for each variable, we will have 2^N combinations to test. When N is large (i.e., many different variables to explore), 2^N becomes very large. The task becomes even more daunting when instead of two options per variable we have three options. Thus, we might have three colors for a package, three different labels, three different pictures of a food, three different sizes, etc. For N variables, each with three options, we have 3^N combinations.

For the past 70 years or so, statisticians have been quite involved in this issue of multiple stimulus testing, especially when the test stimuli are systematically varied (Box, Hunter, and Hunter, 1978). Although statisticians can provide valuable advice about testing the many different combinations (i.e., allocating a fractional part of the total set to each individual in a systematic but efficient way), the statistician really shines when it comes to designing the combinations in the first place.

Most statisticians working in the field of product and packaging development are familiar with the methods of "experimental design" (see Montgomery 2005; Ryan, 2007). Experimental design is a branch of statistics that lays out the different combinations. Few rational researchers are so daring and oblivious to cost when it comes to testing many combinations when they can get by with fewer stimuli, better varied, so their research is more cost-effective. Experimental design provides just such a solution. Indeed, we might say that experimental design actually "saves the day" and moves beyond finding answers through testing to uncovering rules that makes the developer, the designer, and the marketer far smarter. The actual evaluations look like tests, and they should because that's what they are. It's the disciplined thinking and disciplined experimentation that creates the true base of knowledge.

Beyond Tables to Models

You will see that we have progressed from considering one variable that is "continuous," to considering several variables all at once. Furthermore, if you are like most people, you probably get a bit overwhelmed by a table of numbers. This is to be expected—people are not constructed to look simply at numbers, but rather to look for patterns in the numbers. We tried to find some patterns in the previous paragraphs, such as "Do the more compound pictures with multiple variables perform better than the simpler pictures having only one variable?"

There must be a better way, and there is. We don't have to stay with columns of numbers in a table, looking for a pattern that nature is trying to reveal. Statistics can help here. Let's introduce the notion of "regression analysis" (Wikipedia, 2008), also commonly referred to as curve fitting, although in the case of an on/off relation, the idea of a curve doesn't really fit, but the approach of curve fitting actually does quite well (Arlinghaus and Arlinghaus, 1994).

Regression analysis is a branch of statistics, often called model building. Regression analysis looks for the relation between one or more independent variables, and a dependent variable. Those of you who have taken a statistics course probably will remember the relatively

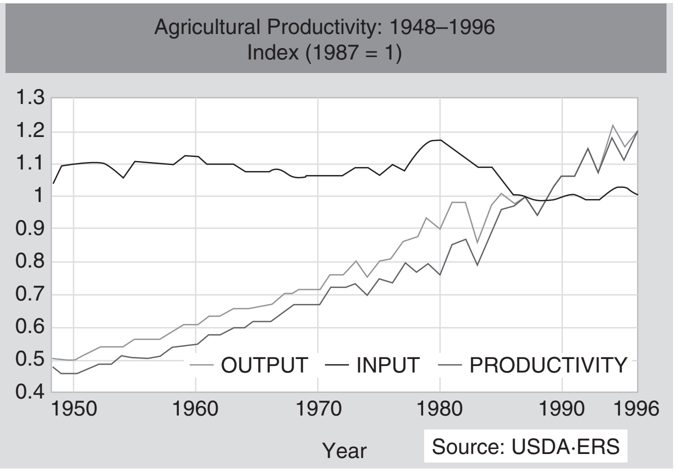


Figure 5.2 How agricultural productivity changed over a 48-year period, from 1948 to 1996

simplistic yet instructive example that most introductory statistics courses give to explain the idea of regression. Let's look at an example of regression, this time from the U.S. Department of Agriculture. We see the data in Figure 5.2. The independent variable is year, starting in 1948. The dependent variable is a measure of relative productivity, with 1987 normalized to 1.0.

Once we plot the data, how then do we make use of it? What type of question should we ask? The figure itself simply retells the table of data. There is a bit more, however. When we plot the data, we can see the nature of the relation. We see that over the passage of time, starting in 1948 there is a systematic rise in the productivity of agriculture. We could look to any year and find its relative productivity simply by keeping our finger at the year (abscissa or x-axis), and moving upward until we find the data, and finally moving leftward to the ordinate (y-axis) to discover the relative productivity.

We want to go further, however. We want to create a model or equation that shows us the numerical relation between the year and the agricultural productivity. To do this, let us move out of the world of plotting data and into the world of regression.

First, let us look at the actual data from which the curves in Figure 5.2 are drawn. Happily for us as readers and analysts, the U.S. government, specifically the Department of Agriculture, publishes these numbers. They can be found at the same website as that from which Figure 5.2 is taken. We see some of these data for the first five and the last five years in Table 5.1.

Table 5.1 Data about agricultural inputs, outputs, and productivity. The data are shown for the first five years and the last five years only.

Trends in U.S. Agriculture, published by United States Department of Agriculture—National Agriculture Statistics Service

Index of Agricultural Productivity: 1948–1996

Source: USDA—ERS

Year	Output	Input	Productivity
1948	0.507	1.035	0.490
1949	0.507	1.097	0.462
1950	0.503	1.094	0.460
1951	0.527	1.108	0.476
1952	0.540	1.107	0.488
1992	1.137	0.991	1.147
1993	1.071	0.997	1.074
1994	1.217	1.025	1.187
1995	1.153	1.038	1.111
1996	1.202	1.009	1.191

We can learn a lot by plotting the data, but there is more. Suppose we want to develop a model showing the expected change, say in output, as a function of the number of years since the analysis began. Let us call 1948 year 1, 1949 year 2, etc. Now, looking at the data in Table 5.1, let us relate the number of years to the output, by the simple equation: Output = k_1(Number of years) + k_0. This is a simple linear equation. The results appear in Table 5.2. It says in words:

Table 5.2 "Linear" regression analysis that fits a straight line to the relation between agricultural output (dependent variable) and number of years since 1948 (independent variable)

Dependent Variable: OUTPUT N: 49
Multiple R: 0.983
Squared multiple R: 0.967
Adjusted squared multiple R: 0.966
Standard error of estimate: 0.038

Effect	Coefficient	Standard Error	t statistic	P(2 Tail)
Additive Constant (k_0)	0.436	0.011	39.353	0.000
Years since 1947 (k_1)	0.014	0.000	37.168	0.000

1. The output is a linear function of the number of years.
2. Furthermore when year = 0 (i.e., 1947), we can expect an output value of k_0.
3. Finally, for each year, we expect a constant increase in output equal to k_1.

We will use the standard statistical packages for regression. Let's unpack the figure to understand what the statistics mean. Our analysis will be helpful in the future when we look at the effects that different features exert on the perception of packages.

1. The dependent variable is "output." The economists measured the agricultural output, in relative units, and gave that data in Table 5.1.
2. The number of "cases" or observations is 49 (N = 49). In Table 5.1, we see only 10 of the 49. However, when it comes to analyzing the data and building a model, we use all 49 observations.
3. The goodness of fit is shown by the multiple R. The multiple R shows the degree of linearity. The multiple R ranges from a low of 0 to a high of +1.00. We have a very good fit, indeed, almost a perfect fit. The multiple R is 0.983.
4. The square of the multiple R shows the proportion of the variability in the dependent variable (output) that can be accounted for by knowing the value of the independent variable (number of years). The squared multiple R, 0.967, means that almost 97% of the variability can be accounted for by knowing the number of years. In other words, output has been steadily increasing. In consumer research, we will see lower values for the squared multiple R.
5. The additive constant is 0.435. This value can be found in Table 5.2 in the column marked "coefficient." We interpret this constant to mean that at time 0 (i.e., 1947) we expect the agricultural output to be 0.435. Of course, we did not measure the output then, since the data start at 1948. Nevertheless, because we have a linear equation, we can estimate the value of that equation when time is 0 (i.e., when the year is 1947). Notice that the regression analysis comes out with a coefficient value with three significant digits. This is purely mathematical. The regression modeling can estimate the data to 20 or more significant digits. However, the reality is that for most cases we would use at most 1 significant digit.
6. The coefficient for the single independent variable, k_1, is 0.014. This means that the output increases by 0.014 units for each year since 1947. Thus, if we look at a four-year period, from 1947 to 1951, we can expect (0.014 = coefficient) × (4 = number of years since 1947). This is 0.056 units. Notice that once we have this coefficient, we have a sense of the rate at which agricultural output increases for each year. The goodness of fit need not be so high. Even if the multiple R^2 were lower, say approximately 0.60 (i.e., 60% of the variability in the output accounted for by the number of years), we would feel comfortable that we somehow have a "handle" on how fast the output grows for each year. It is this sense of learning, of rules, that makes the analysis so gratifying, and leads to an increased satisfaction that we know what is really occurring, rather than just plotting the data.
7. The standard error tells us the variability of this coefficient or additive constant. If we were to run the study again, and do the analysis, then based on these data, we would expect the coefficients of the equation to vary. About 68% of the time we would expect to see a coefficient between the mean ± 1 standard error. The standard error is 0.011 for the additive constant, so that about 68% of the time we would expect the additive constant to lie between a low value of 0.425 and a high value of 0.447 (corresponding to 0.436 ± 0.011). For the coefficient for "years," the standard error is almost infinitesimal, so the computer output shows it to be 0.000. Of course, if we were to extend the results to, say, 10 decimal places, we would see a non-zero value for the standard error.

8. The "t" value is the student "t" statistic. The t value is defined as the value ((coefficient − 0)/standard error). The t value has a sampling distribution. That is, for any t value, we know the probability of getting that t value if the coefficient were truly 0 rather than what we observe. The probability that the constant or coefficient is really 0, rather than what we observe, is infinitesimally small. The "t" is very high, so the probability is virtually nil that we are seeing a random fluctuation from a true mean of 0.

Extending Our Approach to the More Simple Case—Present/Absent

Let's now move forward with our understanding of modeling. We will move out from the world of continuous variables such as year, which take on a stream of values such as 1–48, and move into the world of "on-off" or "yes-no." This world is more appropriate for package design, where we deal with the presence/absence of features on a package. It's a rare case when we can systematically vary one variable over a wide range, to look at the equation relating the size of that variable (i.e., size of logo) to the rating (i.e., interest in buying the product based on the package).

The more typical situation is a package that comprises several silos (variables), with each silo comprising, in turn, several options (elements). The number of silos may go from as few as one (i.e., presence/absence of a logo) to a dozen or more (logo, color, burst to show "new/improved," price, color of background for price, etc.). The types of silos are endless, limited only by the imagination of the designer and, of course, the space on the package itself. The more complex case, but not necessarily more difficult in the long run, will comprise several silos, and different numbers of options for each silo. We will look at an approach to solve the problem of "How does each element in each silo drive the response?" in this more complicated situation. And, as a bonus, this straightforward approach that we outline will be used in the rest of this book to help us learn rules about package design.

Arrays of 1s and 0s—A Useful System to Represent the Combinations

When we deal with these complicated problems of many different variables, each with different combinations, a good strategy begins with the ultimate analysis in mind. Our analysis will be the so-called "dummy variable regression." Dummy regression refers to the nature of the independent variables, which take on only one of two values. If in a test stimulus (i.e., package design, concept, etc.) the element is present, then the element is represented by the value 1. In contrast, if the element is absent from the test stimulus, then the element is represented by the value 0.

The representation of 1 and 0 is not done simply as a way to show presence/absence. Rather, the representation will allow us to use these two numbers as the values of the independent variable. There is a simple logic operating behind the scene here. Let's return for a minute to our example about agricultural output versus year. The equation is written as:

$$\text{Output} = k_0 + k_1\,(\text{year})$$

Recall that the coefficient k_1 shows us the expected change in output for each change in one year. So when $k_1 = 0.15$, we expect output to change by 0.15 units when we go from year 1 to year 2, and the same 0.15-unit change when we go from year 2 to year 3, etc.

Now imagine that we are dealing with package design, rather than with agricultural output. We have a database like we had for Table 5.1, but this time the independent variables are design elements. These are the elements A, B, and C. The three design elements can either be present or absent. We see the coding of the eight different combinations, as well as the percent of respondents who rated each package design as communicating "healthful" (rating of 7–9 on the 9-point healthfulness scale). See Table 5.3.

Table 5.3 Dummy variable, systematically coding eight package designs, and the rating for healthfulness associated for each

Package Combination	Element A	Element B	Element C	% Top-3 Box on Healthfulness
1	1	1	1	46
2	1	1	0	24
3	1	0	1	38
4	1	0	0	37
5	0	1	1	62
6	0	1	0	46
7	0	0	1	39
8	0	0	0	53

Now that we have represented the stimulus in a simple format, with 1s and 0s, let us apply the method of least-squares regression that we used previously for the agriculture data. This time, however, we are going to use three predictors, not one. They are the three design elements. These predictors (i.e., design elements A, B, and C) take on only two values: 1 if present in the package design and 0 if absent. Our computer program for regression doesn't know that. It only knows that it is dealing with three "continuous" variables as independent variables and one continuous variable as the dependent variable. Actually our independent variable takes on only two values, 0 and 1, respectively, but there is no way for the regression program to know that.

When the regression program runs the data, it comes up with the equation that is very similar to the equation we saw in Table 5.2. That is, we have an additive constant (k_0). This constant corresponds to the rating that we would get for the combination if we were to work at the 0 or "absent" level for each of the three variables.

Let's now go a step further. Let us look at the three coefficients, one for each design element. Remember that the coefficient tells us the effect on the dependent variable (% top-3 box for the rating of healthfulness) corresponding to a 1-unit change of the independent variable.

But, just exactly what is a 1-unit change of the independent variable in this case? Well, it is simply going from 0 (the element not present in the package design) to 1 (the element present in the package design). *So, by coding the elements as binaries (0, 1), and by using ordinary least-squares regression, along with the proper experimental design (combinations of factors), we can determine what each of the three design features contributes to the rating of healthfulness!*

Let's look at what we discover when we do this more complex case. We will see the results in Table 5.4. It's always easiest to list the discoveries in some type of block order, rather than try to piece together prose. So, without any further ado, let's proceed.

1. *Baseline:* The additive constant is a measure of basic interest. It is the *expected %* top-3 box for healthfulness in the case that none of our three design elements are present. Clearly the concepts had elements, for the most part, but there is one combination that has none of the added elements. The additive constant shows the expected rating of that combination. That is, the constant shows what that single combination with no design element should achieve, based on the pattern from all eight package designs. The value from Table 5.4 is 45.5, meaning that without any of our three design elements, about 45.5% of the ratings for that particular design will be 7–9. Of course, this means that the remaining 54.5% of the ratings will be 1–5.
2. *Elements analyzed for separate contributions:* We are looking for the separate contribution of each of the three elements. We treat these elements as being statistically independent of each other, which they are by virtue of the experimental design. When we do, we discover that each element makes its own contribution. The elements each do some work, with element A subtracting from the perception of healthfulness, and elements B and C adding to the perception of healthfulness.
3. *Creating a combination and estimating the total healthfulness:* The impact values or utilities show us the expected rating. If we want to, we can take any combination, add together their separate coefficients, and then add in the constant. The resulting sum is our best guess about the rating of that particular combination. Of course, you might say "Why bother, when we just tested ALL COMBINATIONS!" And of course, you would be correct. However, this is the simplest of cases. What happens when we have six categories, each with six elements, as we often do? We have 6^7 or 279,936 combinations. The reasoning is simple. Each variable has six nonzero options and a seventh zero option. The total number of combinations that a person would test would be the 6^7. In such a situation we will use a fraction of these combinations. The full set of combinations would be impossible to examine in a reasonable time frame.
4. The dependent variable can be either ratings or some derived measure from the ratings. Here we used the

Table 5.4 Results from analysis of a package design comprising three silos, each with an element that can either appear or not appear (be absent)

	Model for percent respondents rating the package as 7–9 on the 9-point scale (top-3 box)
Additive constant	45.50
Element A	−13.75
Element B	2.75
Element C	5.25

top-3 box, a derived percentage. We could have used the mean rating of healthfulness instead. In the end, the patterns are pretty much the same if we use a continuous scale versus percent at the top of the scale.

5. There will be some occasions when the scale points themselves are not on a continuum, but represent different qualitative responses. We will see such a situation in the case of emotions, when we use seven scale points to denote seven different emotions. In that case, we will see an entirely different treatment of the data. We present that approach later in Chapter 23.

Lest we overlook the importance of this simple approach, it's best to return to it right now and summarize the major points. Understanding those major points will let you understand the rest of this book, and show you how to develop a deep insight into what the features of package designs contribute to consumer perceptions.

1. *The variables and levels:* We code the elements as either present or absent. When present, we put a 1; when absent we use a 0. These 0 and 1 values represent the components of the concept, and are used as numerical predictors in the regression analysis.
2. *Recipe:* Each package design is a formula or recipe with 1s and 0s.
3. *Elements are important:* In the analysis, it will be the specific elements, not the combinations, which will be the focus of our attention. That is, in the modeling, the elements will be the independent variables. The test stimuli (i.e., package designs or concepts) are used as a means to embed the individual elements, in a layout dictated by the experimental design.
4. *Incomplete designs are a definite advantage when combined with regression modeling:* A specific package design can have, at most, one element from any silo. However, in many situations, the package design will be absent all elements from a silo (if physically possible). Thus, when we deal with health messages, some package designs will be absent any element from the silo of health messaging. This "true" zero condition is important. By having combinations that are absent, it will be possible to estimate the absolute utility of each element.

References

Arlinghaus, S.L. and Arlinghaus, W. (1994) Practical Handbook of Curve Fitting. CRC Press, Inc., Boca Raton, FL.

Box, G.E.P., Hunter, W.G., and Hunter, J.S. (1978) Statistics for Experiments: An introduction to Design, Data Analysis, and Model Building. John Wiley and Sons, NY.

Economic Research Service, United States Department of Agriculture —National Agriculture Statistics Service, Washington, D.C. Publications available from http://www.usda.gov/nass/pubs/trends/productivity.htm [accessed 11/22/08]

Montgomery, D.C. (2005) Design and Analysis of Experiments. John Wiley and Sons Inc., New York.

Ryan, T.P. (2007) Modern Experimental Design. John Wiley and Sons, Inc., New York.

Wikipedia, http://en.wikipedia.org/wiki/Regression_analysis [accessed 11/22/08].

Further Reading

Kelton, W.D. (2000) Experimental Design for Simulation, R. R. Joines et al, Eds. In: Proceedings of the 2000 Winter Simulation Conference.

State University of Munich, Euroworkshop on Statistical Modelling—Model Building and Evaluation, 31 October to 3 November, 2002.

Chapter 6

Identify What Works by Letting the Competition Do the Work

Introduction

By now, we have gone through the first part of our journey together, looking at packages and graphics while discovering what wins and what loses, and also delving a bit into the mind of consumers. We found out that it is quite easy to test one stimulus with lots of people and even ask them a lot of questions. Certainly, we learn more by asking one respondent many questions than by asking one question; that's obvious. We get to different aspects of the one stimulus. However, the downside to this effort is that, at the end of the day, we don't learn much.

We learn a lot more by testing several stimuli, rather than testing the proverbial "best shot." That is also obvious or at least will become obvious as we travel together through this book. Only by having respondents assess multiple stimuli on multiple attributes can we uncover patterns in the data that tell us about the mind of the consumer, and in turn, something about how to make a better package or product.

We learned something else as well. Not all attributes are created equal. We learned, or better posited, that some attributes are more "actionable" than others. By *actionable* we mean that the package or product developer knows what the attribute *means* in terms of some physical change to the stimulus. Guidance from the consumer responses (e.g., that a certain "sensory level" is best), translates into a reasonably direct, and intuitively obvious, change in the stimulus. Size of the package is just such an attribute. It's fairly easy to get a sense of what attributes are actionable and useful. Just ask yourself whether someone who works with the data knows "precisely" what to do to the "stimulus" after being presented with the specific attribute and the rating of one or several stimuli on that attribute. If you cannot decide what you must do, even with the data from the attribute, then chances are you are dealing with an *inactionable* attribute.

Unfortunately, in the case of packages and graphics there are not as many actionable rating attributes as we would like to have. Packages vary on a relatively limited number of sensory "continua" such as size, heaviness, darkness of color, size of print. Each of these attributes is meaningful to the package designer. But the language of design that describes the characteristics of a package is not particularly rich in such continua. Perhaps the poverty of the language comes about because people can do a lot with a few sensory continua such as length, width, heaviness, and some appearance attributes. We don't need to build up a rich vocabulary in order to describe our visual sensory experience. All the work is done because we can simply point to a feature of a product. In contrast, we possess a significantly richer language of taste/flavor, with a plethora of descriptors. Perhaps it's because taste/smell don't have many well-defined physical continua, and so we must rely on the sensory perception of a taste/smell attribute to guide us. It might turn out, however, that properly described package stimuli may yield far more of these actionable attributes, beyond those related to package size. Perhaps, we just haven't looked for them.

Category Appraisal—Using the Competitor Features as Guides

Let's move in a new direction. Whereas in a previous chapter dealing with milk we worked with the continuous variable of size, now we move to deal with the *presence/absence of specific features*. That is, we don't concern ourselves with finding the correct "continuous variables" for design, of which there may be very few. Rather, we *deconstruct* the competitive frame into specific classes of features, a task that may be much easier. These features now become the predictors of responses, rather than the harder-to-identify physical variables.

Packaging Research in Food Product Design and Development Howard R. Moskowitz, Michele Reisner, John Ben Lawlor and Rosires Deliza
 ISBN: 978-0-813-81222-9

When we think about the shift in mind-set from continuous variables to features, we will probably soon realize that this shift is intuitively obvious. The designer works with features, not with variables. The features are like the subjects that a painter chooses, the objects that are represented. It is the features that give the package its life, its uniqueness, and the perception of the use for which the package is deemed most appropriate. There are classes of such features, and what could become an infinite variety of these features within any class.

Let's look at a study of margarine packages done in the early 1990s. The objective of the project was simple—identify how different features of a package drive consumer perceptions for a possible margarine package. If we step back almost 20 years, we can get an idea of the thinking that worked then and continues to work. If you had gone to the supermarket, you would have been accosted by many different margarine packages, some classic bars, others tubs, and others in packages that might be called "second generation tubs." It wasn't clear where there were new opportunities or what features of the margarine package were "effective."

We went a bit further in that project. We needed to look at many different features, which at that time the margarine packages simply did not possess. We looked at packages with different features, but the packages were from other products, not from margarine. Once you remove the brand identification and call the package "for margarine," you quickly discover a wide range of alternatives that could do the job of containing the margarine product. But what are the relevant features?

Looking at these data and our conclusions two decades later, we have the continuing benefit of a very rich database that allows us new discoveries, even now. We look for relations between variables, so we can learn about the dynamics of the package.

Our consumers profiled 67 different margarine (and nonmargarine, but appropriate) packages on a set of 29 attributes. It was clear that the packages differed from each other. All you had to do was look at the range and the variability of the 67 packages' attributes to realize that consumers differentiated quite easily (see Table 6.1). The real goal, however, was to discover what drives this differentiation. Specifically, what rules could we discover?

Let's look a little more closely at Table 6.1 to get a sense of what a competitive analysis can provide. The data we present show only the variation across the 67 products for the different attributes. Looking closely at them, you see that the study used the products to create a wide range of different subjective impressions. We have four types of attributes, which we list in the first column:

1. *Sensory:* Amount of an attribute. Here the respondent acts as a measuring instrument, to tell us how much of a characteristic is present in the margarine package. Of course the judgments are subjective, not objective. Yes, the respondent tries to act as a measuring instrument, but we must always keep in mind that the respondent filters the reality through the lens of personal experience. Nonetheless, for the most part, respondents accurately "measure" sensory experience such as the brightness of light, or the perceived area of the circles, etc.
2. *Liking.* The respondent can evaluate the package overall (overall liking), or evaluate different and specific parts of the package (e.g., like the appearance, or even more specifically, like the size). It's not clear from the thousands of research studies conducted every year how much more we truly learn from attribute liking in addition to overall liking. Often when a respondent likes the entire package, the respondent will say that he likes the individual attributes. This is called the "halo effect"—"*If I like it, then I say that I like most things about it*." If the respondent likes the product but dislikes all of the specific aspects, then the consistency of the respondent's data is in doubt.
3. *Performance.* Performance ratings show how the product actually "does" when it is put to the test in actual use, or at least how the respondent feels the product will actually "do." To some extent there may be an evaluative component in performance, but that evaluative component does not predominate. Examples of performance attributes are "*easy to open*" and "*easy to close*." These attributes can be judged by actually using the product, or in the case of package, by opening and closing the package, even in the absence of any product inside. The opening and closing give the respondent a sense of doing something with the package, interacting with it, and allows an experience-based opinion.
4. *Image.* Image characteristics are more complex perceptions. The image characteristic integrates a person's experience with products, along with expectations. Thus, the respondent might look at the package, with the brand name or label removed, and

Table 6.1 Consumers differentiate margarine packages on the basis of perceptions, and rate these packages on sensory and image characteristics. The table shows the average data and range of each attribute, based on responses to 67 different margarine packages.

		Mean	Maximum	Minimum	Range
Liking	Like overall	51	83	11	72
Liking	Like appearance	54	83	14	68
Liking	Like size	56	80	16	64
Sensory	Long	47	80	14	66
Sensory	Wide	41	72	18	54
Sensory	Heavy	26	48	11	37
Sensory	Small versus Large	40	93	7	87
Performance	Easy to open	70	90	22	68
Performance	Easy to close	69	92	24	68
Liking	Like the seal	49	90	7	82
Performance	Will spill	47	86	15	70
Liking	Quality of package	52	87	10	77
Liking	Quality of closure	49	91	9	82
Image	Package appropriate for margarine	53	95	9	86
Image	Package appropriate for ice cream	32	93	2	91
Image	Package appropriate for yogurt	28	80	3	77
Image	Unique	49	90	16	74
Image	Expensive	42	72	21	51
Image	Fresh	54	86	12	74
Image	Easy to read	71	89	37	52
Liking	Like shape	63	84	31	53
Image	Contemporary	47	74	19	55
Image	Caloric	49	61	41	20
Image	Nutritious	46	55	32	23
Image	Easy for a child to use	64	91	20	70
Image	Appropriate at table	59	86	15	71
	Appropriate for microwave	39	71	7	64
Performance	Clumsy to use	56	86	15	71
Image	Expect product to taste good	50	72	21	51

form judgments about how this particular package would fit a yogurt container, instead of a margarine container. The respondent has an idea of what a yogurt container should be. Therefore, it's not particularly difficult for the respondent to assign a rating to "appropriate for a yogurt." Researchers and package designers may have no idea why the respondent feels that one package is more appropriate than another for yogurt or for margarine. We will discover the answer to the "why" question, however, when we create the models relating the different features of the product to this image attribute.

So, Where Do We Now Travel on Our Voyage of Discovery?

If we would just look at some of the source data in Table 6.1, then we would see a rich vein of data to mine in our quest to discover what drives the response to packages. Yet, are we missing anything? Are summary data sufficient? No, they are not.

Let's go a step further, and look for relations between variables (Moskowitz, 1994), as we have done before in the previous chapter. We know that we could learn some things from a response-response analysis, where the independent variable is a sensory attribute (e.g., size of container), and the dependent variable is a liking, image, or performance attribute. We know about that type of response-response (R-R) analysis from the previous chapters. R-R analysis is insightful, but, at the same time, may not be particularly actionable.

Just for fun, let's look at one particular sensory-image relation to get a sense of what we might get, where it leads us, but also what we are missing. Quite arbitrarily, we will look at perceived heaviness on the ordinate, and perceived expensiveness of the package on the abscissa. This is a specific example of an R-R analysis. Parenthetically, the term R-R analysis comes from experimental psychology. The rationale for the name R-R analysis is simple—we are examining the probable relation between two attribute responses. The attributes "heaviness" and "expensive" are responses, even though someone might argue that "heaviness" is an objective characteristic of the product. It is not; heaviness is a response.

R-R analyses make intuitive sense when we plot the data, and more often than not teach us a great deal. We look now at how heaviness drives the perception of expensive. Is there a relation? What is the nature of the idealized relation? Is it straight linear upward at 45 degrees, downward, inverted U curve, or no relation at all? What is the optimum heaviness to create the highest level of expensive?

Anyone who engages in these types of analyses learns far more in doing the analysis than we might be able to express in our book. A word to the student, scholar, researcher—it's worth doing R-R analyses!

We see this sensory-image curve in Figure 6.1. We have seen this type of response-response curve before. The key difference is that now we based our graph on the evaluation of 67 different packages. The curve that we fit through the data comes from the analysis by readily available statistical programs (e.g., Systat, 2007). We fit a curve through the data to give us a sense of what type of idealized relation might exist.

Of course, the best-fit curve (here the quadratic fit) simply describes what might be happening, but doesn't prove anything. To prove the relation, we would have to change heaviness systematically (or some physical correlation of heaviness) and measure perceived expensiveness. Even in that case, we are not sure because we "operationally" do not know what heaviness really is in the mind of the responder. We think heaviness is weight, but it may be density, etc.

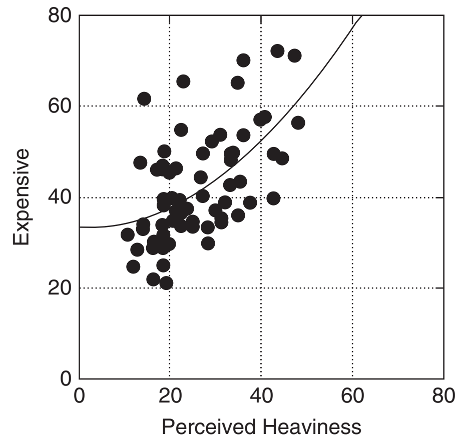


Figure 6.1 How perceived expensive, an image attribute, covaries with perceived heaviness of margarine packages, a sensory attribute. Each circle is one of the 67 margarine packages.

"Inferring Structure"

How does the package designer work with the information? Or, as many would say, is it all just good guesswork? Certainly looking at Table 6.1 and Figure 6.1 do not lead to direct prescription of what to do, no matter how insightful the researcher might be.

Deconstruction and Content Analysis

Another approach might work, taken from the world of experimental design. This other approach "deconstructs" the current product set into a limited set of physical variables. The strategy is known as a *content analysis*, for obvious reasons. We analyze the content of current stimuli. Here the packages are described as having contained margarine, but originally the package might have been designed for another food.

Let's look at Table 6.2, which shows the "content" (i.e., the presence/absence of specific features) for the first four packages. The brand name and other identifying

Table 6.2 Deconstruction of four margarine packages into components and the average ratings of those four packages

Features discovered during the content analysis	Package			
	101	102	103	104
Shape: Round, oval, cylindrical	1	1	1	1
Shape: Square/hexagon	0	0	0	0
Shape: Rectangle	0	0	0	0
Shape: Tube	0	0	0	0
Size: Small	0	1	0	1
Size: Medium	1	0	1	0
Size: Large	0	0	0	0
Material: Paper	0	1	0	0
Material: Plastic	0	0	1	1
Material: Paper/plastic	1	0	0	0
Appearance: Clear	0	0	0	1
Appearance: Opaque	1	1	1	0
Height: Tall	0	0	1	0
Height: Medium	1	0	0	0
Height: Short	0	1	0	1
Lid: Paper	0	0	0	0
Lid: Plastic	0	1	1	1
Lid: Paper/plastic	1	0	0	0
Closure: Cap	0	0	0	0
Closure: Lid	1	1	1	1
Cap/Peel inner seal	0	0	0	0
Liner: Absent	1	0	0	1
Liner: Present	0	1	1	0
Lid Appearance: Opaque	1	0	1	1
Lid Appearance: Translucent	0	0	0	0
Lid Appearance: Clear	0	1	0	0
Examples of Rating Attributes				
Overall Liking	34	31	67	64
Like Appearance	37	35	72	72
Like Size	55	22	63	62

graphics features were removed. In some cases, the actual product name was stamped on the container. These were filled in or rubbed out. Although the aesthetics may have been ruined, the brand identification was deemed to be so strong that the judgment was made to eliminate brand despite loss of attractiveness.

The hardest part of doing a content analysis is the up-front work, about deciding what particular silos of features one will search for, and the criteria for the presence of specific elements in each silo. Remember that the researcher attempts to create a structure out of the natural array of different features. Thus, one might have a continuum of sizes, but in the spirit of the content analysis, one might wish to divide this continuum into discrete elements. In most cases, it will be simple to deconstruct the package into components, but in others it will be more difficult because the criteria are subtle.

In Table 6.2, we have tried to list all of the features that we discovered in our content analysis. We may have missed some. There is the perennial tug between "granularity," where we go into the "innards" of the company, and external analysis, where we just skim and look at the high points. External superficiality does not work as well as granularity, but granularity may not work because some features appear together, so they cannot be disassociated. Choosing which features to include in the model turns into a "judgment call," a subjective exercise.

Each specific package possesses some of these physical features but lacks other features. The researcher should inspect each package, tick off what features are present, and by definition what features are absent. Usually the researcher does not try to ensure that the different features are balanced so that they appear equally often. Typically that "statistically admirable" approach to design simply cannot work and the researcher has to make do with what the different, competing companies have chosen to offer to the market.

When we use the above mentioned approach to deconstruct the packages, we end up with a content analysis that, on the surface, looks surprisingly like an experimental design, except that the elements may not appear equally often, and some elements may correlate with others. Experimental design would take care of both of these issues, forcing elements in the same silo to appear equally often in the full set of combinations, and ensure that the elements do not correlate with each other. Table 6.2 gives you a sense of this deconstruction.

Now that we have bootstrapped ourselves to a structure, it is time to uncover patterns that might be lurking in the data. With the total of 67 products, we should imagine 67 columns of data, not the four columns of package data that we see in Table 6.2. With 29 rating attributes, we should imagine 29 more rows of attribute ratings, not the three that we show (overall liking, like appearance, and like size).

Occasionally some elements in the deconstruction covary so much that knowing one of these elements informs us what the other element will be. For example, if the presence of a specific lid always appeared with one specific shape of the package, then we could not use those two elements together in the analysis. They correlate with each other too much. Knowing that the specific lid is present tells us that the package will have that certain shape. In such cases, the researcher should use only one of the two elements in the analysis and discard the other element as being redundant. We won't get into the specifics here, except to mention that correlated variables are very common when we deconstruct the competitive array. If we don't systematically vary the elements ourselves, then we have no control over what we test. Our only option for control is to select specific packages or decide to look only at specific elements, ignoring the redundant elements in our analysis.

Letting Stepwise Multiple Regression Analysis "*Automatically*" Uncover "What's Working"

Regression analysis plays a large role in this book, and it will play a correspondingly large role in this chapter. Regression analysis, also called curve fitting, reveals the quantitative relation between one or several independent variables simultaneously, and the dependent variable. With this overview in mind, let us apply regression analysis to our data, to see what works.

We first lay out our knowledge base. What specifically do we know from our deconstruction? By this action, and in subsequent steps, we quickly arrive at the answer to our question, not only knowing what works, but the degree to which each element "drives" every consumer attribute. The key here is systematically approaching the data set as a system of variables whose working is to be understood through analysis.

What Variables Do We Select, and Why Can't We Use All the Elements as Predictors?

For our regression, we first try to use the full set of 29 different elements that we "created" during our content analysis. However, we really cannot use all 29 features as independent variables in our regression equation. There are two reasons. One reason is that we do not want to work with "sparse data." In the deconstruction of the features, some features only appeared once or twice. We cannot really create a model using those rarely appearing features. Of course, we don't have to use all of the variables that we deconstruct. So, happily, we can choose the variables that we wish to omit, which we see in Table 6.3. We have deliberately eliminated some of the more unusual features that appeared too infrequently.

The second reason is that the variables that we use are not independent. For example, there are four shapes. Knowing the condition of three shapes (e.g., all are absent) automatically tells us the condition of the fourth. Let's look at what happens when we do. We will attempt to "predict" the rating of "overall liking," from knowing the presence/absence of each of the elements in the 67 containers. The approach makes intuitive sense. We should be able to uncover the relation, and thus discover what particular elements drive acceptance.

Look at the results of this very first regression analysis in Table 6.3. We used one popular statistical package, Systat. We would get the same results if we used other regression programs as well. Table 6.3 shows us that we are able to relate overall liking to features, but only somewhat. We can get "contributions" to liking, but we see something strange. A number of the elements "fail to enter the equation." We are left with the cryptic message ".", meaning that the regression modeling could not estimate the contribution. That's nothing to worry about. Systat estimates the contribution of what it can estimate in a statistically valid way and leaves the remaining features "unmodeled."

Once we understand how to read the results of the modeling, we quickly discover what every element contributes, as well as what we cannot learn from this particular analysis. We now go step by step to unpack the insights that the regression provides. We will refer to the results in Table 6.3, but the same type of interpretation will apply when we look at the results of any regression modeling that uses this approach.

1. *Dependent variable*. This is what we try to predict. In our case, we try to predict the rating of overall liking, which respondents rated using an anchored, 0–100 scale (0 = hate → 100 = love).
2. *N*. This refers to the number of cases or observations on which we are basing the analysis. For our margarine package study, we deal with 67 different packages, so N = 67 cases.
3. *Multiple R*. This is a measure of goodness of fit. The multiple R ranges from 0 (no fit at all) to 1.0 (perfect fit). We have a high multiple R (0.869), meaning that we have good prediction of overall liking knowing

Table 6.3 Regression results, relating overall liking to the presence/absence of the 27 elements

Dependent variable = overall liking N = 67 Multiple R = 0.869	Coefficient	Standard Error	T	P(2 Tail)
Additive constant corresponding to: Shape = tube; Size = medium; Material = plastic; Appearance = clear; Height = medium; Lid = paper/plastic; Closure = cap; Liner = present; Lid Appearance = clear	44.10	12.79	3.45	0.00
Shape: Round, oval, cylindrical	11.42	7.90	1.45	0.16
Shape: Square/hexagon	4.71	7.28	0.65	0.52
Shape: Rectangle	8.61	7.07	1.22	0.23
Shape: Tube	.	.	.	.
Size: Small	−8.25	4.23	−1.95	0.06
Size: Medium	.	.	.	.
Size: Large	−7.41	6.55	−1.34	0.19
Material: Paper	−14.72	4.17	−3.53	0.00
Material: Plastic	.	.	.	.
Material: Paper/plastic	−0.56	4.45	−0.13	0.90
Appearance: Clear	.	.	.	.
Appearance: Opaque	−17.39	6.53	−3.14	0.00
Height: Tall	−1.42	6.59	−0.25	0.80
Height: Medium	.	.	.	.
Height: Short	−7.68	3.87	−1.99	0.05
Lid: Paper	−6.59	7.85	−0.71	0.48
Lid: Plastic	9.22	6.53	1.67	0.10
Lid: Paper/plastic	.	.	.	.
Closure: Cap	.	.	.	.
Closure: Lid	6.47	6.94	0.79	0.44
Cap/Peel: Inner seal	−36.10	11.74	−2.99	0.00
Liner: Present	.	.	.	.
Liner: Absent	−2.50	4.62	−0.54	0.59
Lid appearance: Opaque	3.10	4.79	0.65	0.52
Lid appearance: Translucent	13.79	6.76	2.04	0.05
Lid appearance: Clear	.		.	

the features of the package. This will be an important consideration in studies. We rely on the model to teach us about "what works." In turn, we expect that the model will fit the data reasonably well so that we can be confident about the conclusions that we draw. Keep in mind that although the computer prints out the results to three decimal places, our data is not that precise. The computation can estimate the R statistic to many more decimal places. We should be happy to know the value of R to the first decimal place (i.e., 0.1, 0.5, 0.8, etc.).

4. Now we come to the essence of the model, *the statistics*. For each element in the model, we have four statistics that tell us about how the element performs. We will go through each of these statistics, one at a time for two variables—the round/oval/cylindrical shape and the small size, respectively.
5. We begin with the *additive constant*. The additive constant corresponds to the expected value of the rating in the absence of any variables. That absence occurs when the variables can truly be "absent" from a package. For example, if we deal with graphics,

then we could eliminate the logo. This would be a true absence.

> *For some absolutely necessary "structural features" of a package, we cannot allow any absence. One element from the silo must appear in every package. There is no package without a shape. This specific requirement that a package must have one element, a silo, means that in the modeling we must arbitrarily omit one element from that silo. It doesn't matter which element we omit from that silo. We just have to omit the element.*

6. For our package study, we put all of the variables into the equation. The computer's regression program automatically "kicked out" one element from each silo. This is why you see the ".". These are effects that cannot be estimated. Actually they should be called "0" effects. As we just noted, we could choose one element from each silo, which we will do in the next section. Now that we kicked out one element, actually we don't get rid of the element at all. Rather, all of these reference elements combine to generate the "additive constant." The additive constant corresponds to the nonseparable contribution of the following combination:

   Shape = tube
   Size = medium
   Material = plastic
   Appearance = clear
   Height = medium
   Lid = paper/plastic
   Closure = cap
   Liner = present
   Lid Appearance = clear

7. The additive constant is 44.10. This means that when we evaluate a package with the aforementioned set of nine features (tube, medium size, etc.), we expect a liking rating of 44.10. We cannot separate out the contribution of the various features to this 44.10. This "absence of separability" is important. It is inevitable when we deal with actual physical manifestations of a package. Actual packages must obey the laws of physics; the package has to either have a liner or not have a liner. If we deal, instead, with descriptions of the packages, then we are in a different world altogether. We can have silos of features that can be absent and no one would miss them. In such a case, the different elements are separable.
8. When we deal with the *individual elements*, all of the impact values in the same silo will be relative to the feature present in the additive constant. These features in the additive constant are the "reference" features. When we work with reference features in a model, the element utilities or impacts do not have absolute meaning. They have relative meaning. This will become a bit clearer later on when we talk about interpreting the model.
9. We now move to the four statistical parameters, which will tell us how the particular elements "drive" the response, and the degree to which we can believe these elements (versus randomness).
10. *Statistical parameter #1—Coefficient:* The coefficients are the multipliers, $k_1 - k_{27}$ in the equation:

$$\text{Rating} = k_0 + k_1(\text{Shape} = \text{round/oval/cylindrical}) \\ \ldots k_{27}(\text{Lid} = \text{clear})$$

   The interpretation of the coefficient depends upon the type of scale that we choose to be the dependent variable. We chose to use the mean or average on the 0–100 point scale. Thus, we interpret the coefficient or impact value to be the number of rating points (e.g., out of 0–100 scale) that are added when the particular feature is present.

   Looking at the table, we see that the coefficient for round/oval/cylindrical is 11.42. This coefficient tells us that when the shape is round/oval/cylindrical, we expect to add 11.42 points. When the shape is a tube, we don't add or subtract anything. This comes about because we chose "tube" to be a reference.
11. *Statistical parameter #2—Standard error:* The standard error shows the variability of the coefficient. Although we estimate the value of the coefficient, say, to be 11.42 for the round/oval/cylindrical shape, we know that there is variability in the data, and therefore variability of our estimate of this value 11.42 for the contribution of the round/oval/cylindrical shape. Of course we don't do the study 100 times and determine the coefficient. However, from the variability in the data, we can estimate how variable our estimate of this coefficient would be. The standard error of the coefficient is 7.90. This means that if we were to repeat this study

say 100 times, then 68% of the time we would observe the coefficient to range between a low of 11.42 − 7.90 (i.e., 3.52) and a high of 11.42 + 7.90 (i.e., 19.32). The smaller the standard error the "tighter" the variation around the coefficient that we actually see. Looking again at Table 6.3, we might search for a coefficient with a lower standard error, just to see whether there are other impact values (i.e., contributions) that show less swing. In fact, one feature, "short height," shows a lower standard error, 3.87. Thus, the variability of the coefficient is not fixed, at least for these types of studies where the elements in the model are taken from "what's out there," rather than systematically varied by experimental design.

12. *Statistical parameters #3 and #4—T value and P value, respectively:* These two statistics tell us the degree to which we can believe that the coefficient is different from 0. We already saw above that when we vary the features, we can drive the responses. We also know that the coefficient shows us the contribution of the packaging feature or element, and that when the study is repeated we will get different values. Do these different values of the impact actually mean that we are dealing here with an effect that is truly "0" (i.e., an irrelevant package feature)? To find out whether or not what we observe can be attributed to a natural variation around a true coefficient of 0, we perform a T test. The T value is defined as:

$$\begin{aligned} T = &(\text{Observed Mean} \\ &- \text{Possible True Mean})/(\text{Standard Error}) \end{aligned}$$

Our observed mean is +11.42. Our possible true mean is 0. Our standard error is 7.90. The ratio is 1.45, our T value. So far we don't know anything else. However, statisticians have computed the probability of getting a T value of that magnitude when the mean is really 0 so that what we observe results from random variation. A T value of 1.45 would occur about 16% of the time if the true value were to be 0. Thus, we conclude that the odds of observing our coefficient of 11.42 when it is really 0 are about 16 times in 100.

13. We mentioned above that the utility value is relative. So, for our value of 11.42, we have to keep in mind that this is not an absolute value. It is relative to the impact of tube, the reference value, which we arbitrarily set equal to 0. (Any other shape can be the reference). What does this mean? You start with the additive constant, which defines a specific combination of package features. Then you substitute different options in place of the elements currently part of the additive constant. And, while doing that, simply add or subtract the utility of the "new element" or new feature. The result will be the expected overall liking.

A Better, Broader Picture—Changing Reference Features and Using the Models

We end this deconstruction analysis by looking at how the different elements drive the perception of a number of different attributes, liking, sensory, image, and performance, respectively. This time, however, we are not going to let the computer program take control and "spit out" which elements it puts into the additive constant and which elements will be in the model. We must choose one element from each silo to put into the additive constant. *But we will choose the composition of the additive constant. We will create an additive constant most of whose elements have the lowest impact values in their respective silos. That way, all of the other elements will have positive impact or utility values.* Since all of the impact or utility values are relative, the difference in utility values between pairs of elements in the same silo will be unchanged. That is, the utility value of the element will change, but not the relative utility values within the silo. Furthermore, we won't use all the features as predictors because a number of them are too highly correlated with each other.

Armed with our data, let's see how much additional learning we can extract from the competitive environment. Again we must accept that we rely on the "kindness of strangers," or more to the point, on the packaging efforts of a lot of competitors, both those who sell margarine and those who sell allied products having plastic-type packages. We really don't know the rules of the mind when it comes to packages, but from an analysis of the competitive frame, we might be able to find out.

Let's move to the revised model in Table 6.4. We put in the relevant predictor elements, choosing the references as we want. As we noted above, "relevant" means those features that appear fairly often. We are not going to look at all the attributes and contributions of all features that we selected. It's simply mind numbing to do

Table 6.4 Contributions of the different elements to attribute ratings, showing only those features that are present in the original product (defined by the additive constant), and the contributions of features for a new product. The table is abstracted from a larger table showing how each element in the study "drives" every attribute.

	Like—Appearance	Sensory—Long	Sensory—Heavy	Performance—Easy to close	Performance—Easy to open	Appropriate—Margarine	Image—Unique	Image—Expensive	Image—Keeps fresh	Image—Contemporary	Acceptable—At table	Acceptable—In microwave
Features and total score of the current package (corresponding to the additive constant)	39	34	14	77	75	44	50	30	42	44	52	29
Shape: Tube												
Size: Small												
Material: Paper												
Appearance: Clear												
Height: Short												
Closure: Cap/peel												
Liner: Absent												
Lid: Opaque												
Components of the new package												
Shape: tube												
Size: medium												
Material: plastic												
Appearance: opaque												
Height: short												
Lid: paper												
Closure: cap												
Liner: absent												
Impact of each feature of the new package												
Shape: Tube	0	0	0	0	0	0	0	0	0	0	0	0
Size: Medium	11	16	8	–2	–2	10	–2	7	8	5	3	10
Material: Plastic	16	5	5	0	–4	20	1	10	13	–1	20	11
Appearance: Opaque	–18	–2	1	–5	–6	–19	–2	–6	–23	–3	–13	–3
Height: Short	0	0	0	0	0	0	0	0	0	0	0	0
Lid: Paper	0	0	0	0	0	0	0	0	0	0	0	0
Closure: Cap	–14	14	2	11	9	–26	15	6	6	–20	–16	–36
Liner: Absent	0	0	0	0	0	0	0	0	0	0	0	0
Lid appearance: Translucent	13	–3	0	13	9	19	–23	–9	10	17	5	6
Score of the new package (sum of changes + additive constant)	47	64	30	94	81	48	39	38	56	42	51	17
Change (new package, reference package)	8	30	16	17	6	4	–11	8	14	–2	–1	–12

so. Rather, we will create the model, look at the additive constant (which defines a package), create a new package with different features, and by using the model estimate how this new or at least modified package performs.

Remember that we start with actual packages, taken from a wide variety of products, all of which could be margarine containers. The packages don't have any branding on them, so it's pretty easy to inspect them, rate them on attributes, without being biased by the product for which the package was originally intended.

We will start with the product described by the additive constant, simply as a reference product. The value of the additive constant is how we expect that product to rate on the attributes. We do not know any more closely what contribution each element has on the additive constant.

Now, let us assume that we want to change our package from the current (defined by the features constituting the reference) to a new package. For each attribute (column), the contributions of each new feature from the new package appear as a row in Table 6.4. The sum of the constant (current package) and the contributions of the elements (changes brought about by these elements) represent the score that the new package would achieve. The difference, shown at the bottom of Table 6.4, tells us precisely how the profile of the package will change when we change the package features in a defined way. Keep in mind that this is our "best guess" about what will happen, based on a competitive analysis of "what's out there."

Summing Up

In this chapter, we looked at the competitive frame of packages. Rather than being content to measure which package performs best on a particular attribute, we looked at the ratings of 67 packages on a variety of attributes. This research exercise generated a report card that showed how well each package performs on each attribute.

We then hypothesized the features of a new package. We used the model to show how these features would "change" the profile of ratings from the "reference product." In Table 6.4, we see the new profile. The model allows us to estimate the change of profile from "reference" to "new," and to see the effect of each feature when we change the specific features.

With this type of analysis, it becomes far easier for the package designer to "engineer" a new package, and "guesstimate" with modest precision the sensory, acceptance, and image profile of the package. The key drawback, of course, is that despite having deconstructed the packages into their features, we cannot read the impact of all features because some of the features are not statistically independent of others. Yet it is clear that we are moving toward experimentally designed combinations of features, and the ability to engineer subjective responses by knowing "rules" that relate the domain of features to the domain of subjective impressions.

References

Moskowitz, H.R. (1994) Product testing 1: The distinction between S-R (stimulus-response) and R-R (response-response) models. Journal of Sensory Studies 9, 313–321.

SYSTAT (2007) SYSTAT for Windows. Chicago, IL: SYSTAT Software International.

Further Reading

Box, G.E.P., Hunter, J. & Hunter, S. (1978) Statistics for Experimenters. John Wiley and Sons, Inc., NY.

Khuri, A.J. & Cornell, J.A. (1987) Response Surfaces. Marcel Dekker, NY.

Moskowitz, H.R. (1985) New Directions for Product Testing and Sensory Analysis of Foods. Food and Nutrition Press, Westport, CT.

Moskowitz, H. and Krieger, B. (1998) International product optimization: A case history. Food Quality and Preference, 9, 443–454.

Plackett, R.L. and Burman, J.D. (1946) The design of optimum multifactorial experiments. Biometrika 33, 305–325.

Chapter 7

Psychophysics and the Issue of Price/Value

Introduction

When this book is read, the shock of commodity prices will have been absorbed by the world's economies, for better or worse. We cannot help but notice that the food industry is strained to the limit, coping as it must, on the one hand with the increasing cost of commodities, and on the other hand with consumer reluctance to happily accept the necessary price increases on the other. *Homo economicus* reigns supreme in the world of the everyday consumer. It is that unhappy realization that consumers don't like higher prices that forces the food and beverage companies to think about what they must do.

Of course, in this quickly evolving world of higher ingredient costs, packaging is a key player, along with the deft hand of the talented food technologist. The technologist may be able to modify the product formulation to achieve some of the same acceptance with lower cost of goods. It remains, however, for the packaging designer to create containers that don't overemphasize the increased price or that they at least present the price increase in a palatable form.

When it comes down to basic knowledge, however, exactly how do consumers respond to changing costs of goods? That is, how do consumers react when the package remains the same but the price changes? Or, perhaps even more typical, how do consumers react when the price remains the same but the package weight disappears?

What We Know about the Dynamics of Pricing in the Food Business

We now are going into the world of pricing, first from the world of branded product evaluation (product in its package), and then from the world of package and price changes in the world of commodity inflation.

What's more important, price or product quality? And, how do different people respond to pricing versus product quality when they evaluate the product first by seeing the product in the package and then afterward eating a sample?

The relation between purchase intent and both price and product quality began to be of great interest in the 1990s, with the proliferation of competitors in the deli meat business. The issues facing companies was whether the products that they were manufacturing and selling could sustain higher prices. Company after company focused on providing foods for the "meal out of home." A lot of the issue was whether the salesperson could convince the corporate buyer in the deli that the product being offered was of superior quality and could justify the price. So, in the end, the issue came down to the tradeoff between quality and price.

Dollars and Delis—The Bulk Turkey Problem

Our first case history comes from bulk turkey, a product that most delis feature, and that can be used in many different sandwiches and salads. The issue facing the company was simple—create a "model" showing the relation between purchase intent as the dependent variable and two independent variables. One of these was product quality of the competitive frame, defined as the "liking rating" assigned to a bulk turkey product, with the product shown in the original packaging, brand and all. The second was the selling price to the deli.

Doing the bulk turkey experiment was fairly easy. The respondents came from the world of food-service operators, rather than consumers. The reason for working with these operators is fairly straightforward. It is the food-service professional who first purchases the product and then passes the product to the customer. Typically, vendors representing different purveyors of bulk turkey

 ISBN: 978-0-813-81222-9

call on these operators, usually small businesses, and present their wares, along with a price. The consumer never gets to see the bulk turkey "in bulk" the way the operator does, and probably doesn't even pay attention to brands. But the food-service operator does.

The question was the relation between how good the product tasted when the operator "knew" the brand, the price at which the product was offered, and the purchase intent. To develop this relation is fairly straightforward. Just follow the six steps below, and you can develop the relation yourself, for almost any product.

Step 1—Select the product. For this particular exercise, we selected six bulk turkey products, all well-known brands, familiar to the food-service operator.

Step 2—Select five prices. Since the bulk turkey products were to be presented as similar weight samples, the prices could be the same. There were five different prices, appropriate for the product and weight. These five prices applied to each of the products.

Step 3—Select the appropriate group of respondents, and identify these people in terms of geo-demographics (age, income, gender), job in the food-service establishment, and job history (e.g., years of experience).

Step 4—Evaluate each bulk turkey product on sensory characteristics, using a 100-point scale (0 = very weak → 100 = very strong). Also evaluate each turkey product on overall liking (0 = hate → 100 = love). Make sure to evaluate the six products in a randomized order so that each of the respondents evaluates different products first, etc. This randomization reduces bias.

Step 5—Compute average ratings for the products on the different attributes.

Step 6—Create the model using regression analysis.

Creating the Psychophysical Turkey Model for Liking and Price

Creating a model to describe the results is not just a matter of fitting a curve or a surface to the data. Rather, it's important that the curve be grounded in some type of theory. The theory need not be a massive, all-encompassing one, nor does the experiment have to answer all of the questions about pricing. Rather, it is important that once the equation is fitted to the data, the results tell us "something" about price and liking. That is, we have to be able to interpret the results, not just do the curve fitting.

If we think about price and liking as determiners of purchase intent, then it's clear that we want the two of them to appear in the same equation. It's also clear that price is measured in dollars and cents, and liking is measured on a 100-point scale. Furthermore, purchase intent is measured in terms of a rating scale (0 = definitely not buy → 100 = definitely buy).

To create the model we take our cue from the economists and psychophysicists. The economists look at variables that are measured with different units. If the economist works with pennies, then parameters of the model look a great deal different than the same model, with the money expressed in millions of dollars. The model should say the same thing, or show the same general relation, no matter what the unit of money may be. Therefore, our first consideration should be that what we learn from including "price" as a variable should not be affected by the way we express price.

The second consideration comes from the fact that we seem to be mixing apples and oranges (i.e., incommensurate quantities). We know that we want to relate purchase intent to price expressed in monetary terms (i.e., dollars and cents), and liking in terms of a 100-point scale. We could just as easily express liking in terms of a 9-point scale or percent top box. Faced with the problem of changing units, psychophysicists suggest that the way to create the model is to use logarithms. The equation they suggest is:

$$
\begin{aligned}
\text{Log Purchase} &= \text{Log } k_0 + k_1(\log \text{ Price}) + k_2(\log \text{ Liking}) \text{ or} \\
\text{Purchase} &= k_0\left(\text{Price}^{k1}\right) \times \left(\text{Liking}^{k2}\right)
\end{aligned}
$$

Let's look a little more at the equation. When we use logarithms, we don't have to worry about the size of the units. Whether we express price in dollars or cents, the exponent k_1 for price remains the same. Furthermore, the exponent shows how ratios of the independent variable get translated to ratios of the dependent variable.

1. If the exponent is 1, then doubling of the independent variable corresponds to a doubling or 100% increase of the dependent variable.
2. If the exponent is 0.5, then doubling the independent variable corresponds to a 1.41 change or 41% increase of the dependent variable.
3. If the exponent is 2.0, then doubling the independent variable corresponds to a four-fold or 400% increase of the dependent variable.

4. Whether the units are in dollars, pennies, 100-point scale or 10-point scale, the same rules hold.

Now it is time to look at the actual data from the bulk turkey study, or more correctly, the models that we created. The models for purchase versus both the branded liking of the product (product tested with package) and versus price appear in Table 7.1.

We could have just looked at the results for the food-service professionals in the row marked "total." There we see a pattern that will repeat itself but in slightly different ways. Recall that the exponent tells us how ratios of the independent variable (liking, price) covary with the dependent variable. Furthermore, we don't have to worry about the unit in which liking or price is measured.

The results are quite clear. Increases in price drive down purchase intent far more effectively than do decreases in acceptance. The exponent 0.90 means that increasing the acceptance by 10% increases purchase intent by 1.09 or 9%. Decreasing price by 10% increases purchase intent by 39%! The same type of change occurs when acceptance drops down and price increases.

When we break out the data, creating models for the different groups, we see these same effects, but occasionally to a greater degree. When the respondent has worked in the food service industry for a long period (5 years+), the price exponent goes from −2.34 to −3.33. To make the number realistic, let us put some numbers to this. Assuming a 10% increase in price, the purchase intent score of the more junior professionals drops 20%, whereas the purchase intent score of the more senior professional drops 46%!

Table 7.1 Parameters of the model for different groups of respondents: Log Purchase Intent = log k_0 + k_1(Log Branded Liking) + k_2(Log Price). (Note: K_0 not shown)

Group	K_1: Log Branded Liking	K_2: Log Price Bulk Turkey	Multiple R^2
Total	0.90	−3.12	0.22
Female	1.16	−1.78	0.35
Male	0.75	−3.96	0.19
Job: Dietitian	1.06	−1.64	0.39
Job: Food Director	0.94	−2.60	0.23
Job: Chef	1.16	−3.63	0.23
Job: Purchaser	0.92	−3.40	0.35
Job: Owner	0.82	−3.40	0.19
Buy a Lot	0.86	−3.57	0.23
Buy a Little	0.96	−2.91	0.22
Loyal	0.78	−3.22	0.19
Switcher	0.99	−3.04	0.25
Experience: Long	0.88	−3.33	0.22
Experience: Short	0.95	−2.34	0.25

Summing Up

Our first experiment with packaging suggests that the brand itself is not particularly strong versus reasonable changes in price that might be encountered by the food-service professional. That is, the exponent for price is far higher than the exponent for product quality, as signaled by product acceptability. The same percent change in price and acceptability (e.g., both increasing by 10%), will send a shock wave, traceable to the response to price, not acceptance. The food-service professional reacts far more strongly to price than he reacts to product quality as communicated by both the product itself and by the product presented in the branded package.

Coping with Increasing Ingredient Price through Psychophysics

We now move beyond price and product to the actual package itself. Many food and beverage companies are realizing that they cannot increase the price of what they are selling, perhaps because they have encountered the dramatic price resistance that comes from acute perception of price changes. This "sticker shock" would be expected, based on the sensitivity to price versus product, clearly evident from Table 7.1.

Another way to cope with increasing cost of goods decreases the size of the package but maintains the price. At least the customer doesn't experience "sticker shock" per se, although the shock will still be there when the customer sees the unit cost. However, this shock may not necessarily be as dramatic because the prices are not in large bold letters.

Cereals, Boxes, Weights, and Perceptions

Our next case history looks at how a company went about investigating the likely consumer reaction to changes in package size, price, and another package feature (side pouring spout). The project was originally done with a cereal manufacturer, which had five brands on the market. The issue was whether increasing the price, decreasing the package size, or removing the special package feature entirely would make a difference to interest in buying the product.

In this second case history, the focus is on specific actions that the company could do in the short term and how those actions would affect the purchase intent of the branded product. The approach is quite straightforward. The company recognized that it would have to take significant action in light of rising prices of commodities. The choices were to increase the price by 7%, deemed the most minimal price increase that would make sense, or else reduce the size of the package and thus the product by 12% but keep the price the same. There was a third cost saving—changing some of the package configuration to remove special features that cost money but were sufficiently novel and functional to reinforce the impression of higher quality.

The precise actions were not clear. Since the company had five brands on the market, some far more popular than others, it also wasn't clear what the effect of these actions would be on purchase intent. Would the actions have the same effect across the different brands? Would the same price increase have the same effect?

Doing the Psychophysical Experiment

The experiment was very straightforward. The company created all eight cereal packages with cereal in them. The eight boxes were created according a full factorial design (2 sizes × 2 prices × 2 conditions of package feature = 8 combinations). The respondents, comprising a general population of cereal users, inspected, held, opened, and poured the cereal from the package, one package at a time, put the cereal in a bowl, added milk as typical, ate a bit of the product, and then rated the product on purchase intent, among other attributes.

Since the company marketed five brands of interest, there were 40 different combinations to test, incorporating brands and sizes. Each respondent evaluated a randomized 8 combinations of the 40. In a study of this type, it's quite likely that a respondent evaluated at least three different brands, with many respondents evaluating at least one variation of each of the five brands. In these types of studies, where a respondent evaluates different types of products, the only caveat to obey is that the respondent accepts all five cereals or tests only cereals that he accepts.

Questions like package size, pricing, and the like are both tactical and strategic. Tactical questions are those that simply need an answer—Do we do this? or Do we do that? There isn't any real focus on the ***why***, other than additional information to help support the specific decision. The project was tactical because the goal was to select the appropriate action for each of the five brands. One brand might do better with a price increase while another brand might do better with a size decrease, etc.

The project was also strategic because no one really knew what to expect. It was clear that price would be important, but no one had mapped out the relation between purchase intent, price of the cereal, and size of the box. There were some bits of information "floating around" from depth interviews about price and cereal amount, but nothing that provided the equation that we saw above for the case of bulk turkey. Corporate history was no help, because corporations, this manufacturer included, don't do such systematic studies to profoundly understand their products and packages. This factorial study would be the first systematic study of its type for the company.

Analyzing the Cereal Box Data by Regression Modeling

Collecting the data was fairly easy, but in actuality, analyzing the data was just as easy, and it produced quite a number of insights that we will look at in a minute. The analysis was done on a brand-by-brand basis. Since every participant in the study evaluated different stimuli from the set of 40 (5 brands × 8 combinations per brand), we didn't have the ability to create a model for each person. We can create models for the total panel, and should we want, create models for each particular user group. For this explication, we will just stay with the total panel, all of whom evaluated samples from many, and occasionally all, of the brands.

A purist consumer researcher would argue that it is vital to have only brand users evaluate their brand. For some situations this is correct. However, when it comes to package, price, volume, and feature, it probably is not necessary, as long as the samples are *randomized across the respondent. In this case history, we are looking for general patterns from the entire world of cereal purchasers, rather than looking for the responses of a specific brand user.*

The analysis is straightforward. There are really three variables of interest: increased price, decreased package size, and eliminated special package feature. Each of these variables is either present or absent in every one of the eight packages. Table 7.2 shows the experimental design. The eight cereal samples are coded in terms of presence/absence of the smaller size, higher price, or elimination of the special package feature. It's important to remember that we are dealing with an actual product so that there are no true "zero" conditions for price and size. A cereal box has to have the current or the increased price, as well as the current size or the reduced price. The only true zero or true absence is the removal of the package feature.

Now that we have the eight combinations, let us relate the presence/absence of each of the three specific actions (smaller size, higher price, remove package feature) to the purchase intent of each of the five cereals. We estimate the contribution of each action, using the regression model below:

$$\text{Purchase Intent} = k_0 + k_1(\text{Smaller size}) + k_2(\text{Higher price}) + k_3(\text{Remove feature})$$

The analysis is straightforward. The additive constant shows the purchase intent for the current product, for each of the five brands. The current products vary in acceptance for the general population. Brand #5, in particular, is a very poorly performing brand in the marketplace, whereas Brand #1 is the market leader. Keep in mind that these five brands correspond to ***different types of cereals*** that the company manufactures. Not all of its entries are particularly popular.

Table 7.2 Example of binary coding of the eight cereal products

	Specific actions taken to modify the cereal as a response to massively increasing commodity prices		
Cereal sample	12% smaller size	7% higher price	Special feature eliminated
101	1	1	1
102	1	1	0
103	1	0	1
104	1	0	0
105	0	1	1
106	0	1	0
107	0	0	1
108	0	0	0

To reiterate our objectives, we want to focus on what happens, in terms of the change in package size, price, and feature for the brands, rather than on how acceptable they are. What do the intended changes produce in terms of a single overall, evaluative response or purchase intent?

We get our answer in Table 7.3. The table shows us the current level of purchase intent for each of the top brands. This current level is captured by the additive constant. To make the comparisons easier, we have listed the expected impact of the current pack, current price, and current package feature, all as being "0," which they are because these three variables never appeared in the model.

What Do We Learn about the Different Cost-Coping Strategies?

Looking at the data more clearly, we can do the analysis two ways. One way looks in depth at each of the five brands, focusing on what exactly happens with the brand as we make the three cost-reduction changes. Keep in mind that the products were evaluated one at a time, rather than compared to each other on a shelf.

Let's look at Brand #1, the company's best performer. First we see that this product has an 87% top-2 box rating, using the traditional 5-point purchase intent scale (1 = definitely not buy → 5 = definitely buy).

1. Reducing package size and increasing price has little effect.
2. Removing the current package feature is irrelevant.
3. The bottom line—This best-selling product can withstand some changes and not lose more than 5% of the purchase intent.

We can follow this analysis for each brand, determining the likely effect on the rating for each proposed change to the product and package.

Summing Up

The message that comes from Tables 7.2 and 7.3 is that one has to do the experiment, to look at what happens for each brand, for each particular change in price,

Table 7.3 How changes in the size, price, and package feature of cereal drives purchase intent (top-2 box) for five brands

	Brand 1	Brand 2	Brand 3	Brand 4	Brand 5
Additive constant, k_0, (captures purchase intent of current size, current price, package feature)	87	79	78	46	16
Current size	0	0	0	0	0
12% Smaller size	–2	4	1	0	1
Current price	0	0	0	0	0
7% More expensive	–3	–9	–5	–3	–3
Current value, add package feature	0	0	0	0	0
Remove feature	0	–2	2	5	–2

package, and feature. There is no single rule, even within a product category. The same proportional price changes exert a large effect or a small effect, depending on the brand for which they are implemented.

At a deeper level, however, we see that the experiment needs to get larger. We have just looked at a few options—one price change, one package design change, and one feature change. We see that these effects vary by the particular brand that we are studying. So, we know that we ought to be looking at several prices and several package sizes, not just one.

It would be nice to look at these types of products on a shelf, so we can see the dynamics of product size and price in both evaluation of single product and choice of a product against others. That, however, is a more complicated approach. The issue really comes down to a choice between two alternatives:

Alternative 1: understand the dynamics of price and package size for a single product, by looking at many combinations of price and package size for a single brand.

Alternative 2: understand the dynamics of shelf selection, by finalizing one price, one package size, appropriate for each of the competing products.

It is probably too difficult for researchers to accomplish both right now.

Changing the Size of the Package—Psychophysics on the Shelf and Psychophysics in the Bowl

Our third case history deals with the systematic change in the amount of product, with no money specified. The question was simple—When we change the size of a can of our product on the shelf (canned meat stew, ready to serve), what happens to the consumer response? And, what happens to rated "purchase intent" when the respondent then actually opens the can, pours out the product, and inspects the amount of stew in the bowl?

Perhaps this is the most classical experiment of all. It does not involve price or anything else. Rather, it concerns only the relation between changes in the volume of the product and an evaluative criterion—here, purchase intent.

Our case history deals with canned beef stew, a popular product, and one that stands in for many other canned products. These products are fairly heavy. They are uniform, easy to handle, fairly dense, with regular volume (cylinder). They present an easily controlled surface on which the label can be attached. Finally, they are easily opened by a can opener, and allow their contents to be poured easily into a bowl for inspection.

Guidance and Insights from Classical Psychophysics

In the introductory chapter, we visited Memorial Hall at Harvard University to see psychophysics in action. As we mentioned in that section, psychophysics would become a foundation for a lot of thinking about packaging, even if the designer or researcher were not cognizant of the fact.

Now we have the chance to use some of that psychophysical thinking. The question is what might we expect if a consumer were to inspect a package whose volume has been reduced and whose weight has been reduced as well. Will the consumer respond more strongly to the visual cues, that is, the package seeming much smaller than it did when full size? Or will the consumer respond more strongly to the sense of weight? How, in fact, will the consumer respond to these package changes that must be done in light of the rapidly mounting cost of ingredients? Psychophysics will give us some answers.

Over the past 60 years, researchers have looked for ways to understand how people transform physical magnitudes to perceptual magnitudes. Starting in the 1950s, S.S. Stevens at Harvard University proposed that researchers might be able to discover lawful relations between what the experimenter does to the stimulus and how the person perceives that change. In fact, through patient, exhaustive experimentation, Stevens reported that it was likely that the relation between physical stimulus P and subjective perception S conforms to a power function, of the form:

$$S = k_0(P^N)$$

or in log-log terms

$$\log(S) = \log(k_0) + N\log(P)$$

We saw this power function (or straight line in log-log coordinates) when we dealt with the issue of pricing and liking, as determinants of purchase intent for bulk turkey. The power function again returns here, when we deal with the change in the size of can.

The important fact to keep in mind is that, according to Stevens and to many of his collaborators, colleagues, and students in the world of psychophysics, the exponent N is a "constant" for a particular stimulus category. Thus, for the loudness of sounds measured in sound-pressure level (dynes/cm^2), the rating of loudness can be described by a function:

$$\text{Rated Loudness} = k_0(\text{Sound Pressure Level})^{0.66}$$

The same type of power function applies when the respondent rates stimuli that have been systematically varied. Table 7.4 lists some of these exponents for power

Table 7.4 Power function exponents for a variety of sensory continua: $S = k_0(P^N)$

Continuum	Stimulus	Power Function Exponent	Sensory Change upon Doubling	Physical Change to Double Perception
Brightness	5 degree target/dark adapted	0.33	1.26	8
Tactile	Thickness/solution	0.50	1.41	4
Visual volume	Sphere	0.60	1.52	3
Loudness	Tone/binaural	0.60	1.52	3
Vibration	Finger/250 hertz	0.60	1.52	3
Visual area	Projected square of light	0.70	1.62	3
Taste	Sweet/saccharin	0.80	1.74	2
Tactile	Hardness/rubber squeezed	0.80	1.74	2
Vibration	Finger/60 hertz	0.95	1.93	2
Temperature	Cold/arm	1.00	2.00	2
Repetition	Light, sound, touch, shock	1.00	2.00	2
Finger span	Thickness/wood blocks	1.00	2.00	2
Visual length	Projected line	1.00	2.00	2
Duration	White noise stimulus	1.10	2.14	2
Pressure	Palm/static force on skin	1.10	2.14	2
Vocal effort	Sound pressure/vocalization	1.10	2.14	2
Lightness	Reflectance/gray paper	1.20	2.31	2
Visual velocity	Moving spot of light	1.20	2.31	2
Taste	Salt/sodium chloride	1.30	2.46	2
Taste	Sweet/sucrose	1.30	2.46	2
Heaviness	Lifted weight	1.45	2.73	2
Temperature	Warm/arm	1.50	2.83	2
Tactile	Roughness/emery grits	1.50	2.83	2
Force	Handgrip/hand dynamometer	1.70	3.25	2

functions, as well as "sensory implications." For example, when the exponent is lower than 1.0, the sensory system "compresses" the physical range to generate a narrower perceptual range. That is, we might measure the stimulus ratio and conclude that "objectively" the ratio is say 2:1. However, the subjective ratio will be smaller. Conversely, when the exponent is higher than 1.0, the sensory system "expands" the physical range, so the 2:1 subjective ratio is actually higher.

Let's keep that power function in mind, but move on to the world of perceived volume and perceived weight. The reason for volume and weight is simple—these are the perceptual dimensions that come into play when the consumer respondent inspects the can of beef stew on the shelf and when the consumer lifts the can, senses the weight, opens the can, and pours out the contents into a bowl.

According to Table 7.4, the exponent for visual "volume" is about 0.60. We say "about" because the actual exponent is subject to measurement error. However, 0.60 is a pretty reasonable estimate. It means that the subjective ratio will be smaller than the physical ratio. When the consumer inspects the product whose visual volume has changed, it's likely that he will notice the difference, but that the difference will be "compressed." We don't expect very severe judgments that the product is "much less than before." Certainly there will be these complaints, but a 20% decrease in visual volume will seem to be a drop, not the 20% to 0.80, but rather $(0.8)^{0.6}$, or about a drop from 1.00 to .87! This 20% decrease in physical volume of the can will be relatively small, perceived as only a 13% drop.

Now let's look at the other side of the coin—the change in perceived heaviness. We deal with an entirely different situation. Table 7.4 tells us that the exponent for heaviness is 1.45, more than twice that of visual volume. Furthermore, we have crossed that magic threshold of exponent = 1.0, where the sensory system neither compresses nor expands the physical range. With the exponent of 1.45, we now venture to a new realm, where the sensory system actually expands the physical change. Indeed, our 20% change, so innocuous in the case of volume, is downright devastating when the consumer has to lift the product and pour it out into a bowl. This 20% physical change corresponds to a $(0.8)^{1.45}$ or a drop from 1.00 to 0.72! This 20% decrease in physical weight of the can will be magnified, corresponding to a 28% drop. We should expect many complaints.

Now let's move on from our brief sojourn back in psychophysics to the consequences of coping with runaway commodity prices by judiciously reducing the amount of the product, and of course, the weight and perhaps the size of the package. We won't do the traditional psychophysics experiment where the respondent judges size or weight. We are really far more interested in the key evaluative criterion of purchase intent. Specifically, at what weight/volume change do we think we will encounter "resistance"?

Unleashing Psychophysics—The Canned Stew Experiment

The experiment comprised systematically varied sizes of cans, with the same circular area, but different heights. The five cans comprised five weights (100% current, 95%, 90%, 85%, and 80%). Respondents inspected each can in a randomized order, rating the can on purchase intent. In a separate part of the study, run afterward, the respondents opened the can with a can opener and poured the content into a bowl, inspected the bowl, and rated purchase intent.

To recap: during the course of the evaluation, the respondent evaluated five cans, one at a time in random order, rated each, waited, and then opened and evaluated the same five cans, albeit in a different order, rating purchase intent after pouring the contents into a bowl. The respondents never saw the ratings assigned to previous cans, whether by inspection or in a bowl. And, of course, the products to be opened had different identification numbers. All in all, 10 samples, two methods of experiencing the product (visual, opening/pouring), and one evaluative criterion (purchase intent on a 5-point scale converted to percent top-2 box (percent rating the product 4 or 5 on the scale).

As hinted above, this type of experiment is reminiscent of classical psychophysical scaling, of the type done thousands of times, in dozens, perhaps hundreds of laboratories. Of course the stimuli are different. We deal here with cans of beef stew rather than simple metal cylinders, our respondents are beef stew purchasers rather than college graduate/undergraduate students, and the rating scale is purchase rather than perceived size of can, heaviness of can, or amount of material inside. That being said, we are still dealing with a classical experiment in psychophysical scaling, albeit one ported from the halls of academe to the aisles of a supermarket, or the warmth of one's kitchen.

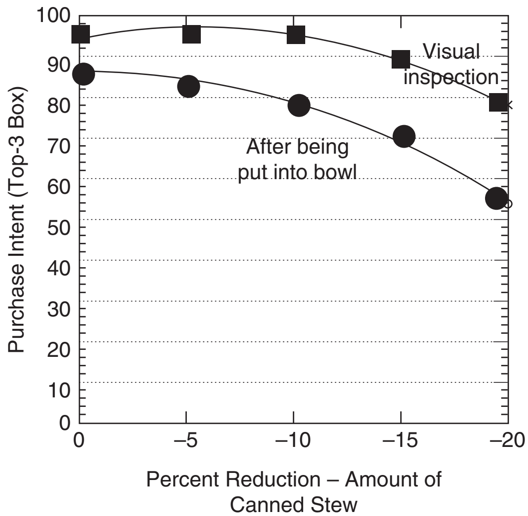


Figure 7.1 Relation between purchase intent and change in the volume/weight of canned beef stew. The experiment was run two ways—visual inspection first, followed in a second phase by opening and pouring the full can into a bowl.

To make a long story short, just look at Figure 7.1. As expected, the visual inspection leaves the acceptance of the cans pretty much unchanged, until we come to a 15% decrease in weight. However, let the respondent lift the can, open it with a can opener, pour the contents into a bowl, and all of a sudden the same change in volume or weight becomes more irritating. Customers who lift the product evaluate the changes far more critically than customers who simply observe the product on the shelf and see that the "size" of the package has changed.

As a final exercise we wanted to look at the mathematical relation between the relative amount of canned stew and the liking rating. Our initial prediction was that the effect would be greater when the respondent actually held the can, opened it, and poured out the contents into a bowl. However, would the relation between relative volume/weight and liking fall into line with what psychophysics might predict?

Our analytical strategy goes like this. Read it, not so much for precision of prediction, as for the fun one might have using basic science in a commercial setting!

1. We know that the exponent for volume is about 0.6 when we relate perceived volume as the dependent variable to the actual "amount" or volume. We also know that the exponent for weight is about 1.45 when we relate rated heaviness to actual weight.
2. We might expect a steeper curve relating acceptance (top-3 box purchase intent) to relative amount when we deal with "evaluation after pouring into bowl," because the judgment involves perceived "heaviness," a steep function. We might expect a flatter curve relating acceptance to relative amount when we deal with "evaluation after looking at the can but not holding it," because the judgment involves perceived "volume" or "amount," a flatter function.
3. We fitted a curve of the form:
   $\text{Log}(\text{Top-2 Box}) = k_0 + N(\text{Log Relative Amount})$
4. Since the Relative Amount is increasing from 0.8 to 1.0, we expect both exponents to be positive, because the amount of stew is increasing.
5. Step 2 above suggests that the slope N (or exponent in the power function) would be higher for the "evaluation after pouring into bowl."
6. The exponent for acceptance (Top-3 box) based on pouring into the bowl is 1.74. This is the steeper function, which makes sense because pouring involves the perception of weight.
7. The exponent for acceptance based on visual evaluation alone is only 0.77. This is the flatter function because looking at the product involves the perception of volume.
8. The ratio of the two exponents is about the same, approximately 2.3 (1.74/0.77 = 2.26)!
9. We conclude, therefore, that respondents are probably evaluating perceived magnitude in two ways, and translating those perceptions to acceptability. The reason that we see such strong effects with the "bowl" is because respondents have had a chance to incorporate a sense of heaviness into their perception, which they cannot do with visual evaluation alone.

Summing Up

This third experiment suggests the value of systematically manipulating the product in the package. It is not enough to present cost reductions and package changes at the conceptual level, or even at the level of visual inspection. Certainly respondents may be able to evaluate graphics and respond to shape. However, for cost reductions through reducing volume and package size we are dealing with price on the one hand, and sensory perceptions on the other. It is important in such situations that respondent move beyond visual inspection on the shelf to actually feeling, holding, and seeing the product.

Further Reading

Jacoby, J., Olson, J.C., and Haddock, R.A. (1971) Price, brand name and product composition characteristics as determinants of perceived quality. Journal of Applied Psychology 55, 570–590.

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

Moskowitz, H.R. (1983) Product testing and sensory evaluation of food. Food and Nutrition Press, Trumbull, CT.

Stevens, S.S. (1975) Psychophysics: An introduction to its perceptual, neural and social prospects. John Wiley and Sons, Inc., New York, NY.

Part II

Ideas and Inspirations

Packaging Research in Food Product Design and Development Howard R. Moskowitz, Michele Reisner, John Ben Lawlor and Rosires Deliza
 ISBN: 978-0-813-81222-9

Chapter 8

Idea Factories: Where Do Packaging (and Other) Ideas Come from?

Introduction

Did you ever wonder: "how did they think of that?" Walk down the supermarket aisle sometime. You might be totally mystified trying to find the tea that you enjoyed just last week, hiding somewhere in the multicolored splash of what must be 200 tea facings. You're probably equally lost trying to find your way around protruding end-aisle displays, beckoning cases of dairy products, or perhaps freezers filled with foods that look as if their packages have been designed by a commercial artist and then put behind a case to languish, saying "take me, buy me, eat me."

So the question remains: How did they come into the world? How do marketers, package designers, product designers, merchandising experts, and the like come up with ideas about the package? What system do they use? Is it serendipity, coming to them as they sit around waiting? Is it inspiration, perhaps because they are the creative group and must fulfill their manifest destiny? Do they work with paper, incessantly drawing and crumpling rejected ideas into a ball of paper, throwing the ball somewhere, so dozens of these rejected ideas sit in the corner, a mound of paper balls and discarded ideas?

Aids to Creative Thought—A View from the Early 21st Century

First, let us look at the world today. Many who are familiar with the inner working of business believe, or at least act as if they believe, that the process of coming up with ideas is a standard, disciplined approach taught to all aspiring MBAs. And, to add to this, many people believe that once a newly minted MBA reaches the company, that person brings with him (or more often now, her) these new methods, which are happily adopted by the company. And of course "the business goes on and on."

The truth is somewhat different. There are no "best practices" for ideation. Each company has its own favorite methods. These methods differ. Furthermore, a company will change methods over time, so that variation in method is cross-sectional (across companies) and longitudinal (across time within a single company). Surprising as this may sound, it comes from the realization that businesses need to compete in a world that constantly shifts. This shift from problems to opportunities, from dominance one day to a fight for survival the next, produces Schumpeter's so-called "creative destruction" in the world of ideation research (Schumpeter, 1975). Businesses try new methods. For a while a new method works, and then the method evolves as a living organism, to address new issues. The competitive environment fosters development of better methods over time. The pace is fast, the evolution of methods is just as fast, and the result is that each company adopts its own methods for short periods of time, exploiting those methods until they no longer work. Then, when the need arises again, the corporation searches once more for the next "technique" that will produce strong ideas.

Ideation and Its Descendants

The old truism "necessity is the mother of invention" is especially true in the commercial world of consumer products. The intense competition among brands you see in the supermarket is the visible outcome of enormous efforts behind the scene. Whereas 60 years ago consumers were happy to get any product that was sufficiently nutritious and tasty as well as affordable, today consumers are bombarded by products. It's estimated that more than 20,000–25,000 products per year emerge on the store shelves, only for most of them to fail. Some of these products fail slowly, others fail quickly, some die out, and some flame out (Mintel, 1998).

 ISBN: 978-0-813-81222-9

Our question is not about the products per se, but about the packaging. How do the packaging ideas come about? Do the packages spring fully formed from the mind of the package designer, as does Athena from the head of Zeus, at least in Greek mythology? Or, are the packages carefully crafted, scientifically developed, tested to within an inch of their inanimate lives by assiduous, knowledgeable, and eminently capable market researchers, who leave no stone unturned to discover just how well the package will perform? Or, more likely does the truth lie somewhere in between?

Packaging, unlike product, combines art, engineering, consumer science, and a bit of showmanship flair. There is an artistic aspect to packaging. Packaging requires an understanding of how to work in space, how to make the package unique, attractive, functional, and stand out to the bombarded, advertised eye. The package designer is an artist, perhaps a commercial artist, but an artist nevertheless. Packaging is also engineering. The package must perform. Unlike advertising, an engineer cannot merely promise. The package has to deliver. In the absence of any other redeeming qualities, the package must store, protect, and even advertise the product that lives within it.

So how does the package designer get ideas? What methods do designers use in companies to create the new packages? Are these ideas primarily for the structure of the package (probably hard to do, sometimes requires engineering breakthroughs), or are these ideas primarily for the graphics (easy to do, requires a flair for graphics design, but not necessarily the technological mastery of materials and shapes).

In this chapter we focus on the source of the ideas. We will look at methods that companies use in order to create new ideas about products. Since this book deals with packaging, we will abstract from the results some illustrative ideas about package. Keep in mind that in most corporate exercises for "idea generation," packaging is only one focus among several foci, which often compete with each other for attention. Usually, the new product exercise involving idea generation treats package as one component of a much fuller development effort.

Observation—The Oldest and the Newest Method

Ideas for packages often come from simply observing people coping with the issues, stresses, and even joys of their daily life. Just look at a child trying to open a bottle of "heart pills" or tranquilizers. An overdose could be deadly. Or, think of a child trying to tip the juice bottle to get more juice, spilling the contents all over the table, over the floor, and inevitably over all of the newly cleaned clothes. Or consider the older person, say one's grandmother, age 75, not so old, but not young, coping with the all too tightly closed jar. How does she get that "darned" jar open?

The package designer learns a lot just by walking around. It doesn't take the genius artist to recognize that the package of pills has to be childproof, that the bottle has to be easy to open for an aging person, that the print on a package in the store should be legible, especially when it's going to have nutritional information on it.

Or imagine the package designer walking the store, looking at the array of products on the shelves, say, for example, tea. One favorite is the increasingly popular chai tea. A few years ago one could try the two or three chai teas, make a decision, and then go back to the store a month later to restock. Those wonderful simple days are over. The tea section is overgrown, sort of like a yard of weeds with a detergent-laden stream running near it. The weeds, or in this case teas, wonderful as they are, spring up in such profusion every week that it's hard to find one's delightful chai tea the next time around. How does the package designer deal with this issue, to make a product memorable and "findable" in what is turning out to be a jungle?

Ethnography Comes to Business

More than a century ago anthropologists studying other cultures began to publish monographs on the societies in which they lived. They called their methods "ethnography," covering a wide expanse of procedures. The notion was that one could best learn about another culture by immersing oneself in that culture. Ethnography, traditionally a research method for anthropologists, revealed aspects of everyday life in other cultures that could not be otherwise obtained.

It should come as no surprise that the business community adopted it. An entire discipline of business anthropologists emerged, some basing their approaches on the traditional anthropology, a social science. Others based their anthropological endeavors on a modification of qualitative approaches in market research. Today, as this book is being written, anthropological approaches are all the rage as ways to understand the consumer. (See for example, Sherry, 1995; Malefyt, 2003; Sanchez and Casilli, 2008.)

Where does this have impact on package designers? It is quite clear that a good observation of consumers can tell you a lot. When it comes to likes and dislikes, scientists studying human behavior find that it's the chemical senses, taste and smell, that rapidly come to the fore and dominate what we think about the product. We don't often talk about liking or disliking the package, but we almost always talk about the flavor of the product, whether it's type, or strength, whether it tasted great or terrible, whether it was too strong or too weak, etc. In a very practical sense, this prepotency or immediacy of the chemical senses means that we have to make more effort, use better methods, and be prepared for harder questions, when we try to understand packaging.

Beyond Ethnography to Structured Methods

Today's world is so competitive that companies cannot wait for the results of ethnography to show them opportunities. They must make these opportunities. Sometimes "opportunity creation" comes from clever observations of daily life. Knowing that it is hard for older consumers to open packages means that one can observe these consumers to see how they cope. Or, in surveys, one can ask consumers whether they have the problem, and if they do, then would they please describe how they cope with the problem.

The rest of this chapter deals with two different approaches to creating new ideas, both in general and for packaging. (See van Kleef et al., 2005; van Kleef and van Trijp, 2007.) The first general approach uses direct interactions among people. Methods such as brainstorming fit into this first approach. Through brainstorming, people come up with ideas, and each individual in the session may attempt to improve the idea, generally at the same time withholding criticism. The key to this first approach is that it relies on the direct interaction of people to spark new ideas.

The second approach uses technology to facilitate the development of ideas. Technology doesn't necessarily come up with ideas. Rather, technology helps the interaction among people, records their ideas, and presents the ideas in new forms to other respondents, perhaps separated from each other. From the interaction of people and technology emerge new ideas. These new emergent ideas are polished by individuals who may live thousands of miles away from each other, who certainly do not know each other, and yet who collaborate to create the new ideas.

Tapping the Consumer Mind in Groups—Ideation to Create New Ideas for Foods and Beverages

Consumers often have good intuitions of what's needed, especially when it comes to packaging and to solving a particular problem. Quite often consumers faced with packaging problems come up with inventive solutions that, in the hands of a package engineer, can create totally new, very strong performing packages.

Ideation for new ideas in the traditional methods occurs either in a group, or in a conversation, or even in isolation with paper and pencil, respectively. All three methods are used widely, each having its own proponents and its own detractors.

When ideation is conducted in a group, the objective is for an individual to come up with ideas, and then to have other individuals build on these ideas. There are many different types of such methods, usually called colloquially "brainstorming." The participants may be encouraged to do homework ahead of time and bring in ideas, and to build upon the ideas of others, and even to modify the ideas in a structured way (i.e., what is the opposite? etc.). Common to these approaches is the use of the person or group as the builder of a complete idea. The objective of the session is to come up with as many ideas as possible, usually unedited, with the recognition that many of these ideas are incomplete, infeasible, redundant, etc.

It is not unusual for companies to hire experts to facilitate these ideation sessions. These experts may not necessarily know a lot about packaging or other specific topics that are the focus of the sessions, or about the current needs of the business. A few briefing documents are all that are needed for many of these better facilitators to jump right into the exercise and begin working with consumers. The experts are called "experts" because they know the process. The specific content is left to the client who hires the experts and to the participants who provide the information. Of course it is vital for experts to listen with the "third ear," to know when they have hit "pay dirt," and to know when they are uncovering new things. The facilitator is often rewarded for this ability by receiving new assignments, not necessarily in the same topic area, since as we just noted, the expert knows and facilitates the process, not the specific topic.

Focus groups (i.e., person-to-person [peer-to-peer] contact) are popular today among facilitators as a way to generate ideas. It's worth noting that much of the facilitation is done for businesses, simply because the goal is to generate product and service ideas, rather than as an academic exercise to increase the sum of our knowledge. This state of affairs holds worldwide. The vast majority of ideation sessions are run by professionals, for companies, under mandate to come up with new ideas. Academic research is left to observe and comment on the process (i.e., write articles about the process of ideation, rather than ideation itself for a problem). Occasionally academic research may use ideation in a limited fashion to deal with relatively small-scale applied problems, the type given to universities to solve as part of a grant or a contract.

One major consequence of the applied nature of ideation is the absence of literature on results. Some journals in product development (i.e., *Journal of Product Innovation Management*) deal with ideation as a business process to be studied. There are also topic-specific journals, such as those dealing with food, which occasionally feature an empirical article, dealing with an applied problem, in which packaging research figures as well. The applied nature of ideation and the relative scarceness of research studies, pure and applied, dealing with packaging mean that there is relatively little in the scientific literature but much more in the trade literature.

What Works Best—In the Box versus out of the Box versus Directed Ideation?

Although idea generation sounds relatively easy, it can be quite challenging. Creativity that fuels ideation is not linear in nature. One consequence is that the logical, ingrained, problem-solving approach to generating ideas, familiar to many concept developers, is unpredictable in the degree to which it can "produce". The traditional approaches may generate a few new ideas, or in the worst case, close-in ideas that represent mere minor extensions, rather than coming up with the often-needed concept innovations. On the other hand, an anything-goes approach to ideation isn't the answer either. The results can be disappointing and even counterproductive. The right balance must be struck between restricting the boundaries of idea generation, thus killing what could be viable ideas, and establishing absolutely no parameters, which results in a barrage of nonsensical ideas.

Decision Analyst, Inc., recently reported that disciplined, guided ideation exercises with a relevant frame of reference in which to ideate produced more actionable ideas than did the corresponding number and quality of ideas generated by the more traditional but apparently less effective "blue sky" approach to ideation (Callahan, Ishmael, and Namiranian, 2005). Ideation sessions were conducted with two groups of consumers each in the United Kingdom and France. Group 1 received specific "directed" instructions and creativity exercises. Group 2 received more "nondirected" instruction, typical of the "out-of-the-box" instruction often given in ideation sessions. Both groups spent the same amount of time generating ideas and had the same general assignment—create a new chocolate product. After the ideation was completed, the effectiveness of each group was evaluated on four criteria:

1. The total number of ideas generated by each group.
2. The number of unique categories represented by the content of the ideas generated—a measure of participants' ability to generate diverse ideas.
3. The relevance of each idea in addressing the task assigned to the group.
4. The originality of the ideas.

Indices to quantify originality and relevance were created by dividing the raw originality and relevance scores by the number of ideas produced. This approach measured how successful each group was at generating creative, yet relevant ideas for the project. Panelists both in France and in the United Kingdom who followed in-the-box directions generated more total ideas, more categories of ideas, and achieved both greater relevance index scores and greater originality index scores than did the corresponding group of panelists who followed the nondirected instructions.

Ideation in Business versus the Idealized Ideation Desired by Academia

Much of today's approaches to ideation occur in companies, usually facilitated by professionals who use methods that they themselves developed or modified. There is a simple reason for the proliferation of new methods, many of which are never reported in the academic literature, nor even known by academics. That reason is the need for results, at any cost, just to make sure one's company survives in today's hypercompetitive environment.

When a facilitator is given an assignment to come up with a new product/packaging idea, it is not acceptable to come up with ideas that are simply restatements of "I want what's currently on the market." Unless the group comprises highly introspective, articulate, vocal individuals, the needs and desires expressed in group discussions typically sound something along the lines of "I want something that is less expensive," or "Make it more convenient." This negative result may be acceptable to academics and represent the absolute truth. Nonetheless, such bland, negative, inactionable results certainly guarantee that the particular facilitator will not get future work with the company. The methods have to work. This demand for results forces facilitators to create methods that deliver ideas, no matter how these methods "stretch" the envelope of credibility.

Another reason for the proliferation of methods in business is that the review process for business is nothing like the review process for academic articles. In academia a great deal of professional legitimacy is conferred on those researchers whose work "fits into" the current mold. These researchers focus on problems that are just slightly beyond today's knowledge, ensuring that they locate their methods in the stream of research published by their fellow academics in the relatively recent past. The research should produce results, measured advances, that are explained for acceptance into the research canon, and otherwise conform to the academic standards that the journal editors are sworn to uphold. If the world of academic publishing could be described to an outsider, it would be described at its surface to be a well-oiled, self-correcting system that takes some risks, but is rewarded to keeping within the limits of what is acceptable as "true science." Such academic work welcomes new approaches only sparingly. The ultimate criteria are continuity, steadiness, and defensibility. There are no economic paybacks to consider that can justify trying new, out-of-the-box thinking. Such conservatism, no matter how effective in protecting against mistakes, is simply not tolerated by companies, at least by those in positions of responsibility who must make something happen or else lose their jobs.

The Value of the Indirect Approach to Ideation

Despite what the marketer may think, consumers do not typically talk about products or features. Rather, consumers are people. They behave as if their world revolves around them, which in fact it actually does. So consumers are more focused about their own needs and wants, and less focused about product features.

Yet even needs and wants are too concrete. Sometimes these needs and wants are best inferred from talking with consumers, rather than forcing consumers into a rational mode where they dig deep into their conscious intellect to pull out aspects that may be artificially constructed in order to satisfy the interviewer for that particular, unsettling, uncomfortable moment of interaction. The interviewer might be better off inferring the needs/wants, products/features, packaging structure/graphics options from what the consumer talks about when not prompted to answer specific questions.

Projective techniques are particularly helpful in this type of indirect situation, where the approach infers the package or product, rather than confronts the consumer with the demand to "produce an idea." Projective techniques attempt to reduce the participants' awareness of self and foster a depersonalized state of mind, by "projecting" their thoughts onto something else. The benefit of this approach is that in this depersonalized state respondents can speak about a topic indirectly, thus weakening self-censorship.

Projective Techniques to Generate Ideas

We begin this section with the caveat that the methods here are part of the evolving core of ideation. They are just a few examples of methods commonly used. Furthermore, the evolution of ideation methods by qualitative researchers is now becoming so rapid and so widespread that when we describe one method we really describe a burgeoning class of methods. The methods we describe come to us courtesy of Ms. Gwen Smith Ishmael, a well-known and accomplished specialist in ideation, who uses these tools regularly.

We begin with "Zoom Out," a projective technique blending visualization and projection. The moderator begins by instructing respondents to imagine themselves in a specific setting, such as opening a soup product. Using their five senses and their imaginations, respondents explore the store, paying particular attention to the person, and what the person does to open the product. Next, respondents are directed to focus on the selected individual's hand and to notice everything about it—skin texture, jewelry, nails, etc. Then participants "zoom out" until they can see the individual's entire arm, and once again note all its detail. This process continues until participants can view the entire individual in his or her immediate surroundings. After briefly discussing who and what they saw, respondents answer questions about the particular individual's behavior when opening the product, why the person opened the product the way he/she did, and what things they might have wanted and why. It many cases, the answers provide insight into their own beliefs, attitudes, and need-states, providing the foundation to identify so-called sweet spots for subsequent development. Although "Zoom Out" finds its greatest use in studying shopping behavior, "Zoom Out" should work in packaging, where the focus is more specific. The trick here is to frame the situation so that the participant's visualization task is easy and intuitively meaningful.

Another way to discover opportunities for package design explores the "jobs consumers need to have done in their lives (Christensen, Cook, and Hall, 2005). The moderator challenges respondents with a specific frame of reference or job to do, such as creating the package for a child's afternoon snack. The respondents verbally walk through the various thought processes, consideration sets, and actions that they would normally undertake in getting that job done. After sharing their accounts, respondents elaborate on the deficiencies that they experienced when trying to get the job done. These limitations, what failed, constitute the jobs that need to be done by the package and suggest new direction.

Word association and metaphor techniques help prompt consumers to "disclose" needs and desires, even when at the start of the process neither the consumer nor the researcher is aware. Metaphor techniques such as Zaltman Metaphor Elicitation Technique, ZMET (Zaltman and Coulter, 1995), allow respondents to express one thing in terms of another. Respondents present their images along with an in-depth, insightful explanation of why the images were selected, and what these images represent. Through their explanations, the respondents divulge insights into their need-states and desires. The interviewer more profoundly understands what the package means and should do. Of course, in ZMET there is no necessary creation of a new package, but rather a deeper understanding of what the package means. The package designer still must interpret the results, by reaching into his own imagination and coming up with the proper package. Nonetheless, ZMET provides a foundation for understanding, if not the ideation of an actual new package.

Other approaches to ideation are more focused, but again they share the common theme of freeing the individual from the constraints of everyday life. Let's look at a few of these, to finalize the list of tools for in-person ideation.

Brain Writing

The session leader begins by giving each participant one sheet of paper, each of which has been primed with a different issue to solve such as, "What is the next big idea in the package for breakfast products?" or a possible solution to a problem such as "Portable breakfast cereals in a pouch." Each participant silently reads what is written on the paper, and then writes in his or her additional ideas, remaining silent. When a participant runs out of suggestions, he or she exchanges papers with a fellow participant. The participant then reviews the ideas on the new sheet of paper and adds new ideas to the page. This process is repeated in silence until the group has no more new ideas to suggest.

Windtunneling

Many times the most innovative ideas are generated near the end of the ideation session, after participants have given the more ordinary ideas. How can these better ideas be generated up front, when there is more energy? Windtunneling (Wenger, 2001) uncovers the more unique ideas earlier rather than later in the process. Windtunneling forces the participants to dig for original ideas right at the start of the session. Ideation partners are divided into pairs, and one partner is assigned the role of the "Windtunneler" and the other the "Listener." The pair is given an opportunity around which to generate as many ideas as possible. For the next six minutes, the Windtunneler says everything that comes into his or her mind that might address the problem in a non-stop flow, and the Listener captures the most interesting ideas he or she hears during the Windtunneler's tirade. The participants then reverse roles and repeat the six-minute process.

Scamper

Often one can create a new idea by deliberately changing an existing product or service. This may be good for packaging, where the respondent can work from an existing product. For example, participants in an ideation session for sportswear might be told to think of a water bottle and then instructed to come up with a new idea using R—reverse. One resulting idea might be clothing that releases moisture onto the wearer's skin to cool the body.

Technical Aids to Creative Thought—Harnessing the Value of the Internet and Discovering Weak Signals

Let us now move from the personal interaction to technology-enabled interaction. The Internet allows many individuals to participate in an ideation session, doing so in many different ways. For example, individuals can participate in a chat session. An adept moderator can create the environment in the live chat session to focus on new product and package ideas. The key difference between the focus group ideation or depth interviews that we discussed previously and the online version is that online, respondents don't see each other.

When a lot of communication occurs subconsciously, then the Internet-enabled technologies may sacrifice the intimacy of personal communication for the efficiency of mass data acquisition. When 6 to 10 people participate in an ideation session, they speak with each other and observe each other. In contrast, 100 people or even more can participate in the Internet-enabled chat session. The respondents write their answers, which are shown to other participants. The respondents cannot see each other, at least using today's technology, although the technology might allow the speaker to be seen by the listeners.

As Internet penetration goes "mainstream," researchers and marketers can effectively reach and interact with consumers in an online environment. Thus, there is the option of conducting extended online focus groups—online message boards where respondents log in and participate in a group discussion that can last for several days.

In a typical extended online session, consumers receive an invitation to join a discussion and log into a protected message-board environment. Once in the message board, a facilitator leads them through a series of posts. These posts comprise direct, iterative questions designed to probe experiences, elicit stories, and prompt responses that ultimately reveal opportunities for new concepts. In an extended online group discussion, participants see the answers posted by other respondents as well as their own answers. The discussions can last several days. On average, the volume and depth of insight gathered from one extended online discussion equals the information from three or four in-person focus groups.

Soliciting Ideas "Online" through Problem-Solution

A good way to get new ideas is to work with people who have had problems and who have solved these problems. Their solutions may provide the nugget of an idea for a new package. To get a sense of what happens in the world of problem-solution, let's look at the results of a project called Innovaid™. The original objective of Innovaid™ was to create a database of ideas that a food manufacturer could use to create new products.

The Innovaid™ project comprises 75 different food/beverage and food/beverage related lifestyle studies. Each study, in turn, comprised 36 elements, combined into small combinations (test concepts), evaluated by respondents. For packaging, the important part of the Innovaid™ project came at the end, when respondents filled out the classification question. Four of these questions dealt with problems and solutions. Here is one set of questions for cookies. Note that Q10 deals with packaging problems, presenting a general problem, and asking for a solution. Q10 focuses the respondent's mind on the specific issue, but does not give any hint to the respondent. About one-third to one-quarter of the respondents gives reasonably insightful solutions to the problem.

Q9: Please tell us what YOU do to select a cookie in the store, with so many varieties to chose from.
Q10: Describe how YOU store cookies to keep them fresh, once you've eaten what you want.
Q11: If YOU tell a friend about a great new cookie you just ate, what would YOU tell them?
Q12: If YOU tell a friend about a cookie you just ate that was disappointing, what would YOU tell them?

With this set of questions in mind, let's see how some of the respondents in the cookie study answered question #10. Keep in mind that the respondents had already chosen to participate in the cookie study, and had finished evaluating a set of 60 concepts about cookies. Now as the respondents complete the final portion of the classification questionnaire, Table 8.1 lists what some of them said when answering Question #10. With hundreds of respondents participating, it is likely that some of these answers will generate new ideas to follow. However, the big problem here, as in all open-ended research, is the reliance on respondents to provide the new ideas from their own experience and imagination.

Table 8.1 Examples of responses to problem-solution questions

Q10: Describe how YOU store cookies to keep them fresh, once you've eaten what you want.
In a plastic bag
I would store it in a glass jar.
Reseal the bag
Zip lock bags
Just re-seal the package
In a can with a lid. They do not stay around too long.
I use a tin
Either in stay fresh bags or put large baggies around them and zip them up
Rewrap in the package.
I usually store the package in a large Ziploc baggie.
I close the bag again.
Reseal, but when I open a pack, I usually eat them until they're gone

Creating Ideas "Online" Using Collaborative Filtering

Basically, the approach presents consumers with "pieces of ideas," on the Internet, through a program that first asks the respondents to choose which ideas are "relevant," then to rate some of these ideas, and then finally add their own ideas. The approach thus automates the process of idea generation.

Let's look at a worked example of this approach, as embodied in a proprietary program known as Brand Delphi™ developed by Laurent Flores in Paris (Flores et al., 2003). We begin with the e-mail invitation, which goes to qualified respondents. Typically these respondents have "opted in" or agreed to participate in online surveys. They may be users of specific products or of a general sample.

The study begins with an orientation screen telling the respondents what is expected of them. We see that orientation screen in Figure 8.1.

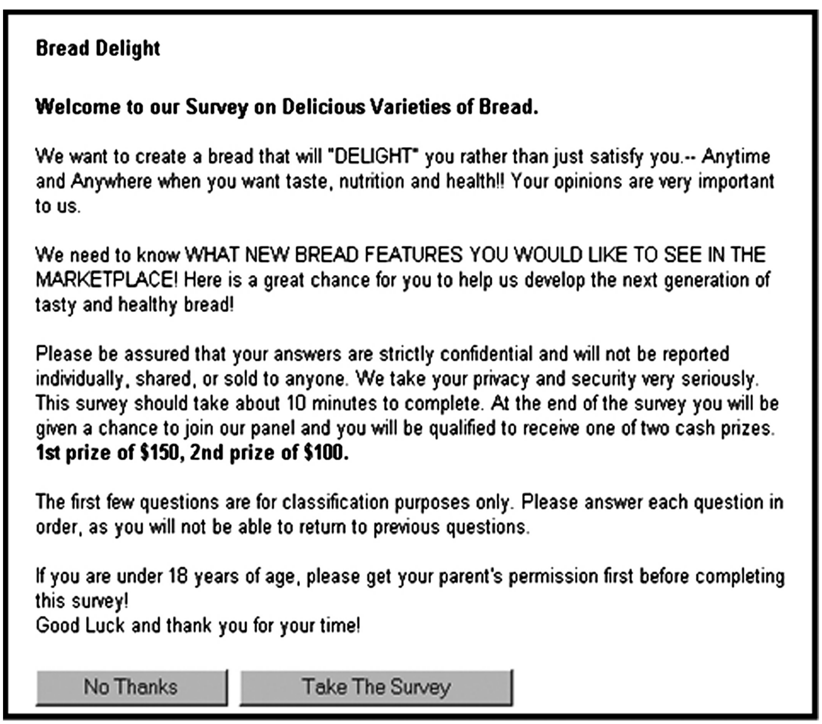


Figure 8.1 Orientation screen.

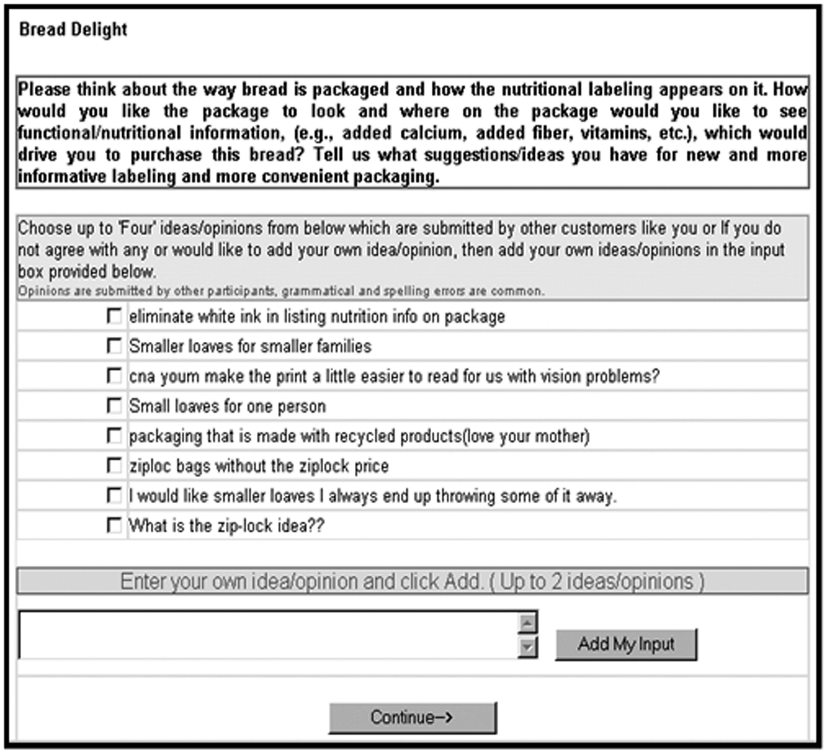


Figure 8.2 Showing how the computer program takes previously submitted ideas, in their raw form, and resubmits them to new respondents. The task is to select up to four ideas from the set that appeal to the current respondent.

After the respondent logs in, the actual survey begins. The survey is divided into three parts as follows:

Part 1: The respondent is shown a set of elements submitted as answers to the introductory problem. Look at Figure 8.2 to get a sense of the elements that were submitted by previous respondents as answers to the question of packaging. The computer program contains a heuristic that looks through previous results, finding which ones are selected as "relevant" and which ones are rated as "important." From these two responses to the elements, the heuristic "kills" ideas that it discovered as not being chosen when they appear to new respondents, or ideas that are rated as not being important.

Part 2: In this part, the respondents rate a randomized set of four elements previously provided by other respondents. Only the elements that are not "killed off" by the program survive to be included in this portion of the interview.

Part 3: In this final part of the interview, the respondent is instructed to submit an additional four ideas to be evaluated later on. At the end of the interview, the respondent is shown the most popular ideas. This type of information, although not generating new ideas or asking for other respondent actions, gives the respondent feedback about the progress of the interview (see Figure 8.3 for an example).

A potpourri of results from collaborative filtering and Brand Delphi™ appear in Table 8.2 for health bread and Table 8.3 for bottled water. The results of these exercises generate a wealth of ideas for packages. The tables show the analysis of these elements by the heuristic, which generates three numbers corresponding to the right-most columns:

1. *Selected:* When the element appears, it must be selected a minimum number of times. The analysis we did for these two studies presents elements that were selected a minimum of 15 times. Elements that appear late in the interview don't have a chance to be selected sufficiently often, and so they are penalized. Nonetheless, by focusing on those elements that are frequently selected as important, we can be sure that the elements are perceived as relevant to packaging.

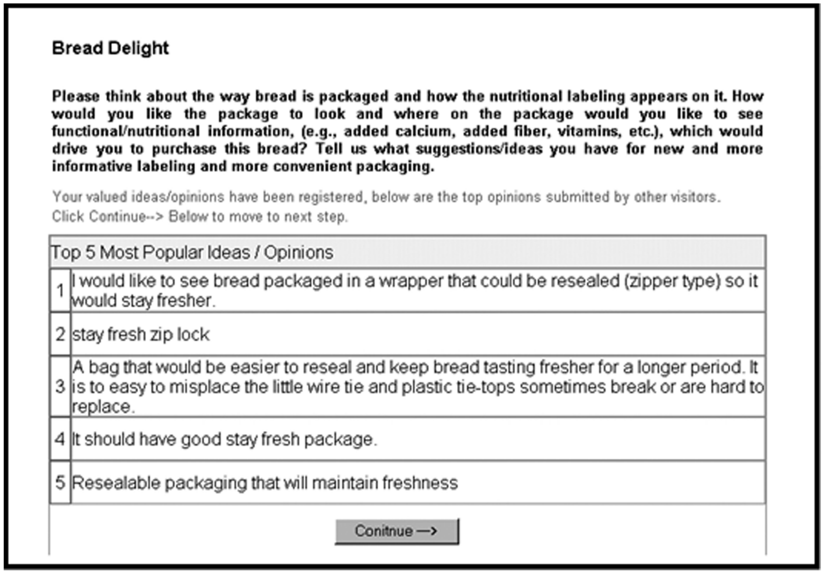


Figure 8.3 Feedback given to a respondent at the end of the interview.

Table 8.2 "Surviving & popular" elements provided by Brand Delphi™ for packages, when the topic is health bread. Elements are unedited, and so represent the actual inputs of respondents that have been presented to and voted on by subsequent respondents.

Health Bread	Times selected	Proportion of times selected	Average rating
It should always have the delivery AND expiration dates.	26	79	8.4
I would like to see bread packaged in a wrapper that could be resealed (zipper type) so it would stay fresher.	40	78	8.6
Zip lock package instead of the tie wrap to keep bread fresh	21	75	8.8
Stay fresh zip lock	76	75	8.5
A bag that would be easier to reseal and keep bread tasting fresher for a longer period. It is too easy to misplace the little wire tie and plastic tie-tops sometimes break or are hard to replace.	60	74	8.6
It should have good stay fresh package.	56	73	8.6
Easily resealable packaging that helps bread stay fresh longer	40	70	8.8
A good price	21	68	8.3
Make the package a little stronger so when you get home its not crushed	29	67	8.1
Resealable packaging that will maintain freshness	43	67	8.7
Be nice to have bread that doesn't mold in like 3 days.	61	67	8.5
An easy way to determine how long the bread has been on the shelf	27	66	8.4
Resealable bags would be great	35	65	8.7
Love the zip lock idea. Bread needs to stay fresh longer.	29	64	8.7
I would prefer that the freshness date be much easier to find and read on the package.	18	64	8.1
The bread stays fresher	81	64	8.5
Bread that can be fresh-frozen and will still be good when it thaws.	19	63	8.4
Simple and informative packaging with a zip-lock bag	15	63	8.5
We do need a Ziploc bag instead of the twisties	41	62	8.5
Sealable with something that collapses as the bread is used and a closure other than a twist tie that is secure and you can't lose.	62	61	8.4
A clear expiration date in a clear package so you can see the bread.	44	60	8.2
Packaging that will keep bread fresh longer	21	60	8.4
A bread that stays fresh because of quality packaging	46	60	8.5
Make package a zip lock type bag	27	59	8.4

Table 8.3 "Surviving and popular" elements provided by Brand Delphi™ for packages, when the topic is bottled water. Elements are unedited.

Bottled water	Times selected	Proportion of times selected	Average rating
Smaller size bottles that aren't double the price	25	61	8.8
Cheaper priced bottled water, they charge entirely too much.	26	63	8.7
I would love a bottled water which is convenient to carry and to consume on the run	23	29	8.7
Cap that seals well enough to contain spills if bottle is tipped over after it has been opened.	28	42	8.6
I prefer the sport top which allows you to drink the water without removing the cap—it minimizes spills.	30	54	8.5
lightweight containers—no glass	30	42	8.5
sippy tops for children	18	29	8.4
I dislike the regular screw-on cap. I think they should all have the sippie tops that you pull up to sip, and push down to close. I that easier, especially if you are working. Also, you don't drop or misplace the cover.	36	41	8.4
I don't like the way plastic containers in 20oz or so collapse in on the sides when you drink from them	15	31	8.3
Easy on off cap for use while on the treadmill	23	33	8.3
Spout that doesn't leak	37	46	8.2
Easy to hold	29	39	8.2
I want fluoride available in my bottled drinking water.	17	22	8.2
A large amount of water in a bottle easy to carry…	17	38	8.2
I'd like to know what chemicals and minerals are in my water, even if they occur naturally. They should show up on the label.	32	48	8.2
Bottles with a small enough base to fit standard holders in cars.	27	55	8.1
I would love a product design and labeling for drinking water presented in such a form that gives me the impression of purity and health	25	33	8.1
Smaller, easier to take bottles	27	34	8.1
Offer different sizes of bottles.	17	41	8.1
I love to have an easy and handy closure.	32	36	8.0
Easier to hold on to. Like have an indented area where the hand could easily grip the bottle	36	44	8.0
Easy to carry container.	37	43	8.0
Easier to open and drink right from the bottle for people on the go.	17	50	8.0
I would like to see caps attached to the bottles.	15	41	7.9
I think they are too expensive, and if they are cheap they taste like chlorine	29	47	7.8
I would like to have an easy and convenient bottle container for transportation in the car	26	35	7.8
Clearer Labeling so it is easy to tell what the water has to offer	26	30	7.7
They all pretty much look the same - hard to tell one brand from another.	15	19	7.6
Some are too wide at the bottom to fit into the holder in a car	34	40	7.6

2. *Proportion of times selected when presented:* Here we focus on the proportion of times an element was selected when presented. Ideally this proportion should be as high as possible, indicating that respondents who saw the element selected it.
3. *Average rating:* Here we look at the average of the 9-point rating scale dealing with importance, when the element was presented.

Summing Up

Creating "concept elements" for packaging has become a big business. With the increasing competition for the consumer's wallet, it is important for companies to create better packages and better graphics on those packages. In response to this need, a large "ideation" industry has grown up. For the most part, practitioners in the "ideation" business industry do not specialize in packaging alone. On the other hand, many design specialists, limiting their practice to packaging, do incorporate one or another form of ideation.

We have presented the two main schools of ideation. One school uses face-to-face, that is, personal interviews, or even observation. These personal approaches try to pull out needs and wants from consumers by interacting with them, questioning them, and listening with the "third ear." This first group considers itself to be experts, using consumers are sources of ideas, but not necessarily as good judges. The other school uses technology to identify new ideas, with consumers acting as both generators of ideas and judges of the ideas. This second school considers itself to be facilitators of the process, but not necessarily subject matter experts.

In the end there is room for both. Packaging is becoming increasingly important in the wildly competitive environment. Both schools of ideation have a great deal to contribute, albeit with different talents and different proclivities for the way they get ideas and serve them up to the eagerly waiting clients.

References

Callahan, R., Ishmael, G., and Namiranian, L. (2005) The Case for In-the-Box Innovation. Paper presented at ESOMAR Innovate 2005, 2/27/05.

Christensen, C., Cook, S., and Hall, T. (2005) It's the Purpose Brand, Stupid! The Wall Street Journal, 11/29/05.

Flores, L., Moskowitz, H.R., and Maier, A. (2003) From weak signals to successful product development: using advanced research technology for consumer driven innovation. In: Proceedings of the Second ESOMAR Conferences: Technovate, Cannes, France.

Malefyt, T. (2003) Ethnography and Focus Groups: Mutually Exclusive? Brandweek (Letters) 43, 14.

Mintel. (1998) High Failure Rate Poses Challenge to New Novelties. Frozen Food Age, April 1.

Sanchez, S. and Casilli, A.A. (2008) Status and use of food products with health claim (FPHC) in the USA, Japan and France an anthropological perspective. Food Quality and Preference 19, 682–691.

Schumpeter, J.A. (1975) Creative Destruction. In: Capitalism, Socialism and Democracy. 1975. Harper, New York (orig. pub. 1942), pp. 82–85.

Sherry, J.F. (1995) Contemporary Marketing and Consumer Behavior. Sage Publications, Inc.

van Kleef, E. and van Trijp, H.C.M. (2007) Opportunity identification in new product development and innovation in food product development. In: H. MacFie (ed.) Consumer-led Food Product Development. CRC Press, pp. 321–341.

van Kleef, E., van Trijp, H.C.M., and Luning, P. (2005) Consumer research in the early stages of new product development: a critical review of methods and techniques. Food Quality and Preference 16, 181–201.

Wenger, W. (2001) Windtunnel!!! A new creative problem-solving procedure. Home page, http://www.winwenger.com/part55.htm.

Zaltman, G. and Coulter, R. (1995) Seeing the voice of the customer: Metaphor-based advertising research. Journal of Advertising Research 35, 35–51.

Chapter 9

Defining the New Package: Specifying the Package at the Concept Level

Introduction

Packaging is just one part of the new product. Marketers like to talk in terms of the five Ps: product, price, positioning, placement, and of course packaging. Product includes the physical product that is being bought, price is what is being charged, positioning is how the company "talks about the product" in advertising and public relations, placement is where on the shelf the product is located in the store, and of course packaging is what holds the product.

It's a rare case that the focus of any project remains strictly on packaging. Yet quite often, new product research involves packaging, and a lot of the focus may be on the features of the new package. With experimental design of ideas, it is possible to work with packaging as a silo or set of silos, comprising different elements or options. There will be other silos, of course, dealing with topics beyond packaging.

The Story Behind the Data

With this introduction in mind, and recognizing that packaging is just one part of a new product, let's see how one manufacturer dealt with the opportunities for liquid margarine. Whereas the notion of liquid margarine is not particularly revolutionary today, it was when companies realized that the very nature of "liquid" denoted health, and that they could get an added "something" by offering a liquid margarine rather than a solid margarine. Focus group after focus group suggested that "liquid" connoted healthful, perhaps because there was an unexploited connection between oil (especially olive oil) and health. Oil is liquid, and liquid is health. In contrast, people have a feeling that the solid margarine product is somehow "artificial," "chemical," and thus not so healthful.

Like most "early-stage" projects in new product design, the underlying need was really not particularly focused. That is, when we began the project we really didn't concentrate on any one topic. The idea of converting from a solid to a liquid margarine was new, but not outrageously so. You might think that with companies being in the business for dozens of years, they would know the things to say about the liquid margarine, what they should promise in the way of better health and, since the margarine is liquid, what type of package is most suitable.

Nothing could be further from the reality. Yes, there was an opportunity, and as it turned out, a very big opportunity in many countries. It became clear that although there were lots of reports from small, unconnected tests dealing with people's preferences for packaging, actually there was no comprehensive database about what the package could be. Furthermore, it again became clear that this project would have to create its own database, with many hundreds of different ideas, each fighting against the others, to drive ultimate acceptance.

With this story in mind, let's see how the company solved the problem, identified what the product should be, and of course in doing so, identified the specifics of the packaging. We will see the breadth of information that can be created and then quantified through today's research methods. At the end of this chapter, we will see how this database identified the impact of literally dozens of package ideas, as part of an even larger-sized study dealing with the many facets of the liquid margarine.

Formulating the Problem—What Exactly Should the Product Features Be?

In the early 1990s, the company had developed liquid margarine. It quickly became clear that this product could provide additional and probably pretty significant money to the company, especially in Europe. The question was simply what to do with this product.

 ISBN: 978-0-813-81222-9

At that time the use of experimental design to create new product ideas was becoming increasingly accepted. Companies in the 1990s recognized that they would have to compete on knowledge, not on simple hope. It wasn't sufficient to have insights alone. In the food business, especially, competition was heating up. We see a lot of the same situation today, 15 years later. The cost of entry is low. It doesn't take much money to put a new margarine on the shelf, if you can pay the "slotting fees" that the stores ask (almost rent for their space, to be occupied by your product). If the product can be shown to be reasonably unique and consumer-acceptable, then you have a chance to get onto the crowded shelves and fight it out with competitors, at least for a little while.

With this in mind, let's travel back to those years. You can imagine a group of six people. Our two key players are the 37-year-old brand manager who just received her latest promotion the year before for a successful launch of a new soup, and a colleague, a young, 29-year-old market researcher anxious to try new methods and "push the business ahead."

The questions facing the group were very straightforward regarding this new liquid margarine. Top management had dictated a pan-European launch. The research goal was to nail down the many different aspects of the margarine—from what to say about it, to how to package it. Most importantly, the fact that the product was to be a liquid rather than the conventional solid meant that there were issues of packaging. Margarine has many functions, from helping with cooking to being used as a spread, etc. What should the packaging form be for this new liquid?

Exploring Many Ideas—Strategies, Hazards, Remedies

At the end of the day the product developer, package designer, marketer, and advertising agency have to agree on the product, its features, its packaging, and the other Ps that we discussed in the introduction. Of course, the task is pretty straightforward when one is constrained to one of a few product forms, a few shapes, a few containers, and a few legally approved messages. Such is the case with the pharmaceutical industry, where most of the work concentrates on messaging, not on product form and certainly not on packaging. Turn that set of constraints firm on its head when it comes to products in a new form, which can redefine a product category. Certainly there are a limited number of basic packages, but the number of alternative ideas can be staggering. The imaginative package designer can work with many different basic containers, and for these containers work with as many different features on the container. In the end, only imagination limits the opportunities.

Such was the case with our margarine. The young team came up with literally hundreds of ideas about the margarine, of which several dozen dealt with the package. The ideas were created by corporate teams, helped along by experienced "creativity" moderators, individuals who specialized not so much in packaging per se, as in the creation of new ideas for foods.

Dealing with "Very Many" Elements for a Product Concept

Researchers who work with product concepts have to make choices in what they do, especially when they work in a business environment that demands decision and action. When there are many aspects to a product, such as the liquid margarine we deal with in this chapter, the researcher can opt to select some promising ideas, and after refining these ideas in focus groups, test the ideas for appeal and for possible market success. In such a case the research deliberately chooses to focus on a few, promising ideas. Testing the concepts for potential performance is like running a "beauty contest." The objective of the test is to decide which concept is the most promising for the marketplace. In such tests typically the researcher screens relatively few concepts (i.e., 1 to 20 or so), instructing respondents to rate each test concept on a number of different attributes. The output is a report card of the type we will see for margarine packages later in this chapter.

But what happens when there are dozens of elements, or perhaps even hundreds, as there were for the liquid margarine product? We are referring to many elements that, together, cover a wide range of alternatives. Such an abundance of ideas for a product is not new, and the fact that of these 69 were package-related is not particularly unusual. The reality is just the opposite. Most product initiatives could be opened up to hundreds of ideas, some constituting radically new directions, some simply variations of the current or variations of these new directions. More often than we care to admit, the decision to limit the number of options to a testable set is made through judgment, and all too often in light of financial constraints.

Now let's return to the margarine example, where the foregoing constriction of vision was absolutely not the case. The project team decided to launch a full-scale evaluation of all the ideas. It's not that the team needed to learn from the "ground up." Certainly that was not the case here. Rather, the team recognized that in reality they did not know as much as they had assumed. It became clear that when talking about packages for this new liquid product, no one really could predict what would win and what would lose. Ideas sounded good on the drawing board and in the ideation session. But, and a very strong but, would they "fly" when put to the test?

A First Foray into the Day—Looking at Some of the Elements "by Country"

As researchers, package designers, product developers, and marketers, we are accustomed to thinking in our own categories or ways of dividing information. One of these ways is by country. It is "common knowledge" (although rarely attributable) that people differ in their preferences by countries. It's obvious that there are differences in countries and cultures because the world is not yet a homogenous whole. Yet, to describe people of different countries as wanting "different things, such as packages" begs the question as to what specifically are these different things that they want. We can test the hypothesis that there are different country-to-country preferences. Our data with 833 respondents from four European countries with different heritages, languages, etc., let us do that. Let's see how far we get!

First, look at the orientation page to the study, in Figure 9.1. The orientation page tells the respondent what is expected of him. The respondents in this study were recruited to participate for two hours, during which time each respondent evaluated 100 concepts on the computer screen and rated each concept on three attributes. When a respondent knows that the study will take a certain period of time and comes into a central location, he takes the task seriously. The respondent just can't "whiz through" the evaluation, because there are "monitors" or interviewers in the test room watching the respondent, albeit discreetly.

We begin by looking at the data by total panel, and the same data by respondents in four countries. From the set of 316 different elements, we selected one silo of elements, the package plus benefit. It was in this silo, shown in Table 9.1, that we see differences among the package alternatives. Table 9.1 shows us the results from 27 different elements in the study (a little more than 1/12 of the elements). We can see modest differentiation by total panel, and some evidence of differences by the countries. However, we would be hard pressed to find a pattern. Furthermore, the concept elements do not really

All of the concepts you are about to see refer to a

LIQUID PRODUCT TO BE USED
IN THE KITCHEN FOR COOKING (OR BAKING)

Please take your time and read each concept (screen) thoroughly. Once you have read the concept, please enter your rating based on the following question. The entire concept should be rated as a whole.

How interested are you in using this product?

← NOT AT ALL INTERESTED VERY INTERESTED →
1 2 3 4 5 6 7 8 9

How well does this product fit with BRAND X?

← DOES NOT FIT AT ALL FITS VERY WELL →
1 2 3 4 5 6 7 8 9

Would you use this product in place of oil or in place of butter/margarine?

← IN PLACE OF OIL IN PLACE OF BUTTER/MARGARINE →
1 2 3 4 5 6 7 8 9

PLEASE USE THE ENTIRE 1 TO 9 SCALE.
It is not necessary to press the <ENTER> key after entering your rating.

Figure 9.1 Example of the orientation page for the study

Table 9.1 Impact or utility value for 27 packaging elements (element + benefit) for total panel and for four countries

	Total	Germany	Netherlands	Sweden	UK
Additive constant (baseline—no concept elements)	44	46	50	36	43
The packaging allows easy opening and reclosing	3	2	4	2	2
The packaging has a transparent stripe at the side of the pack to show how much is left	3	3	1	1	0
The packaging is foldable, thus minimizing space in the waste-bin	3	1	4	2	3
The product stays fresh for longer, because the packaging is resealable	3	−1	4	3	2
The packaging allows controlled dosage, which gives you more value for money	2	3	2	2	0
The packaging allows dosing of just the right amount	2	2	3	1	5
The packaging allows easy dosing in an instant	2	3	1	1	3
The packaging can be stored outside of the refrigerator at room temperature	2	3	2	1	1
The packaging has a good grip which makes it handy to use	2	1	2	0	2
The packaging has a transparent stripe and measuring marks at the side of the pack for quick and controlled dosing	2	3	1	4	4
The packaging has a handle for easier carrying and pouring	2	4	0	2	3
The packaging has a handy and nonmessy spout for pouring out the product perfectly	2	3	2	1	0
The packaging is environmentally friendly	2	1	0	4	3
The packaging is designed for cleaner handling, preventing greasy fingers	2	2	3	1	2
The pack is fully transparent so you always see how much is left	2	2	2	2	3
The pack allows even coverage of food or pans	2	0	5	1	6
Can be used till the last drop of the product	1	−1	1	−1	4
The packaging is tamper-proof to guarantee it hasn't been opened before	1	0	1	2	0
The packaging is economic, clean and handy to store	1	−2	5	0	2
The packaging is sealed to guarantee freshness	1	−3	2	2	3
With this kind of packaging you won't need a knife or spoon anymore	1	2	2	−1	3
The packaging is entirely recyclable	1	−3	0	3	1
The packaging stays clean—every time. So does your fridge. And so do you.	1	5	−2	1	2
The packaging allows one-hand use	0	0	1	−1	1
In a modern packaging for today's people	0	1	−1	−1	3
The packaging is ideal to take away, like on holiday	0	1	2	−2	5
The packaging is refillable	0	1	−6	1	4

perform strongly in any country. We would like to have impact or utility values of eight or higher to say that an element drives interest in the product. We cannot find that high-performing element in any country.

The lack of any pattern whatsoever across the four countries comes across even more clearly when we plot the 27 individual impact values for pairs of countries, as we see in Figure 9.2. A good strategy in such cases plots the data in a scatterplot matrix, commonly available as a graphing procedure in many statistical packages for the personal computer (i.e., SYSTAT). Plotting the data often shows quite clearly that there is no pattern that can be discerned, a discovery just as important as one that shows there is a pattern.

Transnational Segments—Where the "Packaging Action" Lives

If packaging elements do not perform strongly for the total panel or by individual country, then perhaps it is because the 833 respondents in this study comprise different groups of individuals with different preferences. The segmentation of people into different mind-sets constitutes an ongoing organizing principle—a *leitmotif*, a

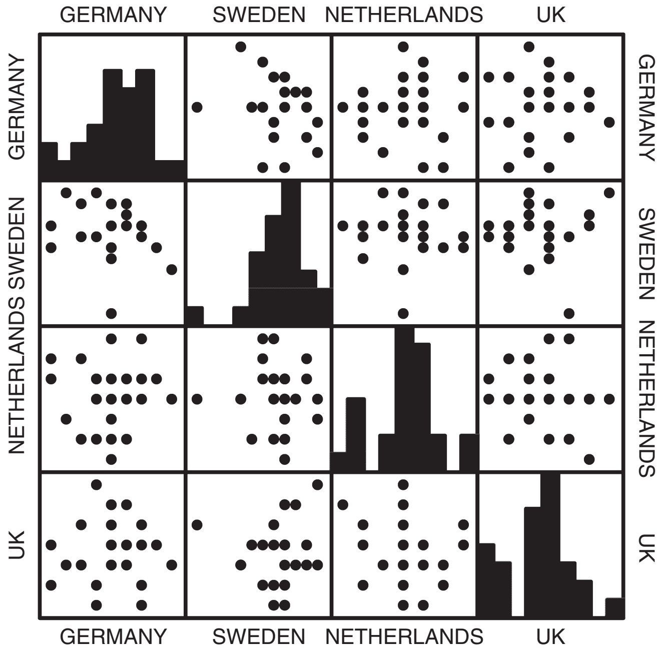


Figure 9.2 Scatterplot matrix for the impacts or utilities corresponding to the 27 package elements (element + benefit). Each filled circle corresponds to an element. The bar graphs show the distribution of the 27 utilities, allowing a comparison across countries.

recurring and organizing theme of this book (Jacobsen and Gunderson, 1986; Green and Krieger, 1991; Moskowitz, 1996; Qannari et al., 1997; Tang et al., 2000; Vigneau, et al., 2001; Westad et al., 2004; Moskowitz, 2006; Sahmer, et al., 2006).

We know that people differ in the foods they eat, the flavors they like, the perfumes they accept. We also know that although the segmentation is very strong for taste and smell, which are the chemical senses, we don't see such strong segmentation for other sensory attributes such as appearance (vision), texture/shape (touch/vision), and sounds (hearing).

Let's look at the segmentation of our 833 respondents. We're going to segment them by the pattern of the 319 utilities—one utility or impact value for each element—for each of the 833 respondents. Segmentation generates groups of people who exhibit different preference patterns.

When you read the description, keep in mind that these respondents did not "describe themselves." Actually, *we as researchers* described them by looking at the elements to which they responded most strongly and at the elements to which they responded most negatively. The elements spanned the range from sensory descriptions to product use to packaging, so the study that encompassed packaging was not limited to packaging alone. From the utility values for these 319 elements, we were able to craft the story for each segment, or at least give a précis of how these segments reacted.

The three segments that emerged can be described as follows:

1. *Convenience Seekers (30% of the total sample).* Convenience Seekers exhibit a number of clear themes in their response to the 319 elements. They want a product for busy, contemporary people. Convenience seekers want cleaner products, quick and easy, a packaging design that makes it cleaner, quicker, and easier. They want a product that won't spit, spatter, burn, or cause a mess or skin burns. *They want a liquid or a spray*, a modern product for a busy life. Overall, this segment is a very good target for an innovative cooking product. They will use a liquid for ease, speed, and cleanliness. *They will accept a contemporary package as a contributor to convenience.*

They can be convinced that the product will not burn like butter or spit like oil.

2. *Health Seekers (29% of the total sample)*. Health Seekers exhibit five themes. The product should (1) contribute to a healthy lifestyle, including specific health benefits in terms of (2) calories, (3) cholesterol, (4) vitamins, and (5) additives, respectively. Health Seekers respond to some kind of oil mix, likes a product that tastes good and a product that is versatile. Health, however, is predominant and the first three themes outweigh the two latter. Although this Health Seeker segment is similar in size to the other two, it is less motivated by the element list. Health Seekers are motivated by an oil/margarine combination, but not specifically a liquid. *They are not receptive to innovative form or packaging. They are not receptive to sprays.*
3. *Carers (27% of the total sample)*. Carers exhibit six major themes. The product should (1) come in a bottle like olive oil, (2) be made from vegetable oil. The product is (3) all about good home cooking. It's (4) the best, (5) can be used for baking, and finally, it (6) contributes to a healthy lifestyle. Caring in cooking, however, is predominant, and the first four themes outweigh the latter two. This caring segment is very responsive to the element list. *Carers are not receptive to innovative form or packaging*, but an oil-type product in a prestige package has appeal to them. They *are not receptive to sprays*. They will respond to expert and chef-based elements, and are motivated by special and family appeals.
4. The remaining 14% of the respondents did not fall into any group. Not everyone need fall into a segment. Sometimes some small proportion of respondents in a study cannot be classified.
5. Each segment represents a viable but quite distinct target. These segments exist in each country and across all user groups. These opposing minds cancel each other out. Their opposite ways of thinking drive the average toward 0, and explain the relatively neutral data—the cross-sectional data across countries averages out three quite diverse attitudinal segments.
6. It is impossible to reach everybody with a single product. Convenience Seekers and Carers tend to be diametrically opposed in terms of preferences.
7. A sense of these three segments as they respond to packaging comes from the three right-hand columns of data in Table 9.2, one column for each mind-set segment. We see in the mind-set segmentation that only the Convenience segment responds strongly to the new package features.

Exploring the Whole Gamut of Package Features

If you search the available knowledge bases, you won't find any sense of what packaging ideas work in different countries. The knowledge bases don't really provide any idea of "what works" in terms of packaging, although there is a lot of information locked up in the mind of the package designer. We can create the beginnings of this knowledge base by looking at the different ideas about packaging, at least within the realm of the study we ran.

We looked at the different elements for packaging in Table 9.1 and Table 9.2, respectively. These were only a modest fraction of the different ideas that the company had developed. One of the key benefits of experimental design of ideas is the richness of the element results. We can get a better idea of this richness when we look at Table 9.3, where we have combined the remaining two silos of package features into one group, and sorted the utility or impact by the total panel.

Looking at Table 9.3, we see that package ideas by themselves do not "sell" the respondent. However, package ideas can "unsell." There are a number of ideas that do poorly. The worst ideas vary by country. Thus, the notion of a *portionable pouch* performs very poorly in the Netherlands (impact = −13) and in Sweden (impact = −10), but is irrelevant in the United Kingdom.

Summing Up

Our study with 833 respondents provided a massive amount of data about how package elements perform in the four countries where the study was run (United Kingdom, Sweden, the Netherlands, and Germany, respectively). Let's summarize what we learned:

1. When it comes to packages for the total panel, no element really does well. Respondents have some packages that they absolutely do not like (i.e., resealable plastic pouch) (impact value = −6).
2. With respect to the additional package specifics, again we see that resealable is a strong negative.

Table 9.2 Performance of 27 different package "features and benefits" among three mind-set "segments"

Package Features and Benefits	Total	Convenience	Health	Caring
Base size	833	253	244	223
Additive constant	44	41	44	42
The product stays fresh for longer, because the packaging is resealable	3	5	3	2
The packaging allows easy opening and reclosing	3	9	4	–2
The packaging is foldable, thus minimizing space in the waste-bin	3	5	3	2
The packaging has a transparent stripe at the side of the pack to show how much is left	3	8	1	1
The packaging has a transparent stripe and measuring marks at the side of the pack for quick and controlled dosing	2	8	2	–1
The packaging has a handy and nonmessy spout for pouring out the product perfectly	2	8	2	–2
The packaging has a handle for easier carrying and pouring	2	8	2	0
The packaging is designed for cleaner handling, preventing greasy fingers	2	9	0	0
The pack allows even coverage of food or pans	2	4	3	1
The pack is fully transparent so you always see how much is left	2	8	–1	1
The packaging allows easy dosing in an instant	2	9	1	–1
The packaging can be stored outside of the refrigerator at room temperature	2	6	1	2
The packaging allows controlled dosage, which gives you more value for money	2	6	2	0
The packaging is environmentally friendly	2	3	3	3
The packaging allows dosing of just the right amount	2	5	2	–2
The packaging has a good grip which makes it handy to use	2	5	1	0
The packaging is economic, clean and handy to store	1	5	2	0
The packaging stays clean—every time. So does your fridge. And so do you.	1	9	–1	–2
The packaging is entirely recyclable	1	2	3	1
Can be used till the last drop of the product	1	2	1	1
The packaging is sealed to guarantee freshness	1	3	1	0
The packaging is tamper-proof to guarantee it hasn't been opened before	1	1	0	3
With this kind of packaging you won't need a knife or spoon anymore	1	7	1	–3
The packaging is ideal to take away, like on holiday	0	3	1	–2
The packaging allows one-hand use	0	6	–2	–2
In a modern packaging for today's people	0	7	–1	–3
The packaging is refillable	0	4	–1	–2

3. For package benefits, the impacts are also low, with a high of +3 and a low of 0. This range is important. It means that the benefits that we want to embody through packaging do not themselves have strong appeal, at least when they are described in a concept. Perhaps seeing these benefits in a three-dimensional way might change a respondent's mind. We don't know. But, there is also one bright spot. The benefits are all either neutral or positive. We have not created a "straw man" benefit that actually repels a respondent.
4. There are country-to-country differences in the responses to packages. However, there is no clear pattern. Nor, in fact, is there a database that one could use to discover general rules about "what works" in different countries.
5. Segmentation helps to reveal some stronger opportunities. There are some positives when we segment the respondents by their "mind-sets" (i.e., the pattern of utilities). One concept-response segment is open to new package ideas—the *Convenience Seekers*. The remaining two segments are less open to ideas. The *Convenience Seekers* are present in all countries, albeit to different proportions.
6. All in all, the data suggest that many of the package ideas are at best neutral to slightly positive. There are a few package ideas that are moderately negative, at worst. We would conclude that for the total panel

Table 9.3 How the different ideas for packages performed in the study, for total panel and across four European countries

	Total	Germany	Netherlands	Sweden	UK
Additive constant	44	46	50	36	43
Transparent plastic bottle	1	0	3	3	–1
Plastic bottle (oil)	1	1	0	1	1
Head stand bottle	1	2	1	1	–1
Glass bottle	1	3	1	–1	2
Dispense pump bottle	1	0	1	1	1
Comes in plastic bottle that can be folded away after use	1	3	–1	1	–1
Comes in an up-side-down squeeze pack	1	–1	1	2	1
Comes in a bottle similar to olive oil	1	0	1	0	3
Collapsible plastic bottle	1	–2	3	3	0
Better plastic bottle (smaller cap)	1	–1	2	2	0
Plastic bottle (handle)	0	2	1	–1	0
Comes in a very lightweight plastic bottle	0	–1	1	2	–2
Comes in a squeeze tube	0	–2	0	–1	3
Comes in a squeezable plastic bottle	0	–3	1	1	1
Comes in a plastic bottle with a handle	0	–1	–2	2	2
Comes in a plastic bottle	0	–1	–2	3	0
Comes in a bottle similar to “seed or cooking” oil	0	1	0	0	1
Tin can (round ribs)	–1	–2	2	–3	–2
Spray dispense aerosol	–1	–2	–4	–1	2
Reclosable stand pouch	–1	–1	0	–2	–1
Reclosable gable tetra	–1	–2	0	–2	0
Plastic bottle “Milda”	–1	–2	–1	–2	–1
Gable tetra (without cap)	–1	–2	–2	1	–1
Existing bottle	–1	–2	–2	–1	2
Comes in a plastic bottle with a cardboard-label wrapped around it, which can be recycled separately	–1	1	–4	1	–2
Comes in a plastic bag within a cardboard-box, which can be recycled separately	–1	0	–6	2	–2
Comes in a glass bottle	–1	–2	–2	–1	0
Comes in a dispenser pack, which gives a fixed amount of product each time the top is pressed	–1	2	–4	–3	1
Bottle-in-Box	–1	–4	0	0	0
Bag-in-Box	–1	–1	–1	2	–3
Tetra plus cap (UHT)	–2	–2	–4	–1	–2
Squeeze tube	–2	–4	–3	–2	0
Comes in a tetra-pack as used for milk cartons	–2	–5	–2	1	–2
Pump spray packaging	–3	–1	–5	–6	2
Dosing tube	–3	–2	–7	–4	1
Comes in an aerosol spray can	–3	–8	–3	–5	3
Wall dispenser	–4	–1	–7	–6	–3
Spray	–4	–4	–8	–7	1
Comes in a wall dispenser (i.e., pushing the pan against the dispenser gives a fixed amount of product	–4	3	–8	–7	–4
Comes in a stand-up pouch as used for refill packs	–5	–4	–12	0	–3
Comes in a non-aerosol pump spray pack	–5	–4	–11	–5	–1
Comes in a metal/tin can	–5	–5	–5	–6	–4
Comes in a box of many single-use sachets	–5	–4	–9	–5	–2
Re-sealable bag	–6	1	–14	–4	–5
Comes in a re-sealable plastic pouch (bag)	–6	–3	–12	–3	–4
Portionable pouch	–7	–5	–13	–10	0

packaging is not particularly important when combined with other elements of a different nature (i.e., flavor promises, health benefits). Look to the segment for the big opportunities.

Technical Appendix—How to Run Very Large Studies with Many Elements

How to deal with 316 different elements is a complete technology unto itself. We can explicate some of the approach here, in the technical appendix. The goal is to create individual-level models, one model for each of the 833 respondents, showing the part-worth contribution of each of the 316 different elements. At the end of the day, we know what every element contributes to interest, and to the two other attributes.

The basic idea behind running large-scale studies with conjoint analysis was first proposed by Moskowitz and D.G. Martin in 1993. The approach began with the following facts and constraints.

Fact 1: Experimental design of ideas (conjoint analysis) reveals how elements "drive" responses. Thus, if we are dealing with 300+ elements, we want to determine how each element drives the end response. The response can be interest, uniqueness, end use, etc.

Fact 2: It is ideal for each person to evaluate as many concept "elements" as possible. Ideally when the number of concept elements is limited, each respondent may evaluate all of these elements. On the other hand, it is clearly not possible to evaluate all of the elements with large numbers of elements, because there may be hundreds of these elements.

Fact 3: A person can participate in a conjoint study, evaluating test concepts. However, the person's ability to concentrate on the task and to give honest answers is certainly not infinite. Any normal respondent will become bored, sooner or later. The limited ability of a person to sit through a long interview means that either the number of elements has to be shortened, or the individual respondent can test only a fractional portion of the elements. If the individual tests only a portion, then we can either choose to average across individuals whose results we recognize to be incomplete. Or, we can develop a method to "complete" the data from any individual.

Fact 4: To build a model for a given individual, that individual must test a specific number of concept elements, and test each concept element against different backgrounds. Only by systematically varying the backgrounds in an experimental design is it possible to estimate what each element contributes, and "partial-out" that contribution from the contribution of the remaining elements.

Fact 5: The individual can test only a partial set of elements through experimental design. The impact or utilities of these elements can be estimated through standard regression analysis (also called ordinary least squares). The independent variables are the elements that appeared in the test concepts. These elements are coded "1" if the element appeared, and coded "0" if the element did not appear. The dependent variable is the rating assigned to the test concept. The rating may either be a scale (i.e., 1 to 9) or a binary (0, 100).

Fact 6: After an individual evaluates his full set of elements, and after the regression analysis estimates the utilities for the tested elements, it is necessary to estimate the utility of untested elements. The estimation of utility or impact values is akin to estimating the value of "missing data."

Fact 7: One way to estimate the "missing data" is by means of "dimensionalization." Briefly, dimensionalization works by having a small group of respondents "profile" each concept element on a set of nonevaluative attributes or dimensions. Then for any individual, an algorithm estimates the value of "untested elements" by replacing the not-as-yet estimated utility of the untested element with the average of the closest elements whose utilities are known. By "close" we mean those elements whose dimensional profiles are similar to the dimensional profile of the untested element (Moskowitz, Porretta, and Silcher, 2005).

Fact 8: When all of the elements are dimensionalized, it becomes a straightforward computational task to estimate the utilities or impacts of the untested elements. These utilities are continually refined, by repeating the algorithm, always getting a better estimate of the utilities that are estimated, but never touching the utilities of the tested elements. These utilities of tested elements are established through research and are never modified.

For further information on the use of dimensionalization in IdeaMap®, and the results in large-scale studies, the reader is referred to Moskowitz, Porretta and Silcher, 2005.

References

Green, P.E. and Krieger, A.M. (1991) Segmenting markets with conjoint analysis. Journal of Marketing 55, 20–31.

Jacobsen, T. and Gunderson, R.W. (1986) Applied cluster analysis. In: J.R. Piggott (ed.) Statistical Procedures in Food Research. Elsevier Applied Science, London and New York, pp. 361–415.

Moskowitz, H.R. and Martin, D.G. (1993) How Computer Acided Design and Presentation of Concepts Speeds Up the Product Development Process. Proceeding Of The 46th ESOMAR Conference, Copenhagen, Denmark, pp. 405–419.

Moskowitz, H.R. (1996) Segmenting consumers on the basis on the basis of their responses to concept elements: an approach derived from consumer research. Canadian Journal of Consumer Research, 15, 38–54.

Moskowitz, H.R. (2006) Clustering people, defining segments and the limit to product "knowability." Food Quality & Preference 17, 661–663.

Moskowitz, H.R., Porretta, S., and Silcher, M. (2005) Concept Research in Food Product Design and Development. Blackwell Publishing Professional, Ames, IA.

Qannari, E.M., Vigneau, E., Luscan, P., Lefebvre, A.C., and Vey, F. (1997) Clustering of variables, application in consumer and sensory studies, Food Quality & Preference 8, 423–428.

Sahmer, K., Vigneau, E., and Qannari, E.M. (2006) A cluster approach to analyze preference data: Choice of the number of clusters. Food Quality and Preference 17, 257–265.

Tang, C., Heymann, H., and Hsieh, F-H. (2000) Alternatives to data averaging of consumer preference data. Food Quality and Preference, 11, 99–104.

Vigneau, E., Qannari, E.M., Punter, P.H., and Knoops, S. (2001) Segmentation of a panel of consumers using clustering of variables around latent directions of preference, Food Quality & Preference 12, 359–363.

Westad, F., Hersleth, M., and Lea, P. (2004) Strategies for consumer segmentation with applications on preference data. Food Quality & Preference 15, 681–687.

Chapter 10

What Should My Package Say?

In companies, the designer doesn't just go off, willy-nilly, to create the package design, no matter how talented the designer may be or no matter how excited the external design company says they feel. Rather, a lot of what can be said has to motivate the customer to buy the product. Just exactly how do you discover what motivates?

Well, the answer can be obtained either by guessing (happens more often than you might like to believe), by promise testing (which of these statements do you like?), or by experimentally designed concepts where the messages are mixed and matched. We're going to illustrate the homework by looking at experimentally designed concepts for cereal. We'll find that the designer has a lot to choose from, once the experiment is run, and that the winning ideas "float to the top."

A lot of what we will present comes from the world of concept testing. In fact, we are going to focus this chapter on getting the right language, before we go to the design aspect. As you will see when you read further, the "right language" doesn't just mean the correct topics to put on the package. Rather the "right language" means the right way of saying what you want to say, efficiently, and persuasively.

Talking about Cereal

Where did this ever-so-popular staple found in our cupboards originate? Well, it has a place deeply embedded in our history. For example, in Eurasia, remains of domesticated cereals (barley, einkorn, and emmer wheat) as well as pulses and flax have been dated at various sites to 8 millennium BC (Harris, 1996). Cereal has been around a long time and from all commercial activity around it, it looks like cereal is here to stay. The word cereal itself comes from CERES, the Roman goddess of harvest and agriculture. Cereal grains are grown in greater quantities worldwide than any other crop. In some developing nations, cereal grains constitute nearly a person's entire diet (Source: Wikipedia.org).

Let's move forward 2,000 years to today and explore the origins of our favorite breakfast food. The first modern and commercial cereal foods were created by the American Seventh-day Adventists. A century and a half ago these Adventists formed the Western Health Reform Institute in the 1860s. The Institute was later renamed the Battle Creek Sanitarium after its location in Battle Creek, Michigan. The Adventists manufactured, promoted, and sold wholesome cereals. In 1906, now just a century ago, Will Keith Kellogg established the W.K. Kellogg Foundation. Twelve years before that, in 1894, Kellogg had been busy improving the then rather miserable diet fed to hospital patients. Kellogg believed that he could develop an easily digestible substitute for bread. The answer, he surmised, lay in some variation of the process of boiling wheat to improve its digestibility. Like so many inventions, it began with a simple error. In his search for this digestible bread substitute, Kellogg accidentally left a pot of boiled wheat on the oven. During the time Kellogg was busy with other things and neglected the wheat. In turn, the wheat in turn followed nature and in the presence of hot water the wheat softened (i.e., became tempered, in the language of the food industry). When Kellogg played with this wheat by rolling it, he deliberately left some grains to dry after rolling. When dried, the grain of wheat found a new life, emerging as a large thin flake. The taste of the wheat was simply delicious. And so corn flakes, our first commercially processed cereal, entered America and then the world (Source: www.inventors.about.com).

The foregoing story about a now well-known, indeed iconic product, paints a nice background for our project. There are so many variations of cereal products that knowing "what to say" requires some experimentation. Cereals provide more food for human consumption than does virtually any other crop. We're all familiar with the

 ISBN: 978-0-813-81222-9

now-common cereals made from rice, wheat, barley, oats, maize (corn), sorghum, rye, and certain millets, with corn, rice, and wheat being the most important. A new cereal, triticale, adds to the list, but triticale is a man-made, genomically created product. Triticale comes from crossing wheat and rye, thus doubling the number of chromosomes (Source: AskDrSears.com).

Today, cold cereal, also called RTE or ready-to-eat, is a booming industry, constantly growing to keep up with the ever-changing demands of today's savvy consumers. As recently as 10 years ago, the industry presented consumers with noticeably fewer choices. Consumers typically didn't demand much. Observational research, so-called "shopper insights," revealed that consumers generally homed in immediately on what they wanted. They typically went directly to the familiar location on the shelf to find and select their family's cereal.

Now, a walk down the cereal aisle can be quite long, filled with offerings, amazing in the variety to be found. And so looking at how companies market their cereal products provides us with a nice study topic to introduce to our world of design. But first, let's deal with the ideas, the hot buttons that ultimately the designer will have to incorporate in the package.

With such a history and pedigree, cereals have lots of things that can be said about them. All cereal grains have high-energy value, coming mainly from the starch fraction but also from the fat and protein. In general, cereals are low in protein content, although oats and certain millets are exceptions. Whole-grain foods are valuable sources of nutrients that are lacking in the American diet, including dietary fiber, B vitamins, vitamin E, selenium, zinc, copper, and magnesium. Whole-grain foods also contain phytochemicals, such as phenolic compounds, that together with vitamins and minerals may play important roles in disease prevention (see Burkitt and Trowell, 1975; Jacobs et al., 1998; Slavin et al., 2001; Liu, 2003; Koh-Banerjee et al., 2004; Flight & Clifton, 2006; Kochar et al., 2007).

So, with so many good things to say, what works? Specifically, how can we get winning ideas about what we might put on a package?

Cereal—A Matrix on Which to Develop New Ideas

We, in the early part of the twenty-first century, no longer think of cereal as merely a breakfast item. The magic and innovativeness of marketing changed all that. Cereal is eaten as a snack, for lunch, and even for dinner. Cereal has even penetrated the out-of-home-eating situations. Cereals can be found in the form of snack bars—grab, eat, and run. Marketers relish such expansion of uses, and the sheer number of opportunities (also called meal occasions) sparks the corporate imagination. Just think of the young brand manager who is handed the job of growing a company's cereal business, or as it happens, to grow the use of a specific cereal brand.

In the increasingly competitive food industry, this opportunity could mean a new product, a new form, even a hybrid product that combines cereal with "something else." Think of breakfast bars, which 20-plus years ago were a great new, innovative idea, allowing people to eat cereal "on the run." You can scarcely go to a convenience store or a health-and-wellness store like Whole Foods, Inc., without being assaulted with the latest innovations in cereal bars.

Cereal is also finding its place among established comfort foods such as ice cream and chocolate. Simply stated, in some circles, cereal also has turned into a mood food. Think of the days your mother would make you a bowl of cereal in the morning before you went off to school. The old familiar flavor and aromas flood your memories with each spoonful to bring you back to simple pleasures.

How the Company Gets Ideas for Cereal

To prepare for this chapter, we took a trip to three supermarkets and observed what people did when they shopped the cereal aisle. You can do this yourself; it doesn't require much expertise, just a bit of discretion, although probably not even that. In the cereal aisle, we watched how people shopped the cereal (also called shopped the cereal category). Shoppers picked up the box, and after about 2–5 seconds (i.e., quickly but not immediately), they turned the box to the side and the back, read the panel, and then replaced the product or, in some cases, put the cereal into their shopping carts. For the most part, shoppers looked as if they were weighing some aspects of the cereal, which was quite surprising to us who had been schooled in the fact that today's shopper hardly pays attention to the product anymore.

Our research question was fairly simple. What are these consumers looking for? What drives their choices? Why in a world where time is so precious and where the minutes in a supermarket seem to be spent scurrying about, do people who pick up cereals look so carefully at the package? Is it taste, nutrients, brands, respectively? Or, perhaps, is it the eye-catching glitzy package design and graphics? Does it differ among males and females, or differ between younger and older ages? What differences can we find between people who consume cereal on a frequent basis as opposed to those who don't? What's going on in the consumer's mind? We set out to test these observations of consumers with consumers. Who could tell us better?

We wanted to test a wide range of stand-alone messages or elements. Elements are pieces of information, communicated as text statements, which define a feature, benefit, claim, idea, etc. We tested these elements among consumers to see which are most appealing and resonate best, what they don't really care about, and what turns them off altogether. We're not yet at the stage of dealing with packages, but rather just with the information that might be contained in the graphics. So let's look at what the results offer the product manufacturer or graphic designer, who then has to create the most effective package for the cereal product.

Setting the Stage for Research—Systematic Combination of Ideas

Keep in mind the systematic experimentation that we describe extensively in the section on "Tools." This tool, RDE (Rule Developing Experimentation) works in a straightforward manner. To reiterate a couple of salient ideas:

1. People don't know what they want until they see it.
2. You get more realistic results when you present combinations of ideas together. That's the way nature works anyway, so let us replicate nature rather than present people with one idea at a time.
3. If you systematically vary the combinations or vignettes so the respondent sees a set of these combinations, then you can identify what each cereal element "brings to the party." You use regression analysis, a standard statistical method found as a feature in most spreadsheet programs, but almost always available in common statistical packages. We talked about regression analysis in Chapter 5 on systematics.
4. We had a choice of using the ratings or dividing the ratings into classes (accept versus reject the combination of ideas). We chose to work with the binary world of accept/reject. This leads to a specific interpretation of the numbers we get from regression, namely *the proportion of our respondents who accept a specific idea or element*.
5. We worked with 36 elements. The full set of elements to be tested divided into six logical silos or buckets, each silo comprising six different, related, or similar types of elements. The respondent evaluated 48 different combinations, with each element appearing three times in the set of 48 combinations. Each respondent evaluated a separate and unique set of 48 combinations, although from time to time a respondent might evaluate a concept that another respondent had seen. For the most part, however, the concepts were unique, avoiding the bias that would occur if the same combination were to be evaluated by everyone.
6. Each element was a free agent in the set so that we could estimate the impact or contribution of each element to the rating. We performed this analysis at the individual level.

Finding Ideas to Test—Sometimes Easier but Also Occasionally Harder than You Might Think

When we were brainstorming the categories that would best describe our topic, we simply thought about what first comes to mind when we think about cereal. Although people might feel that coming up with these ideas is difficult, the truth of the matter is quite different. Think about a product for a long time (e.g., 5–10 minutes!), and you are likely to come up with the silos and the elements. Let's follow this train of thought:

Silo 1: Texture

We selected this category to represent the various "mouthfeels" of a cereal. It's well known that cereals are prized for texture. We wanted to look at textures that were crunchy, those that were light and crispy, those that were hearty and dense, and even those that were soft and chewy.

Silo 2: Flavor

Cereal flavors can be traditional or niche. Examples of traditional are Brown Sugar and Cinnamon. Examples of niche flavors are Vanilla Bean and Milk Chocolate. Packages often feature graphics to support the flavors, but we merely worked with flavor names.

Silo 3: Ingredients

What do consumers really think is important, when it comes to ingredients or fundamentally the content of the cereal? Do they really care about fiber, whole grains, reduced sugar, or fat content? Do these ingredients drive a person to buy?

Silo 4: Nutrients

We hear so much about consumers wanting their food choices today to be nutritious, especially in a food that typically starts the day. How strongly do claims such as "All natural, No artificial flavors, No preservatives, or Provides 100% of the daily value of 10 essential vitamins and minerals" resonate among consumers?

Silo 5: Benefits

The cereal category is often driven by slogans or by images that the phrases conjure up. There are "benefits" to the consumer. We tested ideas such as "Great for a late night snack; A mid-day snack that will hold you over until your next meal; and An alternative to lunch."

Category 6: Brands

With more and more brands appearing on the shelves each year, we selected a mix of some of the top-selling brands to specialty brands to measure the overall influence of the impact *brand* plays in purchase.

From "Germs of Ideas" to Composing Elements

A question that often arises is "How do we get elements? After all, we are not professional copywriters. How should we know what to say? Isn't that the job of a professional?" To deal with the issue of "writing elements," we went to the Internet for assistance and looked up cereal websites of leading manufacturers to find meaningful elements to best fit in these six aforementioned categories. We placed what struck us as relevant elements into judgmentally the appropriate categories. The final set of elements appears in Table 10.1. We will talk about how to get these elements, how to test them, and what the data look like. For right now, just look at Table 10.1 to see the 36 different elements that we developed for this cereal test.

> *Parenthetically, such "competitive analysis" is often the best way to jump-start one of these exercises. The competitive frame often provides great examples of ideas, both in word and in picture. Thumb through a dozen or so websites in a product category and you will definitely come away with many new ideas.*

Getting People to Participate in Interviews

When the Internet first began in the late 1990s, it seemed that the supply of participants for a study was virtually almost inexhaustible. In fact, Dennis Gonier, then CEO of Digital Marketing Services, Division of AOL, coined the phrase "stream of respondents," almost reminiscent of a stream filled with fish. In those early, pioneering days it was simple to put out a notice of a "survey," and get dozens of people to participate.

A lot has happened in 10 years. It's harder to get people to participate. The novelty of Internet-based surveys has worn off. To get our panels, we sent out an e-mail that told them a little about the study and offered them a chance to win money in a sweepstakes. Prizes are the norm for today, and this study is no different.

Most researchers who do these types of studies now work with so-called Internet panel providers (i.e., companies that specialize in getting panelists to participate). We did as well. The participants were to be between the ages of 18 and 65, who had purchased and eaten cereal in the past three months. They had to be primary household grocery shoppers and could not work in certain industries such as marketing research, advertising agencies, or cereal-marketing companies.

When the panelist received the email, all that was necessary to participate was to click on the embedded link in the survey, or paste the link into the browser. The panelist was directed immediately to the "welcome page," shown in Figure 10.1. Welcome pages are exactly what you might think they are—pages that tell the panelist what the study is about, what the rules are, and of course, what the prizes are. As in most research, the less said about the study the better. In that way we don't bias the respondent. To this end we kept the introduction fairly simple. Rather than telling the respondent about cereals in general, we simply said, "We are interested in your opinions about cereal." The respondents were then told they would evaluate concepts and would rate each

Table 10.1 Getting the language right. The table shows six silos of cereal elements, each silo having six elements or options. The elements are shown in rank order according to the performance among the total panel of 446 respondents.

	Data from the total panel of 446 respondents	Total
	Additive constant	41
	Silo #1: Texture	
A4	The crunchy texture of thick flakes and big clusters of nuts	3
A6	Crunchy cereal that never gets soggy in milk	2
A5	A crunchy rice and oat clusters cereal	1
A2	Crunchy on the outside filled with soft, melt-in-your-mouth filling	0
A1	The thin, light and crispy texture of your favorite flakes	−1
A3	Hearty, dense texture of only the finest ingredients	−2
	Silo #2: Flavor	
B1	Plain and simple, made with brown sugar and cinnamon for a traditional taste	3
B4	Made with imported dried fruits that plump up nice and juicy when adding milk	2
B6	Enjoy the simple combinations of apples and cinnamon	2
B5	The old time favorite Honey Nut	0
B2	Now in Vanilla Bean and Milk Chocolate. An indulgent cereal	0
B3	For the sophisticated taste buds …flavored with Hazelnut, Amaretto and Kahlua	−4
	Silo #3: Ingredients	
C2	Made with whole grain, a good source of fiber, important in reducing your risk of chronic diseases like stroke and diabetes	6
C1	Only 100 calories of wholesome goodness per serving	3
C6	Helps you lose weight the safe, healthy way	3
C4	Low in sugar	0
C3	Low fat, only 1 g fat per serving	−2
C5	Use as part of your points system	−3
	Silo #4: Nutrients	
D5	Provides 100% of the daily value of 10 essential vitamins and minerals	2
D4	Contains essential omega-3 fatty acids, which may reduce your risk of heart disease	1
D3	Full of antioxidants and phytonutrients that help you to maintain your heart health	1
D6	Helps you maintain a healthy lifestyle	0
D2	All natural, no artificial flavors, no preservatives	−1
D1	100% organic	−2
	Silo #5: Benefits	
E6	Fills that empty spot in you, any time of the day	1
E1	A quick and easy breakfast	0
E4	Great for a late night snack	−1
E2	A midday snack that holds me over until my next meal	−2
E3	A wonderful alternative to lunch	−2
E5	A food you feel good about feeding your family	−3
	Silo #6: Brands	
F4	Made by Quaker	2
F5	From your local supermarket	2
F2	Made by General Mills	0
F3	Made by Post	−1
F1	Made by Kashi	−4
F6	From your local specialty/gourmet store	−8

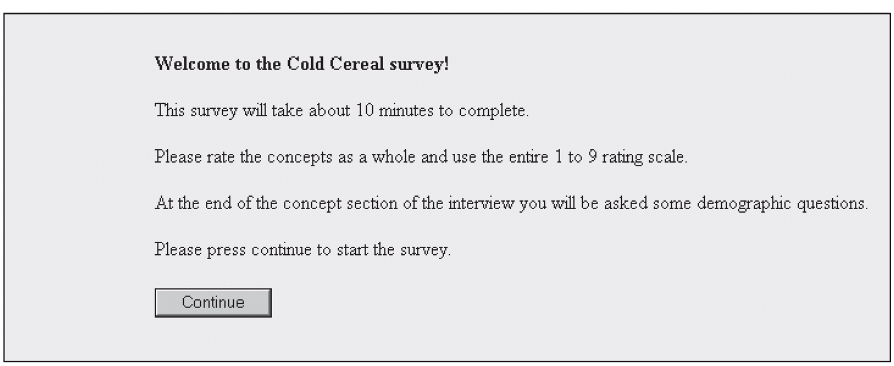


Figure 10.1 The welcome page for the cereal study.

concept on a simple 9-point scale: "How likely would you be to purchase this cereal?" 1 = Not at all likely, 9 = Very Likely.

What Test Concepts Look Like

Right now we're dealing with text-based concepts, although most of the rest of this book will deal with graphics. Text-based concepts are fairly straightforward. The elements appear as "bullets," or short stand-alone phrases, one stacked up on the other, as we see in Figure 10.2. Quite often purists state, sometimes quite vehemently, that the concepts have to be written out in full paragraph form. Actually, the form that we see in Figure 10.2 does just fine. Panelists have no problem reacting to this type of disjunctive set of elements. They simply read it and respond, much as they would do for complete paragraphs. The only difference is that this format is quite a bit easier on the eyes and on the mind. Furthermore, for research purposes, the format of concepts in Figure 10.2 is easy when one wants to work with new ideas. One need not spend hours trying to find just that "proper bridge" to link two ideas. The respondent's mind does all the work.

A Look at the results—What Should We Say?

Our analysis relates the presence/absence of the 36 elements to the individual respondent's ratings. But before we did the analysis, we looked at the rating for each concept, and followed the conventions of market research:

1. We divided the 9-point scale into two locations, to represent accept or reject.
2. A rating of 7, 8, or 9 represented "accept" the concept (i.e., would buy the cereal). We coded that acceptance as 100.
3. A rating of 1, 2, 3, 4, 5, or 6 represented "reject" the concept (i.e., would not buy the cereal). We coded that rejection as 0.
4. The specific "cut-point" (i.e., 1–6, 7–9) was arbitrary. We had used this cut point many times before, and found it to work in terms of predicting real world behavior. So, we used it again here.
5. Each concept tested by each person was thus recoded, with the rating on the 9-point scale replaced by the binary, 0 or 100.

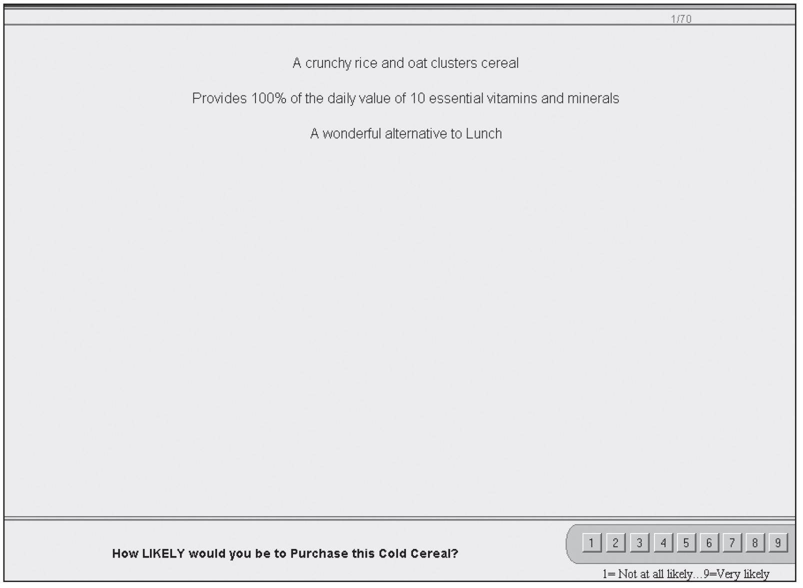


Figure 10.2 Example of one concept, comprising three elements. The rating scale is at the bottom of the concept.

Much of the story appears in the data tables. Let's return to Table 10.1, which shows the results from the total panel of 446 respondents. Parenthetically, these types of studies often provide very strong data and are stable by the time we get to 50–75 respondents. The base size of 446 is somewhat "overkill" but will allow us to look at subgroups or different sets of people in the population.

Now we will go through the results in Table 10.1, which come from the total panel. Each row has meaning.

Let's look first at the additive constant. The constant is 41. That means that 41% of the respondents are interested in cereal if there is no element present. Clearly all concepts had elements, so this number "41" is an estimated parameter. Still, it's an important number to keep in mind. It's a baseline value. You can compare this number across different groups of respondents in the set of 446 individuals, or across studies. Just for comparison purposes, credit cards have a baseline value or additive constant of only about 15–20, meaning that only 1 person in about 5 or 6 is interested in credit cards to begin with, whereas for cereal we already start with 2 in 5.

The story starts to get more interesting when we delve into the elements, although we wouldn't know that from the total panel. We'll see the interesting findings when we get to segments. Right now, however, let's just look at what the average data from our 446 respondents tell us:

1. For this table we show the six different silos separately. Within each silo we sort the elements from best performing to worst performing, based on the impact or utility value.
2. There are some rules of thumb for interpreting the impacts of the different elements. First, the number is the proportion of respondents who would change their rating from not buy (1–6) to buy (7–9).
3. Second, there are some norms:
   a. Impact >15 = Extraordinarily strong performer, keep this element
   b. Impact 10–15 = Strong performer, keep this element
   c. Impact 5–10 = Significant performer and relevant
   d. Impact 0–5 = Marginal performer
   e. Impact –5–0 = Poor performer, detracts, eliminate unless absolutely necessary
   f. Impact less than –5 = Causes damage, avoid
4. Our top scoring element, "Made with whole grain, a good source of fiber, important in reducing your risk of chronic diseases like stroke and diabetes" will push an additional 6% of respondents from voting "not interested" to voting "interested."
5. We see a few more elements that generate modest positive values. They mention crunchy texture, low calories, and weight loss. But none of these elements by itself comes across as "Wow, this is what we need to capture for our product!"
6. For the most part, the elements hover around 0, so they neither drive acceptance nor drive rejection. We already know that values ranging from 0–5 add little value to interest. This is the landscape where the majority of the data falls. There isn't much in the data that will add to purchase interest among total sample.
7. Surprisingly, cereal from the local specialty or gourmet store is a turnoff (–8).
8. Trusted brands don't bring much to the story. Surprising, yes! But this answers one of our ongoing questions: What drives consumer behavior when purchasing cereal?
9. Where are the healthy promises an organic cereal will deliver, or the innovative flavors for the sophisticated taste buds? Weren't any of the consumers interested in these areas?

Looking at the Frequent Cereal Eater

What's happening if we look at the data based on the amount of cereal a person says he eats? We asked respondents to tell us how frequently they eat cereal. Do they eat it everyday as part of their daily routine, or occasionally, more like when the mood hits them? Here an interesting story appears among heavy cereal consumers (once a day or more often) versus moderate consumption consumers (several times a week). We didn't put in all of the frequency subgroups in Table 10.2, which shows the results.

Our "heavy consumption" consumers who ate cereal frequently had a constant of 43, similar to the total panel. This group was motivated by a cereal made with whole wheat goodness and fiber with traditional taste and flavors that were plain and simple with cinnamon and brown sugar. They wanted a crunchy texture of thick flakes, big clusters of nuts, and lower in calories. Enhancing the taste and texture profiles with imported plump fruits also attracted them. This group wants a traditional taste and flavor with the rewards of health benefits. They are similar to those people we mentioned

Table 10.2 Best performing elements based on how much cereal a person eats or purchases. We show only the extremes of the frequency subgroups, not all the frequency subgroups.

		Total Sample	Frequent Cereal Eaters	Frequent Purchases of Cold Cereal	Infrequent Purchases of Cold Cereal
	BASE SIZE:	446	100	240	84
	CONSTANT:	41	43	46	22
	Heavy Eaters				
C2	Made with whole grain, a good source of fiber, important in reducing your risk of chronic diseases like stroke and diabetes	6	13	6	5
B1	Plain and simple, made with brown sugar and cinnamon for a traditional taste	3	9	2	1
B4	Made with imported dried fruits that plump up nice and juicy when adding milk	2	8	3	–1
F5	From your local supermarket	2	7	4	1
	Frequent Purchasers				
C2	Made with whole grain, a good source of fiber, important in reducing your risk of chronic diseases like stroke and diabetes	6	13	6	5
	Infrequent Purchasers				
C2	Made with whole grain, a good source of fiber, important in reducing your risk of chronic diseases like stroke and diabetes	6	13	6	5

in the beginning of the chapter—those shoppers who walk down the cereal aisle and know exactly where they're going—to their old-fashioned favorite with no bells and whistles, just trusted goodness with a familiar taste.

We now know what to put on a package to attract them if they do the shopping for cereal! Let's find out more about shoppers to see whether we can continue with our discovery. Let's also look at those who buy the cereal. A frequent buyer of cereal may buy it for someone else in the household. These may or may not be the ones who actually eat the cereal.

This time we divide our 446 respondents into two groups, those who are frequent purchasers and those who are infrequent purchasers. We use the classification data at the end of the interview to find out where a person fits. The bottom line here is that no element really pops. The only element to do anything is the whole grain, probably because we've added in the risk reduction for stroke and diabetes.

If we were to summarize, we would say that consumption frequency probably affects the impact of the concept elements, but buying pattern does not. The pattern is not clear why winning elements for frequent eaters really win at all. We really don't know why. It's hard to uncover the underlying reasons why the four elements that strongly appeal to the frequent cereal eater do so. But hold on—there's more to come in the next section!

Different Mind-Sets Lead to Stronger Messages!

Since we don't see anything of significant interest among the total sample, and only slight effects when we look at eating or purchase frequency, we should look somewhere else. Fortunately, researchers have recognized the fact that people profoundly differ from each other in what they like, but these differences may not manifest themselves in how people describe their behavior. These different groups are mind-set segments. We have seen the power of such segmentation again and again in these chapters, and we will continue to do so. These segments are "real," and profoundly different from each other.

We grouped our 446 consumer respondents into three "mind-set segments", according to what elements "drive" their stated purchase intents. The segmentation is straightforward. We used the statistical method of clustering, a set of well-established approaches, to group

Table 10.3 Winning elements from the three mind-set segments evaluating cereal

	Total	S1	S2	S3
Base Size	446	80	287	79
Additive Constant	41	43	40	43
Segment S1: *Conventionalists*				
Enjoy the simple combinations of apples and cinnamon	2	15	–2	2
Plain and simple, made with brown sugar and cinnamon for a traditional taste	3	13	–1	6
Made with imported dried fruits that plump up nice and juicy when adding milk	2	12	–3	14
The old-time favorite Honey Nut	0	10	–3	0
Segment S2: *Reassurance*				
The crunchy texture of thick flakes and big clusters of nuts	3	–1	5	3
Made with whole grain, a good source of fiber, important in reducing your risk of chronic diseases like stroke and diabetes	6	4	4	16
Only 100 calories of wholesome goodness per serving	3	3	4	1
Segment S3: *Indulgents*				
Now in Vanilla Bean and Milk Chocolate, an indulgent cereal	0	7	–9	23
For the sophisticated taste buds, flavored with Hazelnut, Amaretto and Kahlua	–4	14	–15	16

together individuals with similar patterns of what they liked. People with similar patterns (i.e., similar utility values for their elements) fall into the same cluster or segment.

You can see these results in Table 10.3. Here we simply sort the elements from best to worst for each segment and show the results for the three segments. Let's look at the segments one by one. We'll only look at the winning elements for each segment. Otherwise the analysis will become unduly complicated. If you look at Table 10.3 you will see how each of the "winning elements" scores in both the segment it appeals to and how well it scores among the other two segments. Look closely and you will see that a winning element in one segment may be a losing element in another segment. This countervailing force of one segment versus another may be one reason why the elements don't score very well among the total panel. The segments neutralize each other. The same neutralization story occurs again and again, in product category after product category. And so it occurs here, as well.

Now let's delve into the data more deeply. The best way to follow along is by looking at the best scoring elements for each segment that we show in Table 10.3. Once you understand that, step back and try to develop a "story" about each segment, using as cues the elements to which they strongly respond.

Segment 1: Conventionalists. These consumers look for the old tried and true: simple combinations of apples and cinnamon, plain and simple, made with brown sugar and cinnamon or a traditional taste. To keep them happy, make it simple and keep it traditional. There is only one element that appeals both to *Conventionalists* and to *Indulgents*. This element is *made with imported dried fruits that plump up nice and juicy when adding milk*. It could well be that the same element conveys two different messages, depending on the segment. To *Conventionalists* the hot button is "plump up nicely when adding milk." To the *Indulgents* the hot button could be "imported dried fruits." Of course this is only a hypothesis, but it could be tested easily by dividing the element into both halves, and testing each half again, but as a separate element.

Segment 2: Reassurance. Our largest segment looking for a brand they know and trust, a healthy cereal that's good for them, one that they can rely on to deliver whole grain goodness, fiber, and a crunchy texture.

Segment 3: Indulgents. Here is strong opportunity for a product developer looking to carve a niche appeal in the industry. *Indulgents* may be a group of savvy consumers with sophisticated palates. They're looking for innovative flavors such as Kahlua, Hazelnut, Amaretto, Vanilla Bean, and Milk Chocolate.

You can entice them with imported dried fruits in a cereal, which will also deliver whole grain goodness and fiber. This group wants a new twist to an old staple. They're looking for something new in the cereal aisle.

How Many Segments Are Best?

Throughout this book we will be referring to segments and using segmentation as a way to better understand the business opportunities. Sometimes we will present two segments, sometimes three, sometimes four, rarely if ever five. The operative question here is "How many segments and why?"

First, keep in mind that segmentation is not an exact science. You can't just approach the data and expect it to give up the segments in an absolute sense, just like you cannot divide a group of people and say that the division is the "absolute best" way of partitioning the group. The reality is that segmentation is subjective. Segmentation depends on the criteria or variables on which you segment, the statistical methods for dividing the respondents, and of course on the degree to which you want to have many segments versus a few segments.

Each of these issues deserves a short comment here, as preparation for our upcoming chapters.

1. Segmentation depends on the variables you select and the criteria you use to segment. We will do three types of segmentation in this book. The first way will be dividing people by what we know about them. Thus, we might divide the people on the basis of gender (male versus female), by income, education, etc. The second segmentation will be on the basis of what they tell us about themselves, their attitudes, their body state, etc. We just saw a few examples of this when we divided the people by their self-stated frequency of buying cereal or consuming cereal. The third way will be on the basis of the response to concept or package design elements (so-called mind-set segmentation). We just saw an example of this for our mind-set segmentation.
2. Segmentation depends on the statistical methods you choose. This is important. We could divide the respondents using a procedure that puts people into different clusters so that the "distance" between people in a cluster is "small," and the distance between the centroids or center points of different clusters is "large." Distance itself can be measured in different ways. Depending on the measure of "distance," we will create different looking clusters. However, for the most part the results will be reasonably similar across segments.

Wrapping It Up—Now What?

So, what have we learned from this exercise and how can a product developer use these findings to create a package design for a new cereal?

1. The key findings of this research lie in the impacts of element utilities (i.e., the power of messages to motivate interest or detract from interest). You need to know both the positive performing and negative performing elements when you develop visuals and messaging for packages.
2. No one silo of elements piqued respondent interest in and of itself. The story wasn't that cut and dried.
3. The heart of the learning emerged from the mind-set segmentation. Here the product developer has opportunity for diversity in both product offerings and visual communications.
4. The *Conventionalists* segment is motivated by images that convey simple combinations of apples and cinnamon, plain and simple, made with brown sugar and cinnamon or a traditional taste. A package that when you look at it reminds you of a taste of old-fashioned goodness filled with childhood memories. Don't miss the opportunity to add extra value to the customer experience with these nostalgic images. This group is not motivated by health benefits or brand names.
5. Or you can target the *Reassurance* segment with your trusted brand name and messages which reassure them of health and that this is a cereal that's "good for them." This is a very strong segment of people that can't be overlooked and don't need much to keep them as customers.
6. And, last but not least, grab that very unique opportunity that exists within the *Indulgents* group. The more exotic the images, plump imported fruits in a new and exciting flavored cereal, entice these consumers. They are not turned on by brands but by the excitement of something new and innovative. Make the package splashy and sassy. Entice this group with innovative designs and flavors of cereals that awaken the taste buds with a new cereal experience.

References

Burkitt, D. and Trowell, H. (1975) Refined Carbohydrate Foods and Disease: Some Implications of Dietary Fibre. London, Academic Press, London, UK.

Cereal grains (Source: www.Wikipedia.com).

Flight, I. and Clifton, P. (2006) Cereal grains and legumes in the prevention of coronary heart disease and stroke: a review of the literature. European Journal of Clinical Nutrition 60, 1145–1159.

Harris, D.R. (1996) The Origins and Spread of Agriculture and Pastoralism in Eurasia: Crops, Fields, Flocks and Herds. Routledge Taylor and Francis Group, London and New York.

Jacobs, D.R., Jr., Marquart, L., Slavin, J., and Kushi, L.H. (1998) Whole-grain intake and cancer: an expanded review and meta-analysis. Nutrition & Cancer 30, 85–96.

Kochar, J., Djoussé, L. and Gaziano, J.M. (2007) Breakfast cereals and risk of type 2 diabetes in the physicians' health study I. Obesity 15, 3039–3044.

Koh-Banerjee, P., Franz, M., Sampson, L., Liu, S., Jacobs, D.R., Spiegelman, D., Willett, W., and Rimm, E. (2004) Changes in whole-grain, bran, and cereal fiber consumption in relation to 8-y weight gain among men. American Journal of Clinical Nutrition 80, 1237–1245.

Liu, S. (2003) Whole-grain foods, dietary fiber, and type 2 diabetes: searching for a kernel of truth. American Journal of Clinical Nutrition 77, 527–529.

Slavin, J.L., Lavin, R.D., Jacobs, D., Marquart, L., and Wiemer, K. (2001) The role of whole grains in disease prevention. Journal of the American Dietetic Association, 101, 780–785.

SYSTAT. (2004) SYSTAT for Windows, version 11. SYSTAT Software Inc., Chicago, IL.

Technical note: All segmentation was done using SYSTAT.

Triticale (Source: AskDrSears.com).

W.K. Kellogg (Source: www.inventors.about.com).

Chapter 11

What Concepts Tell Us versus What Packages Tell Us for the Same Product—Case History: Pretzels

Introduction—Concepts Versus Visual Designs

Let's review for a moment the difference between product concepts and package designs. Product concepts tell the developer or consumer about the product. Product concepts come in two forms: concepts that tell the developer what is present in the product and concepts that tell the consumer why to buy it. We see these two examples in Figure 11.1, which shows a product concept (for the developer) and in Figure 11.2, which shows a positioning concept (for the consumer).

As we see in Figure 11.1, the product concept has information about what the product is in a material sense, whereas the positioning concept has information about the benefits of the concept, and may have information about the product itself as well. In both cases the concept itself, with words, carries the day. We are left to think about the product. There may be a picture in the concept, but the chief way of communicating is by text. The concept must, therefore, paint a "word picture" in the mind of the reader.

Let's contrast this way of communicating with what might be the case for two different pretzel packages actually on the shelf. You can see these two in Figure 11.3 and Figure 11.4.

Keeping in mind the stark differences between the product/positioning concepts and the actual packages, let's go more deeply into concepts for pretzels versus package designs. First, however, let us take a short excursion into the history of this popular snack food.

Pretzels—The Two-minute "History"

Pretzels were invented by monks and used to teach religion, to feed the poor, and to symbolize the marriage bond. Look at the pretzel's evolution.

> "This humble food comes in a variety of shapes, flavors, and with coatings that would have amazed the humble monk who invented the pretzel sometime between the fifth and seventh centuries. Idling with leftover strips of dough, the monk-baker supposedly twisted and turned them until they resembled a person's arms crossed in prayer, traditional posture for prayer in those days. The brother monks approved the tidbits, and began using them as rewards for the children under their tutelage.
>
> Despite their royal status, pretzels were a convenient way to hand food to the poor, and became a typical alms for the hungry. Apparently the homeless did not line up for soup or a sandwich, but for their daily pretzel. And those who gave the pretzels away were considered particularly blessed. Indeed, pretzels became such a sacred sign that they were often packed into coffins of the dead, no doubt replacing the jewels that were buried with the rich." (Pretzel history)

We see here quite an amazing evolution, from a background steeped in religion and goodness to an all-time favorite snack found on the shelves in supermarkets, school cafeterias, and vending machines. And so we chose this interesting popular snack food. Let's look at ideas about pretzels versus what you can actually show on a package. And, most important, what works!

Although that method of package creation is long gone, we are still faced with the same issue, namely what do designers say on the package and how do they design a product's package that will motivate consumers to buy? What do you see about pretzels, and what do you show on packages?

Before we design the package, however, let's think about how idea concepts for products differ from design concepts for packages. Concepts for products are

Packaging Research in Food Product Design and Development Howard R. Moskowitz, Michele Reisner, John Ben Lawlor and Rosires Deliza
 ISBN: 978-0-813-81222-9

Product concept without picture

Pretzel Sandwiches. Imagine real, creamy peanut butter sandwiched between two bite-size Pretzel Snaps. They're a delicious snack-on-the-go for kids and grown-ups.

Figure 11.1 Product concepts for pretzels

Positioning concept without picture

New Design···Same Great Pretzels!

We have always made our delicious Bachman Original Rolled Rod Pretzels the old-fashioned way—by actually rolling them! This special method, created and developed at our Heritage Bakery, helps ensure their exceptional flavor and extraordinary crispness and crunchiness. And because they are Brick Oven Flame-Baked,® they have the unique Bachman taste advantage. We know you will love them!

We start with only the finest, natural, wholesome ingredients. Then we knead the dough to create the very best "pretzel texture." We roll and cut the dough into rods, and top them with a sprinkle of salt! They are then Brick Oven Flame-Baked® in our vintage ovens for that extra special flavor. The result is the Crispiest, Crunchiest, Best Tasting, Real Rolled Rod Pretzels!

Figure 11.2 Product concepts for pretzels

somewhat richer. Concepts paint word pictures in the mind. People can fill in the missing pieces in a concept (Hippel, 1986; Gibbs, 2000). In contrast, package design is limited to specific images that can be placed on a package. Of course you might say that a person "fills in the blanks" with his mind, even when he sees a package. At the end of the day, however, the mental image we get from a concept is often richer than the mental image we get from a package, simply because there are so many more nuances in language.

To demonstrate this, we begin with the concepts that one could use for pretzels. We have selected the data from a large-scale project on the healthful aspects of food called the HealthyYou! (Beckley and Moskowitz, 2002). In this study respondents selected a food that interested them from a "wall of food names." They were led to the proper study and evaluated 60 different combinations of elements, with these elements selected from four silos. The method is precisely like the concept evaluation that we did for cereals (see Chapter 7), so we don't need to repeat it. All you need to know is that the dependent variable was the percent of respondents who rated each concept that they read as "interesting" (rating 7–9 on a 9-point scale). We created separate "models" for each of the 239 respondents, relating the presence/absence of the 36 different elements to this interest model. The respondents knew that they were evaluating concepts about pretzels in a study that dealt primarily with healthfulness. This knowledge means that the respondents would not be shocked by reading information about nutrients in the pretzels and other health information.

With this in mind we now look at Table 11.1. The table shows summary data for the 36 elements, from the total panel, from males/females, and from the two concept-response segments that emerged from analysis of the data. We sorted the data by the impact value for the total panel so that the most impactful or strongest element is at the top and the least impactful element falls to the bottom. We're not so much interested in the precise results here as with a broader overview, to sense the richness.

We have ranked the 36 elements by the utility value for the total panel. When you look at these 36 elements, try to imagine them first on a piece of paper as part of a concept and then afterward as part of a pretzel bag or box. When you do this mental exercise, you should come up with a very startling realization. The elements that do well in Table 11.1 are those that paint word pictures, of the types you might read about in a story. There are a few elements that would make good label copy, such as "Thick, crisp pretzels, toasted to perfection."

We see a couple of patterns emerging from these data, when we look at the total panel, at genders, and then at segments. Let's just go over these from the viewpoint of 20,000 feet in order to get a sense of what is going on.

1. The additive constants are high (51 for the total panel, 57 for males, etc.). From looking at the concept work for pretzels, we thus conclude that there is a basically high interest in the pretzel category. Even without elements, about half of the respondents are ready to score the concept as 7–9 on the 9-point scale.
2. There is certainly a large range of utility values. Some elements perform quite well among the total panel

Product with picture

Pita Pretzel™ Squares

Unlike any other pretzels, Pita Pretzel™ Squares are the newest addition to our Heritage pretzel line. Baked in our Classic Brick Ovens, they're great for snacking and dipping. No artificial colors, flavors or preservatives, and 0 g trans fat.

The Bachman Company is family-owned and operated with a long tradition of product excellence. Headquartered in Reading with manufacturing facilities in both Reading and Ephrata, we've been committed to quality pretzels and snacks for over 120 years.

11.3

Positioning with picture

The New Puzzle Pretzel!

Our new pretzel in the shape of the Autism Speaks puzzle piece logo arrived in mid-March 2008, and is currently available in grocery stores in the Northeast and online at www.bachmanpretzels.com. The launch coincides with Autism Awareness Month in April, and April 2, 2008, which will be the first U.N. sanctioned World Autism Awareness Day. Five percent of the proceeds from the sale of the Puzzle Pretzel will be donated to Autism Speaks.

The Puzzle Pretzel with its unique, fun shape continues the Bachman tradition of quality pretzels and snacks. Puzzle Pretzels are low in fat, cholesterol free and a good source of calcium. Baked for a crispy, crunchy bite, they're perfect for snacking and dipping.

The Puzzle Pretzel shape represents the logo of Autism Speaks, the largest non-profit organization dedicated to increasing awareness of autism and raising money for research. In early 2007, Bachman formed a partnership with Autism Speaks, and launched an awareness campaign with the placement of autism information stickers on the bags of their most popular pretzels and snacks. The new Puzzle Pretzel was created to help reach millions of parents to continue to promote awareness of autism spectrum disorders (ASD).

11.4

Figures 11.3 and 11.4 Two current (2008) commercial packages for in-market pretzels

Table 11.1 Utility values for the 36 "pretzel elements" from the Healthy You! Study on pretzels. The elements are ranked by total panel. Data courtesy of It! Ventures, Ltd.

		Total	Male	Female	Taste	Health
	Base Size	239	45	194	143	96
	Constant	51	57	49	44	60
A4	Tender, soft pretzels, rolled by hand and baked to a golden brown	12	8	13	19	2
A2	Classic, baked pretzels with a lightly browned taste and just the right amount of salt	9	8	10	15	0
A5	Thick, crisp pretzels, toasted to perfection	6	10	5	10	0
A7	Lowfat, only 1 g of fat per serving	6	–1	8	8	2
A8	All natural, no MSG, artificial flavors, colors, or sweeteners	6	6	6	7	5
D1	From Rold™ Gold	5	6	5	7	2
D3	From Snyder's of Hanover	5	10	4	5	4
B2	Provides essential minerals your body needs, including potassium, magnesium, and zinc	4	–4	5	–2	12
B9	Contains 8 essential vitamins and minerals	3	–1	4	–2	10
C1	A quick and easy snack	3	4	3	4	0
C8	As part of a low-fat, low-cholesterol diet, may reduce the risk of some forms of cancer	3	2	3	2	4
C9	May reduce your risk of high blood pressure and stroke	3	1	4	3	4
D5	Endorsed by the American Heart Association	3	7	2	3	2
A1	Healthy eating that tastes great	2	5	2	4	1
A6	Made with the finest ingredients	2	2	2	5	–3
B1	Provides essential vitamins your body needs, including A, B12, C, and E	2	–6	4	–5	13
B4	A good source of fiber, important in reducing your risk of chronic diseases like heart disease and diabetes	2	–4	3	–2	8
C2	A food you feel good about feeding your family	2	2	2	2	2
C3	Fills that empty spot in you, just when you want it	2	5	2	4	0
C6	Even better for you than you thought	2	3	1	2	1
D6	Endorsed by the American Diabetes Association	2	4	2	3	2
C7	May help you reduce your risk of Type II diabetes	1	–1	1	–1	3
D7	Endorsed by the American Dietetic Association	1	7	–1	0	1
A3	Indulgent, savory pretzels with flavors like buttermilk ranch, honey mustard, and cheddar cheese	0	–9	3	12	–17
D4	From Auntie Annie's	0	–5	1	1	–2
D9	Recommended by nutritionists and dieticians	–1	–2	–1	–4	3
B3	Full of antioxidants and phytonutrients that help you maintain a healthy heart	–2	–7	–1	–8	7
C4	Such pleasure knowing you're eating something healthy	–2	–1	–2	–3	0
D8	Recommended by your doctor	–2	–3	–2	–2	–2
B6	Contains the essential nutrient choline shown to improve memory and learning	–3	–8	–2	–10	7
C5	Calms you down, just what you need when you're feeling stressed	–3	–8	–2	–4	–2
D2	From Bachman	–4	–4	–4	–4	–4
A9	100% organic	–5	–7	–5	–1	–12
B5	With inulin, known to improve calcium absorption and improve digestion	–5	–10	–3	–11	4
B8	Contains soy protein, clinically proven to reduce the risk of heart disease	–5	–11	–4	–10	2
B7	With soy isoflavones, shown to moderate symptoms of menopause and decrease bone loss	–6	–19	–3	–12	3

(e.g., Tender, soft pretzels, rolled by hand and baked to a golden brown). Other elements perform poorly, especially the very strong "health/good for you/ingredient elements" (e.g., With soy isoflavones, shown to moderate symptoms of menopause and decrease bone loss). These results don't surprise. They simply reiterate the fact that a good, delicious description of a product can go a long way toward making it desirable. And they reinforce the fact that perhaps a lot of the health messaging doesn't necessarily convince the customer to buy the product, no matter how "in" or "trendy" the ingredient may be.

3. Gender makes a difference. Thus, we see a five-point spread between women and men for the winning element (Tender, soft pretzels, rolled by hand and baked to a golden brown: women utility = +13, men utility = +8). When it comes to females, ingredients linked to female conditions score higher among women than men, but for many of the health issues, the impacts are higher for men than for women.
4. There are clear segments, again along the lines of taste versus health. All in all, it's the word pictures that talk to sensory satisfaction that do best in total, and exceptionally well among Segment 1, the "taste" segment.
5. Health messages strongly polarize consumers. Some of the simpler health messages do very well among Segment 2, the health oriented (e.g., Provides essential minerals your body needs, including potassium, magnesium, and zinc: utility = +12 for Segment 2).
6. Brand names do not perform particularly well in these types of studies. This will be important for our packaging study, where brand name is a major visual element.
7. A lot of the short messages, of the type that might be flagged (e.g., low fat, only 1 g of fat per servings), do modestly well at best. Some of these short messages do poorly.

Figure 11.5 Example of a pretzel bag, showing the different elements, and the limited "real estate" on which to put brands and messaging

From the Pretzel Concept to the Pretzel Package

Let's now move to the next stage. We've just discovered that the concepts for pretzels do well and that a number of the elements perform superbly, especially those with strong taste promise. When it comes to packaging, however, we don't have that much "real estate" to play with. We don't have endless space on the package. Rather, we are dealing with a package that looks something like what we see in Figure 11.5.

Working in the "Visual World" of Limited "Real Estate"

Let's now move to the complementary world, this one dealing with the package design of pretzels. We're going to move out of language and description where mental pictures are created, into the world of immediate impressions. We are moving from a world of descriptions that might make a writer proud to the world of artistic flair, where less is often more, where impact is critical, and where there is not enough room to present one's message in the most eloquent form. And yet, we move to a world where everyone fights, with limited attention. Truly we move from the world of Shakespeare and masterworks, to the world of Malthus and Darwin.

How Elements Get Created—Comparing Concepts and Package Designs

One of the important topics for this chapter is how we came up with these specific elements for pretzels, first for the concept study, and later on for the package study. Let's answer the first part of the question here—for the concept study. When we think about the kinds of elements that we might use, most of us would automatically revert back to "ad speak," to the language of advertisements. That's the initial type of element that comes out from creative sessions, and no wonder. We're surrounded by advertisements.

When it comes to thinking about a product, it's natural for us to recall what we have heard from others and then replay that (Batra, Lehmann, and Dipinder, 1993; Golden and Johnson, 1982). Afterward, when the ad-speak elements were finished, we came up with ideas from describing the product in our minds.

Finally, because the project focused on health, we looked on the Internet to discover language that would be relevant. We finished the exercise with specific "factoids" that we knew to be relevant from our professional expertise. This sequence of ad speak/mental rehearsal of experience/Internet search/supplemental expertise is the common approach to concept research where the "real estate" is free and where there is no discipline necessary that includes limits on space.

Let's move forward now to creating the elements for the pretzel package. It should be obvious from Figure 11.5 that the commercial pretzel package does not have a lot of space on it for different messaging. There's the logo, there's some messaging, a few pictures, nutrition information, and of course, behind it all there is the package shape and the color. Certainly there is no room for the fanciful language that creates word pictures. Space on the package doesn't allow it, and furthermore, no one will read the information even when the information is actually present and clearly spelled out.

So the question remains. We have the opportunity to do a designed study with the package. We have room to test a lot of ideas (here 36 ideas). How do we come up with the ideas that we wanted to test? One thing was certain. In the quest for the better package design, we did not want to "reinvent the wheel." Our web-based tool allowed us to investigate different combinations. Our job would be to come up with these elements. It's not really all that difficult once you get going.

Let's look at our train of thought, and unpack the set of ideas with which to develop the package graphics. Most important, it helps to know what we were thinking as a group when we began the exercise. We were interested in relevant stimuli that could be put on an actual bag. The stimuli had to be simple, easy to absorb, short, and impactful. None of the long, fanciful text and word pictures of concepts. The operative words were "strong" and "insistent," followed by "relevant."

Our ideas came from Health and Nutrition, Flavor, Types of Pretzels, Various Logos, Colors and Miscellaneous Images, which would pull the bag together. Let's take a look at our "raw material"—what we came up with.

Silo A: Logos

We know every package has to be identified by a brand, but exactly what is the impact of the brand? Where does its domain fall in order of overall importance in the design? We fabricated the names to avoid the bias that might come from well-known brands. Our four "invented" names were Golden Sun, Gregory's, Wholesome Foods, and Orchard Farms.

Silo B: Health and Nutrition Claims

We hear so much today about consumers wanting their food choices today to be nutritious (Costa and Jongen, 2006). Do these claims drive a person to buy? Do consumers really care about a product that is baked, all natural, has 50% less fat than potato chips, or has 0 grams of trans fat? We know that at the time of this writing, in late 2008, trans fat has been banned in many foods, and in many states. However, when this information is presented, but not trumpeted, does it make any real difference?

Silo C: Style

We all know that pretzels come in various types so we chose the classic and a few different versions such as Homemade, Original Style, Classic Style, and Braided.

Silo D: Images

When you look at bags of pretzels, you will typically see pictures. We wanted to simulate reality. There are thousands of images to choose from that might be relevant. To give some life and reality to the mid-section of the pretzel bag, we chose the following four pictures to "romance" the bag: (1) Sun Burst Over the Mountains, (2) Tractor, (3) Old Fashioned Plane, and a (4) Golden Field of Wheat.

Silo E: Flavor

Here we offer the consumer choices that range from the traditional flavor of Homestyle to niche flavors providing opportunity for new snacking experiences such as Homestyle, Honey Mustard, Parmesan Garlic, and Honey Wheat.

Silo F: Color

Here we provide the product developer with the option of different vibrant colors for the bottom portion of the bag.

For this study we wanted a simple, uncomplicated evaluation of the different bags. In the very short, 10–15 minute interview, each respondent rated 40 different "combinations" (i.e., bags of pretzels), created on the computer, using this straightforward scale: "How much do you like the Overall Packaging of the Pretzels?" 0 = Not at all, 9 = Very much.

We get a sense of these different elements by looking at the telescoped version of these 24 elements in Figure 11.6.

Test Design—Putting It All Together

Now that we have the elements, how do we put them together? When we dealt with the word-based concepts for pretzels, matters were simpler. We had a template with rows. We simply populated the rows with the different elements. The computer program had the template prestructured. The experimental design dictated which element would go into the first line of text, which into the second line, etc. The respondent then looked at the sequence of lines, not knowing that the combination had been "arranged" by a computer program, or if the respondent did guess what was happening. It probably made no difference.

Design is a bit different. It has to "look right." Let's return to our notion of a basic "template" to be filled out by the design, stimulus after stimulus to create the different concepts. But look at the template as a set of transparencies or overlays. The basic idea is that the package design concepts are a set of layers (silos), each layer having several alternative options (elements).

Figure 11.7 shows the approach schematically. The template can be thought of as a structure on which transparencies are superimposed. When we think of graphics design in this systematic fashion, we are immediately struck by the possibility of using experimental design to mix and match features, test full combinations, and then deduce what every graphics element contributes to the respondent's rating (Moskowitz and Gofman, 2007).

Let's look at this approach in action, again with pretzels. Figure 11.6, laying out the elements, shows a schematized drawing of the different silos and elements within the silos. The bag itself without any elements presents the template. As we look around the template, we see the different silos and the various elements in each silo. These elements are typically developed as single options by a graphics artist. There is no reason to

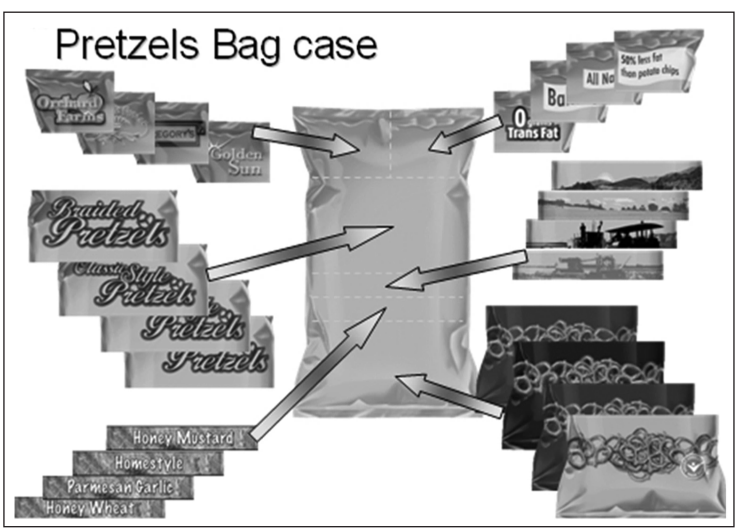


Figure 11.6 The elements for the pretzel bags study

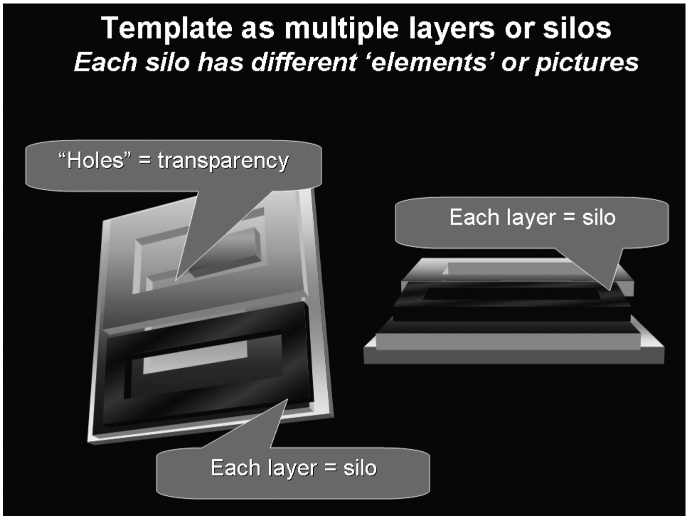


Figure 11.7 Graphics design schematized as a template, with silos and elements

assume that the elements need to be "high art." They simply need to have some verisimilitude. Look closely and you can see that the texts can emphasize health, brand, etc. Like concept work, graphics design does not limit the investigator; this is in contrast to limitations on the physical characteristics of the pretzels themselves in product tests.

What We Learn from the Package Study

By now it should become clear that the majority of the learning comes from looking at the data in two ways. The first is total panel. We can learn a lot from the 200 respondents who participated. We discover what wins, what loses, in what marketers call the "general population." However, we get a lot of information from dividing the respondents into different populations of consumers—our recurrent theme of "mind-set" segments. These segments are defined by the pattern of their responses to the different elements, whether these elements be text or pictures (Hammond, Ehrenberg, and Goodhardt, 1996; Moskowitz, Krieger, and Rabino, 2002; Neal, 2003; Wright, 2003).

As designers, marketers, researchers, or just plain scientists or business people, we always look for stories in the data. Nature tries to talk to us in these designed experiments. We may not decipher the message, but we can at least make a very good guess by looking at which silo of elements do well and which do poorly, or perhaps which are even irrelevant. The whole story will come out of the results in Table 11.2.

Just to reiterate the approach, all of the analysis is done at the individual respondent level, an appropriate analysis given the fact that the ingoing experimental design was done so that the ratings from each respondent sufficed to create a model. The analysis begins by converting the rating to a binary (0 = reject the concept, corresponds to the ratings 1–6; 100 = accept the concept, corresponds to the ratings 7–9). We run ordinary least-squares regression on the individual respondent's data to generate an additive constant and an individual utility (coefficient), that is, one utility value for each element. Only when we segment the respondents do we work with the original 9-point ratings and only for the specific clustering analysis. The reason for working with the rating data rather than the transformed (0, 100) data is straightforward. We wanted to work with more fine-grained, detailed data when dividing people by the pattern of their utilities. The rating data is better for that.

Now let's plunge into the results, which we see in Table 11.2. To make things easier, we have sorted the elements by the impact or utility value within each silo. This type of bookkeeping, while seemingly minor, actually helps us to understand what's going on.

1. The additive constant. Recall that this tells us the basic interest in the package. The additive constants are quite low. We have seen this pattern before—the

Table 11.2 Impact of the 24 elements for the pretzel packages

	Total	Segment 1 Health Seeker	Segment 2 Reassurance and "splashy design"
Base Size	200	61	139
Additive constant	–20	29	–31
Silo A: Logo			
Gregory's	6	6	6
Orchard Farms	6	4	7
Wholesome Foods	5	4	5
Golden Sun	4	4	4
Silo B: Nutrition/Health Claim			
50% less fat than potato chips	8	11	7
0 grams trans fat	8	11	7
Baked!	6	9	5
All natural	6	6	6
Silo C: Type			
Original style pretzels	15	6	17
Homemade pretzels	14	5	16
Braided pretzels	14	4	17
Classic style pretzels	13	6	15
Silo D: Middle Image			
Field	11	–2	14
Sun Burst	10	–1	13
Tractor	10	–3	12
Plane	10	0	12
Silo E: Flavor			
Honey Wheat	8	4	8
Homestyle	6	0	7
Honey Mustard	5	3	5
Parmesan Garlic	5	3	5
Silo F: Basic package color			
Blue	19	–13	27
Dark brown	15	–14	21
Light brown	14	–13	20
Gold	11	–6	15

additive constants tend to be much lower when we deal with packages and pictures, and higher when we deal with concepts. The reason is fairly simple. With concepts, there is a basic interest in the idea that the constant measures. With packages, we are much more responsive to the physical stimulus on the screen. That is, "what we see is what we get," rather than "what we see is part of the story about the product, with the rest in our mind."

2. Looking at logos, which you recall we made up, we see that there are actually differences in the logo. Gregory's is a +6 and Golden Sun is a +4. This difference is minor. The four names in the silo do only modestly well. But, at least they don't detract from interest in the pretzel! One wonders whether we could find names that could, and if we could do so, then what might be the rules or patterns that we could learn? Is there a type of name that is actually nega-

tive? What are its properties? Interesting questions, but we cannot answer them here.

3. We had expected nutrition and health claims to be important. They are only modest players (utilities 6–8). They are positive, but not the "grand slam" we expected them to be, although to be fair we know that these simple claims don't appeal to everyone. When we come to the segments we will see more of an effect, especially when we specify the nature of the nutritional claim.
4. Pretzel type exerts the greatest effect, along with color of the package. Utility scores ranging from 10–15 mean this will bring an additional 10–15% of respondents from voting "not interested" to voting "interested."
5. Looking at type (a text message on the package) and color (a purely visual stimulus that is probably not even verbalized), we are struck by the fact that the response to the package need not be intellectual. We suspected all along that there would be different dynamics at play than in concepts. In concept, everything is intellectual, everything is filtered through the language used. In package, however, sheer color alone without ancillary messaging, suffices to create interest in the package. This is an important result. Furthermore, we can trust the results. All of the elements of the package design varied in a way that eluded any detection of underlying pattern. Yet, basic package color was such a strong driver. The fact that there is an 8-point difference between the best performing color (blue: +19) and the worst performing color (gold: +11) suggests that color is both important and differentiated. Some colors simply perform better than other colors.
6. We have seen and will see again that the segmentation by mind-set really tells a better "story." Let's see whether that segmentation works for pretzel packages. (Hint—It does, just as it did for pretzel concepts.)

Different Mind-Sets Expand the Opportunity

When we look at the mind-sets, we see something different. There are two segments. They differ from each other.

Segment 1—Healthy

These consumers, approximately 30% of the sample, look for a healthy pretzel. They most strongly respond to labels that talk about "baked" and "50% less fat than potato chips." What is important, however, is that even with such health-oriented labeling, the Segment 1—Healthy group generates utility values of 11, and not utility values of 15–20. These healthy people are not our "died-in-the-wool health seekers" that would generate these "over-the-top" values of 15–20. Rather, they simply prefer the more healthful product. And, one other thing, they want the specific claim, not the health that is communicated by general words such as "baked" and "all natural." They are not interested in colors.

Segment 2—Reassurance and Splashy Designs

Our larger group, approximately 70% of the sample, responds strongly to type of pretzel and to the middle image or picture on the middle of the package. Since the experimental design comprised some combinations in which either type or middle image was missing, the utilities or impacts are meaningful in absolute terms. It doesn't matter which type—the segment responds to all of them. Nor does it matter which image is chosen. Segment 2 most strongly responds to color, and with color it matters which colors are presented. All colors perform well, but blue does exceptionally well (impact = 27). Even gold does well (impact = 15). By designing the bag with vibrant colors and images, along with emphasis that the product inside is of any of the specific pretzel types, Segment 2 will be satisfied.

Now What?

What have we learned and how can what we learned help a designer create a package for pretzels?

1. The key findings of this research lie in the utilities (i.e., the power of messages to motivate interest or detract from interest). Both positive-performing and negative-performing elements need to be known when developing visuals for packaging.
2. Overall, three silos were highly motivating: Type, Middle Image, and Color.
3. The heart of the learning occurs in the mind-set segmentation. Here the product developer has opportunity for diversity in both product offerings and visual communications.
4. We learn different things from working with the design features than we learned working with ele-

ments. We would not understand how to design the packages if we were to remain in the world of "word pictures." We must work with the visual elements themselves.

5. We can make two specific proposals for design, based on what we learned.

Entice the *Healthy* segment by images that convey that they are eating a snack that is good for them (that is, a package that when you look at it reminds you that the snack inside is the same as what you have been eating but now healthier). This group is not motivated by splashy images or colorful packaging.

And, last but not least, delight the *Reassurance* segment with your trusted type of pretzel and then hit them with a splashy and sassy package. Entice this group with innovative designs. This is a very strong segment of people that can't be overlooked and don't need much to keep them as customers.

References

Batra, R., Lehmann, D.R., and Dipinder, S. (1993) The Brand Personality Component of Brand Goodwill: Some Antecedents and Consequences. In: Aaker, D.A. and Biel, A. (eds.) Brand Equity and Advertising. Lawrence Erlbaum Associates, Hillsdale, NJ.

Beckley, J. and Moskowitz, H.R. (2002) Databasing the consumer mind: The Crave It!™, Drink IT!™, Buy It!™ & Healthy You!™ Databases. Institute of Food Technologists, Anaheim, CA.

Costa, A.I.A. and Jongen, W.M.F. (2006) New insights into consumer-led food product development. Trends in Food Science & Technology 17, 457–465.

Gibbs, A. (2000) Developing Profitable Concepts—Our Invention Marketing Expert shows you how to make your bright idea a reality. Accessed from http://www.entrepreneur.com/Your_Business/YB_SegArticle/0,4621,278253,00.html

Golden, L.L. and Johnson, K.A. (1982) The impact of sensory preference and thinking versus feeling appeals on advertising effectiveness. In: Bagozzi, R. and Tybout, A. (eds.) Advances in Consumer Research 10. Proceedings of the Association for Consumer Research, pp. 203–208.

Hammond, K.A., Ehrenberg, A., and Goodhardt, G.J. (1996) Market segmentation for competitive brands. In: European Journal of Marketing 30(12):39–49.

Hippel, E.V. (1986) Lead users: source of novel product concepts. In: Management Science 32, 791–805.

Moskowitz, H. and Gofman, A. (2007) Selling Blue Elephants: How to Make Great Products That People Want Before They Even Know They Want Them. Wharton School Publishing, Upper Saddle River, NJ.

Moskowitz, H.R., Krieger, B., and Rabino, S. (2002) Element category importance in conjoint analysis: Evidence for segment differences. In: Journal of Targeting, Measurement and Analysis for Marketing 10, 366–384.

Neal, W.D. (2003) Principles of Market Segmentation. Accessed from http://www.marketingpower.com/live/content.php?Item_ID=1006

Pretzel history. Accessed from http://www.differentpretzel.com/pretzelhistory.htm

Wright, M. (2003) RESPONSE—Competences/Capabilities for Market Segmentation. Accessed from http://www.marketingpower.com/live/content.php?Item_ID=17124

Chapter 12

"Closing in on the Container"

Introduction

Most of what we cover in this book deals with the various aspects of design. But what about the actual package itself, the three-dimensional container, stripped of pretty designs, and simply a functional product? Can we apply the methods of systematics to understand what features people want? And, even more importantly, can we do this in an efficient way so that early in the design phase the consumers can "guide" the packaging engineer?

In this chapter we will do just that. We will take the principles of experimental design that we used for graphics and bring it to the representation of packages, of containers to hold food (Green and Srinivasan, 1987; Moskowitz, Gofman, Katz, Manchaiah, and Ma, 2004). The design elements that we deal with comprise features that are functional, but also some that used for aesthetic reasons.

Today's computer technology allows the designer to present many different features of packages to consumers. At the early stages, the packaging work can be done in the virtual world of the two-dimensional computer screen. After this initial work is done, the data from the study should point toward the specific product features that the three-dimensional container should comprise. At that point, it becomes important to create the physical package in three dimensions, using the proper materials fabricated to represent the final packaging.

A Case History—"Next Generation" Clear Plastic Containers for Food

Let's now move out from the world of pure package design into the more functional world of plastic containers for food. When we talk about design in this world, we are talking about the way the container works, as well as the degree to which the container pleases. In other chapters we have talked about design as a tool by which to help sell the product. Here we talk about design as the input to making the package actually "perform its job."

Our case history deals with the systematic analysis of features for the next generation food container, to be used at home. Most people are familiar with clear plastic containers into which one can put leftovers. These have different physical features. Systematic exploration can help us identify what "works" and what doesn't.

As we progress through this chapter we should keep in mind that the criteria for "winning ideas" are not just "liking." Rather, the respondent must evaluate the designs on a number of different attributes. Beyond interest, the main evaluative attribute is the degree to which the product would perform its function. The product may be attractive, but if it doesn't appear to hold promise that it will be a good storage container, then chances are that it won't be purchased.

Setting Up the Stimuli

We move now to the test stimuli themselves. It is difficult to create actual physical prototypes for this type of study, or at least to do so in an economical way. Instead, we looked at the different silos and elements of the new container and put together test concepts that look like Figure 12.1 and Figure 12.2. In a few minutes we will move into the actual elements and the nature of the test design. Right now it's more instructive to look at a combination that we made for this study.

The actual study comprised 29 features, arranged into 70 different combinations, each combination on its own "concept board." The combinations were created ahead of time. There were only a limited number of these combinations.

There is a reason underlying this pre-creation that wasn't operating before. In most concept studies and in many package studies, all of the features may, in theory,

 ISBN: 978-0-813-81222-9

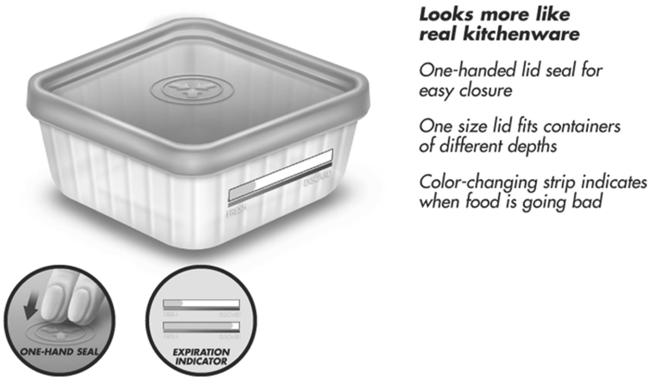


Figure 12.1 An example of the container, deconstructed into its components.

Figure 12.2 Another example container, deconstructed into its components

"go with each other." That is, the combination may not have been very acceptable, but the elements that were combined could actually coexist in the same test stimulus. Here we deal with containers. The actual structure of the container means that we have large numbers of pairs of elements that cannot appear together in the same package design. In light of these restrictions (i.e., elements X and Y can never appear together), it was simply easier from the standpoint of executing the study to create the allowable combinations ahead of time. Precreating combinations allowed us to make sure that the combinations were both realizable and that the elements in the combinations acted as independent agents (i.e., were statistically independent of each other).

That being said, let's look at the range of elements that the container study actually explored. We see this list in Table 12.1. What should strike us here is the nature of the elements. These are options that comprise all types

Table 12.1 The 29 features that were combined into 70 package concepts, along with their utility values for purchase intent, uniqueness, and fulfillment of a need. The dependent variable is the percent top-3 box on the 9-point scale. The additive constant or baseline corresponds to the expected sum of impacts (utilities) for the three shaded elements.

	Buy	Unique	Fulfill
Additive constant (corresponding to a package that has a "see-through appearance," is "square-shaped," has a "standard seal")	41	28	37
Silo A: Visual appearance (must appear)			
Color, Clear	1	1	2
Color, See-Through	0	0	0
Embossed	–1	–1	–2
Color, Bright	–8	–4	–8
Silo B: Visual shape (must appear)			
Round	1	1	1
Square	0	0	0
Silo C: Visual closure (must appear by "fiat")			
One-Hand Seal	3	3	2
Visual Indicator of Seal	2	8	2
Hinges	2	8	1
Latches	1	7	–1
Easy-Grip Rim	0	3	1
Tabs	0	1	1
Standard Seal	0	0	0
Silo D: Visual Modularity/Lid (can be absent)			
Reversible Lid	6	14	5
Compartments	6	9	5
Secure Stacking	6	4	4
One-size lid fits containers of different depths	4	5	3
Save space by stacking efficiently in the refrigerator or freezer	4	2	3
Mix and Match	3	9	4
Lidded serving plates and bowls for saving and serving any meal	3	4	3
Silo E: Visual Refrigerator/Freezer/Microwave (can be absent)			
Expiration Indicator	12	29	12
Vented Lid	8	16	7
Day of Week Indicator	6	19	6
Vacuum Sealed	5	21	1
Freshness Dial	4	21	2
Insulating Barrier	3	7	2
Silo F: Cleaning Benefits			
Odor-resistant barrier keeps your containers smelling new	3	5	3
Stain-resistant barrier keeps your containers looking new	0	–1	0
Dries completely in the dishwasher	0	–1	–2

of ideas, not just simple graphic ones. In this and other three-dimensional package work, it is important both to show the respondent what the feature looks like, as well as tell the respondent what the feature does. The description can be short. Its job is to explain what is being seen, what is the raison d'être of the particular features.

Doing the Study

This study is similar to the other projects we have discussed in this book. The respondents are recruited to participate. In this particular study, the respondents showed up in a central location in a shopping mall and took the study in a room to the side of the main mall. These rooms are often rented to market research companies, who run "taste tests" and other studies in them. The rooms are equipped with computers.

The sessions run in a straightforward manner. The respondent is recruited, usually from the list of individuals. The list is developed by a local market research agency, called a "field agency," which contracts to provide respondents of a specific type. Here the respondents had to be users of plastic containers for food. The field agency finds out additional information about the respondent in the course of preparing for the project. Such information might include brands purchased and the like—a level of specificity not typically necessary for general recruiting but of vital interest to the package designer working on this specific container (Lappin, Figoni, and Sloan, 1994).

The actual study took about 90 minutes to run. There are a few reasons for this extended time, which is far longer than the typical study that is reported here on the web, which takes only about 15–20 minutes. These reasons are instructive, helping us understand how to run such studies in the future, and also explaining the particular why's:

1. Many test stimuli: There were 70 regular concepts about the container, created by experimental design, as well as an additional 30 package concepts that are not considered here in the analysis. With 30 additional package concepts, the packaging engineer could "have her cake and eat it." That is, with these additional concepts beyond the experimental design, the designer could venture into new areas, without necessarily creating an extensive test array for each new container concept. However, the downside to this stimulus array is either to work with many respondents, each of whom tests a few combinations, or to work with fewer respondents, each testing many combinations. We chose the latter strategy, requiring so-called pre-recruited respondents who participate for an extended period of time (hours). They are paid to participate.
2. Many attribute ratings. We report on three ratings here, specifically purchase intent ("buy"), uniqueness, and satisfies (fulfills) a need. There were two other ratings, a frequency rating, and a fit to a specific, relatively new brand name. This means we have 100 concepts, profiled on five attributes, for a basic set of 500 ratings. That number of ratings is quite a lot and requires that the respondent be motivated to participate and complete the study. That motivation was accomplished by recruiting the respondent to participate in a supervised, central location, for which the respondent was compensated (Schiffman and Kanuk, 1999).

Creating the Model—Revisiting What to Do When the Product Must have a Specific Feature, or the Silo Must Appear

In much of the design research we discuss in this book, silos may be absent from a concept without doing much damage. Of course, we want to make the stimulus realistic, but by having one silo occasionally absent from the design we can better estimate the impact value of all of the elements in that silo. In fact, the impact values have absolute meaning, when we have "true zeros" or true absence of a silo from a concept. A "true zero" (i.e., absence of a silo from a concept) allows us to estimate the true impact of the element versus the element being absent completely from the concept. The benefit is that we can compare the impact value of a feature in one study to the other impact features in other studies. A strong property, indeed!

Sometimes, however, we simply cannot do without a silo in a stimulus, because then we have no real stimulus. The combination is impossible to realize, physically. This unhappy necessity is always the case when we deal with physical stimuli that must have a shape. Just look at Figure 12.1, trying to imagine the stimulus without the container having a shape, or having a shape but no visual appearance. It is logically impossible. You cannot have physical stimulus with no shape. Certainly you can have a verbal description of the stimulus without mentioning shape, but it is logically impossible to show that stimulus unless it looks like something.

This issue of "necessary silos" is not a theoretical one. It generates specific steps in both the experimental design, which lays out the specific combinations, and in analysis, which identifies the contribution of each element to the rating (Gustafsson, Herrmann, and Huber, 2000). Our goal is to create a model that relates the presence/absence of the different elements to the dependent variable (e.g., rating the concept 7–9 coded as 100 or percent of respondents rating a specific concept as 7–9).

The presence of a silo, one of whose elements must be present in every concept, is addressed in a rather straightforward, if not entirely satisfactory way. We did this before in the deconstruction of the margarine containers, so the process is not altogether new. Here are the steps:

1. Place each element into its proper silo.
2. Identify the silo or silos where there are "true zeros" (i.e., the silo can be legitimately absent from the concept).
3. Identify the silo or silos that MUST be in the test concept or visual design. Call these silos the "forced silos."
4. Code all of the elements as 1 or 0 for the different stimuli, with 0 denoting absent and 1 denoting present.
5. For all of the silos that have a true "zero" condition, let all of the elements be predictors in the regression model.
6. For all of the silos that are "forced silos," choose one element from each silo deliberately to leave out of the equation. It doesn't matter which element is going to be left out, as long as one and only one element is selected to be absent from the model.
7. In our project here on the container, we have three silos that must be forced in. We can choose whichever element from the silo we wish to "leave out." That set of "left-out" elements will show up in the additive constant, as we see below.
8. The three forced-in silos are visual appearance (leave out color—see through), visual shape (leave out square), visual closure (leave out standard seal).
9. Now run the regression model on the data. You will get the results in the conventional format, comprising an additive constant, and impact or utility values, corresponding to the elements that acted as predictors.
10. The additive constant in conventional research throughout this book corresponds to the estimated utility or impact, if no elements are present. Ordinarily this "no elements present" would be the case for these container data as well. Three of the silos can be absent (Silo D: visual modularity; Silo E: appropriateness and action in refrigerators, freezers, and microwaves; Silo F: cleaning). Three of the silos cannot or should not be absent (Silo A: visual appearance; Silo B: visual shape; Silo C: visual closure).
11. The additive constant corresponds to the zero cases for all six silos. For three of the silos (D, E, F), there are true zeros. For the remaining three silos, the zero cases are specific conditions "left out of the model." Keep in mind that it doesn't really matter which particular element in each silo is left out.
12. Once we decide about true zeros and forced elements, we can estimate the impact or utility values for all elements. The impact values have true, absolute meaning for those silos that allow "true zero" (i.e., those three silos that can be legitimately absent from the package concept without affecting the reality of the package).
13. In contrast, the utility values are relative to the particular element that is selected as the reference for those silos that we force to be in the stimulus. Change the element, and the impact values change. However, the differences among elements never change; only the impact values do.

What Do the Data Look Like for the Modeling and How Are They Used?

We can get a sense of the data for modeling by looking at Table 12.1 for the full set of 29 elements. The table gives you a sense of how to arrange the data.

Now, let us go further. Let us create some new combinations, and estimate how they perform. We arranged Table 12.2 to show three different combinations from the 70. These are the combinations that are most interesting versus least interesting, and most fulfills the respondent needs versus least fulfills the need. We see the composition of the test concept on the left, and the percent top-3 box on the right for purchase interest, fulfillment, and uniqueness, respectively.

Table 12.2 Elements and three test concepts for the plastic containers, showing what is in the concept, and the percent top-3 box for purchase interest, fulfillment, and uniqueness

	Appearance Visual	Appearance Text	Shape Visual	Closure Visual	Closure Text	Modularity/Lid Visual	Modularity/Lid Text	Refrigerator/Freezer/ Microwave Visual	Refrigerator/Freezer/ Microwave Text	Dishwasher Saver/ Reusable Text	Interest	Fulfillment	Unique
Most interesting	Color—See Through	See what you have inside without	Round	Latch	Latches ensure your food stays locked in	Compartments	Dividers keep food from	Expiration Indicator	Color-changing strip indicates	Stain-resistant barrier keeps	64	53	72
Least interesting	Color—Bright	Available in Contempo	Round	Hinge	A container with a hinge	—	—	—	—	Odor-resistant barrier	36	32	41
Most fulfills	Color—Clear	Clear like your real glass containers	Square	Easy grip rim	Special rubber rim ensures a tight lid fit	Compartments	Dividers keep food from	Expiration Indicator	Color-changing strip indicates	—	59	55	70

What Works in the Container … and Where?

Our container project provides a rich source of information to the package engineer about what works and how it is perceived. We thought about the various features in a storage container and came up with what we considered to be the key categories to explore:

- Visual appearance
- Visual color
- Visual closure
- Visual modularity/lid
- Visual refrigerator/freezer/microwave
- Cleaning benefits

Let's go step by step into these results to see what we can extract from the modeling. Refer back to Table 12.1 for the actual numbers.

Let's begin by looking at the additive constant to assess consumers' in-going interest in this new storage container. Remember, we begin with a basic, see-through, square-shaped container with a standard seal. The reason for this basic product is simple—the physical product must have an appearance and a shape. Furthermore, by fiat we decided that it should have a seal.

Not surprisingly, we see strong interest in this basic container. Remember, our target audiences were consumers who used storage containers on a regular basis.

As with all research, we see that all of our thoughtful groundwork when setting up the test design enriches our outcome. Our stimuli enable us to "learn" from the consumers themselves how to make an optimal food storage container that will fulfill not only their needs but is unique to the marketplace and, in the end, will sell more. Sounds great. Let's go.

Each silo writes its own story. We have some silos that comprise both motivating elements along with some that actually demotivate, some with no impact so they don't offer much "jazz." Some elements dance right off the page. Since our goal is to design an optimal storage container that delivers all three above-mentioned perceptions (i.e., consumer attributes), we'll look at how each element performs on the three attributes as we build our new container.

Silo A: Visual Appearance

We began with a simple see-through container. Remember that this is a reference level. We had to have some visual appearance. We wanted to identify the best visual. It didn't matter which of the elements we chose to be the reference—everything would be relative, and the "differences between the impact or utility values" would not be affected, whether we were to choose clear see-through, embossed, or bright colors, respectively. So, when we look at the data we want to choose the most positive scoring element. When we talk about "embossing" or talk about "bright colors," we see negative impact numbers.

Looking at the data, a designer knows immediately that the container should be clear. It's important to know what not to put in, as well as know what to put into the container. You don't need to go for bells and whistles here. Whether it is clear or see-through doesn't really matter. The choice is up to the designer. What does matter is that the contents within should be seen.

Silo B: Visual Shape

In this category we kept it simple by offering one of two standard container shapes, square and round. Here, utility scores reveal that either shape is acceptable. Furthermore, you don't score many points here by choosing one option over the other.

Silo C: Visual Closure

In this silo we get to see whether or not consumers respond strongly to new and exciting ways to close our container. Is there just one way to close a container, or are there some "unique" opportunities here? Keep in mind that we "forced-in" a closure. We didn't have to; the container might have "made sense" without a closure, but the closure itself is important. So, having forced-in a standard seal as the reference, let's see what else "pops." In this silo of visual closure we begin to see some interesting results. Consumer respondents begin to wake up, even though the underlying design was not apparent. Consumers responded to the totality of the container, but the design lets us parcel out the contributions of different closures, relative to the standard seal. From there, we branched out. How about a container with hinges, or a visual indicator of a seal, or latches, an easy-grip rim, a one-hand seal, or one with tabs? We see positive utility scores for all closure elements (relative to standard seal), across our three rating questions with only one minor exception—"latches" with an impact of −1. Our strongest container enhancements were the mentions of a "visual indicator of seal" or a container that seals with "hinges." Not as strong, but still relatively impactful, is the mention of a "one-handed seal." In this silo of "closure," therefore, we begin to see the basis of some innovation for the new container.

Silo D: Visual Modularity (can be absent from our concept)

What additional features would we like from a storage container? We varied our features to cover a wide range of tasks and asked consumers to rate how they felt about "reversible lids," "compartments," "secure stacking," a "one-size lid fits containers of different depths' or ones that "save space by stacking efficiently in the refrigerator or freezer" to name a few. Keep in mind that we tested visual designs without the silo of visual modularity. Thus, the impact or utility values have a true, meaningful zero value. Visual modularity was, overall, a motivating category with strong impact. "Reversible lids" and containers with "compartments" to separate foods along with those that offer "secure stacking" are a few areas of opportunity for further exploration.

Silo E: Visual Refrigerator/Freezer/Microwave (Can Be Absent)

Again, this silo can be absent from the stimulus. Therefore, the impact or utility values are absolute. We don't need a reference here. This silo generates the strongest impact values. We see high, positive utility scores among nearly all of the elements, enriched and wonderfully unique ideas that add to purchase intent and are fulfilling as well. The reason is simple. In this silo we look at different ways to ensure freshness. After all, isn't that a key feature in a storage container? An "expiration indicator" or a "vented lid" or a "day of the week indicator" will be important to our optimal concept.

Silo F: Cleaning Benefits

Here we have a silo that contains three simple elements dealing with cleaning. Important? Yes. However, the elements perform only moderately. One element has a positive impact value. The element "odor-resistant barrier that keeps your containers smelling new" adds value. In contrast, mentioning "stain resistant" and "dries completely in dishwasher" are not effective.

Table 12.3 Optimal combination of elements for a package

	Purchase intent	Uniqueness	Fulfills needs
Additive constant (corresponding to a package that has a “see-through appearance,” is “square-shaped,” has a “standard seal”)	41	28	37
Clear like glass (optional substitution)	1	1	2
Round (optional substitution)	1	1	1
One-Hand Seal (probably a good substitution)	3	3	2
Reversible Lid/Domed lid can be flipped upside-down for smaller-sized containers (definitely yes)	6	14	5
Expiration Indicator (definitely yes)	12	29	12
Odor-resistant barrier keeps your containers smelling new (optional inclusion)	2	5	3
Predicted sum of components plus constant	67	80	62

Creating an Optimal Container

How do you create an optimal container from these data? Perhaps the simplest way is to identify the elements that win. Start with the additive constant (which corresponds to a specific package that is “see through appearance,” “square shape,” “standard seal.” Then look at each of the silos, and identify elements that perform well. If the element has an appreciably high utility (>5), then by all means include this “winning element.” If the element does not have an appreciably high utility (0–5), then you may or may not want to include it. It’s an option. The best thing of all is that the basic product is already “taken care of” by the additive constant. The utility or impact values are substitutes for the basic elements included in the additive constant.

We see the optimal concept for total sample purchase intent in Table 12.3, along with the estimated uniqueness and “fulfills needs.”

Can You Believe the Results? How to Demonstrate Statistical Validity of the Model

Let’s take a short trip into the world of validity and reliability (Moskowitz, Beckley, Mascuch, Adams, Sendros, and Keeling, 2002). Much of this book looks at equations or models that relate design variables to subjective responses. We use regression analysis to relate the two worlds, design features under the engineer’s control and perceptions under the respondent’s control.

The immediate question that comes to mind is “does the regression model really represent the ratings in an accurate way?” That is, should we believe the equation?

A standard way to establish the validity of an equation uses a statistic, called the multiple R^2, which shows the percent of variation in the ratings (dependent variable) accounted for by the equation. A lot has been written on statistical analysis by regression (Batchelor, 2001; Hosmer and Lemeshow, 1989; Louviere, 1988). We don’t need to review it here. All we need to do is use the approach to assess whether or not we can believe our model or equation.

Look at Figure 12.3. The figure presents a scattergram. The y-axis locates the actual percents for purchase interest (top-3 box). The x-axis locates the estimated percents of purchase intent (top-3 box), using the equation that we developed and whose parameters appear in Table 12.2. Each darkened circle is one of the 70 concepts. We know from the experimental design precisely what elements are present in each concept. We know from the model what value each element brings, including the additive constant. So we can predict the top-3 box rating (the dependent variable).

All of the foregoing is a way to show that the modeling is adequate and that it represents the dependent variable, top-3 box, quite adequately. Of course there could be other models that we might use, but the model we chose does a good job. The model is simple; it is the weighted contribution of the different elements so that we can predict the top-3 box from knowing the components of the stimulus concept. In our model for purchase intent, the R^2 is a very high 0.88, meaning almost 90% of the variation in the top-3 box purchase intent can be accounted for by the model. In other words, we can believe the model. See Figure 12.3.

Summing Up

What have we learned here? And can we apply it when designing a new and hopefully improved food storage container? As with nearly every area of consumer goods, there are always ways to improve upon what currently is in the market. Did we uncover any? In a word, think “freshness.” We saw this theme carried throughout our summary of results. Remember, we began with a basic, see-through, square-shaped container with a standard seal.

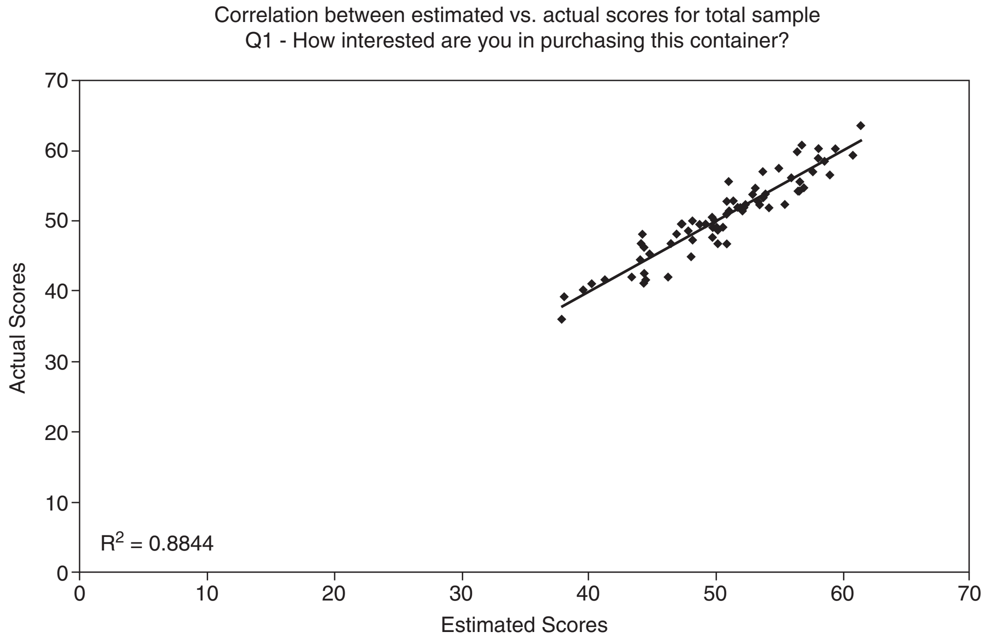


Figure 12.3 The figure presents a scattergram. The y-axis locates the actual percents for purchase interest (top-3 box). The x-axis locates the estimated percents of purchase intent (top-3 box), using the equation that we developed, and whose parameters appear in Table 12.2.

Typically when we create a new appearance for a product, don't we think we can motivate the consumer with splashy new colors or perhaps funky shapes? Absolutely. But not in this category. What we have here is a functional product, one that doesn't need to scream at you from the shelves—Pick me, pick me!—but will do just fine with the standard shape and see-through color that allows the consumer to store and see. Or as the truism goes, "If it ain't broke, there's no need to fix it."

We definitely ought to include the freshness feature of an "expiration indicator." Here we saw our highest utility values across the board.

As we have read in previous chapters, building an optimal concept is done by selecting elements or features that will give us a high-predicted sum.

In ending this chapter, keep in mind the valuable realization that some things may be standard and need not be improved upon; they rest on their own merits. However, as researchers, we know that with the ever-changing tides of today and the necessity to stay ahead of the competition, there are always demands for improvement. But do the research first, however. You don't want to change something that doesn't need changing.

References

Batchelor, R. (2001) Logit regression and neural networks. City University Business School, London, UK.

Green, P.E. and Srinivasan, V.S. (1987) A general approach to product design optimization via conjoint analysis. Journal Of Marketing 45, 17–37.

Gustafsson, A., Herrmann, A., and Huber, F. (2000) Conjoint Measurement—Methods and Applications. Springer Verlag, Berlin, Germany.

Hosmer, D.W. and Lemeshow, S. (1989) Applied Logistic Regression. John Wiley & Sons, New York, NY.

Lappin, J., Figoni, P., and Sloan, S. (1994) A Primer on Consumer Marketing Research: Procedures, Methods, and Tools. National Transportation Systems Center, Cambridge, MA.

Louviere, J. (1988) Analyzing decision making: Metric conjoint analysis. Sage Publications, London, UK.

Moskowitz, H.R., Beckley, J., Mascuch, T., Adams, J., Sendros, A., and Keeling, C. (2002) Establishing data validity in conjoint: Experiences with Internet-based 'mega-studies' In: Journal of Online Research. Accessed from http://www.ijor.org/ijor_archives/articles/stablishing_data_validity_in_conjoint.pdf—size 165810 bytes—9/19/2002 6:31:25 PM GMT, 11 pp.

Moskowitz, H.R., Gofman, A., Katz, R., Manchaiah, M., and Ma, Z. 2004. Dynamic Package Design & Optimization in the Internet Era. In: Proceedings of ESOMAR Conference: Technovate2, Barcelona, Spain.

Schiffman, L. and Kanuk L. (1999) Consumer Behavior, 7th Edition. Prentice Hall, Upper Saddle River, NJ.

Chapter 13

Action and Reality: Using Video for the Package Experience

Introduction

We experience packaging in a world of three dimensions, in a world of action, and in a world of consequences. Scientists and market researchers are interested in the well-tempered laws of reality. Often this focus on solid results leads investigators to make the packaging experience static, whether the experience is text concepts, graphics designs, or actual packages themselves.

We don't necessarily think about the package in action, or if we do, then more often than not we leave some of that thinking to the designer. After all, the reasoning goes, scientists try to measure what is, try to figure out the "rules" of reality. It is up to the designer to integrate these rules into the package or its graphics.

In recent years, researchers have begun to realize that it is important to capture the package experience. In one of the earliest examples of the application of video, Mike Gadd in Toronto developed an Apple Quick-Time simulation of the shopping experience, for which he won a well-deserved award at an ESOMAR Conference on innovation. Gadd's work presaged some of the later work on the shopping experience. Through simulation, he was able to capture some of the actual "footage" of a shopping experience, and bring that footage into an experiment where respondents could "go shopping" in what would turn out to be a virtual environment.

Academics and then practitioners soon recognized the value of video, or perhaps it would be more correct to say that at the same time that Gadd was doing his work in Toronto, others in the academic world were looking at the potential of video to provide deeper knowledge about the shopping experience. Starting in the mid 1990s, for example, Professor Raymond Burke and his colleagues at Harvard and then at Indiana University started to study how consumers responded to the shelf. Rather than providing the consumer respondents with simple "shelf shots," which are static, Burke presented the respondents with shelves stocked with products, representing what would be seen at a store. By using the computer, respondents could then pull out a product, "turn it around," zoom in, inspect it, and then select it if desired (Burke, 1991). Burke's pioneering work was incorporated by Steven Needel into a commercial venture called Simulation Research, Inc., using Burke's program, Visionary Shopper (Needel, 1998; Treiber and Needel, 2000). By this writing, the use of simulations in the study of shopping is well established. The important lesson here is that it is not the static stimulus, but the experience of the stimulus, that can teach a lot (Deliza and MacFie, 1996).

With this introduction, let us move out of the static world of packaging into the more dynamic world. In this chapter we focus on incorporating video into the experimental design. With video the respondent can get a more immediate sense of the action.

Both of our case histories go back to the end of the 1990s. First, a few reminiscences from Alex Gofman, who developed the technology with his colleagues:

> *In the development history of the class of technologies we now know as Ideamap®, it is sometimes difficult to pinpoint the cause-effect relation between experimental psychology, client needs and enabling technology (Moskowitz and Gofman, 2007). Normally, one would expect it to develop from business needs to science to technology (or science-business need-technology sequence). Whereas this is absolutely true for the general underlying architecture of Ideamap®, some specific implementations followed a more peculiar route.*
>
> *The adage "If you only have a hammer, then every problem starts to look like a nail" is exactly what happened to Ideamap®. The paradigm itself and general architecture of Ideamap® were so robust and universal, that many problems unnoticed or unsolvable before got their second chance. This was the case for testing packages with video.*

Packaging Research in Food Product Design and Development Howard R. Moskowitz, Michele Reisner, John Ben Lawlor and Rosires Deliza
 ISBN: 978-0-813-81222-9

I remember my trip to a Microsoft conference, where the new video technology for Windows was being shown. A natural reaction to seeing the first implementations of the emerging multimedia capabilities of Windows was enabling short video clips (with sound) to be used as stimuli instead of a static picture. A huge step ahead but we did not stop there. A question naturally emerged. I'm not exactly sure how it came out, but during the discussion the notion came up … "Could we put two videos, side by side?" This was a technical challenge then, but the discussion triggered a new and unheard of before idea—why not sequence the clips, the same way it's done in commercials, but this time do it according to an experimental design?

The rest is the typical story. Thousands of lines of C-language code developed in just the first eight days, followed by the inaugural test in a facility in southern New Jersey for which we had to rent minivans and load them to the top with several desktops to run the project. As discussed here, the project was FunFeast. We didn't know it at the time, but this work would open up the door fifteen years later to using video for many more things. But then it was just that first FunFeast project, followed a few months later by the towelette project, and then a few more.

Swanson's "FunFeast"

In the early 1990s, Swanson Foods, manufacturer of mid-priced frozen foods, began a program of aggressive development. One of the key issues at the time was to identify specifically what type of frozen food "dinner" would appeal to moms and kids. Swanson's research director at the time, Alexis Fried, recognized that the design of the specific frozen meals might be improved if the marketers and product developers at Swanson could get an idea of what the product should contain, and how some of the packaging might work.

Packaging was an extremely important issue for this particular project and brand, because the frozen meal was, in essence, a combination of differently packaged foods, with the opportunity to use some new packaging technologies, incorporate some different types of products within it (i.e., Kool-Aid packet), and, finally, to use graphics design (characters on the product, which were taken from familiar as well as unfamiliar sources).

With this introduction to the project, let's see how the packaging research dealt with the problem of what should be in the meal, and how to represent different alternatives in a way that children and mothers would find interesting. The objective was not so much to entertain the respondents as to create a method by which to test a package-related "experience" among both children and adults.

As we begin, let's go back a moment to some memories of that study, provided especially for this book, by Alexis Fried.

When we initially designed and launched FunFeast, we conducted focus groups (Casey and Kreuger, 1994), with children and tested many different concepts with moms, all of which led to a wonderful product with many unique features, including the only lift-out frozen ice cream dessert to come with a microwaveable dinner, a cast of unique and fun cartoon characters, and collectable fun prizes inside the box. The product launched in the early, 90's to great success, during a period when the kids frozen dinner segment was under-performing due to economic factors and a battle within the adult frozen dinner category resulted in those products often selling for less than the kids' meals.

As a result, we needed to validate the importance of the products' most expensive components, and understand which elements were required for market success, and enable a better product margin. To accomplish this, we needed to isolate the value of each individual package element. The challenge, of course, was you couldn't just ask a 6-year-old child if a lift-out dessert made them more or less likely to ask their mom to buy the product! Instead, we hoped showing them the product elements in a way that simulated watching a TV commercial would enable us to identify the true value of each component.

When the results came back, we found that the methodology really worked. We had identified clear preferences among both kids and moms, and identified those elements that increased product value for moms as well. The results answered questions regarding which food components were most important (the main!; sides could be standardized around most popular/low cost options), if we should invest in character licenses (not required), and if we could eliminate the prizes (not if you wanted to appeal to younger children!).

As an added bonus, not only did the method work, but it was a lot of fun to create the video elements or "snippets" used as stimulus for fielding the study.

Working with Kids and Adults

We began this work on FunFeast with a bit of trepidation because we were exploring new lands, *terra incognita*, unknown territory. For many years, there had been a great reluctance to work with children as if they were young adults who understood scaling and evaluation in the way that adults do. We were faced with the problem of what to present to children, and how to get the children to react to the test stimuli. There are many in the research community who believe that children cannot really understand the meaning of numbers, and, therefore, it's best to use some type of pictorial scale, such as a so-called Smiley scale. That scale has faces corresponding to different levels of liking, ranging from a frown to a wide smile (Kroll, 1990; Chen, Resurreccion and Paguio, 1996).

We chose not to follow that point of view, and simply to let children rate the test concepts as if they were young adults. We reasoned that these children are exposed to numbers in school, and they seem to do perfectly well. Since we were about to work with children ages 6–12, we simply required that the children have finished First Grade. This stipulation was not hard to fill since we were working with 95 children, approximately equally distributed by ages (95 children all together, 45 in the 6–9 age bracket, 50 in the 10–12 age bracket).

What We Wanted to Do

Our project was to provide the child and the adult, separately, with a set of test stimuli on the screen. The stimuli would show different FunFeast packages. The package was a full dinner comprising six silos. The silos were main course, side component, dessert, a surprise pack, a cartoon character, and a price. Each silo comprised several options, from the smallest silo of side components (only two elements, French fries/corn; mashed potatoes/corn), to the largest silo of six different cartoon characters. Video was critical to show that the dessert component could be removed.

Ordinarily, we might present these components as words or as still pictures. This time, however, we changed the approach somewhat. We created small 2–4 second "clips"—one clip of each of the product elements (mains, sides, desserts) in each silo. Look at Figures 13.1 and 13.2 to get a sense of the clips for the brownie, showing how the product was actually "lifted out of the package for eating." These two clips, part of the 3-second video clips, made the experience come alive for the respondents.

13.1

13.2

Figure 13.1 and Figure 13.2 Two portions of the 3-second video clip for the brownie dessert.

The FunFeast Stimuli

Creating the test stimuli was straightforward. If you look at the different silos and elements in Table 13.1 then you will see that the sizes of these silos are different. There are a total of 25 different elements. Since experimental designs are analyzed by regression, we need more combinations than we have predictor elements. For the

FunFeast project, we created 40 different combinations. Elements in the same silo appeared equally often. However, when a silo had relatively few elements (i.e., side components), its elements appeared far more often than when a silo had more elements.

Creating these combinations is not particularly difficult. One needs only to set up the combinations so that the individual elements from different silos are statistically independent of each other. It will be impossible to make all of the elements in the same silo completely independent of each other because there are constraints (some items cannot appear with others). However, for the most part, the elements are reasonably independent of each other in a statistical sense.

To make the study interesting for the children, we presented each FunFeast as a sequential video, comprising components that were spliced together. Some parts of the video were simply "still shots" of the meal components in a package, along with the name of the item. We did this for the main and side dishes, as well as for the price. The remaining parts of the stimuli were video clips, selected in order to make the package experience come alive to the respondents, especially the younger ones.

The actual stimuli were 40 different test videos, constructed combinations, with identifying labels when necessary. The labels identified the different silos and elements, along with the appropriate video. By presenting the stimuli as videos with clearly marked labels when necessary, the research engaged the children, especially those ages 6–9, who might not have been able to comprehend the test stimuli were they to be strictly text descriptions. The video helped to bring to life the unique removable nature of the frozen dessert, which could NOT have been done as easily with a photo and text.

What We Learned about "Process" from the FunFeast Study

The bottom line in all of this work is what we found and, of course, what we learned about the process. Could the children, ages 6–9, participate in this type of project? The answer is yes. Would they discriminate? The answer is yes. Could their data be used in a meaningful, quantitative way? The answer is yes. Of course, we did not do the control experiment with these stimuli as words alone, so we can't be sure that it was the video clips that helped the project to generate the insights it did. We are fairly certain that the video helped a great deal to generate strong interest, as we will see.

Even before we walk away with the "findings," the real question is whether or not the project actually worked! All through this book, we have been dealing with the responses of adults to these stimuli, rather than the responses of children. We know from everyday experience that kids are attracted to packages (Robinson, Borzekowski, Matheson, and Kraemer, 2007). What mother or father, dutifully shopping in the supermarket, trying to buy the products written down on a carefully prepared shopping list, has not during the course of reading, been importuned by their child to "pick up that package, ***please***, Mommy (or Daddy)!" Those may not be the exact words, but they get the message across. Kids are attracted by packages, by stuff that is interesting, and when they get a bit older, by packages that are "cool," or whatever word is, "in."

One interesting point worth remarking on before we actually go into the data is that the 95 kids who participated in this study had a wonderful time. The senior author was at the session with the kids. No kids complained. In fact, the kids were fascinated by the task and paid attention. Of course, part of the reason was that an attending (female) interviewer, running the session with something like 12 kids, treated the situation like school. Kids liked the discipline, felt that they were doing something important, and after a few moments of the typical raucous behavior that kids show, they all settled down to do their task, just as they would do in school. In fact, in two of the sessions, the children were so occupied with the task that you could "hear a pin drop."

What We Learned about What Kids Want

Let's look at the results in more depth. We see the results laid out for adults, all kids, and the two ages of kids, in Table 13.1. We also see the results for the rating of "value for the money" at the right side of Table 13.1. As is our custom, we recoded the 1–9 points into a binary, with ratings of 1–6 recoded to 0 (denoting not interested, or not a good value), and ratings of 7–9 recoded to 100 (denoted interested, or a good value, respectively).

We begin by looking at what wins, but just as importantly, how strongly the item wins. Our first look focused

Table 13.1 Performance of the different video elements in the FunFeast project

	Kids (6–9)	Kids (10–12)	Adults	Adults
	Buy	Buy	Buy	Good Value
Varieties				
B1 Chicken Drumlets	10	14	34	13
B4 Fried Chicken	13	16	32	17
B3 Pizza	13	16	16	4
B2 Fish Sticks (Reference)	0	0	0	0
Side Components				
C2 French Fries/Corn	1	2	8	6
C1 Mashed Potatoes/Corn (Reference)	0	0	0	0
Dessert				
D3 New brownie no lift-out	12	6	8	13
D1 Cold lift-out	6	8	6	14
D2 New brownie lift-out	13	9	3	14
D4 No dessert (Reference—True Zero)	0	0	0	0
Surprise Pack				
El Current surprise pack	5	–3	9	3
E5 No surprise pack (True Zero)	0	0	0	0
E3 Flavored Drink Mix (unbranded)	9	1	–1	2
E4 Collect and save for free prizes	7	–1	–1	1
E2 Kool-Aid packet	9	4	–4	6
Characters				
F2 Garfield	11	3	10	5
F1 FunFeast Shark	4	–1	10	8
F5 Campbell Kid	6	11	3	8
F3 Animaniacs	7	–1	3	7
F4 Tintin	8	–3	0	5
F6 No character on box (True Zero)	0	0	0	0
Price				
G1 \$1.89 (Reference)			0	0
G2 \$1.99			1	–9
G3 \$2.19			–17	–41

on the four varieties. Recall that there were four of these main varieties and that each of the 40 test videos had to begin with one of these varieties. The consequence of that particular decision is that the four numbers are relative to each other. We can choose any one of the four varieties to act as a reference, which reference we give the value 0. It seemed logical to give the lowest scoring variety the position as reference. This is B2, fish sticks. Compared to that, the winning elements are very strong performers (fried chicken, chicken drumlets). Even among the children, these are the strong performers. However, kids also like pizza a lot more than adults do. Finally, if we look at "good value," we see that adults don't feel that pizza is good value. We don't really know whether the adults "down rated" pizza because they didn't like it, they didn't want to buy it for their children, or they thought it was poor value. All in all, it is gratifying to see, however, that everyone discriminated among the different varieties, especially the younger children, ages 6–9.

When we move to the side dish, we again see discrimination by adults, but no response by children. So now we see that the side dish is not particularly relevant for the children, although it is for adults.

Let's move on to dessert. It was in the dessert portion of the video where we tried to emphasize the packaging of the product, especially the brownie (see Figures 13.1 and 13.2). For desserts we created videos that had no desserts, so when it comes to reading the impact, we have a true impact value. Kids like the video of the brownie lifted out from its package in the FunFeast; adults are not as excited, however. The excitement about the "action" of "lifting out" portrayed by the video is greater among the younger children than among the older children.

So what did we learn? First, this portion of the FunFeast project teaches us that it's possible to engage younger children, show them an aspect of the packaging, and excite them. We also learned that the younger children preferred what they knew, not what the brand manager thought would be best. The younger children preferred the warm chocolate brownie, a more traditional Swanson dessert item, over the novelty frozen dessert the product line had originally been launched with!

Children like toys and extras. That's the reason behind the fourth silo. The difference between children and adults becomes clear here. Kids want a Kool-Aid packet; adults don't. Yet, when it comes to ratings of good value, adults believe that the Kool-Aid packet is a good value, even though they don't like it. We also learn here that adults clearly distinguish between what THEY like and what is good value.

We get a good sense of what kids want, how ages differ, and how they differ from adults when we look at characters in the fifth silo. Again there were some video clips without characters, so the numbers have absolute value. Adults differ from children, but the children differ among themselves by age, as well. No surprise here, at least not to a parent. Younger children are a lot more excited by the characters than are older children. Surprisingly, adults feel that the character adds value.

We finish our analysis of the FunFeast project by looking at the economics (i.e., the pricing). Only the moms saw the pricing. Economics plays a strong role among adults, with interest and value going in the expected direction, only at different rates. As price increases from $1.89 to $2.29 (in 1993 U.S. dollars), interest drops down, but good value drops down far more precipitously. We have seen the sensitivity of interest to pricing before. Now we see that the value for money relation is even more steeply related to price.

Summing Up—What Did We Learn from FunFeast to Help Us Work with Packages and Kids?

The original objective of the FunFeast project was to identify the "hot buttons" for moms and kids. That is, what particular elements of the packaged dinner appealed to both groups? The particular findings are interesting, but limited. What we discovered is that the children react, occasionally quite strongly, to packaging of actual products with which they are familiar (i.e., frozen dinners). The real discovery is that video may be a way to deal with children in packaging studies, not to mention an additional option for studies among adults.

On to Towelettes, Experimental Design of Ideas, Applications, and Conferences

Our second case history came about a year later, in the latter part of 1994. One of our clients at the time, a packaging specialist at a paper goods company, wanted to commission work on the development of some new ideas in towelettes and wipes. The issue was to find out what type of product packaging looks best.

The guiding aspect when studying towelettes is the person-package interaction. It was quickly realized by everyone that describing the action of using the towelette might produce a reaction, but would it be the reaction from a picture or would it be better from an actual experience? Another issue, less important but still operative, was that it was difficult to get some of these new packages in sufficient quantities for people to experience using them. A final issue, always important in companies, was confidentiality. It was one thing to talk about a new package, another to show a video of a new package, but it was quite out of the question to let the package outside of one's direct control. Readers who work in companies recognize these different levels of concern, sometimes more intensely in certain projects, sometimes just as latent "hygiene" practices that simply are ever-present, albeit at a low intensity.

Some of the results of this study appeared previously in 1996, in an ASTM technical publication "Creative applications: Sensory techniques used in conducting packaging research" (see Moskowitz, Krieger, and Tamber, 1996). The middle and late 1990s, in fact, saw the start of a number of conferences focusing on package design, perhaps because the Microsoft Windows platform as well as the Apple Mac allowed for better graphics. With those graphics came a generation of researchers ready to exploit the expanding technology, this time to understand visual perception in the same way that they understood the perception of ideas in a text format.

With the development of video capabilities in the computer program coupled with high-resolution monitors, we found it quite possible to introduce video and pictures as "visual elements" into concept studies. It didn't matter to the respondents who were looking at the screen whether they were looking at text alone, at text plus picture, or text plus video. One of the silos was a "visual silo," present on the screen for about 5 seconds if a video, or present for the entire concept evaluation if a simple picture. The respondent could replay the video by a simple keystroke.

Table 13.2 The best and worst performing elements for each silo on "interest." The data come from the total panel.

Silo	Element	Interest
Opening/ Closing	The snaps are easy to close and open, even when the baby doesn't hold still	4
	Opening these wipes is the easy part	–5
Threading	You're in control of how big or small a wipe you get	3
	Now you decide what size wipe you need	–7
Baby's Comfort	Added aloe makes them so soothing	7
	Moisturizes baby's skin	–3
Sensory Attributes	Soft and moist	2
	Smells like a baby SHOULD smell	–4
Uses	Keep some in the car for the inevitable…	4
	One of Mommy's essentials	–4
Disposability	Biodegradable because we care about your child's future too	1
	Gentle for baby AND Mother Nature	–4
Packaging	Easy to remove with one hand	4
	Better packaging for better moisture retention and product sterility	–2
Video clips	Opening yellow box, and showing towelettes	6
	Opening green lid, taking towelette out	–8

What the Towelette Study Taught Us

We begin with a look at the different silos, and what scored best and worst. Look at Table 13.2. There were eight silos in the study. One silo was video/visual stimuli, with a total of 32 different elements. Of these visuals, 11 were video clips to show the various aspects of person-package interaction. The remaining 21 elements in the silo were visual stills, showing the package, but not showing any person-package interactions.

The eight silos had different numbers of elements, so we used a version of experimental design in which each respondent evaluated 50 different combinations of the elements. At the end of the evaluation, we had enough data to create a model showing how every one of the 117 elements performed.

Packaging studies using experimental design can generate a lot of different issues to be answered (Gacula Jr., 1993). Since the respondents were looking at the different videos, we wanted to answer several questions. At the end of each "concept" on the computer, the respondent rated the concept on four questions, one after another, and each on a 9-point scale. The questions were:

1. How interested are you in buying this premoistened towelette?
2. How different is this premoistened towelette from others in the market?
3. How easy is this product to use?
4. How effective is this premoistened towelette in doing the job?

When we look at Table 13.2, we see that the average utility values for the best and worst performing elements lie within a reasonably narrow band, about +8 to –8. The visual elements show the widest range, but the truth of the matter is that the range, a low of –6 to a high of +8, is not particularly large. So, our first piece of learning is that although video clearly shows the person-package interaction, in actuality it doesn't make that much of a difference to ratings of interest in the towelette product.

What We Learn from Instructing Respondents to Evaluate Concepts on Different Scales

By instructing respondents to rate a concept on four different rating scales, we get a sense of the different perceptual dimensions that packaging, as well as other elements, can generate. We can get a sense of the performance of some of these elements from Table 13.2. Look closely at the interest impacts, which in this table show some that range from a high of +6 to a low of −8. This is a good range, not particularly high, but sufficient to demonstrate that the video clips generated different reactions, not all positive.

Beyond the simple question of "how do visual/video elements" perform is the question of the impression that these same elements make on uniqueness, ease of use, and effectiveness. Does it look like the highest "interest" elements are also the easiest to use? Are these high scores also the most efficient at doing the job? The answer to this question is important for two different reasons:

1. *Substantive.* Of course, it is important for the package designer and engineer to know how the different packages perform. Performance here is not just acceptance, but also how the product does when the respondent has to judge other, specific things, like apparent ease of use. Do the different videos, portraying different packages and package-related activities generate different impacts? If they do, then we conclude that the respondent rates the attributes differently. The results don't tell us whether one attribute is more "valid" than another. That is, perceived ability to do the job, the fourth attribute, may not be rated validly because the respondent cannot easily judge that attribute from the video. This is empirical. Yet, perceived ease of use may be simpler to assess by video and thus, the ratings are more "valid," or at least the respondent has sufficient information to make a judgment.
2. *Methodological.* Consumer researchers are accustomed to asking lots of questions about a few stimuli. In the most laughable of cases, this proclivity to lots of questions generates a so-called laundry list of attributes, because later researchers are reluctant to discard scales used by previous researchers, and, in fact, retain the old attributes but add their own, new attributes, to the mix. Respondents have a limited amount of patience, a limited attention span. When confronted with a laundry list of rating attributes, most respondents will not get up and walk out of the interview. A few do, but that's rare. Instead, the respondent will "tune out," stop paying attention, and start rating the stimuli in a more random fashion. Perhaps the respondent will not tune out completely, but the odds are that the attention that a respondent would pay when asked one question is quite diluted with four questions. Bearing this in mind, do we see evidence that the respondent simply repeated the same rating for all attributes? The answer is no, at least based on the different impact values. They are different from attribute to attribute. We are not sure that the respondent was paying attention to any of the attributes, but we can feel assured that the respondent wasn't "straight-lining" the answers, copying the answer given to the first question, and using that answer for the remaining questions or attributes.

How do we proceed to make a story out of the results? We find the story embedded in the data by looking at what video-clip elements "float to the top" and what elements sink down. Let's look at a few highlights:

1. *Wiping hands reduces interest*. Looking at the table, we see that the elements showing the towelette actually being used are negative. This is important. Showing the product removed from the package and used for wiping hands is less effective than simply opening the package and showing the towelette.

   Removing towelette from yellow box and shutting lid 5
   Removing towelette from yellow box, wiping hands, and shutting lid −2
   Removing towelette from yellow box and wiping hands −3

2. *Uniqueness does not show a pattern*. There are two elements that show a positive uniqueness impact. They don't seem to have anything in common with each other. Both show the yellow box, but it is not the yellow box alone since, in another case, the yellow box is not associated with uniqueness. The ability to look at a number of different elements to understand what "drives" a response comes in valuable here, because we have the ability to look at several "executions" of the same basic idea.
3. None of the video clips portrays a product that is easy to use (see Table 13.3). This is one of the most important findings. When we presented person-package interactions, we did not present an execution that made the product seem "easy to use," even though some of the text elements suggested "easy to use." Is it that the reality of a video clip is greater than the reality of a text element so that the video-clip leaves less to the imagination? This is an area worth exploring in the future.
4. Effectiveness in doing the job shows very strong impact values. The video clips drive home the message of effectiveness, albeit to varying degrees. However, there is no clear story. One possible key to effectiveness is to show the product being opened from the correct package. The key here may be the yellow box (a strong performer for effectiveness), coupled with a strong protective element, such as the aluminum seal.

 Tearing aluminum seal and showing
 towelettes (yellow box) 18

Need States—How Having Children "Drives" What's Important to a Respondent

Let us end this case history on towelettes by comparing two groups of respondents, those who have young children (ages 4–7) and those who do not have young children. We see that these respondents show dramatically different impact values for interest. We looked at the 117 elements and selected those elements with an impact value of +10 or higher. When looking at the total panel and at those respondents who say that they have no young children, we fail to find any strong performing elements, whether text elements or visual elements. On the other hand, those respondents who have young children are quite attuned to the elements. Look at Table 13.4. We see 12 elements that perform quite strongly. (An impact of +10 means that when the element is present in the concept about towelettes an additional 10% of the respondents say that they would be interested in the product (i.e., would switch their rating from not interested to interested). Three of these elements are exceptionally strong performing video elements.

Table 13.3 Performance of the 11 video-clip elements. Each element generates an impact value on the four attribute ratings.

	Interest	Unique	Easy to use	Effective at doing the job
Opening yellow box and showing towelettes	6	2	–2	5
Removing towelette from yellow box and shutting lid	5	4	–3	6
Showing plain, plastic yellow box	4	–3	0	8
Remove cardboard wrap and cellophane from yellow plastic box	1	0	–7	6
Tearing aluminum seal and showing towelettes (yellow box)	1	2	–3	18
Opening and removing top of "snap" style, yellow container with aluminum seal	0	–4	–6	3
Remove cellophane and open yellow plastic box	–1	–2	–1	1
Removing towelette from yellow box, wiping hands and shutting lid	–2	–3	–7	2
Removing towelette from yellow box and wiping hands	–3	6	–3	4
Tearing aluminum seal from yellow "snap" style container	–6	–1	–3	–4
Opening green lid, taking towelette out	–8	–1	–3	1

Table 13.4 Having children changes one's reaction to messaging about and person-package interactions of towelettes.

	Total	No Kids	Have Kids
Winning elements – Respondents with children			
Video: Opening yellow box and showing towelettes	6	1	20
Just pop open the top and pull out as many as you need	–2	–8	16
Gentle enough for frequent changes	2	–1	16
Video: Opening and removing top of "snap" style, yellow container with aluminum seal	0	–4	15
Video: Remove cellophane and open yellow plastic box	–1	–6	13
You're in control of how big or small a wipe you get	3	1	13
Opens easily with one hand	–1	–4	12
Added aloe makes them so soothing	7	4	12
They're not for bottoms only!	1	–4	11
You shouldn't have to wrestle with a package of wipes AND your bundle of joy!	2	0	11
Showing plain, plastic yellow box	4	0	10
Not for babies only	–1	–4	10
Winning elements—Respondents without children			
No element had a contribution of 9 or more	…	…	–

Video—Opening yellow box and showing towelettes
Video—Opening and removing top of "snap" style, yellow container with aluminum seal
Video—Remove cellophane and open yellow plastic box

Summing Up—What Did We Learn From Towelettes to Help Us Understand Person-Package Interaction?

This early foray into the study of towelettes shows us that it is quite instructive to insert into concepts various videos of a person using the product. Many of these uses are hard to convey verbally, yet very easy to show with homemade video clips. Furthermore, the increasingly powerful technologies available on a personal computer allow the researcher to move far beyond text, and show many different alternative executions of the package "in action." Although the conclusion may sound trite, especially to those who write research grants, far more research is called for to understand the dynamics of interacting with a package. The executions need to be varied to look at the different dimensions, but as this towelette study suggests, the video-clip executions need to be "there," rather than need to be "perfect."

Acknowledgments

The authors are grateful to both Alexis Fried and Alex Gofman for providing first-hand accounts of their experience with using video in the early days of systematized concept research.

References

Burke, R. (1991) Comparing Dynamic Consumer Decision Processes in Real and Computer Simulated Environments. Marketing Science Institute Report (No. 91–116).

Casey, M.A. and Krueger, R.A. (1994) Focus Group Interviewing. In: MacFie, H.J.H. and Thomson, D.M.H. (eds.) Measurement of Food Preferences. Blackie Academic & Professional, p.77–97.

Chen, A.W., Resurreccion, A.V.A., and Paguio, L.P. (1996) Age appropriate hedonic scales to measure food preferences of young children. Journal of Sensory Studies 11(2):41–163.

Deliza, R. and MacFie, H.J.H. (1996) The Generation Of Expectation By External Cues And Its Effects On Sensory Perception And Hedonic Ratings: A Review. Journal Of Sensory Studies 11, 103–128.

Gacula, Jr., M.C. (1993) Design and Analysis of Sensory Optimization. Food and Nutrition Press, Trumbull, CT.

Kroll, B.J. (1990) Evaluating rating scales for sensory testing with children. In: Food Technology (4411), 78–80, 82, 84, 86.

Moskowitz, H. and Gofman, A. (2007) Selling blue elephants: How to make great products that people want before they even know they want them. Wharton School Publishing, Upper Saddle River, NJ.

Moskowitz, H., Krieger, B., and Tamber, N. (1996) Incorporating consumer feedback into package description and presentation: Multi-media approach techniques. ASTM special technical publication ISSN 0066–0558. American Society for Testing and Materials, Philadelphia, PA.

Needel, S. (1998) Understanding response to category management through virtual reality. Journal of Advertising Research 38, 61–67.

Robinson, T.N., Borzekowski, D., Matheson, D.M., and Kraemer, H.C. (2007) Effects of fast food branding on young children's taste preferences. Archives of Pediatric Adolescent Medicine 161(8):792–7.

Treiber, B. and Needel, S.P. (2000) New ways to explore consumer point-of-purchase dynamics. ESOMAR, Impact of Networking September 2000:263–275.

Part III

Health and Hope

Packaging Research in Food Product Design and Development Howard R. Moskowitz, Michele Reisner, John Ben Lawlor and Rosires Deliza
 ISBN: 978-0-813-81222-9

Chapter 14

Do Labels Make a Difference?

As consumers become increasingly health conscious, we know that they say they pay more attention to what's on the label, such as nutritional information (Korver, 1997; Kristal et al., 1998; Costa and Jongen, 2006). However, we don't know what they react to and how unless they tell us, or unless they go through an experiment (Deliza, MacFie, and Hedderly, 1999; Pieters and Warlop, 1999; Roe et al., 1999; Smith et al., 2000). In this chapter, we are going to do both. First, we are going to ask respondents what they pay attention to, or better, we are going to look at how they rate their "label-gazing" behavior. Then, in the second part of the chapter, we will look at what types of label messaging consumers react to. With these two pieces of information, we can start to understand what is important about labels, and what is not.

What Types of Labels Make a Difference?

We began this particular investigation by asking the question "What types of labels do people say they pay attention to?" By the very way we framed the question, we really wanted a self-report from the respondents. Of course, people may not be able to articulate exactly what they pay attention to during the course of shopping. Most people have a difficult time providing these "unaided recalls," wherein they are asked the following: "List what you pay attention to when you look at a label." It is likely that when a person is asked to list things, the person will remember what he has heard, what is "au courant," what is of general interest, but which may not have anything to do with what is the actual case.

A different way of asking this question presents respondents with the specific topic of the label, and gets them to rate the importance of different types of information. In a large-scale study with over 1,000 respondents, we followed this approach for the different types of information listed below. Let's see what happens when we ask the respondents the following question, and have them rate the importance on a 9-point scale. On the scale, 1 = not important at all to 9 = extremely important:

When you read the label, how important is …

Freshness
Lack of preservatives
Organic
No artificial sweeteners
Natural
Number of servings
Antioxidants
Ingredients
Number of calories
Number of ounces

With this type of information, let us now look first at the total panel, at the gender breaks, and at three different ages. Keep in mind that our respondents are rating importance. By convention, we break the 9-point scale into two portions. Ratings of 1–6 are coded as 0; ratings of 7–9 are coded as 100. The coding allows us to separate those respondents who think that the information is important for a label versus those who think that the information is not as important. In this analysis, and others in this book, we use 1–6 versus 7–9 as two regions. It is easy to deal with unimportant/important as a binary variable, rather than dealing with the set of 9 points. We could have broken the scale in another area (e.g., 1–5 is unimportant, 6–9 is important). We chose the 1–6 and 7–9 because we have worked with that break many times and, thus, have a sense of what the numbers mean from many different projects.

What do people look for when they are asked about labels? If we look at Table 14.1, we get a pretty good picture of what's going on.

1. Freshness is the key piece of information that most people want from the label. The label is the messenger

 ISBN: 978-0-813-81222-9

to tell the consumer whether the food is fresh or not. We see the importance of freshness for every group. Thus, 86% of the respondents rate freshness 7–9 in terms of what's important on the label. As the respondent gets older, freshness becomes a more important factor on the label, but even the younger respondents under 30 consider freshness important (82% of younger respondents, 90% of older respondents).

2. Beyond freshness we have two other factors that are about equally important: the ingredients and the number of calories, both at 62% of the respondents. Ingredients and number of calories show similar patterns across genders and ages.
3. Lack of preservatives occupies a lower rung, with 53% of the respondents feeling that it is important. More women than men, and more older than younger respondents, look at preservatives as important.
4. At the next lower rung, with fewer than half of the respondents, we have the following four factors: natural, no artificial sweeteners, and the two numerical facts, number of servings and number of ounces.
5. At this fourth rung, we see a new pattern emerging. Age is irrelevant to a person's sensitivity to "natural." Approximately the same proportion of older respondents as younger respondents look for "natural" on a label.
6. The number of ounces in a product as shown on the label is substantially more important for the older respondents than for the younger ones. Half the older respondents, 50%, look for the number of ounces. Only a third of the younger respondents, 33%, look for the number of ounces. This is an important change, one worth investigating more. It suggests that older respondents need information about the amount to be consumed, as if they are looking at the amount that they are buying, perhaps because they look at the product with a time horizon in mind, and need to know how much they are "stocking."
7. At the very bottom we have "organic," which is of interest to only one-quarter of the respondents, independent of gender and age.

We can do the same type of analysis, this time for different ethnic groups. Table 14.2 shows the results for five key groups. We have White, Afro-Americans, Asians, and two groups of Hispanics. The first group of Hispanics comprises those who were born in the United States, and thus have lived with the American food culture their entire lives. The second group comprises Hispanics who immigrated to the U.S., and represent a group who is acculturating to U.S. food norms.

Table 14.1 Proportion of respondents from different genders and ages (columns) who rate each label factor as important on a 9-point scale

	Mean	Male	Female	Age Under 30	Age 30–50	Age Over 50
Freshness	86	83	89	82	87	90
Ingredients	62	56	67	51	60	76
Number of calories	62	52	70	52	60	75
Lack of preservatives	53	47	58	45	53	61
Natural	45	43	47	44	44	47
No artificial sweeteners	43	39	46	35	45	48
Number of servings	43	34	50	38	47	41
Number of ounces	42	41	44	33	43	50
Antioxidants	38	37	40	32	36	48
Organic	27	25	28	25	27	28

Table 14.2 Proportion of respondents from different ethnic subgroups (columns) who rate each label factor as important on a 9-point scale

	Mean	White	Afro-American	Asian	Hispanics who were born in the U.S.	Hispanics who immigrated to the U.S.
Fresh	86	85	90	87	88	88
Number of calories	62	62	58	57	68	67
Ingredients	62	61	65	70	60	70
Lack of preservatives	53	51	55	74	56	64
Number of servings	45	42	42	39	45	55
Natural	43	41	52	87	49	70
No artificial sweeteners	43	40	48	65	48	64
Number of ounces	42	40	50	35	45	52
Antioxidants	38	37	44	26	36	55
Organic	27	24	31	43	31	39

Rather than repeating the detailed factor by factor discussion, we can look at Table 14.2 to get a sense of the "bigger picture." There are six results that tell us something about ethnicity and labels.

1. There are some differences by ethnicity. For example, Asians look for no preservatives, no artificial sweeteners, and organic. It's clear that Asian respondents are more focused on "health" and "good for you" characteristics, and look for these in the food labels.
2. Afro-Americans focus more on number of ounces, anti-oxidants, organic, and lack of sweeteners than do Whites, but not dramatically so. The difference is only about 8–10% more Afro-Americans than Whites focus on these label factors.
3. Hispanics provide an interesting finding in terms of the effects of acculturation. Those Hispanics who were born in the U.S. show patterns similar to Whites. However, Hispanics who immigrated to the U.S. show patterns that are quite different. These first-generation Americans, who are acculturating, show far greater interest in the label factors than do the Hispanics who were born in the U.S. However, the change is not in freshness or the number of calories. With those factors, there is not much effect of acculturation. The effect comes with the other label factors, the less important ones. The biggest one is antioxidants. This is relatively unimportant to Hispanics born in the U.S. (36%) but much more important to Hispanics who are acculturating (55%).

People Vary When It Comes to What's Important on a Label

We just saw that there are differences in the importance of different label factors. We were able to look at the different groups and see some variation among groups. A more insightful analysis can come from taking a granular view. In this view, we look at the distribution of ratings on the 9-point scale for all 1,022 respondents. The question we should ask is very simple: Does each of these factors show most of the respondents clustering around the mean, or is there a distribution?

The answers won't be particularly surprising. Look at Figure 14.1. We see two types of patterns. The first pattern shows most of the respondents clustering around a small region on the 9-point scale. We see this for the label factor "fresh." We see this pattern but a little less striking for the label factors "ingredients" and "calories," where an increasing number of respondents rate the factor high.

We see a different pattern for the other label factors. The importance ratings distribute. For example, when it comes to organic, we see no clear pattern or, for that matter, when it comes to antioxidants.

We conclude from this graphic analysis that there are different types of respondents, with different mind-sets. The most important finding is that almost everyone thinks that the role of the label is to tell one about product freshness and perhaps about ingredients and calories. The remaining factors are idiosyncratic, depending upon the individual's particular interest in health and wellness.

What Interest in Ingredients Drives a Person to Look at Label Factors?

We conclude this section on labels with an attempt to discover a linkage between what's important to a respondent in terms of "ingredients in a food," and what the person looks for in a label.

The same respondents who rated importance of label factors also rated the importance of different ingredients in foods that they buy. The respondents were not asked whether they looked for information about these ingredients, but rather simply to rate the importance to them of the ingredient in store-bought foods.

As one might expect, there is variation among respondents as to what is important (Grunert, Brunson, Bredahl, and Bech, 2001; Saba, 2001). However, we are more interested here in linking what people think is important to what they look for on the label. In Table 14.3, see the correlations between what's important in terms of ingredients (columns) and what people look for on the label (rows). The numbers in the body of the table are Pearson correlations, the Pearson R statistic. This statistic tells us the strength of a linear relation between the importance rating of fiber (specific feature) and the importance of the ingredient (general factor looked for).

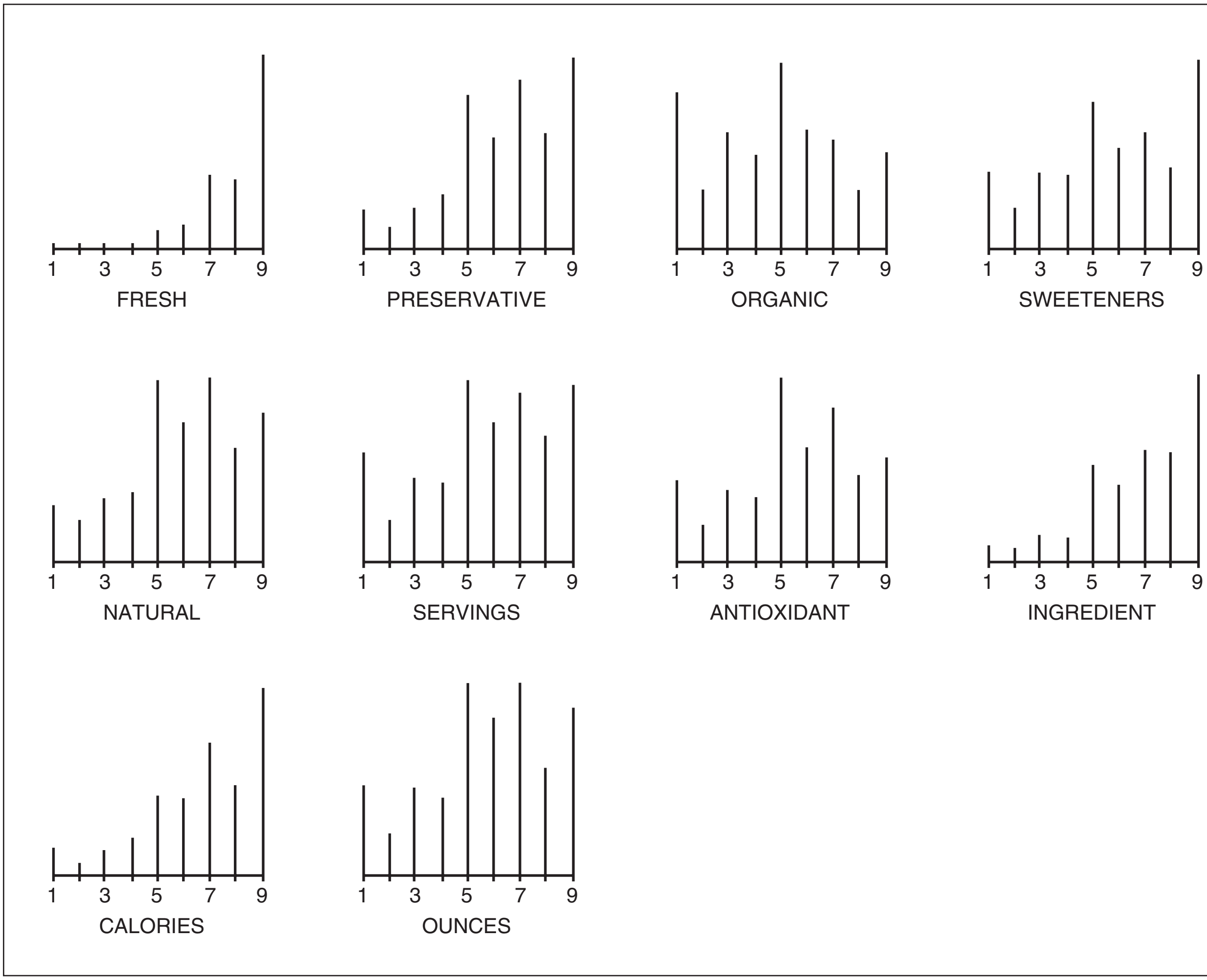


Figure 14.1 Distribution of the ratings for importance on the 9-point scale. Each label factor was rated in importance by 1,022 respondents.

The Pearson R or linear correlation ranges between a low of –1 (perfect inverse relation), to a middle value of 0 (no linear relation at all), to a high of +1 (perfect linear relation). With 1,022 "cases" or observations, it is highly unlikely to achieve a high Pearson R if there is no underlying relation (Box, Hunter, and Hunter, 1978).

We have boldened in those correlations above 0.40. We can feel confident that those correlations denote a high degree of association between what ingredient is important to the respondent when shopping, and what the respondent looks at when inspecting the label.

1. Our results suggest that if a person is interested in any of the ingredients (all columns), then the person will look at the ingredient list, and most likely the number of calories, antioxidants, and lack of preservatives.
2. Being interested in calories as "an ingredient" translates primarily to being interested in label statements about ingredients and number of calories, respectively.
3. The remaining ingredient interests translate to specific interests in label factors. For example, being interested in "sugar" translates to thinking that "natural" is important on a label. Or being interested

Table 14.3 How interest in specific ingredients (columns) correlates with what is stated to be important when one reads a food label (row). The numbers in the body of the table are Pearson correlations (R-statistic).

	Fiber	Protein	Sodium	Vitamins	Trans Fat	Sugar	Saturated Fat	Carbohydrates	Calories
Ingredient (in general)	**0.53**	**0.50**	**0.49**	**0.51**	**0.53**	**0.55**	**0.53**	**0.48**	**0.45**
Number of Calories	**0.48**	**0.47**	**0.47**	**0.38**	**0.54**	**0.55**	**0.57**	**0.50**	**0.83**
Antioxidants	**0.56**	**0.54**	**0.51**	**0.60**	**0.51**	**0.46**	**0.47**	**0.46**	0.36
Lack of Preservatives	**0.50**	**0.48**	**0.53**	**0.48**	**0.50**	**0.48**	**0.47**	**0.44**	0.35
Natural	**0.41**	**0.45**	**0.44**	**0.50**	0.39	**0.42**	0.38	0.37	0.26
Ounces	0.38	0.37	0.36	0.37	0.33	0.35	0.35	0.38	0.37
No Sweeteners	0.38	0.39	**0.43**	**0.40**	0.39	0.37	0.36	0.33	0.20
Number of Servings	0.37	0.39	0.32	0.31	0.35	0.34	0.33	0.33	**0.46**
Organic	0.38	0.38	0.37	0.36	0.35	0.35	0.32	0.32	0.27
Fresh	0.24	0.23	0.25	0.24	0.23	0.25	0.21	0.24	0.20

in calories translates to "number of servings" being important on a label.

Summing Up

This first study suggests to us that respondents find different factors important when they read the labels. It shows differences in label factors by age more than by gender. It also reveals that ethnicity is important, with Asians far more interested than Whites, Afro-Americans, and U.S.–born Hispanics. Hispanics who immigrated to the U.S. differ from U.S.–born Hispanics, with the former, emigrating group, more sensitive to factors such as artificial sweeteners. U.S.–born Hispanics look more similar to Whites. Finally, respondents are very similar to each other when it comes to freshness but differ substantially on the importance of other label factors, such as organic. It's important to most respondents to understand product freshness from reading the label. However, people differ from each other in the importance they attribute to "organic."

What Label Information "Works"

Now that we have explored what is important "in general," let us look at how respondents react to specific messaging. We are still going to work at the "general" level, rather than at the level of a specific food or beverage. With the increasing interest by government agencies on the best way to communicate healthfulness to consumers through food labels (Goldberg, 2000), we now move to a study that addresses that problem.

The origin of the study may be of some interest. In the summer of 2007 ILSI (International Life Sciences, Inc.) of Southeast Asia sponsored a three-day symposium on nutritional challenges in Southeast Asia. The senior author of this book attended and presented a paper. At the end of the conference, he went out for a tour of Singapore with a number of the other participants. During a refreshment break, the group visited a store, looked at some of the products, and talked to some locals. Coming out of that was the realization that it was possible to use some of these scientific methods to explore responses to labels, worldwide. No one at the conference was able to articulate exactly what a Singaporean consumer needed to see on the label. And so was born the experiment—to identify what particular messaging on the label would attract consumers to feel that the food was healthy. The actual experiment we report here was to be the first of a series of studies, conducted for a number of foods, both inside the U.S,. as well as in Southeast Asia and Western Europe. We report on the results of the first study that emerged from the trip. So, in a way, this half of the chapter is our first fruit, in which we look at the specific messaging on the label.

We begin with the nature of food claims, the topic that concerns the Food and Drug Administration in the U.S. The FDA has established language guidelines for characterizing the level of nutrients in food (see Table 14.4). The terms "good source of" or "excellent source of" can only be used to describe protein, vitamins, minerals, dietary fiber, or potassium. Good sources of a given nutrient provide 10–19% of the daily value (DV) per reference amount, whereas excellent sources provide 20% or more. Conversely, such terms as "free," "low," or "reduced/less" apply to calories, total or saturated fat, sodium, sugars, or cholesterol. These, too, are based on reference amounts or on 50 g if the reference amount is small. Clearly, consumers cannot deal with many of these issues. Most consumers pay attention to words like "light," "fat free," etc. They do not pay attention to the legal definitions.

Now that we know the legal definitions, let's see how consumers respond to different test labels. For this experiment, we selected different phrases that would ordinarily appear on the package. This approach makes sense—we are investigating how nutrition labeling affects consumer responses, so we work in the realm of what is legal, rather than exploring close in versus far out ideas.

Table 14.4 Nutrient content claims as defined by the Food and Drug Administration (Source: Drewnowski et al. 2008, unpublished manuscript)

Term	Nutrient content per reference amount
Good source of; contains; provides	Contains 10–19% of the DV. Used to describe protein, vitamins, minerals, dietary fiber, or potassium, but not carbohydrate.
Excellent source of; high; rich in	Contains 20% or more of the DV. Used to describe protein, vitamins, minerals, dietary fiber, or potassium, but not carbohydrate.
More; added; extra; plus	Contains 10% or more of the DV. May only be used for protein, vitamins, minerals, dietary fiber, or potassium,
High potency	Describes individual vitamins or minerals that are present in food at 100% or more of the RDI per reference amount.
Free; zero; no; without	Calories <5; total fat <0.5 g; saturated fat <0.5 g; trans fat <0.5 g; cholesterol <2 mg; sodium <5 mg; sugars <0.5 g (per reference amount and per labeled serving)
Low; little; few; low source of	Calories <40*; total fat <3 g*; saturated fat <1 g[†]; cholesterol <20 mg*; sodium <140 mg*; sugars: not defined
Reduced/less; lower; fewer	Calories >25% less; total fat >25% less; saturated fat >25% less; cholesterol >25% less; sodium >25% less; sugars >25% less[†]
Light	Percent reduction for both fat and calories must be stated

*Per 50 g if reference amount is small

[†]With <15% calories from saturated fat

[†]Relative to appropriate reference food. May not be termed "low."

In order to explore a reasonably wide range of elements, we looked at six silos, each of six elements, which comprise a total of 36 elements. The elements appear in Table 14.5. Note that we have divided the set of elements into three sets of silos. The first set comprises 12 elements that are desirable. We wanted to look at two ways to express these nutrients, because it wasn't clear whether or not the way of expressing a positive nutrient made a difference. No one really knew whether, for instance, there would be much difference between two similar elements with different emphases, such as "This product is a good source of protein" versus "This product is an excellent source of protein."

Our second pair of silos, C and D, covered the nutrients that the FDA advises limiting. Finally, the third pair of silos, E and F, covered emotional messages that aren't technically part of the nutrition label itself, but could be put on the package as additional "wellness-related" information.

We recruited the respondents to participate from a local American panel, specializing in e-mail recruiting of respondents. The key to the entire study comes from correctly orienting the respondents. The respondents received an email invitation, which told them just a little about the study, not enough to tell them exactly what was being tested, but enough information to intrigue them. Of course, it is necessary to reward people for participation. To make this study, and other studies like it, more cost effective, we ran a sweepstakes so that respondents who participated had a chance to win a cash prize. These incentives make the respondents feel that their time and efforts are appreciated.

Respondents who chose to continue (i.e., participate), clicked on the embedded link in the e-mail invitation. This led them to the orientation page that told them about the study. The orientation page set up the respondent's expectations about the study, while providing relatively little detail about the stimuli, the rationale for the study, etc. The goal of the orientation page is to make sure the respondents understand what is expected of them. Here is what we used for this study. If you read it carefully you will see that there is a great deal of

Table 14.5 Elements tested in the nutrition label project and their impact values for the total panel

	Additive constant	**22**
	Silos A and B—Desirable nutrients	
A1	This product is a good source of protein.	17
A2	This product is an excellent source of protein.	16
A3	This product is a good source of vitamin C.	9
A4	This product is an excellent source of vitamin C.	11
A5	This product provides vitamin A.	6
A6	This product is high in vitamin A.	9
B1	This product contains fiber.	9
B2	This product is high in fiber.	14
B3	This product contains calcium.	8
B4	This product is rich in calcium.	11
B5	This product provides iron.	4
B6	This product is rich in iron.	5
	Silos C and D—Nutrients to limit	
C1	This product is low in total fat.	7
C2	This product is fat free.	10
C3	This product contains little saturated fat.	3
C4	This product has no saturated fat.	12
C5	This product is a low source of cholesterol.	7
C6	This product has zero cholesterol.	11
D1	This product is low in total sugar.	8
D2	This product is sugar free.	9
D3	This product is low in added sugar.	4
D4	This product is free of added sugar.	10
D5	This product is a low source of sodium.	9
D6	This product is without sodium.	12
	Silos E and F—Emotional messages and reassurance	
E1	Wholesome food that gives you more nutrition per bite	4
E2	Is a good way to balance your diet to keep it nutrient rich	6
E3	Is a total nutrient package with more nutrients than calories	4
E4	Naturally packed with nutrients for better health	6
E5	Meets your daily nutrient needs without too many calories	5
E6	Puts more nutrient power on your plate	7
F1	Is a great way to enjoy your healthy lifestyle	4
F2	Takes the stress out of healthful eating	4
F3	Lets you eat well to live well, starting today	3
F4	You and your family can eat right, for life.	5
F5	Be at your best—enjoy good taste and good health.	7
F6	You can trust the nutrition label to guide smart eating.	3

general information, but nothing that "gives the game away."

> *Providing nutrition information on food and drink product labels is an important way of conveying the message about diets and health to the consumer. It is important that such information accurately reflect the nutrient composition of the product in a simple manner that is easily understood by the consumer. However, nutrition labels do not always get the right message across. You are invited to review a selection of messages and rate each product on a "healthfulness" scale.*

Participants were told that they would be presented with nutritional information for a hypothetical food product. Based on this information, they were asked to use a 9-point category scale, to rate whether, in their opinion, the product was 1 = least healthy or 9 = most healthy. Each respondent evaluated 48 test concepts, comprising small combinations of the elements. Throughout this book, we use the same type of approach—experimental design, so we will simply point out the highlights of the "method," rather than go into detail.

We ran the study with 320 U.S. consumers. The data speak for themselves. See Table 14.5. Let's interpret the results. The data you will see comes from an analysis of the ratings. The rating scale (1 = least healthy to 9 = most healthy) was converted to the binary scale as we did for the above-mentioned attitude research on labels. That is, ratings of 1–6 for a test concept were converted to 0, ratings of 7–9 were converted to 100. We then developed a model for each of the 320 respondents showing how every one of the 36 elements in Table 14.5 "drives" the less healthly/more healthy response.

With that information in mind, let's now look at the data in a bit more depth. We don't have to probe very deeply—the results are fairly straightforward and, occasionally, surprising. Keep in mind that we did not specify the particular food. We are only looking at the general response to these test "nutrition labels."

1. The additive constant is 22. The additive constant is the percent of respondents who would rate the concept 7–9 if no elements were present. Clearly this constant is an estimated, theoretical value, but it is good as a baseline indicator (Green and Krieger, 1991). Without any additional information, 22% or about one in four respondents would say that the product is "more healthful." That finding provides us with a sense of

how the dynamics will play out. People aren't likely to say that a food is more healthful. It's the messaging that counts, the specifics of the food. We don't yet know what that is, however.

2. All the messaging is positive. There are no negative elements, as there usually are in these studies. Actually, this positivity should come as no surprise. We are working here with government-approved messaging. We are not promising anything new in the test concepts. We are simply looking at the performance of the different, already-approved messages.
3. Talking about desirable nutrients in simple versus more hyperbolic terms occasionally makes a difference, but often does not do very much. For example, talking about Vitamin A in hyperbolic terms (*this product is high in vitamin A*) increases the impact from 6 to 9. Occasionally, some of the positive elements benefit such as those talking about fiber, where "high in fiber" has an impact of 14 (14% more respondents feel that the product is healthy), whereas "contains fiber" only has an impact of 9.
4. Talking about saturated fat in hyperbolic terms really makes a big difference, however. Saturated fat is something to control, preferably eliminate altogether. Saying "contains little saturated fat" has an impact of 3, whereas saying "no saturated fat" has an impact of 12.
5. The emotional and reassurance elements don't perform any better than the nutrient statements. People aren't looking for emotion and reassurance when it comes to "good for you." They appear to look for the facts.

Segmenting the Mind of the Label Reader

We have seen through this book in other chapters that segmentation holds the key to increasing success among consumers (Hammond, Ehrenberg, and Goodhardt, 1996). The segmentation is not necessarily along traditional lines of geo-demographic or even psychological dimensions (so-called psychographic segmentation). Rather, the segmentation emerges from the pattern of responses to the different elements, and may represent profound, hard-to-predict groups. These groups clearly have different and coherent "mind-sets." That is, the segmentation that looks at the pattern of responses to elements comes up with groups that make intuitive sense. What is interesting, however, is that despite the strength of the responses to elements shown by the segment and the coherence of the segments, it is not easy to predict membership in the segments. The segmentation "tells a story."

Let's look at this story for labels. We segmented the respondents by the pattern of their impact values. Now to see the results, which appear in Table 14.6. Rather than showing all of the elements by the segments, which could generate a very large table, we have summarized the results by choosing the highest and lowest performing elements for each segment.

One of the goals of the segmentation was to search for a group of respondents who could be classified as "health sensitive." This segmentation may be especially important for specific categories for functional/enriched foods that will benefit from favorable profiles and health claims.

The segmentation generated three groups that could be interpreted. The respondents in a segment show similar profiles of impacts. That is, the people may differ widely by age, gender, income, even stated interest in health. Yet these people show similar patterns of responses to what they consider to be "healthy." We see these radical differences, and despite the segmentation, we see that the emotional and reassurance elements play small roles.

Now, to label the segments. We choose labels that tell the story. We learn the story by their reactions to the elements, and by the additive constant, or their proclivity to call a product healthy.

Segment 1 can be called "fat is bad."

1. They want low fat and cholesterol.
2. However, they do not feel that sugar and sodium should be limited. Taking away sugar and sodium does not, in their mind, produce a healthful product.
3. They have a high additive constant of 43, meaning that the probability is almost 50% of these individuals calling a label healthful, even without elements.

Segment 2 can be called "extremists."

1. They respond strongly to statements that contain "no levels."
2. They constitute virtually half of the panel (157 out of 320). Importantly, they begin with a low constant, 9. That is, they do not have any proclivity to call a label "healthful" unless they see the messages. The messaging does all the work.

Table 14.6 Performance of the most healthful and least healthful elements for each segment

	Segment	Base / Constant
	Segment 1—Fat is bad: Low fat, cholesterol, don't compromise expected taste (attributed to sugar and sodium)	Base = 58 Constant = 43
Limit	This product has no saturated fat.	**17**
Limit	This product is low in total fat.	**16**
Emot	Be at your best … enjoy good taste and good health.	**16**
Emot	Lets you eat well to live well, starting today.	**15**
Limit	This product contains little saturated fat.	**15**
Limit	This product is a low source of cholesterol.	**15**
Limit	This product is fat free.	**13**
Good	This product is a good source of protein.	**12**
Limit	This product has zero cholesterol.	**12**
Good	This product is an excellent source of vitamin C.	**12**
Limit	This product is without sodium.	–7
Good	This product is rich in iron.	–7
Limit	This product is free of added sugar.	***–7***
Limit	This product is a low source of sodium.	***–7***
Limit	This product is low in total sugar.	***–8***
Limit	This product is low in added sugar.	***–13***
Limit	This product is sugar free.	***–19***
	Segment 2—Extremist: Seek extreme levels of good nutrients and no levels of limited nutrients	Base = 158 Constant = 9
Good	This product is high in fiber.	**22**
Limit	This product is sugar free.	**20**
Limit	This product is without sodium.	**19**
Limit	This product is free of added sugar.	**18**
Limit	This product has zero cholesterol.	**18**
Limit	This product has no saturated fat.	**18**
Limit	This product is fat free.	**17**
Limit	This product is a low source of sodium.	**16**
	Segment 1—Fat is bad: Low fat, cholesterol, don't compromise expected taste (attributed to sugar and sodium)	Base = 58 Constant = 43
Good	This product is a good source of protein.	**16**
Limit	This product is low in total fat.	**15**
Good	This product is rich in calcium.	**14**
Good	This product is an excellent source of protein.	**14**
Good	This product contains fiber.	**14**
Limit	This product is low in total sugar.	**14**
Good	This product is an excellent source of vitamin C.	**14**
Limit	This product is a low source of cholesterol.	**13**
Good	This product is rich in iron.	**12**
Good	This product is high in vitamin A.	**12**
Limit	This product contains little saturated fat.	**11**
Good	This product contains calcium.	**11**
Limit	This product is low in added sugar.	**11**
	Segment 3—Protein is good, and fat is good: Conventional and within reason, don't compromise taste (attributed to fat)	Base = 105 Constant = 30
Good	This product is an excellent source of protein.	**21**
Good	This product is a good source of protein.	**20**
Limit	This product is without sodium.	**14**
Good	This product is a good source of vitamin C.	**12**
Good	This product is high in fiber.	**12**
Emot	Be at your best … enjoy good taste and good health.	**11**
Limit	This product is a low source of cholesterol.	–6
Limit	This product is low in total fat.	***–10***
Limit	This product contains little saturated fat.	***–16***

Segment 3 can be called "protein is good *and* fat is good."

1. They want both.
2. They don't want fat-free foods. They don't see the fat-free foods as being healthful.
3. They constitute about a third of the respondents.
4. They have a moderate additive constant, 30, meaning that they have some proclivity to calling a label "healthful" without the elements, but the messaging has to do a lot of the work, albeit less work than Segment 2 expects from the messaging.

Summing Up

Systematic variation of the information on the label goes a step beyond asking the respondent what is important. We begin to discover how the respondent's mind works. The data presented here suggest that the government regulations about packaging may not affect the respondents in the same way. Although the essence of nutrition labeling is the true "information," and not the reassurance provided by the additional elements, there is now every reason to begin more detailed investigations. We have shown here that there are systematic individual differences in the effectiveness of the label messaging. The same element may affect two segments quite differently.

However, matters are not simply that easy. We do not know, however, how these elements will "play out" when they are associated with individual products. We now have a more complex task but new database tools to handle it. It will be interesting to see the nature of databases that can be created which incorporate product, health messaging, and respondent segmentation. Hopefully this chapter has given some stimulus to that effort.

References

Box, G.E.P., Hunter, J., and Hunter, S. (1978) Statistics for Experimenters. John Wiley and Sons, New York, NY.

Costa, A.I.A. and Jongen, W.M.F. (2006) New insights into consumer-led food product development. Trends in Food Science and Technology 17, 457–465.

Deliza, R., MacFie, H., and Hedderley, D. (1999) An investigation using the repertory grid and focus group methods of the package features affecting consumer perception of fruit juice. Brazilian Journal of Food Technology 2, 63–71.

Drewnowski, A., Moskowitz, H.R., Reisner, M., and Krieger, B. (2008) Concept testing of the FDA regulatory system for nutrient content claims using conjoint analysis. (Unpublished research paper)

Goldberg, J.P. (2000) Nutrition communication in the 21st century: what are the challenges and how can we meet them? Nutrition 16 (7–8), 644–646.

Green, P.E. and Krieger, A.M. (1991) Segmenting markets with conjoint analysis; In: Journal of Marketing 55, 20–31.

Grunert, K.G., Brunson, K., Bredahl, L., and Bech, A.C. (2001) Food-related lifestyle: a segmentation approach to European food consumers. In: Frewer, L.J., Risvik, E. and Schifferstein, H. (eds.) Food, People and Society. Springer-Verlag, Berlin, Germany, 211–232.

Hammond K.A., Ehrenberg A., and Goodhardt, G.J. (1996) Market segmentation for competitive brands. European Journal of Marketing 30:12; 39–49.

Korver, O. (1997) "Healthy" developments in the food industry. Cancer Letters 114, 19–23.

Kristal, A.R., Levy, L., Patterson, R.E., Li, S.S., and White, E. (1998) Trends in food label use associated with new nutrition labeling regulations. American Journal of Public Health 88, 1212–1214.

Pieters, R. and Warlop, L. (1999) Visual attention during brand choice: the impact of time pressure and task motivation. International Journal of Research in Marketing 16, 1–14.

Roe, B., Levy, A.S., and Derby, B.M. (1999) The impact of health claims on consumer search and product evaluation outcomes: results from FDA experimental data. Journal of Public Policy & Marketing 18, 89–105.

Saba, A. (2001) Cross-cultural differences in food choice. In: Frewer, L.J., Risvik, E., and Schifferstein, H. (eds.) Food, People and Society. Springer-Verlag, Berlin, Germany, 233–246.

Smith, S.C., Taylor, J.G., Stephen, A.M. (2000). Use of food labels and beliefs about diet-disease relationships among university students. Public Health Nutrition 3, 175–182.

Chapter 15

Understanding Nutritional Labeling: Case Study—Ice Cream

Introduction—The Importance of Health

Packaging research today needs to consider health. Anyone who works in the food and drink industry realizes more than ever that health considerations are paramount in consumer choice (Vickers, 1993; Westcombe and Wardle, 1997; Prescott et al., 2002; Tepper and Trail, 1998). As the developed economies become increasingly richer, focus moves away from subsistence, toward convenience and, eventually, toward food and oneself. In a more cynical vein one might say that we are in a stage of WIIFM—what's in it for me!

Scientists already know that it is important to put in the necessary nutrients, that consumers are becoming increasingly savvy, and that a generation is growing up that has ever more choices than the consumers of generations before. What is the manufacturer to do? Well, for one thing, beyond putting in the correct nutrients, it is pretty critical to communicate these as health benefits, both by designing the product and by designing the package. When properly communicated, these health benefits can help to sell the product. Furthermore, in an era when the cost of goods is rising, when commodity prices are turning companies topsy-turvy, strong health communications can stave off real problems in the industry. The package is the perfect venue to communicate food "health" (a rather nebulous concept) and nutritional content (a more concrete concept) (Aaron et al., 1994; Bower et al., 2003).

The packaging engineer faces problems, however, because the rules vary. Legislation controlling food package information differs by country. The Food and Drug Administration (http://www.fda.gov/) regulates food label claims in the U.S. The FDA has listed guidelines for food packaging including location of food claims on packages, nutritional facts panel format and guidelines for food claims. In the European Union, food-labeling legislation is detailed in Council Directive 2000/13/EC (see http://europa.eu), whereas in Australia and New Zealand, food labeling is regulated by the Food Standards Australia New Zealand (see http://www.foodstandards.gov.au/). Different organizations control labeling, or at least affect what can and should be said.

In the U.S., nutrition and health-related claims fall into three categories: 1) health claims, 2) nutrient content claims, and 3) structure/function claims (FDA, 2003). Health claims describe the relation between a food, a food component or dietary supplement ingredient, and the reduced risk of disease or health-related conditions (e.g., fat and cancer or calcium and osteoporosis). Nutrient content claims include statements that characterize the level of a nutrient in a food such as "free," "low," and "high," etc. Structure/function claims include statements that describe the role of a nutrient or dietary ingredient intended to affect normal structure or function in humans such as "calcium builds strong bones," etc. In addition, as defined by the FDA, a food described as "healthy" (an implied nutrient content claim) on a food package label must be low in fat and saturated fat and contain limited amounts of cholesterol and sodium (see FDA, 2003, for a full definition). Finally, the FDA has also defined regulations on the use of health symbols (e.g., hearts, etc.). It's clear that there are guidelines that packaging designers have to follow. The real question is: What works within these guidelines?

This chapter explores health and nutrition messages using ice cream as a base product. Ice cream was chosen because it is generally regarded by consumers as an indulgent food and was, therefore, considered an interesting base product to examine health- and nutrition-based messages on. The world's ice cream market is estimated at more than $40 billion U.S., with the U.S. market being the largest per capita consumer of ice cream (dairy foods.com, Feb, 2006). Bech-Larsen and Grunert (2003) reported that consumers reacted differently to health-based messages depending on whether the base product

 ISBN: 978-0-813-81222-9

was perceived as healthy or unhealthy, and that health messages might perform better on "unhealthy" foods compared with "healthy" foods. (See also Kähkönen et al., 1997.)

Going deeper into the chapter, our primary objective is to discover how different types of ice cream package messages (with a primary focus on health- and nutrition-type messages) drive consumer responses. Our secondary objective is to learn how responses to ice cream package messages vary when the "facts" change on the nutritional facts panel.

Running the Ice Cream Study

Choosing the Categories

We looked at a number of ice cream packages. It seemed that, for the most part, the packaging graphics were fairly spare for ice cream. We decided to investigate three silos that typify the standard ice cream package:

1. Brand identification,
2. Ice cream type (e.g., original ice cream, low-fat ice cream, etc.), and
3. Flavor identification with visual.

Our experimental design called for five elements within each silo. We wanted to use actual elements from commercial packaging, in order to discover which ice cream manufacturer "got it right," at least within the confines of this study. The Internet provides an excellent source of such packaging features, as you probably have seen from the other chapters. Indeed, it is fair to say that the Internet continues to provide an extraordinary source of good visual stimuli to test (Moskowitz et al., 2004).

Our three categories of brand, type, and flavor are, by unwritten convention, typically located on the principal display panel (PDP) of food packages. When we looked into what this would entail, we quickly realized that the PDP information was "marketing oriented", and that despite the fact that we're looking at packaging, it would be good to test the more "rational" aspects of nutrition that might not be completely present on the front of the package. That is, we created a hybrid stimulus, with some front-panel information, and some back-panel nutritional information.

We decided to include nutritional facts panels, normally located on the information display panel. This information represents another type of nutrition/health information available to consumers. Nutritional facts panels were part of Silo B, and their content was specific for each ice cream type. In addition, since package size and nutrition facts panels are related via "number of servings per container," package size was also specified. In both of our studies, we investigated the 1.75-quart package.

Silo A: Brand

There are a variety of national brands in the U.S., where we ran our study. Figure 15.1 shows the five brands we considered, to represent a range of existing product qualities and prices. These five brands are Breyers®, Dreyer's, Häagen-Dazs®, Ben & Jerry's, and Stonyfield Farm®, respectively.

Silo B: Type of Ice Cream

We looked at different types of "conventional" ice creams. Since our objective focused on nutrition and health in ice cream as an indulgent product, we looked at different types of messages, trying to make sure that we encompassed the range of standard products. We had to resist temptation to include new wave product ideas or ice cream novelty ideas, which kept cropping up in our search for ice cream products. Furthermore, we

Figure 15.1 The five ice cream brands, shown as brand logos.

	"Original" & "Organic"	"Light"	"Reduced Calorie"	"Heart Smart"
	Nutrition Facts	Nutrition Facts	Nutrition Facts	Nutrition Facts
Serving Size	1/2 cup	1/2 cup	1/2 cup	1/2 cup
Servings Per Container	14	14	14	14
Amount Per Serving				
Calories	140	110	65	90
Calories from Fat	65	30	30	15
	% Daily Value*	% Daily Value*	% Daily Value*	% Daily Value*
Total Fat	7 g 11 %	3.5 g 5 %	3 g 5 %	1.5 g 2 %
Saturated Fat	4.5 g 23 %	2 g 10 %	2 g 10 %	1 g 5 %
Trans Fat	0 g	0 g	0 g	0 g
Cholesterol	20 mg 7 %	10 mg 3 %	10 mg 3 %	10 mg 3 %
Sodium	40 mg 2 %	50 mg 2 %	50 mg 2 %	50 mg 2 %
Total Carbohydrate	16 g 5 %	16 g 5 %	6.5 g 2 %	16 g 5 %
Dietary Fiber	0 g 0 %	0 g 0 %	0 g 0 %	0 g 0 %
Sugars	16 g	16 g	6.5 g	16 g
Protein	3 g 6 %	3 g 6 %	3 g 6 %	3 g 6 %
Vitamin A	10%	10%	10%	10%
Calcium	10 %	10 %	10 %	10 %
Vitamin C	0%	0%	0%	0%
Iron	0%	0%	0%	0%
	*Percent Daily Values are based on a 2,000 calorie diet.	*Percent Daily Values are based on a 2,000 calorie diet.	*Percent Daily Values are based on a 2,000 calorie diet.	*Percent Daily Values are based on a 2,000 calorie diet.

Figure 15.2 The different nutritional labels for Study 1, with a 140-Calorie Original.

	"Original" & "Organic"	"Light"	"Reduced Calorie"	"Heart Smart"
	Nutrition Facts	Nutrition Facts	Nutrition Facts	Nutrition Facts
Serving Size	1/2 cup	1/2 cup	1/2 cup	1/2 cup
Servings Per Container	14	14	14	14
Amount Per Serving				
Calories	210	145	95	100
Calories from Fat	125	65	55	15
	% Daily Value*	% Daily Value*	% Daily Value*	% Daily Value*
Total Fat	14 g 22 %	7 g 11 %	6 g 9 %	1.5 g 2 %
Saturated Fat	10 g 50 %	5 g 25 %	4 g 20 %	1 g 5 %
Trans Fat	0 g	0 g	0 g	0 g
Cholesterol	50 mg 17 %	25 mg 8 %	20 mg 7 %	10 mg 3 %
Sodium	50 mg 2 %	50 mg 2 %	50 mg 2 %	50 mg 2 %
Total Carbohydrate	18 g 6 %	18 g 6 %	7 g 2 %	18 g 6 %
Dietary Fiber	0 g 0 %	0 g 0 %	0 g 0 %	0 g 0 %
Sugars	18 g	18 g	7 g	18 g
Protein	3 g 6 %	3 g 6 %	3 g 6 %	3 g 6 %
Vitamin A	10%	10%	10%	10%
Calcium	10 %	10 %	10 %	10 %
Vitamin C	0%	0%	0%	0%
Iron	0%	0%	0%	0%
	*Percent Daily Values are based on a 2,000 calorie diet.	*Percent Daily Values are based on a 2,000 calorie diet.	*Percent Daily Values are based on a 2,000 calorie diet.	*Percent Daily Values are based on a 2,000 calorie diet.

Figure 15.3 The different nutritional labels for Study 2, with a 210-Calorie Original.

included the nutritional facts panel as well. You can see these different types in Figures 15.2 and 15.3.

1. *Original Ice Cream:* The basic ice cream.
2. *Organic Ice Cream:* A "healthful" method of production.
3. *Light Ice Cream:* An example of an FDA-oriented, nutritional content claim. The panel content for nutritional facts was based on the FDA definition of nutrient content claim "Light" (21 CFR 101. 60[b]).
4. *Reduced Calorie Ice Cream:* Another example of an FDA-oriented nutritional content claim. The nutritional facts panel was based on FDA definition of nutrient content claim "Reduced Calorie" (21 CFR 101.60[b]).

5. *Heart Smart Ice Cream:* An example of an FDA-approved health claim. Nutritional facts panel content was based on FDA definition of health claim "plant sterol/stanol esters and risk of coronary heart disease" (21 CFR 101.83).

Silo C: Flavor

Again we resisted the temptation to explore the unusual flavors, concentrating on the more conventional flavors to which most individuals would respond "knowingly," with at least some experience. Ice creams come in both single and mixed flavors, and often have pictures associated with them to make the flavor more "real." We used a picture with each flavor name in order to make a more realistic package design for the test. Our flavors were vanilla, strawberry, and chocolate, the three most popular single flavors, and two of the many possible mixed flavors, strawberry-banana and vanilla-chocolate chip. We see these flavor labels and pictures in Figure 15.4.

Looking at the Total Calorie Content

A recurrent issue in these types of studies is background context against which the decision is made. We wanted to separate out the two levels of calories (140 and 210 calories) so that we could treat the former as reduced fat/calorie and the latter as the more traditional. This consideration led us to divide the ice cream evaluation into two studies, which differed in the nutritional facts panel (i.e., calories, total fat, cholesterol, and total carbohydrate) (Figures 15.2 and 15.3).

Study 1 worked with an "original ice cream of 140 calories per serving" with nutrition facts panel of the ice cream types (Silo B) including organic, light, reduced calorie, and heart smart, changing *appropriately* for a 140-calorie original.

Study 2 worked with an "original ice cream of 210 calories per serving" with the nutrition facts panel of other ice cream type messages (category 2) including organic, light, reduced calorie, and heart smart, again changing *appropriately* for a 210-calorie original.

Running the Ice Cream Study

E-mail Invitation

We used e-mail invitations, with each respondent getting one of the studies. Respondents did not have a chance to participate in both. Instead, any respondents were randomly allocated to the study. The invitations for both studies were identical. The invitation did not mention the specific content of either of the studies and simply included the following phrase: "We would like to find out what consumers like YOU think about different types of ice cream." We see this invitation in Figure 15.5. It's important to keep in mind here, and in the subsequent studies in other chapters, that a well-written invitation substantially increases the likelihood that an individual will accept the invitation. The text in Figure 15.5 can be used for future studies as well. All that one needs to do is change the topic and some of the information.

One important thing to keep in mind is how to offer incentives to get the respondents to participate. When the Internet first appeared on the scene in the mid- to late 1990s, it was an exciting place to visit. People liked to participate. It was fun, novel, and interesting. As a result, it was pretty easy to find respondents. Over time however, the Internet became simply another vehicle by which to connect with the outside world. Many respondents stopped participating so actively. Response rates dropped. What looked like a wonderful world in the late 1990s and early 2000s soon evolved to the typical research situation, with a lot of people rejecting the offer to participate.

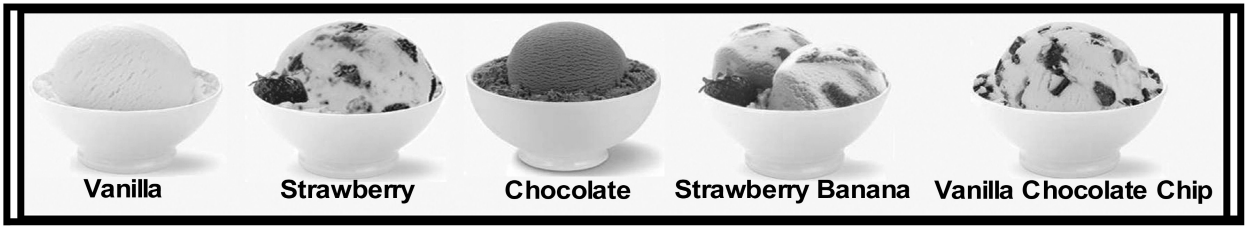


Figure 15.4 The five different flavors and pictures.

i-Novation Inc.
I-Novation, an independent research company, has been asked to find out what consumers like YOU think about different types of ice cream. Your opinions are very important and will help us design the next generation of ice cream products. Here's your chance to tell us what you think! Simply click on the link below (if your email does not support hotlinks, cut and paste the link into your browser) and complete the short, easy-to-answer survey. http://12.109.160.59/NJL734/NJL7344091 Front. asp Depending on your connection speed, the survey should take about 15 minutes to complete. As our way of saying "Thank You" for your input, everyone who completes the survey before 9 PM Eastern Time on Thursday, April 27th, will be entered in a prize drawing featuring a first prize of $100 and a second prize of $50. Because this is a Web-based survey, you will be able to take it when and wherever is most convenient for you, as long as you have access to a Windows-based computer with an Internet connection. Unfortunately, the survey software will not support Mac or Web TV. Please be assured that any information you provide will be held in the strictest confidence. You will not be contacted by an sales or other research organization as a result of your participation in this survey. Thanks in advance for your input, and good luck! The I-Novation team. We respect your privacy. If you feel that you received this message in error, or no longer wish to receive invitations to participate in market research surveys from our compay, please click on the "Unsubscribe" link. I-Novation, 1025 Westchester Avenue, Suite 444, White Plains, N.Y. 10604

Figure 15.5 Text of email invitation for the ice cream studies.

What is the best way to write the invitation? If you look closely at Figure 15.5, you will see a couple of key things. The first is an engaging, respectful opening. In the invitation we give a legitimizing reason about why the study is being done: *I-Novation™, an independent research company, has been asked to find out what consumers like YOU think about different types of ice cream. Your opinions are very important and will help us design the next generation of ice cream products*. The second is WIIFM (what's in it for me). We make the survey attractive by promising a possible reward: *As our way of saying "Thank You" for your input, everyone who completes the survey before 9 PM Eastern time on Thursday, April 27, will be entered in a prize drawing featuring a first prize of $100 and a second prize of $50*. Over the years, the nature of the reward has changed. But one thing has remained certain—it is important to give the respondent some tangible evidence that there is something in this survey for them, even if it is only a chance at a sweepstakes.

Study Welcome Page

When the respondent clicked on the link to participate in the study, the respondent was guided to the study welcome page. Keep in mind that the welcome page does not "tip the research hand." That is, from reading the text in Figure 15.5, one does not know what to expect in terms of the politically correct or appropriate answer for any test package design. All the respondent knows is that there will be a 15-minute interview. The respondent was told that they would evaluate a new set of ice cream package labels, each label on two scales (purchase interest from not interested to interested; description from very indulgent to very healthful).

The Actual Evaluations

Now, let's get into the heart of the study. The respondents each evaluated 35 different combinations, rating the combination on both purchase intent and on a scale of indulgence—healthfulness. The rationale for the two scales is simple. One gets at overall interest but not why. The other on healthfulness gets at the nature of the "type" of product that is being communicated.

We have an example here of two things at play. On the one hand, the experimenter is varying the stimulus. That way, using regression analysis is very straightforward to figure out just which of the different package features drive the response. We have seen that and will see that approach all through the book. On the other hand, we have the respondent doing two things, one after another, namely shifting point of view. The respondent first chooses the appropriate scale point for evaluation, and then chooses the appropriate scale point on another scale for description.

Researchers do this type of multiple ratings quite often. If we were to work with only one picture, then we probably would ask many questions, not just one. After all, from one question and one package design you don't get very much, just a single point, and certainly no pattern. With many questions and one package design you get a little more. You get different angles from which to view the single stimulus. You may find out that the package is perceived as indulgent. Another package might be perceived as less indulgent.

Continuing this train of thought, you can see that with many package designs, preferably varied systematically, and with a number of rating questions, you get a lot more. You get the patterns from the different products from the experimental design. You get the different points of view from the various scales. What one should guard against, however, is the desire to have it all, to ask one respondent many questions about many systematically varied stimuli. After two or three ratings of the same package on different scales, a normal respondent stops paying attention. With two scales we are probably safe; with three, four, five or more scales for that same stimulus, we are okay on the first test stimulus in the interview because the respondent is still excited to participate. By the second or third test stimulus, it becomes quite boring to answer the same three-to-six rating scales, UNLESS the respondent is paid, and thus, highly motivated!

The Respondent Experience

Let's now move quickly into what the research discovered. Keep in mind that the respondent evaluated two scales (purchase interest and indulgent-healthful). Each respondent evaluated 35 combinations, about the right number when we keep in mind that each package gave us the opportunity to perform two ratings studies. On a practical note, it's a lot easier to evaluate 100 combinations of pictures with one attribute than the 50 combinations of pictures with two attributes. Both are far easier than 33 combinations of pictures with three attributes, and so forth. This practical piece of information is important for Internet studies. It suggests that the variation from package to package is less onerous than the evaluation of the same package on multiple attributes.

One other factoid is important to keep in mind because it is the basis of a lot of potential analyses. That is, each respondent saw different combinations of package designs, rather than the same set of combinations. There is a reason for this. Many times the researcher creates a limited set of combinations to test and tests the SAME combinations among many respondents. The replicate or repeated evaluations ensure that we get a good, reliable measure of the mean and top-3 box values. However, what happens when there is a hidden bias in the combinations (e.g., some combinations work unusually well, although we don't know it)? We simply propagate that hidden bias across all of the respondents. So, by giving each person different combinations of the same elements, and with each person getting unique sets of pictures, we minimize this bias.

As always, we followed this set of 35 screens with a self-profiling classification. We ended knowing a lot about each respondent, with information ranging from standard demographics (age, gender), to attitudes to ice cream and health. Finally, the study was only modestly time consuming, taking about an average of 20 minutes to complete.

Looking at Data—The Heart of the Matter and What We Learn about Ice Cream

If you have been following the chapters in this book, you will find that we deal with the data in a number of different ways. Although we use experimental design, the topics we deal with change, the questions change, and what we look at sometimes involves aspects other than the data alone. And so that will be what we do here.

We'll look at topics other than pure ice cream and ice cream labels. These experimental designs are rich with information about people, if we only know where to look. So without ado, let's jump into the data and what we learn about ice cream, about interviews, and about the minds of people.

Who Logs in and How Many Complete?

If a topic is interesting to respondents, then we expect to see a high number of log-ins and a high completion rate. Study 1 had 278 total log-ins and 199 completes. This gave a completion rate of 72%. Study 2 had 301 total log-ins and 205 completes. This gave a completion rate of 68%, and so we can say that 70% of the respondents who start the study complete it. Ice cream is clearly a popular topic. For both studies 1 and 2, once a respondent agreed to participate in the topic of the study (ice cream), the content of the elements held interest for the majority of respondents. Let us put this into perspective with some other data. Andrea Maier and her colleagues reported completion rates of 50% (fruit smoothie), 46% (flavored water), 55% (yogurt beverage), and 47% (flavored tequila) for similar types of conjoint analysis studies (Maier, Moskowitz, and Ashman, 2008; Rabino et al., 2007).

In these studies typically more women than men participate. Females comprised 74% of respondents in the first study (reduced calorie) and 81% of respondents in the second study (full calorie). Let's put this into perspective, however, because we're going to see that the ratio of women to men in the study changes, depending upon the food. Look at Figure 15.6 to get a sense of the proportion of women to men in different types of studies. These were run in the It! studies (Beckley and Moskowitz, 2002).

Food	% Female
Chocolate Candy	86%
Cheese cake	84%
French Fries	81%
Tortilla Chips	80%
Cinnamon Rolls	80%
Taco	79%
Potato Chips	76%
Olives	76%
Gravy	76%
Pretzels	75%
IceCream	74%
Cola	74%
Cheese	73%
Peanut Butter	69%
Coffee	69%
Nuts	67%
Pizza	67%
BBQ Ribs	62%
Chicken	62%
Hamburger	60%
Steak	56%

Figure 15.6 Percent female respondents for concept studies, using IdeaMap.net®. Data courtesy of It! Ventures, and taken from the 2002 Crave It!® database.

What Do We Learn about Interest in the Ice Cream?

Our analysis followed the modeling approach discussed in previous chapters on modeling the results. We focused here on measures of interest, or membership in the group of respondents interested in the ice cream, based on what they saw. We used the binary transformation, which we discussed throughout this book. Thus, ratings of 7–9 for purchase get transformed to 100, and ratings of 1–6 for purchase get transformed to 0.

At the individual level, we begin with 15 independent variables corresponding to the graphical 15 elements (recall 3 silos, 5 elements per silo), and 35 cases corresponding to the 35 test concepts or graphics combinations. For each of our respondents, the ordinary least-squares regression analysis creates an individual model or equation showing the contribution of each ice cream element to the binary rating of not interested or interested (Fox, 1997; SYSTAT, 2004). We write the straightforward equation as the sum of the additive constant (k_0) and the part-worth combinations of the 15 elements, Element 1 to Element 15.

$$\text{Binary Rating} = k_0 + k_1(\text{Element \#1}) + \\ k_2(\text{Element \#2}) \cdots k_{15}(\text{Element \#15})$$

We can see the parameters of this model in Table 15.1. We will spend the rest of the chapter discussing these results, with some more detailed analyses as well. Let's move into the data now.

Table 15.1 Base size, additive constant, and utility values for percent top-3 box purchase interest (1–6→0; 7–9→100), and for indulgent (– numbers) versus healthful (+ numbers) for ice cream positioned as having 140 calories or 210 calories

	Purchase	Purchase	Indulgent versus Health	Indulgent versus Health
	140 Calories	210 Calories	140 Calories	210 Calories
Base Size	199	205		
Additive Constant (k_0)	25	22	44	45
Silo A: Brand				
Häagen-Dazs®	4	2	0	0
Ben & Jerry's	3	3	0	0
Breyers®	3	2	0	0
Dreyer's	2	1	0	0
Stonyfield Farm®	1	0	2	0
Silo B: Line (type of product)				
Reduced-calorie ice cream and nutritional facts panel	6	12	18	17
Heart smart ice cream with nutritional facts panel	4	16	20	23
Ice cream with nutritional facts panel	–1	–4	–2	–8
Light ice cream with nutritional facts panel	–3	0	12	13
Organic ice cream with nutritional facts panel	–11	–12	3	–3
Silo C: Flavor				
Strawberry flavor	6	5	2	–1
Vanilla chocolate chip	6	8	0	–1
Strawberry banana flavor	5	6	0	–1
Chocolate flavor	4	4	1	–2
Vanilla flavor	3	0	2	–1

Baseline Results or the Additive Constant, k_0, for Ice Cream

The additive constant (k_0) tells us the estimated probability that our concept about ice cream (in truth, our package design) will be interesting (i.e., get rated 7, 8, or 9 on the 9-point scale), if no elements are present. All concepts in the study contained elements so that there was no concept comprising zero elements. Our additive constant is simply a construct, an estimated parameter. Nonetheless, despite being an estimated parameter rather than being a real number, the constant tells us a great deal, because it shows the respondent's predisposition to be interested in buying the ice cream.

Previous experimental designs with package and with communications, using this specific type of experimental design (main effects, with categories absent) generate very powerful data, usable in a database, and comparable values of the constant (Moskowitz, Porretta, and Silcher, 2005b). Let's look at some rules of thumb that appear from project to project. They will give us the structure within which we can interpret out data:

1. Constants above 50 represent a high degree of basic interest. The respondent really likes the packages, in general, and is predisposed to buy.
2. Constants lower than 30 represent a low degree of basic interest so that, on the average, the respondent doesn't like what he sees.
3. Constants around 20 or lower represent a very low degree of basic interest, with the product or package almost a commodity. A good example of this low constant is the value for credit cards, which makes sense. If you tell someone about credit cards, you may or may not get much interest. Credit cards are common, a commodity.

Well, when all is said and done, ice cream only scores a 25 on the purchase scale for those ice creams positioned as having 140 calories, and scores a 22 for those ice creams positioned as having 210 calories. We know, however, that people like ice cream, so we have to ask ourselves the simple question: Why does the additive constant for package designs score so low? Is it because we have the wrong group of respondents? Is it because we have selected poor graphics to show? The wrong group of respondents would drop the constant because

Item	Con	Item	Con	Item	Con
BBQRibs	44	Ice Cream	39	Pizza	29
Steak	41	Peanut Butter	38	Pretzel	29
Potato Chips	41	Cinnamon Rolls	38	Tortilla Chips	29
Chicken	40	Cola	37		
Coffee	40	Cheese	36		
Chocolate Candy	40	Nuts	36		
Cheesecake	40	Olives	36		
		Tacos	35		
		FrenchFries	35		
		Gravy	33		
		Hamburger	31		

Figure 15.7 Additive constant (con) for different food studies when the test stimuli are word concepts. Data courtesy of It! Ventures®, and taken from the 2002 Crave It!® database

they might not like ice cream. As Figure 15.7 shows, other concept studies with words and pictures, conveying ideas about products generate higher values for the additive constant (Beckley and Moskowitz, 2002). Indeed, for ice cream, the additive constant is a much-higher value, 39.

A more likely explanation, that will hold throughout this book is that ***graphics design may generate a lower constant because we are dealing with pictures not with ideas***. Pictures have to be more concrete and more subject to "immediacy," whereas ideas can be supported with less information and, perhaps, leave more to the imagination. This has been shown in studies ranging from fragrance to teas (Moskowitz, et al., 2005a). We will be able to follow the results of a relatively large number of studies to check this hypothesis that the additive constant is lower.

What Design Features Drive Purchase Interest?

The coefficients for our 15 elements in three silos tell it all. We see these coefficients in Figure 15.5, with the first column of data pertaining to the impact or contributions of the elements when the ice cream is positioned at the low 140-calorie level, and the second column of data pertaining to the impact when the ice cream is positioned at the high 210-calorie level.

When it comes time to learn about what elements drive the responses, remember that the statistical design enables the researcher, and indeed anyone using the table, to compare across the elements, across the three categories, and across the two product positions. That is, the data are completely comparable, so you can draw conclusions about how well a brand performs in two different positioning conditions, as well as how much negative impact a brand name can counteract!

The coefficients show the contribution of the element to purchase interest (the response) and can be positive or negative. Positive coefficients indicate that when the element is present in the concept, the probability that a person will be interested in the concept increases. For example, when a coefficient is +10 it means that an additional 10% of the respondents say that they will buy the product. Remember to keep these rules of thumb in mind for the data:

1. Utility above 15 corresponds to extremely impactful and important elements.
2. Utility between 10 and 15 corresponds to very impactful elements.
3. Utility between 5 and 10 corresponds to impactful elements.
4. Utility between 0 and 5 means that the element adds little to interest.
5. Utility below 0 means that the element detracts from interest and should be avoided.

Armed with this information, let's now see how the three categories of five elements perform, and whether the number of calories promised for the ice cream can affect what wins and what loses.

With these types of data, one might circle around the data, without extracting anything, trying to draw a single coherent "picture." This might, but often does not, lead to success. Another, generally more productive way, is to jot down observations of what is happening, and from these observations synthesize what might be occurring. This second way is more in the manner of trying to get an "impression of the data," a sense of what nature might be trying to communicate. It's not as formal, not particularly structured, but often very productive. Let's try this second or enumeration strategy. We list some of the observations below, after which we'll try to connect the dots and see what emerges.

1. The brands don't do much. They all have utilities around 0–4, which means that when a brand name (and logo) come into the package concept, we can't expect more than about 4% of the respondents to shift their vote from 'would not buy' to 'buy'. However, it's important that no brands are negative. Ice cream brands don't do much good, but they certainly don't do much bad either! The reader should note that although this sounds somewhat "down," it really is important learning. We often think of brands as carrying the product. The brand name doesn't, even when the logo is clearly on the package.
2. We see a little more of a range when we talk about the specific type of product (i.e., the line). Let's first look at the results when we position the ice cream as having 140 calories. We see a +6 when the package says specifically "reduced calorie ice cream" and "features a nutritional facts panel." Now let's jump to the 210-calorie product. Calling a 210-calorie product "reduced calorie ice cream" convinces an additional 12% of the respondents to say that they'd buy the ice cream, rather than the 6% we had when we had the positioning of 140 calories. Does this mean that the consumers are more responsive to claims of general calorie reduction when they deal with a higher calorie product?
3. The line can really make a difference when we talk about specific health benefits. Talking about "heart smart ice cream" does a lot more with 210 calories than with 140 calories, again suggesting that when the calories are high, health messages may do far better than when the product is perceived to be more healthful in general. One can only speculate about the effect of such health statements in the super-premium category.
4. Flavors are, of course, important, but these are popular flavors. Although vanilla is perhaps the world's most popular flavor, it is the weakest performing, so perhaps vanilla works its magic by being reasonably acceptable to everyone. We can't say anything about the strength of simple versus mixed flavors; they seem to do equally well. There is a very slight hint, but only a hint, that the flavors will do slightly better in the higher-calorie ice cream.
5. Standing back from these results, we see that the simple experimental design of packages with health information, and with positioning, generates significant amounts of information, even for a simple, overall evaluative attribute "purchase intent."

What Design Features Communicate Healthful, and What Communicate Indulgence?

Recall that the respondents evaluated each of the 35 packages on another rating scale: *How would you*

DESCRIBE this ice cream? 1 = Very Indulgent … 9 = Very Healthful.

We can't very well use the simple logic of "purchase/non purchase" for this second rating scale, where we just look at the top of the scale. It doesn't make sense. We move now into a new area, the "communication of the concept." We're going to need the whole scale. We've just turned our interest to what the package design communicates—how much indulgence or how much healthfulness? Note that the rating scale of "healthful" versus "indulgent" does not ask for good or bad.

This distinction is very important. We are interested as to *where* on the scale the package falls, not just whether the package is "healthful" or "indulgent." In communication research, we're interested in nuances. We might, for example, want to "move the package" to communicate just a little bit more health or a lot more indulgence. We'll need to use the scale itself, not the binary equivalent.

However, we also want to stay with the 0–100 scale, perhaps because we all grew up with it and found it is easy to use. We can have our cake and eat it. Let's simply transform the data before analysis, in a simple linear way. The transformation doesn't reduce any information; rather it just expands the scale so that the magnitude of the difference remains the same.

The transformation is quite simple:

1 transforms to 0
2 transforms to 12.5
3 transforms to 25
4 transforms to 27.5
5 transforms to 50
6 transforms to 62.5
7 transforms to 75
8 transforms to 87.5
9 transforms to 100

This is called an affine transformation by mathematicians (Zwillinger, 1995). You don't have to know the technical details. But you should keep in mind that the affine transformation preserves all of the original information that we had acquired using the 9-point scale, but merely stretches the scale, and changes the origin. It's clear from these transformation values that the original 1-point difference on our 9-point health-indulgence scale transforms to a 12.5-point difference, and that the original origin of the scale at point 1 is not changed to an origin at point 0. We do similar types of transformations every day when we convert from Fahrenheit temperature to Celsius temperature.

Armed with this information, let's now look at the results of the ice cream study, this time concentrating on the communications. Low additive constants (below 50) mean indulgence; high constants mean health. Negative utilities push the package toward communicating indulgence, and positive utilities push the package toward communicating health. There may be some overt health messages or indulgence messages. If the research method works, those "blunt statements" ought to come out as very strong communication drivers. Keep in mind that the communication may be subtle, not direct, and is always in the mind of the respondent.

Let's follow the same analytic strategy as we did before. That is, instead of thinking of large-scale, overriding patterns, let's build up our insight by finding specific, noteworthy points that we should remark on. Refer back to Table 15.1, and look at the two data columns on the right, which show the results for the indulgent versus healthful scale. Remember that negative numbers move the impression of the ice cream toward indulgent, whereas positive numbers move the impression of the ice cream toward healthful.

1. The additive constants are 44 for 140 calories, 45 for 210 calories—both below 50. The fact that the constants are low is not surprising. After all, we're dealing here with ice cream, and additive constants below 50 mean indulgence.
2. The five brands neither communicate indulgence nor health. They're all around 0.
3. The line clearly communicates and, in fact, does all the work. It's clear that respondents pick up what the label is saying. For example, when the label is "heart smart," the communication jumps very far over to healthful.
4. Calories can modify the impact of communication. The phrase "heart smart" is more powerful with 210 calories than with 140 calories. Ice cream itself is more indulgent at 210 calories compared with 140 calories.
5. Flavors don't drive communication. For instance, for 140 calories, utilities are between 0 and 2, while flavors in the context of 210 calories are negative and are therefore a little more toward the indulgent side.
6. Overall then, experimental designs of packages can generate differences in how they are perceived, not

just how they are liked. Bipolar rating scales, allowing the respondent to show how the product communicates, do well in picking up these differences.

Two Different Mind-Sets for Ice Cream

People differ from each other. Let's divide them by the pattern of the 15 utilities that we created. Recall that the experimental design we selected allows us to create an individual model for each person for purchase intent. Originally we were just interested in whether a person would buy or not buy the ice cream so that we transformed the rating to a binary value (1–6 transforms to 0; 7–9 transforms to 100).

We don't have to do that, however. We can do the same analysis as we did for communication of healthful versus indulgent, where we keep the numerical or metric information. It's a minor issue here. All we have to know is that we can create a model using the original ratings of purchase intent, not the transformed data that took it to a 0 or 100. We ran the regression model on this original data and got a set of utility values, called persuasion values, which we will use to cluster the respondents.

To clarify what we just said, let's go back to our three analyses:

Analysis 1 (Interest): What drives purchase interest? We transform the ratings to 0 or 100, thus putting people into the category of would buy (100), or would not buy (0). We find the key drivers. We've been concentrating on this. This is the Interest analysis.

Analysis 2 (Communication): What drives indulgent versus healthful? We simply transform the 1–9 to 0–100, but keep 9 rating points. This is the communication analysis.

Analysis 3 (Persuasion, used only for clustering or segmentation as we describe now): Putting people into different groups or clusters or segments, based on the patterns of their utilities. We simply transform the 1–9 to 0–100, but keep 9 rating points. We run the purchase intent model. We keep the fine-grained results, because we have the metric information about purchase intent.

What Did We Find When We Divided Our Respondents?

We ran our segmentation twice, once for the data from Study 1 on 140 calories and once for the data from Study 2 on 210 calories. We could have run one large segmentation study; it's a matter of preference. However, if the two segmentations come up with the same set of segments, then we can be sure that we're onto a strong way to divide the people.

Our analysis suggests two key mind-sets in the population, both when the ice cream was positioned as having 140 and 210 calories, respectively. Let's look at them separately, to see whether we can find a story that makes sense to us.

Segment 1: Health Seekers

What happens with 140-calorie products? These are 51% of our respondents in the lower-calorie study. Segment 1, health seekers (our term), are most interested in ice cream lines (elements in Silo B), including reduced-calorie ice cream, heart smart ice cream, and light ice cream (all with nutritional facts panels). Health seekers were least interested in flavor (category 3) including strawberry flavor, strawberry banana flavor, and chocolate flavor.

What happens with 210-calorie products? First, we find almost the same proportion (54%). Again, these health seekers are most interested in ice cream lines and least interested in flavor (category 3) including strawberry banana flavor, strawberry flavor, and vanilla flavor.

Segment 2: Flavor Seekers

What happens with 140-calorie products? Flavor seekers were clearly most interested in Silo 3, specifically strawberry flavor and vanilla chocolate chip flavor. Flavor seekers were uninterested in ice cream lines (Silo B) including the following health messages: organic ice cream, light ice cream, heart smart ice cream, and reduced-calorie ice cream.

Finally, what happens with 210-calorie products? Flavor seekers were most interested in Silo C (flavors), specifically strawberry banana flavor, strawberry flavor, and vanilla chocolate chip flavor. Respondents in this group were again uninterested in ice cream lines (Silo B).

The bottom line is that the same segmentation occurs, and the same winning and losing elements reappear. This gives us confidence that the segmentation is real and independent of the basic calorie level of the ice cream!

An Overview

It's clear from this chapter that the experimental design of packages worked for ice cream, that it differentiated between messages, and that it is the textual messages with "content" that made a difference. For many years, it's not been clear whether the information on a package is what drives the respondent, or whether it is the logo and brand. We see here that information is key, at least in these types of studies.

The segmentation results are also worth noting and should not be surprising. What is surprising is the division of the population into two equal groups, independent of the calorie level of the ice cream, and the fact that almost the same exact elements do well or poorly in each.

References

Aaron, J.I., Mela, D.J., and Evans, R.E. (1994) The influence of attitudes, beliefs and label information on perception of reduced-fat spread. Appetite 22, 25–37.

Bech-Larsen, T. and Grunert, K.G. (2003) The perceived healthiness of functional foods: A conjoint study of Danish, Finnish and American consumers' perception of functional foods. Appetite 40, 9–15.

Beckley, J. and Moskowitz, H. (2002) Databasing the consumer mind: The Crave It!™ Drink It!™ Buy It!™ and Healthy You!™ Databases. Institute of Food Technologists, Anaheim, CA.

Bower, J.A., Saadat, M.A., and Whitten, C. (2003) Effect of liking, information and consumer characteristics on purchase intention and willingness to pay more for a fat spread with a proven health benefit. Food Quality & Preference 14, 65–74.

Fox, J. (1997) Applied Regression Analysis, Linear Models, and Related Methods. Sage Publications, Inc., Thousand Oaks, CA.

Kähkönen, P., Tuorila H., and Lawless, H. (1997) Lack of effect of taste and nutrition claims on sensory and hedonic responses to a fat-free yogurt. Food Quality & Preference, 8, 125–130.

Maier, A., Moskowitz, H., and Ashman, H. (2008) Comparing across categories: Using mind-set information about health, pleasure, and function to understand consumer choices in bottled water. Bentham Open Food Science Journal 2(2008), 131–150.

Moskowitz, H., Gofman, A., Madhu, M., Zhenyu, M., and Katz, R. (2004) Dynamic Package Design & Optimization in the Internet Era. ESOMAR 2004 Technovate Conference, Barcelona, Spain.

Moskowitz, H.R., Gupton, A., and Beckley, J.H. (2005a) Capturing the algebra of the customer's mind for functional fragrance through pictures, text and decompositional research analyses. In: Proceedings of the European Society for Opinion and Marketing Research (ESOMAR) Conference—Fragrance Research: Unlocking the sensory experience, New York, NY.

Moskowitz, H.R., Porretta, S., and Silcher, M. (2005b) Concept Research in Food Product Design and Development. ISBN: 0813824249. Blackwell Publishering, Ames, IA.

Prescott, J. and Young, A. (2002) Does information about MSG (monosodium glutamate) content influence consumer ratings of soups with and without added MSG? Appetite 39, 25–33.

Rabino, S., Moskowitz, H., Katz, R., Maier, A., Paulus, K., Aarts, P., Beckley, J., and Ashan, H.(2007) Creating databases from cross-national comparisons of food mind-sets. Journal of Sensory Studies 22(5), 550–586.

SYSTAT. (2004) SYSTAT for Windows, Version 11. SYSTAT Software, Inc., Chicago, IL.

Tepper, B. and Trail, A. (1998) Taste or health: a study on consumer acceptance of corn chips. Food Quality & Preference 9, 267–272.

U.S. Food and Drug Administration. (2003) A food labeling guide. (Available at www.fda.gov) [Accessed 12/02/08]

Vickers, Z.M. (1993) Incorporating tasting into a conjoint analysis of taste, health claim, price and brand for purchasing strawberry yoghurt. Journal of Sensory Studies 8, 341–352.

Westcombe, A. and Wardle, J. (1997) Information of relative fat content information on responses to three foods. Appetite 28, 49–62.

Zwillinger, D. (Ed.) (1995) Affine Transformations. §4.3.2 in CRC Standard Mathematical Tables and Formulae. CRC Press, Boca Raton, FL: 265–266.

Chapter 16

Healthy Pasta: Nutritional Labeling and the Role of Messages

Introduction

Walk down any supermarket aisle, and you're likely to see people picking up packages, turning them around, inspecting the nutrition label and, perhaps, reading. Sometimes they smile, sometimes they frown in disappointment, occasionally they look a bit confused, and all too often they look like they just don't care.

Over the past decades, we have been inundated with health messages. We can't escape these messages, in papers, on television, on the web, on signs in elevators, everywhere. Most people in the more affluent countries are somewhat aware of nutrition issues, although the increase in obesity would suggest that despite nutrition education there's a lot less willpower than knowledge out there. During the past 20 years, there has been a dramatic increase in obesity in the U.S., and in 2005–2006 more than one-third of U.S. adults (over 72 million people) were classified as obese. The U.S. Centers for Disease Control and Prevention reported through their behavioral risk factor surveillance system that in 1990, among participating states, 10 states had an incidence of obesity less than 10% and no states had an incidence greater than or equal to 15%. Times change, however, and what was true 10 years and 20 years ago is no longer the case today. In 2006, only four states had a prevalence of obesity less than 20%, 22 states had a prevalence equal or greater than 25%; two of these states had a prevalence of obesity equal to or greater than 30% (http://www.cdc.gov/nccdphp/dnpa/obesity/trend/). Such high rates of obesity are alarming, and the prevention of obesity has become a top public health priority (US Centers for Disease Control, Obesity Trends, 2006).

The nutritional facts panel (NFP) was designed to meet the objective of the Nutritional Labeling and Education Act of 1990. The objective was to provide information on almost all packaged foods manufactured after May 1994, in order to help consumers make informed food choices (http://www.cfsan.fda.gov/label.html (Nutritional Facts Panel, 2008)). Not surprisingly, the NFP has been studied in many areas ranging from nutrition and dietetics to marketing to public policy (Satia et al., 2005; Goldberg and Probart, 1999; Garretson and Burton, 2000; Jones and Richardson, 2007).

Some trends have been uncovered based on the NFP. For example, Neuhouser et al. (1999) reported that nutrition label use was significantly higher among women and was associated with lower fat intake. When controlled for demographics, label use was associated with belief in a low-fat diet, belief in the relationship between diet and disease, and reporting oneself as being in the maintenance stage of change for adopting a low-fat diet.

The simple questions we have here follow: What elements drive the perception of "healthful" and "purchase intent", and are they the same? Does the nutrition label more than anything drive both attributes? So, in the spirit of health and wellness, we decided to investigate some labeling "effects" of healthy (or healthful!). We chose pasta.

Why pasta? What were we thinking about when we chose this as our next topic?

For years, pasta has been thought of as an American staple. A Thursday night meal, gathering the family around the table. Easy to make, a real comfort food. Dinner everyone likes and shows up for. However, and to be frank, pasta was never really known for its health merits. Rather, pasta was, in most people's minds, more like empty carbohydrates, but with a great taste and often a heritage to go along with that taste.

Times change. Eating habits have transformed in this new generation. One of the authors (MR) remembers her daughter coming home from college one summer, going out to buy a box of pasta for dinner, and coming homing with a newly introduced whole wheat pasta. Not the regular white/yellow pasta, but pasta that boldly talked health, wellness, good-for-you, with a package that

Packaging Research in Food Product Design and Development Howard R. Moskowitz, Michele Reisner, John Ben Lawlor and Rosires Deliza
 ISBN: 978-0-813-81222-9

wouldn't quit. It was, to use two colloquial phrases "good-for-you … in your face." And so the inevitable question: "What is this and did you get any of the real stuff for me?" The answer was simple. It turned out that she and her friends weren't eating any of the traditional pastas that the authors came to know so well growing up. She, and indeed the rest of us, had slowly changed eating to healthful eating. Everyone seemed to be eating better. And yes, pasta had entered the territory along with everything else on the supermarket shelf. Next stop, Omega 3. Was there any stopping this darned thing!

In order to stay competitive in today's merciless market trends, product developers need to be ahead of the next trend. Hence, our chapter … Healthy pasta: Nutritional Labeling and the Role of Messages. We thought of a catchier phrase … Healthy Carbs, Happy People, but the phrase just didn't have enough gravitas.

The products in the pasta food segment have increased exponentially, propelled by diet and health guidelines from the medical community, health professionals, and the government's Federal guidelines/food pyramid. For over a decade, we have seen the packaging and claims of benefits for various medical conditions (such as the ability to lower cholesterol, prevent clogged arteries, prevent blood pressure problems, lower glycemic levels, prevent colon cancer, improve eye health) appear ever more frequently on food packages in the supermarket. Barilla, Ronzoni, and others have introduced new healthy grains lines. Even the Manischewitz kosher food company now makes egg, yolk-free, and whole grain noodles. The cereal food segment has gone in the same direction (i.e., Quaker Oats® Weight Control oatmeal and snack bars).

These evolutions in product and packaging have had an impact on all those who are the primary purchasers of products for their home, for their own use as well as for family use. In one sense, it has become easier to find what we want on the shelves. Paradoxically, as more products are developed and marketed for these categories, it has become even more confusing for the shopper to decide what is best. Manufacturers continually update the packaging and look of their existing products as well as rolling out the new healthy ones, all to gain share and cash in on the good-for-you trend. This fertile competition creates a quandary for the consumer who is used to reaching for a particular package by force or habit. When the consumer gets home and unpacks the groceries, he or she may find, dismayingly, that they have actually purchased a reformulated or repositioned product that does not have the attributes that motivated the original decision to purchase.

Thinking about What Might Be Important

With the discovery by marketers in the food world that many purchases in the supermarket are not planned, we wanted to see whether we could get people interested by the proper health label. We deal with health and wellness in other chapters of this book, but the field is so vast and the different aspects so beguiling that even a dozen experiments or more can't provide enough knowledge that manufacturers need. But enough of that jeremiad. It's now onto looking at some of the health messaging on packages to see what makes a difference.

We wanted to discover whether or not consumers merely glance at the front of the box to see a burst that says whole wheat, or enriched with 10 essential vitamins, and go on their merry way, or do they take it to the next step? Do they actually look at the nutritional panels and pay attention to what the panels say? And even more interesting to us are two other questions:

1. The all-powerful nutritional label: Can the nutritional label affect how a person reacts to other information on the box?
2. Different drummers: Are there different mind-sets among people—those who say that if pasta is healthful then they will buy it, and those who recognize that the pasta may be healthful, but such healthful properties have little or no influence on the decision to buy?

We designed our own boxes of pasta, and had fun with the nutritional information displays. Did it matter where we placed them? We wanted to know so we put them on the front of the box, on the side panels, and then where you typically find them, on the back of the box. Next, we varied what the panels said about the ingredients. We increased the calories and fat, modified the sodium and a few other items to measure if consumers are paying attention and how this affects their purchase intent. We see the two different nutrition labels (regular versus higher calories) in Figure 16.1.

What Should the Consumer Rate?—Interest versus Health

In this chapter, we focus specifically on two attributes, interest versus the perception of healthful/not healthful.

The reason for this focus is that we will find an interesting dichotomy among respondents. Some respondents will show highly correlated ratings. If they rate a product "high" on healthfulness, then they will also rate the same product high on "interest." It's not that these attributes are correlated but rather that they are virtually redundant. And, then there are other respondents who differentiate between what they find to be healthful and what they like.

Regular Calories

Nutrition Facts	
Serving Size 1 1/2 cups (56 g/2 oz) dry Servings Per Container 14	
Amount Per Serving	
Calories 140	Calories from Fat 10
	% Daily Value*
Total Fat 1 g	2%
Saturated Fat 0 g	0%
Trans Fat 0 g	
Cholesterol 10 mg	4%
Sodium 20 mg	1%
Total Carbohydrate 24 g	8%
Dietary Fiber 6 g	20%
Sugars 0 g	
Protein 12 g	

Higher Calories

Nutrition Facts	
Serving Size 1 1/2 cups (56 g/2 oz) dry Servings Per Container 5	
Amount Per Serving	
Calories 250	Calories from Fat 50
	% Daily Value*
Total Fat 3 g	5%
Saturated Fat 0 g	0%
Trans Fat 0 g	
Cholesterol 10 mg	4%
Sodium 20 mg	1%
Total Carbohydrate 24 g	8%
Dietary Fiber 6 g	20%
Sugars 0 g	
Protein 12 g	

Figure 16.1 Two nutrition labels, one for regular versus one for higher calories

With this short introduction in mind, let's now visit the study. We begin with the elements that we varied. Look at Figure 16.2, which shows a picture of a stimulus, and the five different silos. Four of the silos each comprised six options. These four silos were package (i.e., specific type of nutritional label), banner (i.e., new size), health message (i.e., provides a full day's supply of 12 essential vitamins), and family message (i.e., know you're serving a family favorite made with fresh, quality ingredients). There is a fifth silo, the background image, always present in every stimulus so that when any silo was missing, the background image was still present so that there did not seem to be any "hole" in the package.

Of course, our primary focus was the nutrition label, and, specifically, on the reaction of consumers to two variables within this nutrition label:

1. Information contained in the label regarding calories (two options, regular and higher calories)
2. Location of the label (three options: front of the package, side of the package, back of the package)

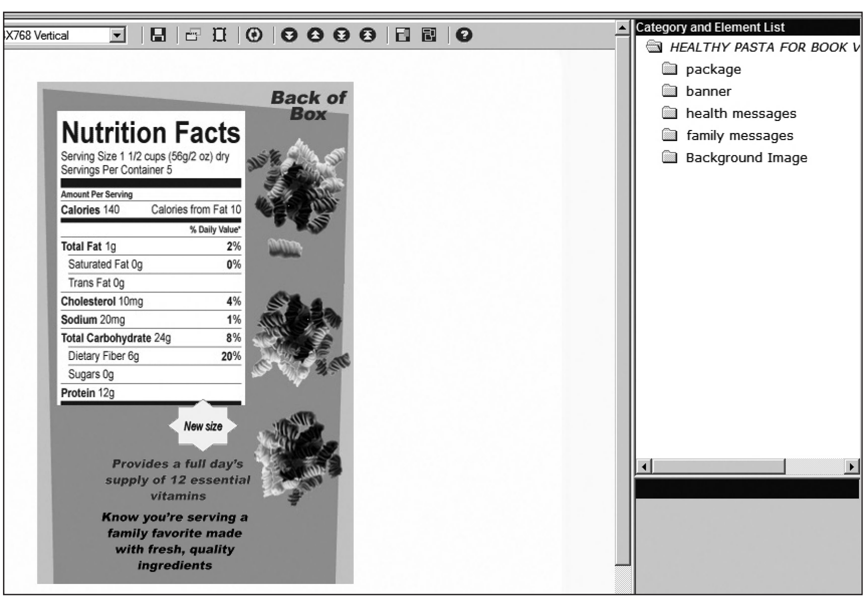


Figure 16.2 A picture of a stimulus and the five different silos

One of the issues that immediately confronted us was just how to represent the front and the back of the package. Pasta packages don't naturally lend themselves to dramatic differences in front versus back. We settled on the simple, perhaps overly expedient solution, of listing the front versus back at the top right of the package. It was easier with the side panel. We were able to draw the package in perspective, but also to maintain the parallel structure, we wrote the word "side" at the top left. (See Figure 16.3.)

Running the Study

We set up the study to generate 54 pasta packages per person, with each person rating the packages on two attributes. The first rating scale was "healthful," and the second was "purchase interest." As interviews go, the interview was a bit longer than typical ones, taking about 20 minutes, rather than 15 minutes. However, these package studies go relatively quickly and, indeed, forcing the respondent to go through the 54 packages ensured that the respondent would probably give a "gut reaction" rather than intellectualizing the entire experience. As we will see shortly, the respondents were typically "dead-on," despite what appears to be an extended interview. That is, the respondents will be seen to have clearly picked up the calorie differences and responded to them appropriately, even though those differences were not even hinted at!

Respondents generated the following results for the total panel. The results come from combining the 54 ratings from 152 respondents into a large data set of 8,208 rows. When we build a single model for "top-3 box healthfulness" or "top-3 box purchase" (percent scoring the package as 7–9 on the 9-point scale), we get the results shown in Figure 16.4. The results are pretty clear and striking:

1. *Similar additive constants.* The additive constants are similar (41 for health, 35 for purchase intent). This means about 35–40% of the respondents would rate the packages 7–9 on the scales (i.e., healthful, probably would buy).
2. *The big action occurs in the nutrition labels.* The strong positives are for the current calorie level. The modest negatives are for the increased calorie level. What's important here is that the positives are a lot stronger than the negatives. Furthermore, there's no

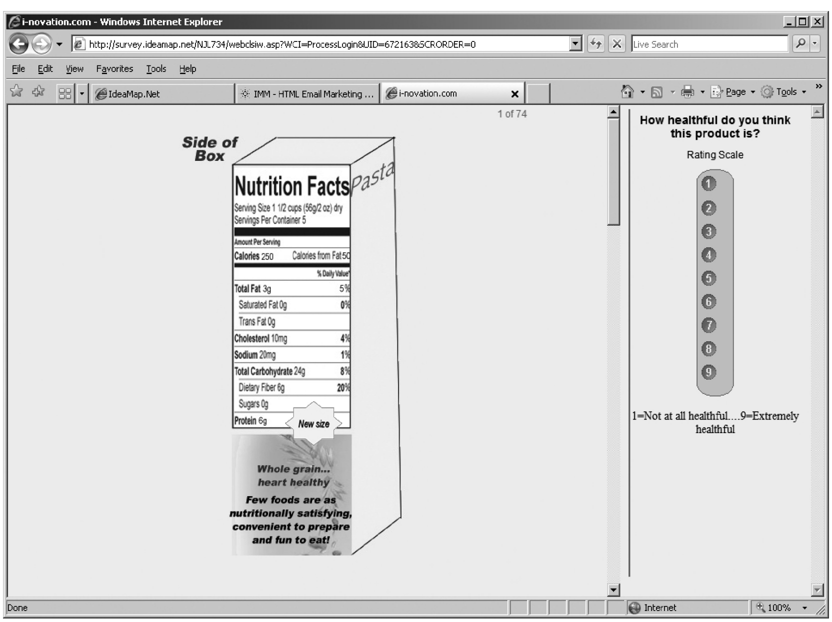


Figure 16.3 Screen shot of a test package, showing the pasta product with the information presented on the side of the package. In addition to the perspective, the word "side" is shown at the top left of the package, to communicate that the package is being shown from the side.

		Health-ful	Purchase
	Additive constant	41	35
	Label location, calorie level		
A1	Nutrition label on front, regular calorie level	21	15
A3	Nutrition label on back, regular calorie level	21	17
A5	Nutrition label on side, regular calorie level	21	15
A4	Nutrition label on back, higher calorie level	–5	–5
A6	Nutrition label on side, higher calorie level	–5	–6
A2	Nutrition label on front, higher calorie level	–6	–8
	Splash on package		
B5	Splash—all natural	3	3
B6	Splash—organic	1	1
B4	Splash—new	1	2
B2	Splash—recipe in box	0	1
B3	Splash—made with farm fresh eggs	0	1
B1	Splash—new size	–1	0
	Health message		
C3	75% less carbs than traditional pasta	6	4
C6	Whole grain. . .heart healthy	6	4
C1	Provides a full day's supply of 12 essential vitamins	5	4
C4	Low in calories and fat	4	4
C2	100% gluten free	3	–1
C5	Naturally low in sugar	2	1
	Family message		
D4	Cholesterol free…ideal for a healthy lifestyle	3	3
D3	A perfect match with your favorite sauce for a wonderful Italian meal	2	2
D1	Know you're serving a family favorite made with fresh, quality ingredients	1	1
D2	Few foods are as nutritionally satisfying, convenient to prepare and fun to eat!	1	3
D5	Feed your family the best	0	1
D6	Tastes like homemade	–1	–1

Figure 16.4 Performance of the elements by total panel on two attributes, healthful and purchase intent. The numbers reflect the percent of respondents who would rate each package element 7–9 on a 9-point scale. The elements are sorted by "healthful."

real difference in response produced by where we put the label; the results are the same whether the label is on the front, on the side, or on the back. Of course, it's hard to know the difference between front and back in these stimuli except for the label, but it is clear when the package stimulus represents the side of the box.

3. *Healthfulness slightly different from purchase intent.* Ratings for purchase intent generate lower utilities than do ratings for healthfulness. It's a lot easier for a person to "up-rate" perceived healthfulness than to decide to buy the product. Ratings of "purchase intent" are "stickier upward." They get stuck at lower levels on the scale. People tend to be more conservative rating purchase intent than rating a descriptor.
4. *Health messaging doesn't do much.* The health messages are also almost irrelevant, except perhaps for two of them, and even these two do only modestly well (+6 for health, +4 for purchase intent):
   - 75% less carbs than traditional pasta
   - Whole grain … heart healthy
5. *Family messaging is also irrelevant.* Just because you tell people something is good for their family, it doesn't necessarily translate to healthy.
6. *Label, label, label.* The bottom line here is that it's the label and, principally, the regular-calorie label that drives the consumer response.

Why Do Healthfulness and Purchase Intent Covary So Much?

There is always the question as to whether or not respondents can actually rate a concept on two separate attributes. That is, can respondents quickly shift their attention from rating "healthfulness" to rating "purchase intent," or do the two attributes influence each other? When we look at the 8,208 "cases," one per test concept and compute the simple correlation, we find that the correlation is 0.75. This is a very strong relation between purchase intent and rated healthfulness, when we realize what we are dealing with given the thousands and thousands of ratings on which the correlation is based.

Fortunately, with the "raw data" file, we can correlate the 54 ratings for healthful (first rating attribute) to the 54 ratings for purchase intent (second attribute) on a person-by-person basis. Do individual respondents all show this covariation so that essentially one of the two attribute ratings is redundant? Is "healthfulness" really good enough to predict purchase intent? It seems so from the summarized data, but maybe we're missing some granularity that is important.

Let's go a bit deeper into the results. Each respondent generates a correlation coefficient, called the Pearson R. The Pearson R (i.e., the correlation coefficient) ranges from a high of +1 meaning the healthfulness and purchase intent are perfectly related, down to a middle value of 0 meaning no linear relation, and down to a low of −1 meaning perfect inverse relation so that increases in healthfulness covary with decreases in purchase intent.

Let's now plot the 152 different individual correlation coefficients, looking at how they distribute. Looking at Figure 16.5, we see that there are a lot of the respondents whose correlation is greater than 0.7, which level we set as a cutoff for "high agreement between healthfulness and purchase." These respondents (100 out of 152) tend to give high ratings for both attributes, or low ratings for both attributes. Whether they feel that they like to buy healthful products, or whether they are just lazy and give the same rating to each attribute, is not clear.

Now that we have plotted the distribution, it's clear that many individuals slavishly follow their ratings of healthfulness with a similar rating of purchase intent. We find that 2/3 of the respondents show this strong relation, whereas 1/3 do not and seem to treat the two attributes independently. However, what does this mean? Is there anything else that we can learn about the respondents that will help us understand whether or not a person will link purchase with healthfulness?

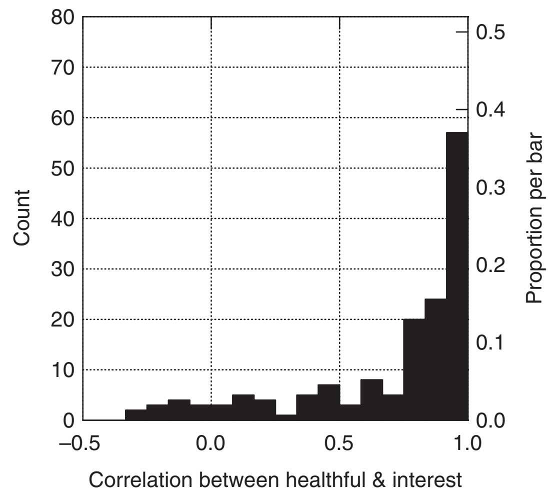


Figure 16.5 Distribution of the Pearson R values across 152 respondents for two attributes, healthfulness and purchase intent, evaluated for 54 visual stimuli.

We can uncover some of the answer by looking at how the respondents describe themselves. At the end of each evaluation, the respondents completed a 19-question profile about themselves, including age, gender, income, etc. A number of the questions went beyond these straightforward "geo-demographics" (who are you, where do you live, etc.), and instead asked the respondent to rate himself on the patterns of behavior when shopping. One question in particular, #8, is important.

Q8: How often do you typically read Nutritional Panels on products?
S1. All the time 41%
S2. Some of the time 49%
S3. None of the time 10%

Let us divide the respondents into three groups, based on how they answered question #8 in the classification questionnaire. For each group, in turn, we now estimate the impact value for the 24 elements, first for health and second for purchase. Once we do that, we can plot the 24 impact values in a scatterplot. We actually create three scatterplots, one for each of the three groups that emerged from question 8 (see Figure 16.6). These are the self-defined subgroups, created according to whether or not

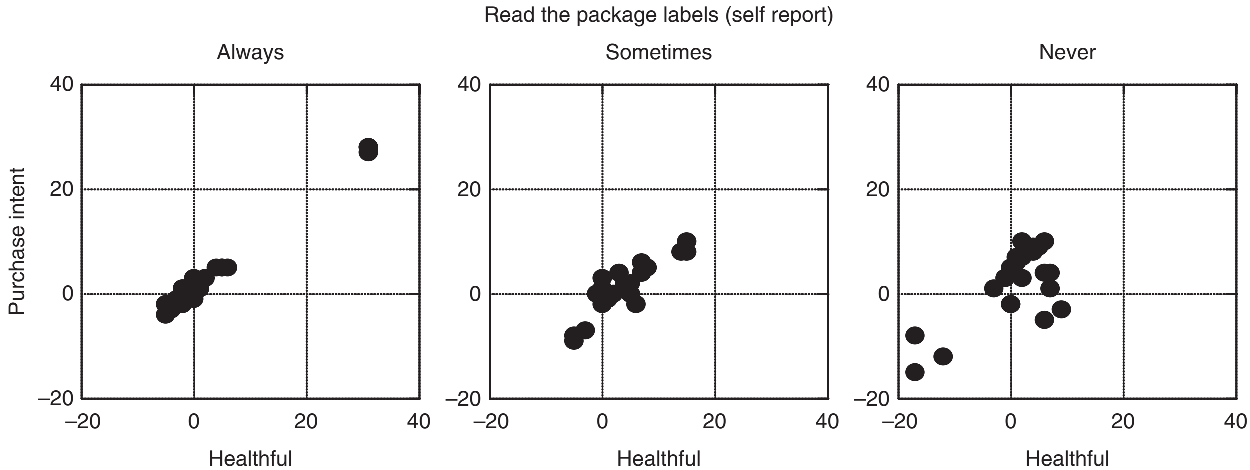


Figure 16.6 The covariation of impact values for "healthful" and "purchase intent" for three groups, defined in terms of how often they read the package labels. The self-definition comes from the classification questionnaire completed in the second part of the interview, after the package pictures had been evaluated.

they read the labels. The filled circles correspond to the 24 different package elements.

It is clear that those who say that they "always" read the package labels seem to "confuse" purchase intent with healthful. Or, perhaps, it's that the "always" group actually makes its decision on healthfulness first and foremost. We have the same relation with the "sometimes" group. However, the health-purchase intent breaks down, not completely, but somewhat with the "never" group.

We might want to ask "why"? Why, in fact, does the "always" group show such highly correlated impact values so that increases in healthfulness are accompanied by increases in purchase intent? Is it that they say to themselves, "It's healthy, I'll buy it"? We really don't know. We cannot answer that question from these data. We only know that we have three different groups and that two of the groups, ***Always*** and ***Sometimes***, show one type of behavior, whereas the ***Never*** group shows another behavior. In fact, the groups may not even be aware that they show this healthful-purchase correlation!

To get a deeper perspective on this difference among the three groups, let's look at how they respond to the different packages, when they are instructed to rate "healthful," the first of the two rating attributes. Look at Figure 16.7, which shows the impact values for the 24 elements, tabulated according to each of the three "label-reading" groups. Look specifically at what happens (i.e., the dynamics). The experimental design lets you do that and provides a rich field of information from which to discern patterns.

1. The additive constant increases as we go from the ***Always*** group through the ***Sometimes*** group and on to the ***Never*** group. Recall that the additive constant is the conditional probability of a person saying that the package is "healthful" (rating of 7–9) if there is no element. Substantively, the pattern means that respondent who says he never reads the label tends to call a product "healthful" more frequently or with greater probability than a person who says he always reads the label. Always reading a label makes a person more cautious.
2. The action is in the label, nothing else. The lower-calorie nutrition label is always the strongest driver of healthful, no matter which group. The higher-calorie nutrition label is always the strongest driver away from healthful. We saw this strong action of the label for the total panel, in Figure 16.4.
3. The difference among the groups comes from the dynamics of the way a concept is built. Let's look at the dynamics of the best combination.

Best combination (using element A1: nutrition label in front, regular calorie level). With the ***Always*** group, having the smallest constant, we have a healthfulness value of 68 (additive constant of 37 plus the

	Impact values for "Healthful" as rating attribute	Always read label	Some-times read label	Never read label
	Additive constant	37	41	55
A1	Nutrition label on front, regular calorie level	31	14	9
A3	Nutrition label on back, regular calorie level	31	15	7
A5	Nutrition label on side, regular calorie level	31	15	6
D4	Cholesterol free…ideal for a healthy lifestyle	6	0	1
C3	75% less carbs than traditional pasta	5	7	7
C6	Whole grain. . .heart healthy	5	7	6
D3	A perfect match with your favorite sauce for a wonderful Italian meal	4	0	2
D1	Know you're serving a family favorite made with fresh, quality ingredients	2	1	1
D2	Few foods are as nutritionally satisfying, convenient to prepare and fun to eat!	1	0	4
C1	Provides a full day's supply of 12 essential vitamins	1	8	5
D5	Feed your family the best	1	–1	6
B5	Splash—all natural	0	5	2
C2	100% gluten free	0	6	2
C4	Low in calories and fat	0	7	4
B3	Splash—made with farm fresh eggs	–1	1	–3
C5	Naturally low in sugar	–1	4	0
B4	Splash—new	–1	3	0
D6	Tastes like homemade	–2	0	–1
B2	Splash—recipe in box	–2	2	0
B6	Splash—organic	–2	5	0
B1	Splash—new size	–3	1	–1
A6	Nutrition label in side, higher calorie level	–4	–5	–12
A4	Nutrition label in back, higher calorie level	–5	–3	–17
A2	Nutrition label in front, higher calorie level	–5	–5	–17

Figure 16.7 Impact values for healthful, for the 24 elements, and three subgroups (label reading behavior) sorted by the "always read group."

impact value of A1 [i.e., 31, sum to 68]). With the ***Never*** group, the same one element concept should score 64 (55 from the additive constant, 9 from the contribution of element A1). The sums are about the same—68 versus 64—but the way the sum is constructed changes. For the ***Always*** group, it is the messaging and label information that drives the perception. For the ***Never*** group, it is their predisposition.

4. The bottom line is these two groups differ in two clear ways. First, their predispositions to call something healthful differ, with the ***Always*** group more skeptical, more reluctant. Second, the reactions to the elements differ. The ***Always*** group is likely to up-rate healthful messaging. The ***Never*** group is more likely to down-rate non-healthful messaging.

Nutrition Labels—Information and Their Role as Guides

We have been thinking about nutrition labels as simply sources of information, which when placed on the front, side, or back of the package don't do much except tell us about the product inside, drive perception of healthfulness and perhaps, in turn, drive the purchase intent rating.

It's clear from our results in this chapter that increasing the number of calories decreases both the perceived healthfulness of the pasta and, in turn, the rated purchase intent.

But, is there more? Are we missing even more important understanding about how design elements "work together"? After all, ask one designer after another about the power of design to drive consumer responses and the almost-universal answer is that "all the parts of the design work together in a harmonious way to drive consumer response." That may not be the exact quote from each designer, but it's pretty close to the sense and sentiments that you will hear. To the designer, the interplay of the elements, the artistic "hand" behind the design is just as important as the elements themselves.

If the interactions among design elements are important, then just how do we discover these interactions? We don't know what to look for, or even how to ask the question. How do you discover "interactions" among these design elements, especially the most important—the nutrition label—if you start at "ground zero"? The last section of this chapter demonstrates how to discover these interactions through a straightforward method called "scenario analysis" that works with experimental design.

Right now we will focus on what we found. Let's look at Figure 16.8, which shows what happens when we relate the presence/absence of the six health messages in Silo C to both health (first set of data columns) and to purchase intent (second set of data columns), respectively. Using regression analysis, we related the six elements in each of three silos (18 elements total, *but not* the nutritional label elements in Silo A) to ratings of "healthful" and purchase intent, respectively. We performed this regression analysis six times, once each for the specific set of test combinations that had element A1 (nutritional label on the front, regular calorie), once for the combinations that had element A2 (nutritional label on the side, regular calorie), etc. Silo A comprises six elements, which is the reason for the six parallel analyses.

Each regression analysis estimated the contribution of all 18 elements in Silos B (splash), silo C (health message), and silo D (family message), in the presence of each specific nutritional label. So, at the end of the analysis, we had six estimates of the impact of each element for "healthfulness," and six for "purchase intent," each estimate taken from those test stimuli with a one specific nutrition label. To reiterate, each separate regression looked at splash messages, health messages, and family messages in the presence of one specific nutrition label.

What's the reason for this? Simply, that when we do this type of "scenario analysis," there is only one thing common to each analysis, namely the specific nutritional label (i.e., where it is and what it says). If the respondent is really paying attention to this label, then the label somehow "sets the scene" for that particular combination.

Let's compare the performance of a specific element across all of these "scenes." That is, let's look at how an element, such as "75% less carbs than traditional pasta," performs when tested in concepts with each of the six different nutritional labels. If there is a difference in performance, a specific, single element across six scenarios, then we can attribute the difference to the "guiding" effects of the six nutritional labels on that element.

If you are following along, then you realize that this type of scenario analysis generates a lot of analyses and a lot to report. Rather than presenting everything, which just may go "on and on" interminably, we looked at one very interesting part of the results, and abstracted this. We present that interesting and rather surprising result in Figure 16.8. It seems to be consistent, and tells us an interesting story. So let's dive in and list what we discovered.

1. Key position—It's the front, and not the side, for the nutritional label: We looked at two positions, front and side. The reason is that these two clearly differ from each other, whereas the pictures of the nutrition label on the front versus on the back are merely differentiated by the phrase "front" versus the phrase "back." We decided not to discuss the messaging on "back" because that location did not add any new learning.
2. Health messaging, where the action occurs: We looked only at the results of one silo, health messaging (Silo C). Silo A set the scenario, Silo B (news splash) did not show anything, nor did Silo D (family messaging). All of the action was in the health messaging. Of course, keep in mind that in the regression analysis leading up to our results, we did include the elements in silos B and D, but choose not to report them. We don't want to weaken the presentation of our discovery by extraneous information which will dilute the impact.

A	B	C	D	E		F	G	H	I
Label location	Front	Front	Side	Side		Front	Front	Side	Side
Calorie level	Reg	High	Reg	High		Reg	High	Reg	High
Attribute rating	Health					Purchase	Intent		
Additive constant	61	36	61	39		50	26	40	36
When presented on front of the package, these messages “compensate” for higher calories									
75% less carbs than traditional pasta	7	14	4	2		4	11	3	2
Naturally low in sugar	1	6	7	–8		–3	3	8	–11
Provides a full day’s supply of 12 essential vitamins	6	8	5	4		1	4	8	1
These messages never “compensate” for higher calories									
Whole grain...heart healthy	8	6	9	4		7	5	7	2
Low in calories and fat	3	3	4	–1		3	4	9	–2
100% gluten free	6	6	3	–1		6	3	–1	–10

Figure 16.8 Results of the scenario analysis. How the combination of nutritional label (regular versus higher calorie) and position on the package (front versus side) drive the response to six health messages.

3. Additive constant or basic effect of the location and type of nutritional label: First, look at the additive constant in Figure 16.8. The additive constant, the baseline, shows the conditional probability or percent of respondents who would call a package "healthful" or be interested in buying the product if there are no elements. With that definition of the additive constant in mind, we see that as we go, stimuli featuring regular calorie to stimuli featuring high calorie we decrease the perception of healthfulness and lower the desire to buy the product. No surprises here, at least not yet.
4. What happens to the impact or utility of the six different health messages when the nutrition label is in front? Look at the impact of health messages when we work with the nutrition label on the front panel, where there is plenty of space, and where we have two options, regular and high number of calories, respectively. These are the two pairs of columns in Figure 16.8, B and C, and F and G, respectively. We see that in both pairs of data columns, the first for "health" and the second for "purchase intent," three of the health messages increase dramatically in terms of their impact when we start out with regular calories and then move to the label showing increased calories. *That is, for the three following messages, something unexpected happens in the respondent's mind.* The elements become more important, almost as if in the presence of "higher calories," a turn-off, the respondent looks for a compensating reason to call the product healthy or to buy it. That compensating reason is found in the increased importance given to three clearly "health oriented," "specific," "tangible" elements:
   75% less carbs than traditional pasta
   Naturally low in sugar
   Provides a full day's supply of 12 essential vitamins
5. The opposite pattern, ***no compensation***, occurs when we put the nutritional label on the side of the package. Now these three previously compensating elements that had somehow ameliorated the health problem, no longer do so.
6. Putting the nutritional label on the side of the package drops interest and diminishes the effect of health messaging! Let's look at columns D and E (health) and columns H and I (purchase intent) to see this opposite effect.
7. Location has two effects, therefore. When the label/calorie information is presented in front of the package, respondents "compensate" for high calories by giving more importance to some of the health messages. The health messages "compensate" for the high calories. However, when the label/calorie information is presented on the side of the package, respondents don't compensate at all. Rather, the opposite happens. Now, surprising, change in calories actually reduces the impact of the three, previously compensating health messages, making their utilities actually lower, and in fact sometimes negative.
8. To summarize: We see that nutrition labels can drive down interest in the healthful pasta, when the label informs that the pasta is higher calorie. However, when other "health elements" are present at the same time, they may, but not necessarily, take on more importance. As the higher-calorie nutrition label drives down healthfulness and interest, the respondent pays more attention to some of the other health message, which in turn become more positive, and compensate for the high calories. *It is as if these other health elements come in to give permission for a respondent to buy the pasta, even though the nutrition label says it is high calorie!*

Summing Up—What We Learn from the Healthy Pasta Case History

In this chapter, we covered a variety of issues. We began with a relatively simple issue: What elements drive the "perception of healthful"?, What elements drive "purchase intent?", and "Are they the same?" We discovered that it was the nutrition label more than anything else that drove both attributes and that respondents were sensitive to the calorie numbers, even without those numbers being pointed out.

Going a bit further into it, we discovered that for the total panel, the two rating attributes—healthful and purchase intent—were fairly correlated over all of the 152 respondents, using the 54 ratings. Thus, the correlation holds, even with 8,000+ data points. However, we also discovered that the correlation between healthful and purchase intent doesn't hold for everyone. The high correlation holds for about 2/3 of the respondents, who treat healthful the same as they treat purchase intent. We also discovered that there are self-reported groups who always, sometimes, or never read the labels. The majority of the "always read" fall into the high correlation group. If a person says he always reads the label, then we discovered that that individual's ratings of purchase intent highly correlate with ratings of healthfulness. For the two remaining groups who say they sometimes or never read labels, there is an increasing proportion that fall into the uncorrelated group, which treat healthfulness differently from purchase intent.

Finally, by looking at a finer-grained analysis of the data, we found that the nutritional labels "set the stage" for the performance of the health messages. The respondents recognize the increase of calories when it is put on the nutritional label and down-rate healthfulness and purchase intent. When the nutrition label is put in front of the package, respondents look for a health message that can compensate for the impact of higher calories. Three health elements do the work of compensation. The compensation dynamic does not occur when the nutritional labels and health messaging are placed on the side of the box, however.

Technical Appendix

Discovering Interactions among Pairs of Elements

In this appendix we present a straightforward way to find interactions among pairs of stimulus elements, even if one doesn't suspect that these interactions exist (see Gofman, 2006). Many researchers who work with experimentally designed stimuli have to limit themselves to main effects or the separate contributions of the different elements. There is a simple reason for the limits. Think about the number of possible interactions among two elements from different silos. In this study we have four silos (nutrition label, information splash, health message, and family message). Each of the silos, in turn, comprises six elements. Therefore, for any particular pair of silos (i.e., nutrition label × health message), there are 36 different combinations (A1 × C1 … A6 × C6 = 36). In turn, with four silos, there are (4 × 3/2) or 6 pairs of silos. With 36 pairs of elements, this means 216 possible combinations, an altogether impossible task if we work in the conventional way, with a limited number of combinations tested by many people. Of course, if ahead of time we know the combinations of elements that we think will interact, then we can test that hypothesis by creating those specific combinations.

In this appendix we outline the approach, and then apply it to the role of nutrition label. We will find these labels both to provide information and to act as guides or conductors, influencing the effect of other "health elements," at least in some situations. Before we jump into the approach, the reader can find a worked example in the author's book, Selling Blue Elephants: How to make great products that people want before they even know they want them (Moskowitz and Gofman, 2007).

1. We begin by preparing the data for a specific type of analysis called "scenario analysis," and then move on to make our discoveries.
2. Look at Figure 16.9, which shows part of a very large Excel file comprising 8,208 rows, one for each package "concept" for each respondent. Recall that each of our 152 respondents evaluated 54 different combinations of concepts, which generates the 8,208 rows of data. Thus our Excel file comprised all the data that we will need for the analysis.
3. Each row shows the 24 elements, coded 1 (present) or 0 (absent), as well as four new, derived variables. These four new variables are labeled with simple, easy-to-remember, and intuitively obvious names: ByA, ByB, ByC, ByD. The "ByA" variable takes on the value "1" if A1 = 1, takes on the value "2" if A2 = 1, etc. There are seven options or levels in the ByA variable, from ByA = 0 (when the stimulus does not have any nutritional label) to ByA = 6 (when the stimulus has a nutritional label, A6, which is the nutrition label on the side, higher-calorie level).
4. We sort the entire data matrix by the newly created variable, ByA. The first rows in the sort correspond to those where ByA is 0; the last set of rows correspond to those where ByA is 6, etc. Sorting is straightforward. The idea here is that there are seven such "layers," each layer with its own specific nutritional label.
5. Now we are going to analyze the matrix. Instead of running one large regression relating all 24 elements (A1 … D6) versus the attribute rating (i.e., binary variable for healthfulness, 1–6 coded 0, 7–9 coded 100), we will run six separate regressions. Each regression analysis will correspond to one of the six layers (i.e., where the variable ByA takes on the value 1, 2, 3, 4, 5, and 6, respectively). We won't analyze those test stimuli where the nutritional label is absent (i.e., where ByA = 0).
6. The dependent variable for the regression analysis is the binary variable for healthfulness (0 if the original rating for healthfulness was 1–6, 100 if the original rating for healthfulness was 7–9). The independent

D	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	DS	DT	DU	DV	DW
UID	Con		A1	A2	A3	A4	A5	A6	B1	B2	B3	B4	B5	B6	C1	C2	C3	C4	C5	C6	D1	D2	D3	D4	D5	D6		BYA	BYB	BYC	BYD
672174	1		0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0		0	1	5	4
672181	1		0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0		0	3	1	4
672194	1		0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0		0	3	1	5
672205	1		0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1		0	2	3	6
672173	1		1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0		1	5	4	0
672192	1		1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0		1	1	2	4
672212	1		0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1		2	3	5	6
672173	2		0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0		2	3	4	5
672172	1		0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0		3	3	2	4
672177	1		0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0		3	1	6	4
672197	1		0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0		3	1	1	1
672203	1		0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0		3	1	1	3
672210	1		0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1		3	2	5	6
672179	1		0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0		4	4	1	5
672180	1		0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0		4	4	1	5
672208	1		0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0		4	2	6	0
672172	2		0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0		4	1	0	5
672177	2		0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0		4	5	4	0
672184	1		0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0		5	1	4	3
672199	1		0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0		5	6	4	1
672215	1		0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0		5	3	6	2
672174	2		0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0		5	2	3	4
672169	1		0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0		6	4	0	3
672186	1		0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0		6	5	1	4
672170	2		0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1		3	2	5	6

Figure 16.9 Example of the database in Excel. The first column is the UID (i.e., the respondent unique identification number). The Con is the concept number, which ranged from 1 to 54. The experimental design comprises 24 elements, four silos (A–D), each with six elements. At the right side are four new variables: ByA, ByB, ByC, and ByD.

variables are the 18 elements in silos B, C, and D, respectively. The elements in silo A, nutritional label and its position, do not appear in the model. As just stated in Step 5, we run a separate equation for those test stimuli where ByA takes on the value 1 (i.e., the label has the regular number of calories, and appears in the front), then a separate equation for those test stimuli where ByA takes on the value 2, etc.

7. The outcome is a model for each layer. The model can be estimated for each attribute (healthfulness, purchase intent) by each option or level of the nutritional label (six in total). There are 12 such combinations. We see only 8 of the 12 in Figure 16.8.
8. The additive constant is the conditional probability of rating a specific pasta package as "healthful" (i.e., 7–9 on the 9-point scale), without any elements. All we know is that the package corresponds to a specific one of the six package alternatives, varying in calories and position of the nutritional label. Again, this additive constant corresponding to the specific package is an estimated parameter. It is a good estimate of the basic interest in the package, given only the label information.
9. The element values correspond to the utility values of the 18 elements. We only present the six utility values corresponding to the six elements of the nutritional message, where the "action occurs."
10. The bottom line here is through the scenario analysis it becomes possible to see how one variable affects another. Scenario analysis looks at one element as a "guide" or "director" of other elements. By partitioning the set of elements in one silo into layers, and by creating the model "layer-by-layer," one layer per element, it becomes possible to identify the interactive effects of two elements, the element defining the layer and the element whose impact or utility is being estimated.

References

Garretson, J.A. and Burton, S. (2000) "Effects of nutrition facts panel values, nutrition claims, and health claims on consumer attitudes, perceptions of disease-related risks, and trust," Journal of Public Policy & Marketing 19, 213–227.

Gofman, A. (2006) Emergent scenarios, synergies, and suppressions uncovered within conjoint analysis. Journal of Sensory Studies 21(4): 373–414.

Goldberg, J.H. and Probart, C.K. (1999) Visual search of food nutrition labels. Human Factors 41, 425–437.

Jones, G. and Richardson, M. (2007) An objective examination of consumer perception of nutrition information based on healthiness rating and eye movement. Public Health Nutrition 10, 238–244.

Moskowitz, H. and Gofman, A. (2007) Selling blue elephants: How to make great products that people want before they even know they want them. Wharton School Publishing, Upper Saddle River, NJ.

Neuhouser, M.L., Kristal, A.R., and Patterson, R.E. (1999) Use of food nutrition labels is associated with lower fat intake. Journal of the American Dietetic Association 99, 45–53.

Nutritional Facts Panel. (2008) Guidance for Industry: A Food Labeling Guide. (http://www.cfsan.fda.gov/label.html).

Satia, J.A., Galanko, J.A., and Neuhouser, M.L. (2005) Food nutrition label use is associated with demographic, behavioural, and psychosocial factors and dietary intake among African Americans in North Carolina. Journal of the American Dietetic Association 105, 392–402.

U.S. Centers for Disease Control, Obesity Trends 2006, (http://www.cdc.gov/nccdphp/dnpa/obesity/trend/

Part IV

Emotions and Experience

Packaging Research in Food Product Design and Development Howard R. Moskowitz, Michele Reisner, John Ben Lawlor and Rosires Deliza
 ISBN: 978-0-813-81222-9

Chapter 17

Emotions and Package Design—Coffee

Introduction

Emotions are at the heart of everything we do. Our emotions guide our everyday choices and determine our biggest decisions. Our lives are often arranged to maximize the number of pleasurable emotions and to minimize those feelings that are less enjoyable (Ekman, 2003; Wood, 2007).

Consumers have always been guided by emotions when making their purchases. Indeed, emotions play an exceptionally large role in everyday life, as most people would agree. It's only been a few years now, however, that researchers working with package design have formalized the study of emotions as a key feature. Of course every package designer is keenly aware of emotion in design, but the research community generally ignored emotion. The reason was simple—they recognized that emotion was important, but they didn't have tools to handle the measurement of emotions, except for large-scale batteries of questions. Furthermore, much of the work on emotions had been done by psychologists and psychiatrists, dealing with the sadder side of life.

After a long period in which consumers were assumed to make largely rational decisions based on product attributes and benefits, marketing scholars have started to study emotions evoked by products and brands (Laros and Steenkamp, 2005). Emotions play an important role with respect to the response to advertising, but also to the sense of satisfaction with a product, with the experience, and with memory.

It should come as no surprise that emotions have been studied by both psychologists and market researchers, interested in their role in everyday life, especially when products and services are involved. Studies have demonstrated fairly convincing data that emotion contributes to brand interest and purchasing attitudes far more than cognition. That should come as no surprise, however, since we are a brand-conscious society, with high-end and luxury brands conferring prestige.

There are a lot of emotions. Lacking any specific vocabulary, but blessed with language and the importance of emotions in everyday life, researchers have not been shy about creating their own language and batteries of emotion checklists. Without yet going into detail about these batteries, the data suggest that feelings play quite a relevant role in the prediction of an ad's effectiveness (Edell and Burke, 1987; Holbrook and Batra, 1987).

Emotions and Consumers

There is a growing body of literature dealing with consumer emotional responses to the experience of consuming a product. The research published deals both with the immediate experience, but also with one's memory about the experience and intentions for buying in the future (Westbrook, 1987; Oliver, 1993; Mano and Oliver, 1993). The data came back showing clearly that emotion was critical, thus, this chapter.

If emotion is so critical, then how do we introduce it into package research in a way that makes emotion more than simply an accompaniment to the experience itself? Can we "get a handle" on emotion through scientific approaches, using experimental design, and if so, then how?

We must begin with understanding the language of emotion. With emotion so important for daily life, what are its dimensions, its descriptions? Our first step is to look at the two major categories—positive and negative emotions. Even before gradations, we need to know whether a feeling is positive or negative. Laros and Steenkamp (2005) classified emotion words as either a positive or negative emotion.

 ISBN: 978-0-813-81222-9

The Next Step Beyond Basic Science—Applying the Thinking to Advertising

We can learn a lot about applying emotion by taking a peek at other applied research, not in packaging science, but in the more widely appreciated area, advertising research. A recent Best Methodology Prize at ESOMAR Congress 2007, awarded to Orlando Wood (2007), described some important new findings for the measurement of emotion in advertising. Wood suggested that all advertisements should generate emotion. Furthermore, emotion can and should be "engineered" in order to maximize the opportunity for different business objectives. Thus, according to Wood:

1. Advertisements to drive sales should generate happiness and emotional intensity.
2. Advertisements whose aim is a call to action should evoke emotions that commonly lead to action or the prevention of action, that is to say anger, disgust, and possibly fear. These advertisements should not lead to sadness, an emotion borne of resignation, and an inhibitor of action.

According to Wood, *the emotional response to stimulus is the most important indicator of future behavior*. Here is the heart of the issue. It may be emotions, rather than interest, which is of importance to our business issues, although truly uncovering the real "mechanism of action" of emotion is going to take many years and a significant amount of research. Nonetheless, it's important to begin exploring emotion, so we can adjust our way of looking at stimuli, basing our insights on both overall evaluation and emotional profile.

Measuring emotions may result in significant cost savings for clients, more creative products, and better advertisements and labels. It is emotions that we should be measuring, because it is our emotions that are the best and most immediate judge of advertising and early-stage creativity. Thus, if it works for ads, it may work for packages and labels as well.

The Existential Leap—Extending Emotion to Packaging

Scientists are already well versed in the study of emotions when it comes to food. The relations between food, eating behavior, characteristics of the individual, and emotions have been studied from various angles and with a wide variety of methods. These studies can be classified into two basic types: studies that focus on the effects of emotion on eating behavior versus studies that focus on the effects of eating behavior on emotion (Desmet and Schifferstein, 2008; Desmet, 2008).

We can now follow a parallel path, focusing our interest on how emotions affect the perception of package design, and in turn, how package design evokes emotions within us. This chapter addresses the latter topic—the nature of the emotions that ordinary people experience in response to looking at instant coffee labels. What do they feel? How many feel that way? What package features drive a particular set of feelings?

How Do We Begin?—What Are the First Steps?

How can we measure emotions? The approach of psychologist Paul Ekman (2003) is especially interesting. He makes a case for a set of seven basic emotions: happiness, surprise, sadness, fear, anger, contempt, and disgust, all of which are universally conveyed by and recognizable in the face. Ekman's research used the rather interesting approach of "reading" emotion in people's faces.

We took the Ekman approach of limiting the number of emotions to seven, which we thought to be manageable and scalable for online research with hundreds and later thousands of respondents. However, we selected those seven emotions in consultation with John Kearon of BrainJuicer®, Ltd., in the UK (see Kearon, 2007). The BrainJuicer® seven seemed more appropriate for our application.

So, what were those seven emotions? We used seven different "single" words: sad, irritated, neutral, calm, joyful, relaxed, and energized, whose meanings are easily understood by participants. We made the job easy. We presented the respondent with a test stimulus shown on a computer screen. We instructed the respondent to select the single emotion that best described how he felt at that moment, after inspecting the test stimulus. The test stimuli were systematically varied coffee labels.

Creating the Test Stimuli

Since we are dealing with design and emotion here, we used coffee as the product. Coffee is interesting because it is typically associated both with "waking up" (i.e., energizing) and with relaxing. We often see or hear advertisements that promote coffee as the wake up drink, yet at the same time, companies like Starbucks have made coffee a product with which one can relax, socialize, or just grab a few private, self-indulgent minutes.

We began with the template, which we see at the left side of Figure 17.1. The template comprises four categories. At the right there is an example of the template "filled out."

Doing the Experiment

Our seven emotion words comprise two negative emotional states (sad, irritated), one neutral state (neutral), and four positive emotional states (calm, joyful, relaxed, and energized).

In any of these applied studies, it is important to ensure that the ratings are easy to assign and that there would not be too many of them. People understand these different emotional states. They are not typical of the types of inventories that a psychologist or a psychiatrist might use, but rather a list that one might use in common parlance.

The study comprised three elements from four silos each, which we see in Figure 17.2. We see the elements, along with the contributions for the seven emotions, which we deal with a bit later. For now, it's important just to see the different elements themselves.

The actual emotion study ran the same way, as did the other studies in this book. That is, the experimental design created the different combinations. In this case, the design of 4 silos and 3 elements per silo (12 independent variables) generated 27 combinations, again with each individual respondent evaluating a different set of these 27 combinations.

Respondents were invited to participate. Those who agreed to participate were led to the orientation page (Figure 17.3). The orientation page tells them that they will evaluate coffee packaging and that their job is to select one of seven different emotional statements for each package. The instructions directs the respondent to identify what the package makes the respondent feel. These types of instructions, to choose one of several different scale points, differ from what we have been dealing with in this book. Most of our focus has been on scales that are more or less "continuous" and unidimensional, so the value 1 corresponds to the lowest level of the attribute and the value 9 corresponds to the highest level. Here we change the rules, so it is truly selection, not scaling.

Setting Up the Data for Analysis—What to Think About and What to Do

Since emotion is a new area of research, one of the first questions to ask is whether the test method really works. That is, we know that respondents participated in the study. Yet, did respondents use the range of different emotions, or perhaps did they just stay at "neutral" and

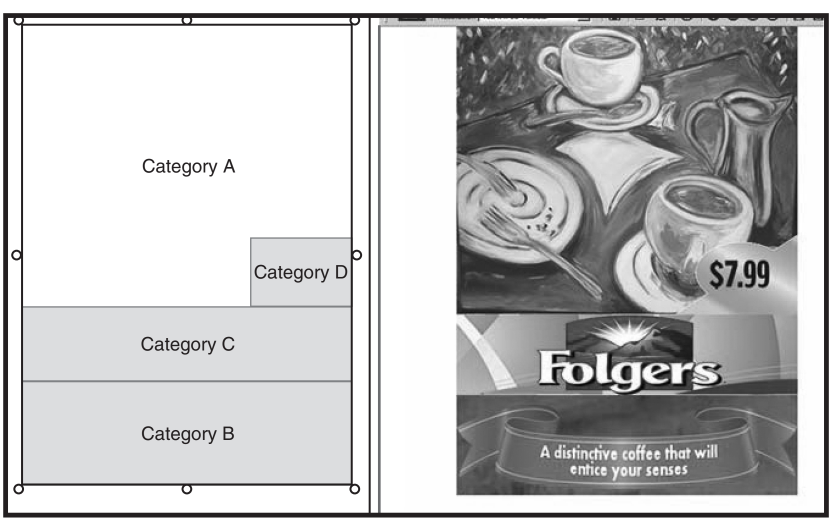


Figure 17.1 The template for the coffee "emotions" study (left side), and an example of a visual concept (right side)

		Sad	Irritated	Calm	Neutral	Joyful	Relaxed	Energized
Silo A: Background								
A1		-1	-3	14	7	2	11	-3
A2		-1	-7	13	2	4	22	-5
A3		1	13	6	2	2	5	-1
Silo B: Label								
B1	Coffee is good for your health. Drinking coffee fights free radicals, aging and degenerative diseases.	1	7	2	9	3	3	4
B2	A real coffee with distinct aroma and flavour	1	6	4	8	2	4	3
B3	A distinctive coffee that will entice your senses	1	7	3	8	2	3	4
Silo C: Brand								
C1	Folgers	1	4	3	9	3	4	4
C2	Maxwell House	1	8	2	11	2	0	3
C3	Douwe Egberts	1	5	3	8	4	2	5
Silo D: Price								
D1	$9.99	2	4	4	10	3	1	4
D2	$7.99	6	14	1	7	0	-2	2
D3	$5.99	4	7	2	11	2	-2	2

Figure 17.2 The four silos and the three elements in each silo, along with the impact values for the seven emotions

Welcome to the **Coffee Survey!**

We are interested in your thoughts and opinions about various coffee packaging.

On the following screens, you will be presented with a series of designs for a new Coffee Package. You will be asked to rate each design on a 1–7 scale by clicking on the number that corresponds to how the packaging makes you feel based on the following statement.

Choose the emotion that best describes how you feel when looking at this coffee package design.

1 = Sad, 2 = Irritated, 3 = Calm, 4 = Neutral,
5 = Joyful, 6 = Relaxed, 7 = Energized

You will also be asked just a few questions afterwards to help us understand your needs.

The survey should take about 15 minutes.

Please press ">>" to continue.

Figure 17.3 The orientation page for the evaluation of emotions.

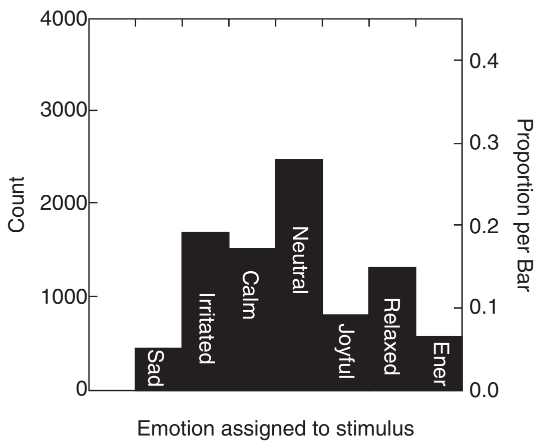


Figure 17.4 Distribution of emotions across 8,856 concepts for the coffee design study.

not sense any internal emotion from looking at the concepts? We don't know yet what the answer will be in terms of how the different design elements "drive emotions," if they even do! Yet we can look at the distribution of ratings. We have a total of 328 respondents, each of whom evaluated a different set of 27 visual designs or 8,856 concepts. Let's see the distribution of the choices for the emotion.

Look at Figure 17.4 to see the distribution of emotions across all concepts for the coffee study. Independently of respondent or concept, we just computed the distribution of votes for the seven emotions. We see that there are four key emotions: neutral, irritated, calm, and then relaxed. There are fewer choices of joyful, sad, and energized. Some of this may be a slight bias to avoid judgments at either end of the scale. Yet we see fewer ratings of joyful, in the middle of the different choices, so our distribution is not purely the result of avoiding the end points.

Figure 17.4 makes us realize that the respondents differentiated the concepts, or at least distributed their choices. We next want to discover how each of the different visual elements drives the different emotion choices. We have a new problem facing us, however. Up to now, we have been dealing with a rating scaling that is essentially continuous. We have chosen to divide the scale into two parts (1–6 and 7–9) based upon success in previous projects.

We can use our accumulated knowledge about "modeling" to solve the emotion problem, or at least move toward an answer. The key to the issue is recognizing that we don't have one scale, but rather seven scales. Each of our emotions is its own scale. When a respondent chooses a specific emotion, we can think of the respondent as choosing that scale (i.e., assign a value of 100 to the scale), and not choosing any other scale (i.e., assign a value of 0 to the other scales).

Let's see what the data look like and what we have to work with for the analysis. Look at Figure 17.5. It shows the first six package designs for one respondent

UID	Concept	A1	A2	A3	B1	B2	B3	C1	C2	C3	D1	D2	D3	Emotion #	Sad	Irritated	Calm	Neutral	Joyful	Relaxed	Energized
667138	1	1	0	0	0	0	1	0	0	0	0	1	0	4	0	0	0	100	0	0	0
667138	2	1	0	0	0	0	1	0	1	0	0	0	0	4	0	0	0	100	0	0	0
667138	3	1	0	0	0	1	0	0	1	0	0	1	0	4	0	0	0	100	0	0	0
667138	4	0	1	0	1	0	0	1	0	0	0	0	0	6	0	0	0	0	0	100	0
667138	5	0	0	1	1	0	0	0	1	0	0	0	1	4	0	0	0	100	0	0	0
667138	6	0	1	0	0	0	1	0	0	0	0	0	1	6	0	0	0	0	0	100	0
667140	1	0	1	0	0	0	1	0	0	0	1	0	0	5	0	0	0	0	100	0	0
667140	2	0	0	0	0	0	1	0	1	0	1	0	0	4	0	0	0	100	0	0	0
667140	3	0	1	0	0	1	0	0	1	0	1	0	0	6	0	0	0	0	0	100	0

Figure 17.5 Layout of data from two respondents, showing the concept order, composition of the concepts, emotion chosen, and the re-coding of the single choice into seven "dummy variable" responses corresponding to the different emotions

(code 667138) and the first three package designs for a second respondent (code 667140). Here we see the information laid out and ready for the regression analysis that will reveal the contribution of each element to each emotion.

To analyze the data for the emotion selection, follow these steps, which for the most part are similar to the other analyses we have run for package designs. There will be one change, however. We will not use the additive constant:

1. Each row corresponds to a specific package design.
2. The first column is the respondent (UID).
3. The second column is the "order number" (1–27, corresponding to the 27 designs evaluated by a respondent).
4. The particular elements are coded as "1" for present and 0 for absent. We see the design for 4 silos, 3 design elements per silo.
5. We then see a column labeled "emotion #." This column shows the selection by the respondent of the "emotion" felt when looking at the design. For the first row, the emotion selected by this respondent for this particular stimulus is "4" (i.e., neutral).
6. We have thus expanded the set of dependent variables, to comprise seven columns, one per emotion. For the first respondent, first concept, we see the column for "neutral" given the value 100, and the remaining emotion columns given the value 0.
7. For the fourth concept, the respondent selected emotion "6" (i.e., relaxed). Thus, in the fourth row, the column for the emotion "relaxed" is given the value 100 and the remaining columns are given the value 0.
8. The data are ready to be analyzed by ordinary least-squares regression. We will relate the presence/absence of the 12 elements to each specific emotion. That is, we will determine how each element drives each emotion.
9. *We will not use the additive constant here. The respondents were required to choose one of seven different emotions. We assume that if there are no elements in the test stimulus, then the respondent will not know what to choose. Thus, it is not meaningful here to use an additive constant, which gives us the predisposition to rate a package design on a unidimensional scale.*
10. We thus express the model as:

$$\text{Emotion choice} = k_1(\text{Element A1}) + k_2(\text{Element A2}) \cdots k_{12}(\text{Element D3})$$

11. The coefficients or impact values ($k_1..k_{12}$) show the conditional probability or proportion of respondents who assign a specific emotion to a visual design if the element is present in the design.

Table 17.1 Key visual elements that drive emotional responses for the total panel and three segments (Segment 1, Segment 2, and Segment 3)

Emotion response	Element that does the driving	Impact
		Total
Relaxed	A2 Table	22
Irritated	D2 $9.99	14
Calm	A1 Cup	14
Irritated	A3 Drinker	13
Calm	A2 Table	13
Relaxed	A1 Cup	11
Segment #1: Responds positively to the drinking "situation" but not to people when inspecting a package		
Relaxed	A2 Table	28
Irritated	A3 Drinker	25
Relaxed	A1 Cup	17
Calm	A1 Cup	16
Calm	A2 Table	15
Joyful	C3 Folgers	11
Segment #2: Price sensitive, tends to be irritable when inspecting a package		
Irritated	D2 $9.99	18
Irritated	A3 Drinker	16
Joyful	A2 Table	10
Irritated	D3 $7.99	10
Segment #3: Strong emotional reactions, emotionally sensitive to brands and prices, positive to pictures of the drinking situation		
Relaxed	A2 Table	28
Calm	A1 Cup	23
Irritated	D2 $9.99	20
Irritated	C2 Douwe Egbert	20
Calm	A2 Table	19
Relaxed	A1 Cup	17
Calm	A3 Drinker	16
Irritated	C3 Folgers	15
Irritated	C1 Maxwell House	15
Irritated	D3 $7.99	13
Irritated	B3 Senses	13
Irritated	B1 Health	12
Irritated	B2 Dist. Flavor	12
Energized	D1 $5.99	11
Energized	C3 Folgers	11
Energized	C1 Maxwell House	10

How Do the Different Elements Drive Emotion?

In order to determine "what works" in terms of emotions, we use the entire data set when we relate the presence/absence of the 12 different elements to the selection of an emotion. That is, we have 8,856 "cases" or observations for our database. Each case, a specific combination of elements, is associated with the selection or nonselection of each emotion. The data format in Figure 17.5 is set up to allow regression analysis for each different emotion.

We see the results for the total panel at the top of Table 17.1. With 7 emotions to select from and with 12 stimulus elements, our results matrix would comprise 84 cells. The number inside the cell would be the conditional probability of a package element "driving" the particular emotion. This number is similar to the corresponding number for the concept, which is the conditional probability of the element for the concept driving the response, whether the response is interest or some other criterion.

Most of these numbers would be small, because most elements do not "drive" the emotions. In the interest of a simplified set of results, we show only the strongest results from this study in Table 17.1. We see only those element-emotion combinations where an element "drove" an emotion more than 10% of the time. Thus, the picture of the table, A2, drove a feeling of relaxation 22% of the time. The price of $9.99 drove the feeling of irritated 14% of the time, and so forth.

Emotional Action Through Mind-Set Segments (Once Again!)

Throughout this book we have found the notion of segmentation to be a very powerful organizing principle. Often when the data for the total panel were "flat" (i.e., with utility or impact values near 0 on both sides, such as, 0 +/– 5), segmentation would reveal that these low values came from countervailing currents. There might be two or more segments having radically different points of view. Together the segments would cancel themselves, but seen alone, the story underlying the elements was far more dramatic.

With such a history of success in segmenting, can we repeat the story here? After all, the scale that we are using is not a simple scale, where one person can love an idea and the other person hate the idea. It's pretty simple with the typical so-called "unidimensional scale," where the respondents simply have to love different topics of the same basic set of ideas. (Just go back to the chapter on cereal to see that notion played out in the mind-set segments. One group wanted health; the other group wanted taste.)

We did the segmentation on emotions using the approach described in the Appendix to this chapter. We won't go into the details of the method here, but rather just point out a surprising and interesting outcome. The story of the segmentation emerges clearly in the bottom of Table 17.1. Again, we only show the key drivers (i.e., those specific elements and emotional responses that exceeded 10, an arbitrary but reasonable cutoff).

We see clear differences in the way these consumer respondents react emotionally to the different stimuli.

1. Segment 1 comprises people who respond to the situation and not to the person, when the situation and the person are shown on the package. That is, the emotions come out for these people when they look at the situation in which they are to drink coffee. They don't like people. The picture of a drinker, even though stylized, irritates them.
2. Segment 2 comprises people who are just more irritable. They are certainly irritated at the prices.
3. Segment 3 comprises people who show strong emotional reactions, who are emotionally sensitive to brands and prices and react in a positive way to pictures of the drinking situation.

Summing Up

In this chapter we learned how to investigate the role of package design as a driver of a respondent's emotion, with the respondent selecting the emotion from a short list of seven. We choose instant coffee as a test stimulus

Table 17.2 Comparison of segmentation methods for standard concepts and for emotions. The methods are quite similar, except for the starting regression models used to relate the package elements to the response

	Rating concepts or designs on an attribute	Selecting an emotion
Stimulus setup	Stimuli systematically designed	Stimuli systematically designed
Respondent task	Respondent selects single "magnitude" on a continuous scale of magnitudes, in order to denote intensity of respondent feeling.	Respondent selects one emotion from a set of emotions, in order to show how respondent feels.
Data set up	Stimuli represented in dummy variable format (absent/present)	Stimuli represented in dummy variable format (absent/present)
New variables developed from scale	Preference or intention to purchase can be evaluated as dependent variable.	Create several new variables, one variable for each emotion presented as a possible response to be selected (e.g., sad, energized).
Recoding depending on the rating	If the rating is 7–9 on a 9-point scale, then code the design stimulus as 100 or else code the design stimulus as 0.	If a specific emotion is selected for a particular test stimulus, then code that emotion as 100, but then code the remaining emotions as 0. For each stimulus, there will always be one emotion coded as 100 and the remaining emotions coded as 0.
Regression	Use ordinary least-squares regression to relate elements to actual ratings (persuasion model). Then relate the elements to the binary response 0,100 (interest model).	Use ordinary least-squares regression to relate elements to each emotion response. For our data with 7 emotion responses, there are 7 equations.
Individual estimation	Estimate the persuasion model for each respondent.	Estimate the 7 emotion models for each respondent.
Factor analysis	Factor analyze the 12 persuasion coefficients (but not the additive constant), extract roots whose eigenvalues >1.5, and then rotate by the quartimax criterion in order to generate factor scores for the individual respondents.	Factor analyze the 7 models × 12 elements or 84 variables, extract roots whose eigenvalues >1.5, and then rotate the solution by the quartimax criterion in order to generate factor scores for the individual respondents.
Clustering	Cluster the factor scores, using k-means clustering	Cluster the factor scores using k-means clustering.
Identify clusters	Do a 2, 3, and 4 cluster solution. Choose the solution whose interest model is most interpretable by segment.	Do a 2, 3, and 4 cluster solution. Choose the solution whose interest model is most interpretable by segment.
Key elements	Identify strong performing positive and strong performing negative elements.	Identify strong performing positive combinations of elements and emotions.

for the package because coffee is both a beverage, and in many cases, a high involvement product.

Our study was very simple, because, as has been the approach throughout this book, we were searching for productive ways by which to analyze data (i.e., methods that are easy to use and whose results are easy to interpret).

Our simple experiment with coffee packages comprised just 12 independent variables (i.e., 4 silos and 3 elements per silo). The experiment generated a set of 27 unique combinations for each respondent.

We used seven emotional attributes, which comprised two negative emotional states (sad, irritated), one neutral state (neutral), and four positive emotional states (calm, joyful, relaxed, and energized).

It is important to emphasize here that the analysis differed slightly from the analyses presented previously in this book, where we dealt with concepts and packages. The scale we used to measure emotion was neither "continuous" to represent "degree of feeling" (value 1 corresponding to the lowest level of the attribute and value 9 to the highest level), nor was the scale "unidimensional." Participants selected one emotion from a set of seven, rather than scaling one emotion from weak to strong. Consequently, the interpretation of the results is new. The emotion results for this selection task are simply the percent of the time that the specific emotion is selected from a set, when the particular element is present on the package.

Appendix—How to Segment Respondents on Emotion Using the Pattern of Their Utility Values

In previous chapters we have discussed concept response segmentation. Table 17.2 below compares the concept-response segmentation here for emotions to the more typical mind-set or concept-response segmentation. The reason for the difference in procedure comes from the fact of modeling for emotion relating the presence/absence of the 12 elements to each of the seven emotions, rather than relating the presence/absence of the 12 elements to a single binary (interested or not interested).

References

Desmet, P.M.A. (2008) Product Emotion. In: Schifferstein H.N.J. and Hekkert P. (Eds). Product experience, Amsterdam: Elsevier, pp. 379–397.

Desmet, P.M.A. and Schifferstein, H.N.J. (2008) Sources of positive and negative emotions in food preference. Appetite 50, 290–301.

Edell, J.A. and Burke, M.C. (1987) The power of feelings in understanding advertising effects. Journal of Consumer Research 14(4), 421–33.

Ekman, P. (2003) Emotions Revealed, Understanding Faces and Feelings, Phoenix Press, London, UK.

Holbrook, M.B. and Batra, R. (1987) Assessing the role of emotions as mediators of consumer responses to advertising. Journal of Consumer Research 14(3), 404–420.

Kearon, J. (2007) Personal communication to Howard R. Moskowitz.

Laros, F.J.M. and Steenkamp, J-B.E.M. (2005) Emotions in consumer behavior: a hierarchical approach. Journal of Business Research 58, 1437–1445.

Mano, H. and Oliver, R. (1993) Assessing the dimensionality and structure of the consumption experience: evaluation, feeling, and satisfaction. Journal of Consumer Research 20(3), 451–66.

Oliver, R. (1993) Cognitive, affective, and attribute bases of the satisfaction response. Journal of Consumer Research 20(3), 418–30.

Westbrook, R. (1987) Product/consumption-based affective responses and post-purchase processes. Journal of Marketing Research 24, 258–270.

Wood, O. (2007) Using Faces Measuring Emotional Engagement for Early Stage Creative. Available from www.brainjuicer.com [Accessed 12/03/08].

Chapter 18

Beyond the Stimulus to the "Experience"

Introduction

Throughout this book we have concentrated on areas that we might call "traditional" for design and packaging. These are the physical layouts of packages and the design of the graphics on the front and back of the package. Of course there are interesting byways to this traditional treatment, such as segmentation of mind-sets, but for the most part, the stimuli we deal with are those that can be inspected, purchased, and used.

Let's move out of packaged goods for a bit and look at the notion of design in the world of food experience. We are living in a time when experience is, or at least soon will be, as important as the product itself. In 1999, Pine and Gilmour coined the phrase "the experience economy" to recognize this shift from the world of products that one owns to the world of activities and situations that one experiences. Look at Figure 18.1 to get a sense of how the world is progressing, from an economy of "stuff" to an economy of "experience (Pine and Gilmore, 1999)."

How can we apply experimental design to this elusive thing called "experience"? Certainly we know that experience is more than the physical stimulus. Experience involves the person who is sensing the environment and the stimulus around him. Which particular tools that we have might we bring to bear? If we are successful in applying design to experiences, then perhaps we will be afforded new tools to better understand the experience of packaging and the person-packaging interaction.

For right now, however, we have to go a bit slowly, since we are entering uncharted territory. Here we cannot easily identify the dimensions, nor of course what we might systematically vary should we be so inclined to do so when we apply the experimental method.

Rethinking Experimental Design for Experience (QSR)

To make our task easy, we move to a somewhat different world, albeit one related to food and beverage.

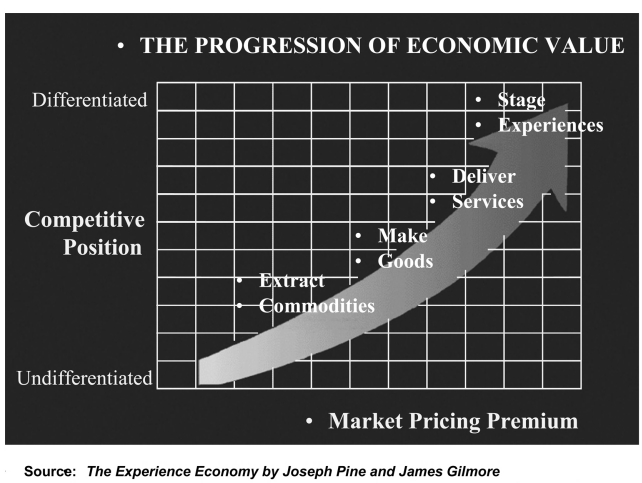


Figure 18.1 The progression of the economy over time

Packaging Research in Food Product Design and Development Howard R. Moskowitz, Michele Reisner, John Ben Lawlor and Rosires Deliza
 ISBN: 978-0-813-81222-9

We move away from pure package goods in the food and beverage industries, and toward restaurants where experience is the key, rather than simple functionality. We may not be able to re-create a particular experience, but perhaps we might be able to present a respondent with a scenario that captures some aspect of the experience. We could then measure how the respondent reacts to a stimulus that represents experience, and systematically study variations in this experience scenario, always a good strategy by which to learn. Of course we haven't captured the experience per se, but we have come closer to experience and moved a bit away from pure product. We will explore this move together in the coming pages.

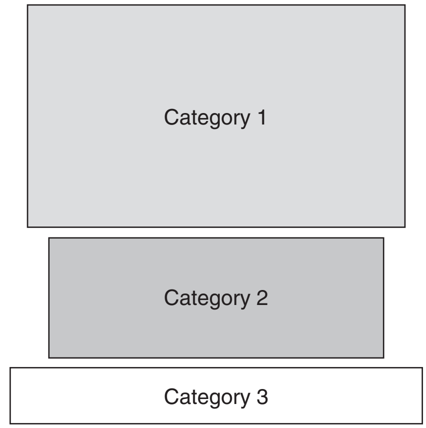


Figure 18.2 Template for the QSR-experience study. The template lays out the location where the elements fit.

A Case History: "Experience" in a Quick-Serve Restaurant (QSR)

Our case history deals with a quick-serve restaurant. The most important part of our work involves the preparation work, rather than the actual data, which we already know how to analyze. That is, we are interested in exploring stimuli that have experience as part of their DNA. Since we are looking at a restaurant, we want to discover three things:

1. How frequently would a person patronize the restaurant as the restaurant experience is described?
2. What emotions are associated with reading about the particular restaurant?
3. What rules emerge that can help us better understand experience?

We have deliberately made the study easier so that we can get a better sense of how experience-based elements perform. We will work with three different silos: a picture of the restaurant as it actually exists, an example of real food that the restaurant might serve, and then a phrase that talks about what the restaurant offers or stands for. When we systematically vary the elements in the three silos, we will have designed different experiences, at least in a virtual world.

Let's see what happens and how respondents assess their feelings.

What Are the Stimuli?

We are not starting here with a well-defined package that has features or with a graphics design that has to have certain types of information. There are no rules for stimuli to represent experience. Since we are at a very early stage, the so-called exploratory stage, it is best to limit our focus. Recognizing we are dealing with new issues and unknown territory, its best to be conservative, limit our focus, and deal with a simple, manageable array of test stimuli. Control here is better than large scope.

Coming up with the "proper" stimulus for an experience required a great deal of thinking. It's pretty straightforward to come up with a package or graphics front of a package. In those cases the goal is to be as realistic as possible, given the fact that the respondent was going to see the stimulus on a computer.

Working with an "experience" posed more problems. How do you represent an experience in terms of a set of visual features? Furthermore, can you even hope to capture the nuances of the experience? After much discussion we realized that we did not have to capture the nuances of the experience. Most print ads about experience (i.e., travel, etc.) must be content to show a picture or several pictures and use a limited set of words. Recognizing that we were modeling ourselves after magazine advertising for experience-oriented topics made our decision easier. We were simply going to show two pictures and one phrase, following the approach of the more sparse type of magazine advertising.

Let's first look at the template in which we are going to embed the silos and elements. Look at Figure 18.2,

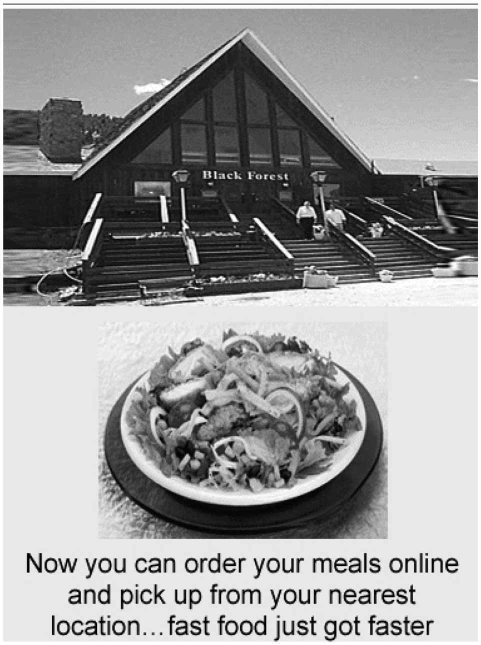


Figure 18.3 Example of a test stimulus constructed according to the template shown in Figure 18.2. The picture of the QSR (quick serve restaurant), the food, and the text are presented as "free agents," in color, on a computer screen.

which shows both the layout and the relative amount of space devoted to Silo A (i.e., category 1) = picture of restaurant, Silo B = picture of a food, and Silo C = text message. A lot of space is devoted to the picture of the restaurant because we feel that this is a key dimension of the experience. Figure 18.3 shows the example, using the template designed in Figure 18.2.

The Relevance of Emotions

We have already introduced the topic of emotions in Chapter 17, when we dealt with coffee packages. The important thing is that emotions are very relevant for food and can be measured in connection with packages as Chapter 17 showed. However, emotions are far more prevalent in situations where the interaction moves beyond one's interaction with an inanimate object and involves situations, people, expected activities, and even expected feelings.

Emotions are set off by situations and constitute an ever-present substrate or foundation to everyday behaviors. Eating has never been and never will be, simply about satisfying physical hunger. We eat not only to quell a rumbling stomach, but also to satisfy the appetite and deal with emotions. Thus, to understand and predict behavior or to have a mechanical robot act in human-like ways, we must have a comprehensive understanding of emotions and how these relate to behavior. According to Richins (1997), however, consumer researchers really have scant information about the nature of emotions in the consumption environment or how best to measure them.

What Are the Emotions?—Lists and Taxonomies of Feelings

There is a huge list of words to describe emotions. Table 18.1 shows the words used by several authors. So, which of these many emotion words should be used to measure consumer emotions in the emotion-rich context of experience? We have already used a battery of seven emotion words in our study of coffee pictures and packages (Chapter 17). It's still worth exploring the world of emotion language, if not for this study on quick serve restaurant (QSR) experience, then perhaps for future studies not even considered in this book.

To address the language of emotion, Laros and Steenkamp (2005) used the important study by Richins (1997). Based on extensive research, Richins had constructed the Consumption Emotion Set (CES). This scale includes most of the emotions that can emerge in consumption situations. The CES Scale was developed to distinguish the varieties of emotion associated with different product classes. In fact, the words included in the CES (see list in Table 18.1) are among the most frequently encountered words in the psychological emotion literature. One can divide them easily into so-called positive and negative affects, which makes intuitive sense since emotions span the range from positive to negative.

Going a little more deeply into emotions, Laros and Steenkamp (2005) proposed the hierarchy of consumer emotions, which consists of three levels: the ***superordinate level*** with positive and negative affect, the ***basic level*** with four positive (contentment, happiness, love, pride) and four negative (anger, fear, sadness, shame) emotions, and the ***subordinate level*** with specific emotions.

Basic emotions are believed to be innate and universal, but because there are different ways to conceive emotions (facial, i.e., Ekman, 2003; biosocial, i.e., Izard, 1992; brain, i.e., Panksep, 1992), there is also the inevitable controversy about which emotions are basic (Turner and Ortony, 1992).

Table 18.1 Emotion words

Negative Emotion Words
Aggravation, Agitation, Agony, Alarm, Alienation, Anger, Anguish, Annoyance, Anxiety, Apologetic, Apprehension, Aversion, Awful, Bad, Bashful, Betrayal, Bitterness, Blue, Bothered, Cheerless, Confused, Consternation, Contempt, Cranky, Cross, Crushed, Cry, Defeat, Deflated, Defensive, Dejection, Demoralized, Depression, Despair, Devastation, Different, Disappointment, Discomfort, Discontent, Discourage, Disenchantment, Disgust, Dislike, Dismay, Displeasure, Dissatisfied, Distress, Distrust, Disturbed, Down, Dread, Dumb, Edgy, Embarrassment, Empty, Envy, Exasperation, Fear, Fed-up, Ferocity, Flustered, Forlorn, Foolish, Frantic, Fright, Frustration, Fury, Gloom, Glumness, Grief, Grouchiness, Grumpiness, Guilt, Hate, Heartbroken, Hollow, Homesickness, Hopelessness, Horrible, Horror, Hostility, Humiliation, Hurt, Hysteric, Impatient, Indignant, Inferior, Insecurity, Insult, Intimidated, Irate, Irked, Irritation, Isolation, Jealousy, Jittery, Joyless, Jumpy, Loathing, Loneliness, Longing, Loss, Lovesick, Low, Mad, Melancholy, Misery, Misunderstood, Moping, Mortification, Mournful, Neglect, Nervousness, Nostalgia, Offended, Oppressed, Outrage, Overwhelmed, Pain, Panic, Petrified, Pity, Puzzled, Rage, Regret, Rejection, Remorse, Reproachful, Resentment, Revulsion, Ridiculous, Rotten, Sadness, Scared, Scorn, Self-conscious, Shame, Sheepish, Shock, Shy, Sickened, Small, Sorrow Spite, Startled, Strained, Stupid, Subdue, Suffering, Suspense, Sympathy, Tenseness, Terrible, Terror, Threatened, Torment, Troubled, Tremulous, Ugly, Uneasiness, Unfulfilled, Unhappiness, Unpleasant, Unsatisfied, Unwanted, Upset, Vengefulness, Want, Wistful, Woe, Worry, Wrath, Yearning
Positive Emotion Words
Acceptance, Accomplished, Active, Admiration, Adoration, Affection, Agreement, Alert, Amazement, Amusement, Anticipation, Appreciation, Ardent, Arousal Astonishment, At Ease, Attentive, Attraction, Avid, Bliss, Brave, Calm, Caring, Charmed, Cheerfulness, Comfortable, Compassion, Considerate, Concern, Contentment, Courageous, Curious, Delight, Desire, Determined, Devotion, Eagerness, Ecstasy, Elation Empathy, Enchanted, Encouraging, Energetic, Enjoyment, Entertained, Enthrallment, Enthusiasm, Euphoria, Excellent, Excitement, Exhilaration Expectant, Exuberant, Fantastic, Fascinated, Fine, Fondness, Forgiving, Friendly, Fulfillment, Gaiety, Generous, Giggly, Giving, Gladness, Glee, Good, Gratitude, Great, Happiness, Harmony, Helpful, High, Hope, Horny, Impressed, Incredible, Infatuation, Inspired, Interested, Jolliness, Joviality, Joy, Jubilation, Kindly, Lighthearted, Liking, Longing, Love, Lust, Merriment, Moved, Nice, Optimism, Overjoyed, Passion, Peaceful, Peppy, Perfect, Pity, Playful, Pleasure, Pride, Protective, Rapture, Reassure, Regard, Rejoice, Relaxed, Release, Relief, Respect, Reverence, Romantic, Satisfaction, Secure, Sensational, Sensitive, Sensual, Sentimentality, Serene, Sexy, Sincere, Strong, Super, Surprise, Tenderness, Terrific, Thoughtful, Thrill, Touched, Tranquility, Triumph, Trust, Victorious, Warm-hearted, Wonderful, Worship, Zeal, Zest

Note: The emotion words come from Richins' CES (1997). Adapted from Laros and Steenkamp (2005).

Listing the emotions is not sufficient. There is much more depth in the emotions, which can be best reached by profiling the different emotions. Because different emotions can have different behavioral consequences, it is important to know, for example, whether a failure in a product or service elicits feelings of anger or sadness. Both angry and sad people feel that something wrong has been done to them; but whereas sad people become inactive and withdrawn, the angry person becomes more energized to fight against the cause of anger (Shaver et al., 1987).

A review of methods used to measure emotion led Wood (2007) to the conclusion that it is necessary to develop a self-report technique (easy to administer and user-friendly) that overcomes a major criticism of self-report. This criticism is that there is often a need for a great deal of cognitive processing on the part of the respondent.

To make things simple, Wood turned to the work of Paul Ekman (2003) who makes a case for a set of seven basic emotions: happiness, surprise, sadness, fear, anger, contempt, and disgust, all of which are universally conveyed by and recognizable in the face to all people and cultures.

Emotions drive all decisions, particularly purchasing decisions, but so much of market research focuses on the rational. Ekman's method investigates the emotional impact of product design on consumers, and to validate the insight that simplicity is the major attribute that people desired in high-tech goods. Wood's adaptation of the Ekman work into an emotional scale is named FaceTrace™.

In the world of business, sometimes simpler and more direct, more targeted, turns out to be better. Thus, Holbrook and Batra (1987) developed their own emotional scale based on an in-depth review of the literature. They believe that the specific needs for advertising research are simply too great to adapt any more general approach to the specific business problem. The list of emotions must be customized for each issue.

In summary, there is wide divergence in the content of emotions studied in consumer research. Studies often use different scales to measure emotions and focus on different emotions. In spite of this, consumer researchers frequently use, or exploratory data analysis yields, a small number of dimensions (Bagozzi et al., 1999). Among these, the classification of emotions into positive and negative affect appears to be the most popular conceptualization.

Capturing the Subjective Response

With all of the foregoing as the introduction, we again opted to use a simplified version of the FaceTrace™, this time with seven words. Six of these were emotion-laden (sad, irritated, calm, joyful, relaxed, energized), and the seventh word was "neutral." We did this with the coffee package in Chapter 17. Let's now move toward experience.

Running the QSR emotions study required that we set the proper mood. We didn't want to tell the respondents too much about what we were after, so we did not use the word "experience." The notion of "experience" in a consumer test doesn't mean anything yet in the course of this writing (late 2008). Consumers are accustomed to answering questions about products, but not necessarily about their "experience," except perhaps if they have just eaten in the QSR. The notion of a prospective estimation of how they would enjoy the experience provides a whole new arena of research to understand the customer's mind.

We settled on a few things to capture the consumer responses, recognizing that the most valuable impression as an immediate one. We instructed the respondent to look at the stimulus and tell us how he felt, that moment, about going to a restaurant like the one shown on the screen. The respondent selected one of seven emotions, as we see both in the instructions (Figure 18.4), and in a screen shot for the test stimulus (Figure 18.5) on the next page.

We then asked a second question for the same stimulus about how frequently they would frequent this type of restaurant. (See Figure 18.6.) This second question is an evaluative question, similar to overall liking or purchase intent, but much more tuned to an experience. The second question assesses the acceptance of the experience by having the respondent tell us how frequently they would like to repeat it. Frequency and acceptance are not the same. Frequency is the more important attribute for a restaurant. It's more important to frequent a restaurant that you like only moderately, than to go to a restaurant on rare occasions, even though you might love it.

Welcome to the **Quick Serve Restaurant Survey!**

We are interested in your thoughts and opinions about various Quick Service restaurants and the meals they offer.

On the following screens, you will be presented with a series of designs for a Quick Service Restaurant and some meal ideas. You will be asked to rate each idea on the following 2 questions.

1) How do you feel about going to a Restaurant like this?

1 = Sad, 2 = Irritated, 3 = Calm, 4 = Neutral,
5 = Joyful, 6 = Relaxed, 7 = Energized

2) How frequently would you go to a Restaurant like this?
1 = Never....9 = Daily

You will also be asked just a few questions afterwards to help us understand your needs.

The survey should take about 15 minutes.

Please press '>>' to continue.

Figure 18.4 Orientation screen showing the respondents what will happen during the interview, and explicating the two rating questions. Note that the first rating question on emotion requires the respondent to select a feeling; the second question requires the response to scale his expected frequency of patronizing the restaurant.

The Overall Evaluation—What Drives the Person to Say "I'll Go Here Often?"

Let's look at the results of our experiment. We have presented each respondent with a unique set of 28 stimuli, each stimulus describing a QSR experience. The respondent rated each combination two ways: telling us how he felt about his emotional experience as he read the concept, and then telling us the frequency of going to the particular QSR displayed on the screen. Before we delve into the emotions involved with the experience, let's see whether we can figure out what drives the desire to frequent a restaurant.

In our previous treatments of these types of experiments, we have typically used a single evaluative

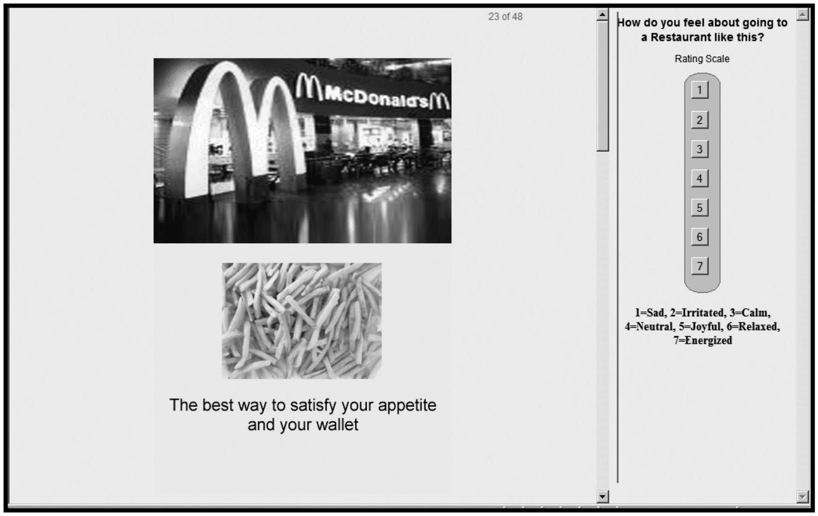


Figure 18.5 Example of a test screen, showing the restaurant (two pictures and text), along with the instructions to select the emotion

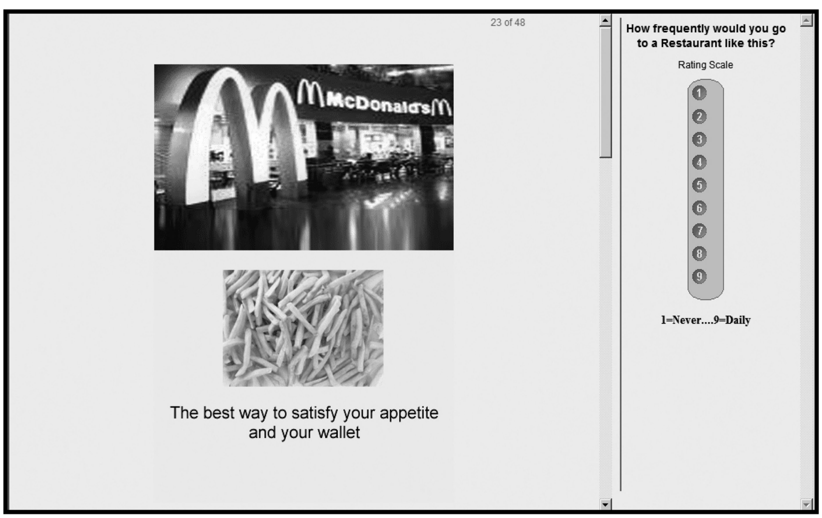


Figure 18.6 Screen shot showing the same stimulus as shown in Figure 18.5, but with the question changed to "frequency"

criterion, such as purchase intent or overall liking. Go to any company where these types of studies are run, and the first question is "how did my product (or idea) perform?" And so it is here, although we asked the frequency question second, after the respondent selected the emotion, we consider the frequency question to be the key rating.

The analysis of frequency follows the pattern we have adopted throughout. We modified the rating so that the high frequency ratings (7–9) become 100, and the remaining lower frequency ratings are transformed to 0. Then we run the equation, relating the presence/absence of the 12 experience elements to the transformed frequency (more frequent versus less frequent).

Table 18.2 How the different elements of the QSR experience "drive" stated frequency of patronizing the restaurant. The dependent variable is the percent top-3 box, or those who say that they will patronize the QSR relatively frequently.

	Total	Gender: Male	Gender: Female	Reason: Atmosphere	Reason: Price	Reason: Food	Reason: Recommendation
Additive constant	13	19	11	16	14	14	21
Restaurant picture							
Black Forest	–1	–2	–1	0	–2	–1	–5
59 St. Diner	–2	–3	–2	–1	–3	–3	–6
McDonald's®	–1	4	–3	2	–2	–1	–3
Burger King®	2	–1	3	2	2	2	–4
Food picture							
Salad	1	4	0	1	1	1	–3
Fries	–2	–2	–2	–2	–2	–2	–1
Cereal Milk	–1	–3	–1	–1	–1	–1	–1
Sandwiches	2	9	0	5	3	2	2
Message about restaurant							
Online and fast pickup	4	5	4	4	5	5	6
Satisfy appetite and wallet	3	1	4	2	3	3	4
Kids' meals	0	0	0	2	0	0	4
Dietary needs	3	4	3	3	3	3	9

Look at the results in Table 18.2. A few things become clear:

1. We can divide the population of respondents in different ways. We will look at gender. We will also identify mind-sets (i.e., groups of respondents who focus on different aspects of the experience).
2. The majority of respondents do not want to frequent a QSR daily or even quite frequently. We see this from the relatively low additive constant. Of course there are some differences. For example, males say that they will patronize the restaurant more frequently than do females. Respondents who describe themselves as making decisions on the recommendation of others feel that they will patronize the restaurant more frequently than respondents who say that they make their own decisions.
3. By and large the elements do not drive the respondents to say that they will frequent the restaurant. There are some exceptions (i.e., elements that do drive the respondent to say he will frequent the restaurant). For example, when it comes to males, one "driver" is a picture of sandwiches. For those who say that they take the advice of others, the message about special dietary needs also works well.
4. When we stop here, at the analysis of "frequency," we might conclude that most of the elements are about the same, performing only modestly. There are a few element-to-element differences and a few reasonably intuitive differences among the respondent subgroups. Yet, we are left with the feeling of "Is this all there is?" We will learn much more when we look at the emotional reactions to these elements.

Uncovering the Emotional Contribution of the Experience

What works when it comes to emotion? What does each element contribute? Are the results reliable? We saw that overall ratings provide us some information about what wins, but not enough about guidance.

Read the appendix to this chapter and you will see that the measurement of emotion is statistically reliable. We know that we can measure emotion with consistent results. Our method to assess emotions instructs the respondent to select the one word for each vignette that best describes the emotion for that vignette. The data reveals a distribution of such emotions across all of the 4,060 different concepts. Now the question is "what more?" Can we link emotions to situations, in the way that we did previously in Chapter 17 on coffee packages? What emotions emerge with our simulated "QSR experiences?"

The Emotions Our QSR Vignettes Create in the Respondent

The real focus of our design study is to identify what emotions we have elicited in respondents when we systematically varied the pictures and the messaging to create different simulated experiences or small telegraphic ads. We allowed a respondent to remain neutral as well, if he didn't feel that he had an emotional reaction to a particular vignette. Just how did the respondent react?

We can learn a lot about the design of experiences by deconstructing the different vignettes into the contributions of the 12 components that we systematically varied. As we have done before, we use ordinary least-squares regression, to create a simple algebra of contributions.

Each choice of an emotion is "all or none." The respondent could only select one prevailing emotion. So, our dependent variable for an emotion, for instance, "sad" is either 0 (*denoting not selected as the emotion for that particular vignette*) or 100 (*denoting selected as the emotion for that particular vignette*).

Following this definition of the dependent variable as yes/no for each emotion, let us create a simple model using our 4,060 vignettes. For each vignette we know which particular elements are present and, of course, which are absent. We had created the 4,060 vignettes by experimental design so that the 12 elements are statistically independent of each other. Also, we see that occasionally a silo is entirely absent from a stimulus or experience vignette.

Now that we have our statistical method to "model" emotions, we move to the actual data from our interviews. We know that we put in 12 different elements (4 restaurants, 4 foods, and 4 messages, respectively). What elements drive which particular emotions?

In the interests of clarity and didactics, we will repeat the steps that we used for the coffee study in Chapter 17, where we also investigated emotions.

1. Lay out the elements as predictors (independent variables).
2. Lay out the emotions as dependent variables.
3. Work with one emotion at a time. We will select the emotion "sadness" as the example. The same analysis will be done seven times, one for each emotion.
4. Relate the presence/absence of each experience element to the choice of sadness for that particular vignette (100 if sadness is chosen for that vignette, 0 if sadness is not chosen).
5. Build a regression model, *but do not use an additive constant*.

Following the five steps above, let us now look at the way each element "drives" an emotion response. The numbers in the body of Table 18.3 show the conditional probability, or percent of respondents who feel that the particular element (shown in the first column) drives the emotion. Thus, only 1% of the respondents felt "sadness" when they saw the picture of the Black Forest restaurant. Yet, 8% of the respondents felt "relaxed."

It's important to note that the respondents never saw an individual element in the experience vignette. The respondent only saw the full or partial vignette, and then selected an emotion. We infer the contribution of each element to each emotion.

Table 18.3 How the different experience elements in a vignette "drive" selection of a specific emotion. Numbers in the body of the table are the impact values or conditional probability of a respondent selecting the "column emotion" in a vignette that contains the "row element."

	Sad	Irritated	Calm	Neutral	Joyful	Relaxed	Energized
Restaurant picture							
Black Forest	1	3	3	16	6	8	1
59 St. Diner	2	2	3	16	7	6	2
McDonald's®	3	2	4	20	3	4	2
Burger King®	2	4	3	15	5	6	2
Food picture							
Salad	0	2	3	17	8	5	2
Fries	1	5	6	17	6	2	1
Cereal Milk	0	3	3	21	5	3	0
Sandwiches	0	2	3	16	7	7	3
Message about restaurant							
Online and fast pickup	3	6	2	8	9	5	4
Satisfy appetite and wallet	1	3	3	14	7	6	4
Kids' meals	4	7	4	18	3	0	2
Dietary needs	1	1	3	10	12	7	4

Following this line of thought, let us look at what emotions emerge from our study. We look here at the results from the total panel:

1. The emotion effects are low for the pictures of the restaurant itself. Two QSRs, neither is a fast food restaurant (Black Forest and 59 St. Diner), drive a sense of "joy" and "relaxation." McDonald's® does not. Burger King®, however, does drive a feeling of relaxation.
2. Pictures of food bring a sense of joyfulness. Fries also bring calmness.
3. Messaging drives the main emotional responses, both negatives and positives.
4. Pictures cannot make one sad or irritated. Messaging can irritate (online and fast pickup, as well as kids' meals irritate some respondents). Messages can also generate joyful and relaxed feelings.
5. Nothing appears to make a respondent either sad or energized.

6. The important thing about these data is that it is now possible to assess emotional potency of different items in a way that does not call attention to the item. By deconstructing respondent reactions to many vignettes, we have a method to measure the emotional reaction of a given respondent, a subgroup, or the entire population.

"Different Strokes for Different Folks"—Do All People React with the Same Emotions to These Elements?

A few paragraphs above we looked at the power of each element to "drive" frequency of patronizing the restaurant, at least based on the experience vignette. We didn't discover very much, except for a few intuitively reasonable exceptions. Most of the elements performed only modestly among the total panel, and the different groups of respondents identified by what they chose to be important to them. Yet, people do differ, and we ought to be able to show differences across people, if not in the overall evaluative criteria (frequency), then perhaps in the emotional response to the elements. It is quite possible that frequency, overall liking, purchase intent, and other integrative criteria miss some of the more nuanced aspects of the experience vignette. Let's see.

We ran the emotion models for all seven emotions, for gender, and for the four attitudinal groups (choose QSR based on atmosphere, price, food, and recommendation, respectively). We know that many of the emotions show little ability to differentiate among the elements. (See Table 18.3 above.) However, the negative emotion of irritation and the positive emotion of joy did discriminate among the elements for the total panel. Let's see how these two emotions emerge in complementary subgroups, in a way that the overall frequency rating does not.

Let's look at Table 18.4. We see the elements on the side, where they usually are. We see three sets of columns—total, the two genders, and the four groups of respondents defined by what they say is important to them when considering and frequenting a QSR. If an emotion is important for one subgroup but not important for the complementary subgroup, then we conclude that the emotion differentiates the subgroups.

To make the comparisons easy to do visually, we have italicized those cells whose values are 7 and above, corresponding to the inferred selection of the particular emotion (irritated or joyful) for a specific element (i.e., picture of McDonald's) by a specific subgroup. The greater the variation in percentages in a row for complementary subgroups, the more the emotion differentiates the subgroups.

We see a lot of alternate shading and nonshading in the rows, especially for the emotion of "joyful." We

Table 18.4 How the different elements in an experience vignette drive the selection of two emotions: irritated and joyful, respectively. The table shows the results for total panel and for various subgroups. Males and females are complementary subgroups. The four remaining subgroups were selected based on what the respondent said was important for selecting a QSR

	Total panel	Gender: Males	Gender: Females	Choose: Atmosphere	Choose: Price	Choose: Food	Choose: Recommendation
Emotion = Irritated							
Kids' meals	*7*	*8*	6	5	*7*	*7*	5
Online and fast pickup	6	*7*	5	5	6	6	4
Fries	5	4	5	*7*	4	5	*8*
Burger King®	4	3	4	7	2	4	–2
Black Forest	3	6	1	1	2	2	3
Cereal milk	3	4	4	5	4	4	6
Satisfy appetite and wallet	3	3	2	1	2	3	1
59 St. Diner	2	*7*	0	2	0	2	0
McDonald's®	2	3	2	3	1	2	3
Salad	2	3	2	3	1	2	5
Sandwiches	2	1	3	4	2	2	5
Dietary needs	1	2	1	–1	2	2	1
Emotion = Joyful							
Dietary needs	*12*	*10*	*11*	*10*	*11*	*11*	1
Online and fast pickup	*9*	*7*	*9*	*10*	*9*	*8*	6
Salad	*8*	*11*	*7*	*11*	*8*	*8*	*14*
59 St. Diner	*7*	3	*9*	*8*	*9*	*8*	6
Sandwiches	*7*	6	*7*	*9*	*7*	*7*	*19*
Satisfy appetite and wallet	*7*	*7*	*7*	*8*	*8*	*7*	*11*
Black Forest	6	4	6	4	6	6	2
Fries	6	*7*	6	*10*	*7*	6	*16*
Burger King®	5	*7*	5	4	5	6	6
Cereal milk	5	*7*	5	*9*	5	6	*16*
McDonald's®	3	4	3	3	3	4	4
Kids' meals	3	3	2	3	3	2	1

conclude, therefore, that emotions may help us better understand the reactions to the experience vignettes, and in turn may help us to design a better restaurant experience.

Taking Stock So Far—What Does "Emotion Modeling" Provide Us When Combined with Experimental Design of Stimuli?

We have just gone through an interesting exercise, trying to figure out the inside of the respondent's mind. As we have seen throughout this book, a productive way to understand the mind is through systematic variation of the test stimuli and the discovery of quantitative relations between what is varied and what is perceived.

Knowing the relation between the test stimulus and the respondent's interest or frequency rating identifies what works and what does not. We can use that knowledge to create new packages, and even new experiences. It's a matter of knowing how the features of the design "work."

Yet, the world of emotions opens up an entirely new realm for us. We saw in our treatment of coffee graphics designs that emotion can play a role. We see here in the presentation of the QSR experience that we can probe far more deeply into the respondent's mind. The results of the probe should help us design the package (a la coffee, in Chapter 18), or even the less tangible but equally important "dining experience," as we see here. And, thinking a little more into the future, we can extend the approach to the "shopping experience" as well. Rather than looking at emotions as responses to the QSR restaurant, we might look at emotions as responses to the store and the store layout itself.

We are only at the beginning of studies of emotion. Most of the work deals with measuring emotion. Our approach in this book is to go beyond measurement, just as we do with the overall evaluative criteria. We aim, instead, to use experimental design to understand and then to engineer.

Statistical Appendix—Analyzing the Emotion Data—A Quick Review

This appendix reviews the approach we used to understand and model emotions. We refer the reader to an earlier treatment in this book (Chapter 17 on coffee packages). We now repeat some of that, but go through the coding in much greater detail, as well as establish the reliability of the technique.

Recoding the Data

Following our strategy for graphics design, we create seven new variables, one variable for each emotion. Now let's go through vignettes or stimuli, one by one, 4,060 times, one pass for each vignette. We have "new" variables, corresponding to the seven emotions (sad … energized).

For each of these seven variables, put in the number "0" when the respondent did not choose that emotion for the particular vignette, or put in the number "100" when the respondent did choose that emotion. The "special sauce" here resides in the way you code the data. The data are ready for statistical analysis by regression. When you stand back and look at the spreadsheet, you will see the data pattern, shown in Table 18.5.

Running the statistical analysis is straightforward. Most off-the-shelf software packages have a regression package. The independent variables are the 12 elements (E01, E12 or Rest1 … Message4). The seven dependent variables are sad, calm … energized.

Run a separate model for each emotion, without an additive constant. The model is expressed by the simple equation:

$$\text{Emotion} = k_1(\text{Rest 1}) + k_2(\text{Rest 2})\cdots$$
$$k_{11}(\text{Message 3}) + k_{12}(\text{Message 4})$$

The coefficients that emerge are the impact values or, more technically, the conditional probability of an emotion being reported in the presence of the particular element. We are looking for high coefficients (i.e., around +7 or greater). These tend to be the predominant emotions for that element.

Establishing Reliability—Is the Emotion Model "Repeatable?"

Before we begin seriously analyzing experience using our emotion data, along with the frequency rating, it is always a good idea to establish that our measuring tool is reliable. By reliability we mean that it produces the same result two or more different times. We only ran the one study here, but we can establish reliability by using the "split half" method. We will determine whether or not people choose the same emotions, in general, in the

Table 18.5 Coding of the data for the first 16 test concepts or vignettes (ConOrder 1–16) seen by the first respondent (UID 669995; UID = unique identifying number). The first set of data columns (Rest1 … Message4) shows the way to code the presence or absence of the 12 elements. The column labeled Q1 shows which emotion was selected. The remaining seven columns show how to code that selection (0 = emotion not chosen, 100 = emotion chosen).

ConOrder	UID	Rest1	Rest2	Rest3	Rest4	Food1	Food2	Food3	Food4	Messge1	Messge2	Messge3	Messge4	Q1	Sad	Irritated	Calm	Neutral	Joyful	Relaxed	Energized
1	669955	0	0	0	0	0	1	0	0	0	0	0	1	7	0	0	0	0	0	0	100
2	669955	0	1	0	0	1	0	0	0	0	0	0	1	6	0	0	0	0	0	100	0
3	669955	0	1	0	0	0	0	0	0	0	0	1	0	4	0	0	0	100	0	0	0
4	669955	1	0	0	0	1	0	0	0	0	0	0	0	4	0	0	0	100	0	0	0
5	669955	0	0	1	0	0	0	0	0	0	0	1	0	2	0	100	0	0	0	0	0
6	669955	0	0	0	1	0	0	0	1	0	1	0	0	5	0	0	0	0	100	0	0
7	669955	0	0	0	1	0	0	0	0	0	1	0	0	4	0	0	0	100	0	0	0
8	669955	1	0	0	0	0	0	0	0	0	1	0	0	3	0	0	100	0	0	0	0
9	669955	0	0	1	0	0	0	1	0	0	0	0	0	4	0	0	0	100	0	0	0
10	669955	0	0	0	0	1	0	0	0	1	0	0	0	5	0	0	0	0	100	0	0
11	669955	0	0	1	0	0	1	0	0	0	0	1	0	2	0	100	0	0	0	0	0
12	669955	1	0	0	0	0	0	0	1	0	0	1	0	2	0	100	0	0	0	0	0
13	669955	0	1	0	0	1	0	0	0	1	0	0	0	5	0	0	0	0	100	0	0
14	669955	0	1	0	0	0	0	1	0	0	0	0	1	6	0	0	0	0	0	100	0
15	669955	1	0	0	0	0	0	1	0	1	0	0	0	6	0	0	0	0	0	100	0
16	669955	0	1	0	0	0	0	0	1	0	0	0	0	5	0	0	0	0	100	0	0

first half of the interview (stimuli 1–14), as they do in the second half of the interview (stimuli 15–28). If we find the same general emotional "profile" for the 12 elements in the first half versus the second half of the interview, then we can conclude that the average emotional profile of an element does not change during the interview, even with a large number of vignettes or concepts to rate. That is, the emotional responses are stable, at least across many respondents.

Look at Table 18.6, which shows the proportion of times that each emotion was selected for the set of 4,060 test stimuli (28 different stimuli tested by 145 respondents). The proportions are approximately the same, suggesting that the general emotional responses did not change across the course of the interview. This gives us some confidence that our measurement of emotions for what we call "experience" is at least stable overall.

So far we can say that overall the selection of emotions for experience vignettes does not show any noticeable drift. That is, during the second half of the interview, respondents are not becoming more irritated or energized, etc. The distributions of emotions are similar.

Table 18.6 Proportion of times each emotion was selected as a response to the concept, for the first set of 14 concepts, and then for the second set of 14 concepts. The patterns of selection are similar, so there is no "drift" in the pattern of emotions during the course of the interview.

	First 14 Stimuli	Second 14 Stimuli
Emotion		
Sad	4	5
Irritated	9	8
Calm	8	10
Neutral	41	43
Joyful	18	17
Relaxed	14	12
Energized	7	5

Let us now create the emotion "profile" of each one of the 12 elements (i.e., the proportion of times each emotion can be linked to each element). With 12 elements and 7 emotions, we have a total of 84 values or percentages that can be estimated. We did this for the total panel in Figure 18.6.

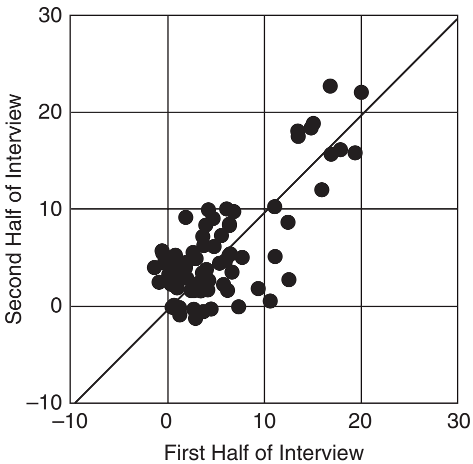


Figure 18.7 Reliability analysis of the 12 elements on the 7 emotions. Each circle is the percent of times that an element is estimated to drive a specific emotion, as estimated from the regression modeling. We plot here the contributions from the first half of the interview and the contributions from the second half. The 45-degree line corresponds to a perfect agreement of the two halves of the interview, indicating that there is no "drift" in the pattern of emotions during the course of the interview.

When we deal with reliability analysis, we want to reproduce Figure 18.7, but do it twice. The first time we estimate the values using only the vignettes 1–14. The second time we estimate the values using only vignettes 15–28. We now have 84 pairs of percents or utility (impact) values. Plotting the data is the easiest way to look at the agreement between the two halves, first half and second half. Let us plot these in a scattergram, to see whether these 84 pairs of elements line up.

Do the patterns remain the same? Do high impact values (a strong link between an emotion and an element) remain high across the course of the interview? If they do, then we can feel confident that our emotion mapping is reliable—we can connect experience elements with emotions. Figure 18.7 shows how the element impacts in the first half of the interview parallel the same element impacts in the second half of the interview. Our method of emotion modeling appears to be quite reliable. We can measure the emotional profile of an experience, use our coding scheme, then do the simple regression modeling, and finally trace that profile to the emotion contribution of the components.

References

Bagozzi, R.P., Gopinath, M., Mahesh, N., and Prashanth, U. (1999) The role of emotions in marketing. Journal of the Academy of Marketing Science 27(2), 184–206.

Ekman, P. (2003) Emotions revealed, understanding faces and feelings. Phoenix, London, UK.

Holbrook, M.B. and Batra, R. (1987) Assessing the role of emotions as mediators of consumer responses to advertising. Journal of Consumer Research 14(3), 404–420.

Izard, C.E. (1992) Basic emotions, relations among emotions, and emotion-cognition relations. Pyschological Review 99, 561–565.

Laros, F.J.M. and Steenkamp, J-B. E.M. (2005) Emotions in consumer behavior: a hierarchical approach. Journal of Business Research 58, 1437–1445.

Panksep, B. (1992) A critical role for "affective neuroscience" in resolving what is basic in basic emotions. Psychology Review 99(3), 554–560.

Pine, II, B.J. and Gilmore, J.H. (1999) The Experience Economy: Work Is Theatre and Every Business a Stage. Harvard Business School Press, Boston, MA.

Richins, M.L. (1997) Measuring emotions in the consumption experience. Journal of Consumer Research 24(2), 127–146.

Shaver, P., Schwartz, J., Kirson, D., and O'Conner, C. (1987) Emotions knowledge: Further exploration of a prototype approach. Journal of Personality and Social Psychology 52, 1067–1086.

Turner, T. and Ortony, A. (1992) Basic emotions: can conflict criteria converge? Psychology Review 99(3), 566–571.

Wood, O. (2007) Using Faces—Measuring emotional engagement for early stage creative. ESOMAR, Annual Congress, September 2007, Berlin, Germany.

Chapter 19

Homo economicus Rears Its Head

Introduction

We are surrounded by economics. Everywhere we look in packaging, we run into one or another influence of economics, whether subtle or direct, whether weak or, more often than not, strong. We need only look at the headlines today, where the talk is about increasing costs of the product, to realize that all too often the product doesn't cost quite as much as the *package* does.

Where Does Price Fit?

What happens when we put price into the equation? When we talk about price, what's the best way to do it? We could put price as one of the communication variables. In such a case, we present different test concepts or packages, some with varying prices, some without prices, and instruct the respondent to rate interest and other attributes as well. We will see that incorporating price into package "videos" makes a difference, with respondents preferring the product with the lower price. (See Chapter 13, Action and Reality: Using Video for the Package Experience.) Increasing the price typically reduces the interest in the concept or package. There are, of course, some individuals who say that increasing price actually "increases" interest in a product because the increased price signals higher quality (Jacoby, Olsen, and Haddock, 1971). This could be true, but it's probably not true in the world of lower- and medium-priced, fast-moving packaged goods.

Price is not a new topic to research. The effects of price as a package or information variable have been investigated many times before (see for example, Deliza et al., 1999, Guinard et al., 2001; Bower et al., 2003; Carneiro et al., 2005; Di Monaco et al., 2005). Here, we wanted to approach the price issue a bit differently, and as it will turn out, in a way that is quite instructive. *Instead of asking the respondent to evaluate a package with a price, let's instruct the respondent to inspect different packages and select an appropriate price!* This approach, of asking respondents to select the price appropriate for a stimulus, tells us about what he values in a package. The respondent may not know what the elements cost, he may like some elements more than others, but when he assigns a price to a package, the results essentially tell us what he values.

Doing the Price Study with Price as a Dependent Variable

We take our cue from the work on emotions. We saw that the respondent could select an emotion to describe how he felt when he inspected a concept. The respondent did not seem to have any difficulty doing this exercise. Although the critic might say that the respondent will do what he is told to do, even if it is incorrect, the emotion responses made intuitive sense.

With that in mind, let's move forward to price as a rating, taking our cues from the research on emotion. We know that the respondent has no trouble rating two attributes, the first being some type of overall acceptance (the key evaluative criterion) and the second being some "selection" act (previously emotion, now price). So, for price, our strategy lays out a number of different price points or ranges, relevant for the product, and instructs the respondent to select the "appropriate" price for each particular stimulus.

Although some experts caution that pricing is a difficult issue to deal with, from an executional point of view, pricing is rather straightforward. Perhaps the difficulty lies more in reconciling the difficult aspects of microeconomic theory with the relatively simple research approach of experimental design. Despite possible problems with expert opinions and theory-based issues, let's

Packaging Research in Food Product Design and Development Howard R. Moskowitz, Michele Reisner, John Ben Lawlor and Rosires Deliza
 ISBN: 978-0-813-81222-9

proceed in the fashion that we have done throughout this book.

Getting Our Feet Wet in Pricing with Combo Meals

Before we launch headfirst into packaging and pricing, let's look at how we approached the problem when we began to investigate pricing (Moskowitz et al., 1995; Ritson and Hutchins, 1995). The initial projects dealt with fast food, specifically combo meals. As we have done before, we systematically varied the different messages. In none of the messages did we even hint at price or value. Rather, we presented the different descriptions of the products and instructed the respondent to select the appropriate price from a scale.

In these types of situations, we didn't know what to expect when we began the studies. No one had had experience with pricing done quite in this way. Although it seems so obvious now, one of the nagging concerns was that the respondent would simply pick the lowest price we offered. Another concern was that the respondent would select one price, the price he was willing to pay, and then not budge from that price, no matter what we offered.

We ran the study, and to our relief, neither of our fears was confirmed. The respondents did not appear to simply assign the lowest price to the fast food "combo meal." Nor, in fact, did respondents select one price and refuse to move away from that price. Certainly when we inspected the "raw" data (experimental design and associated ratings), we found that respondents had a favorite or anchor price. Yet, in each case, there were some offerings that pushed the respondent's price up beyond that anchor price and some other offerings that pushed the respondent's price down.

Let's look at how to run this study. We began with the standard orientation page, shown in Figure 19.1. By now, we have seen many of these orientation pages, so the general form should come as no surprise. The only new thing in this orientation page is an explication of the rating scale. For this combo meal project, we were interested only in the price that someone was willing to pay. Later on in this chapter, when we deal with fish packages, we will incorporate another scale, this one for purchase intent.

When it came time to work with the data, we recoded the ratings. A rating of 1 was changed to 500 (cents), a rating of 6 was changed to 750 (cents), and so forth. We simply substituted the actual dollar and cents value for the rating scale, as the scale had been defined to the respondent. Afterward, we used regression analysis to relate the presence/absence of the 36 different elements to the total price one was willing to pay.

The results should not surprise. People have a good idea of how much they pay for combo meals. We see the

HELP US FIND THE BEST VALUE FOR A FAST FOOD "COMBO MEAL"

When you're hungry and don't have a lot of time, looking for a place to grab a quick bite can sometimes get a little tricky. Looking for a quick meal that's wholesome can even get trickier. With all the advertising and promotional offers the Quick Service Restaurant chains are always presenting to the public it's hard to know which ones are really ok. to have and where you get value for your money and which ones are not so great. And with so many different restaurants to choose from people are becoming more selective for various reasons and smarter when it comes to the price they pay for a Fast Food meal.

Well, we need YOU, the consumer, to help us find the fair price for a "fast food combo meal"!

This survey should take about 15 minutes to complete. You will be asked to rate each screen on a 1-9 scale by answering this question:

What price do you think is 'fair' to pay for this particular 'combo meal' at a Fast Food restaurant?' where

1=$5.00...2=5.50...3=6.00...4=6.50...5=7.00...6=7.50...7=8.00...8=8.50...9=$9.00+

Figure 19.1 Orientation page for pricing study—fast food "combo meal." (All prices in US dollars.)

results from the total panel in Table 19.1. These models will help us understand how the customer's mind works when thinking about value.

When we look first look at the results, we are struck by the fact that the price of the item is realistic ($5.46 for a combo meal, without any other information). This price is in the approximately correct ballpark for the class of combo meals in 2005. We should not be surprised.

What is more surprising is the nature of the elements that drive the price. We see only three elements that add

Table 19.1 Additive model showing how each element of the concept for a "fast food combo meal" drives the price a consumer is willing to pay. Conven = convenience, Descrip = description (All prices in US dollars.)

	Basic price (total cents for fast food combo meal)	546
Features	Generous and accommodating services … stop in before 7 PM on weekends and get $2 off your meal	18
Features	Receive a free movie ticket for every purchase over $20 throughout the summer months	16
Features	Get a free chocolate fudge or strawberry sundae when you buy a kid's meal	10
Features	Over stocked topping bar … load your meal with everything imaginable	9
Features	Karaoke nights every Thursday … stop by and sing a tune	9
Descrip	Golden chicken drumsticks accompanied with a pile of perfectly flavored fries … the ultimate meal to get you going again	7
Features	Classic styled Jukeboxes at each table … so you can have a little entertainment during your meal	6
Descrip	Juicy chicken fingers, toasty hot fries cooked to a crisp golden brown and a cool, bubbly soda to top it off	6
Descrip	A luscious burger … crispy, golden fries and a jumbo soda	5
Features	Food is free of additives, fillers and preservatives of any kind … purely fresh and wholesome	5
Descrip	Indulge in a big delicious burger, steamy fries and a refreshing soda	5
Features	Ask for extra cheese on the house … just the way you want it	4
Features	Extra fresh lettuce, plump bright red tomatoes and real American cheese always available at our toppings station	4
Emotion	A fun, satisfying meal for the whole family to enjoy	4
Emotion	So filling, it's guaranteed to slay your hunger	3
Features	A variety of spices and toppings … to maximize your meal's flavor	3
Conven	Custom built meals to meet your specific taste and nutritional needs	2
Emotion	Fresh and fast … so hard to resist	2
Features	Dinnerware and cutlery made from 100% recyclable materials	2
Features	A variety of flavored sauces … give your meal that extra kick	2
Descrip	A tender burger with a side of piping hot fries and a cool soda to wash it all down	2
Descrip	A savory burger dripping with grilled goodness accompanied by a stack of crispy fries and an icy soda	2
Emotion	Irresistible aroma and bold taste … tantalize your senses	1
Emotion	Fresh and steamy … dig in and enjoy	1
Conven	Prepared when you order with only the finest quality ingredients	0
Emotion	Delicious and mouthwatering … satisfies all your cravings	–1
Conven	Our chain of restaurants are conveniently spread throughout the country … no need to go far to get the food you want	–1
Chains	At Wendy's	–1
Conven	Friendly and efficient staff keep lines minimal … you will be eating within minutes	–2
Conven	Prompt service rapidly delivers exceptional food … puts a new meaning to "fast food"	–4
Chains	At Whataburger	–4
Conven	Made-to-order … get exactly what you want exactly the way you like it	–4
Chains	At Burger King	–5
Chains	At Jack-in-the-Box	–5
Chains	At In-N-Out Burger	–6
Chains	At McDonald's	–13

10 cents or more to the price. The remaining elements add less, and a great number of elements actually detract from the basic price one is willing to pay. These strong performing elements actually have little to do with the product, but rather are simply "stuff"—so-called promotional items or freebees:

Generous and accommodating services … stop in before 7 PM on weekends and get $2 off your meal
Receive a free movie ticket for every purchase over $20 throughout the summer months
Get a free chocolate fudge or strawberry sundae when you buy a kid's meal

Moving into Packaging: Price as a Rating versus as a Stimulus

Now that we have seen how the pricing research works with a simple product like a combo meal, let's move to packaging, using the same approach but adding the overall evaluative attribute as well.

The topic of the research is packaged fish, a topic that today is very "hot" because of the combination of general health concerns, the positive image of fish versus other meat, and finally some of the special nutrients that fish have (e.g., omega-3 fatty acids). That being said, we will confine ourselves to regular fish packages, and not deal with health store packages that emphasize nutrition.

We began this chapter with a case history showing price as a rating. We are going to continue with price as a rating for the first study with fish packaging. Then, we are going to use these prices, originally the rating values, now as elements in the packaging concept.

Our strategy of using prices twice, once as a rating and once as a test element in the package stimulus, will teach us a lot about prices. First we will learn how package elements drive the perception of price. Then we discover how package elements interact with price. The two experiments together will give us a more complete view of pricing.

There will be more, however. "Just for fun" we added a second question in each study:

1. When price is a rating, we instructed the respondent to rate how interested he would be in the fish package. When we finish analyzing these results, you'll learn both about the price of each element, the interest in each element, and the "expected value" of each element.
2. When price is an element, we instructed respondents to tell us the emotion (just as we did before), and then how interested he would be. So, from the results, you'll learn how price drives emotion as well as interest. When the price is high, does the emotion change?

Both of these experiments represent new directions for packaging research. As we have seen throughout this book, and as you will see when you read the business and scientific literature, the study of prices and packages are often left to those who "track" actual store sales, and try to relate sales to pricing. This is the so-called marketing mix analysis, which looks at actual sales history, but not at price as an experimental variable along with package features and communications.

What is the FAIR Price?

Price as a *Dependent* Variable—Fish Study 1

Our first study deals with four silos, each with four elements. We see an example of the package in Figure 19.2 with the pricing question. Figure 19.3 shows a closeup of the four pictures of fish. In this first study, the respondent first selected the appropriate price, as was done with combo meals, but right afterward rated interest. Although the purist might say that a respondent would pay more for a product/package he likes, we weren't sure. This study allows us to test that guess, which sounds right, but needs empirical verification.

Asking the Question—What We Had to Consider

Throughout this book we have given relatively little consideration to the way we asked questions, an attitude that would probably irritate some, and positively drive others into a frenzy. For the most part, the nature of a scale makes relatively little difference. Whether we instruct the respondent to rate interest and use the rating itself, or the top-3 box equivalent of the rating (recode, 7–9 → 100, 1–6 → 0), we will end up with pretty much the same answer. Winning elements will remain winners, losing elements will remain losers. Occasionally the performance of one element versus another may change somewhat, but we're likely not to change the "bigger" picture.

When we deal with pricing, however, matters might become a little stickier, more complex, and we might get the wrong answer. Instructing the respondent to rate price

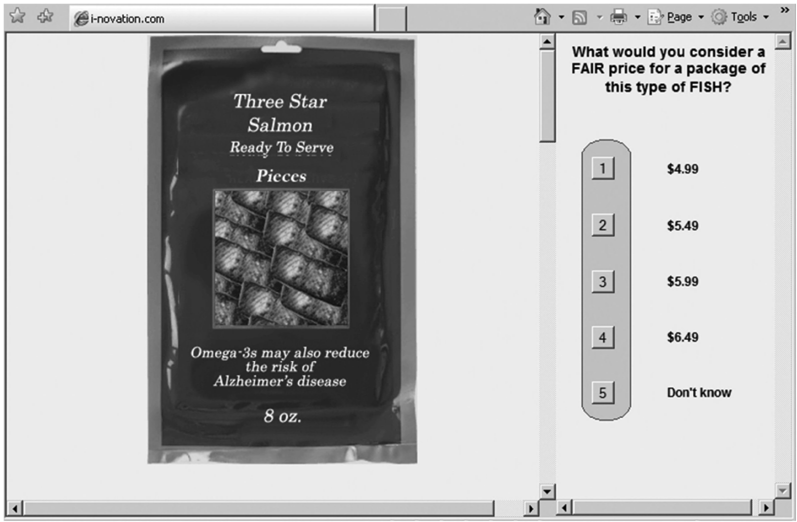


Figure 19.2 Example of a test screen showing fish package and pricing

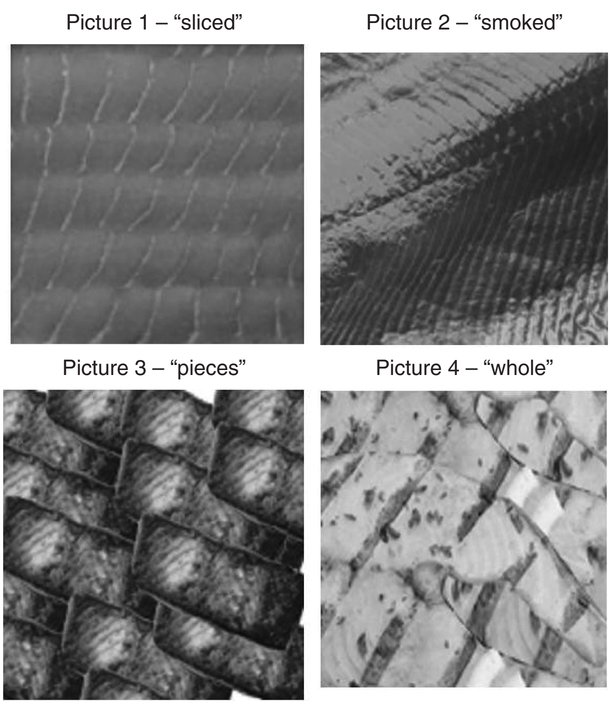


Figure 19.3 Closeup of the four pictures of fish

calls into play a lot of biases that are irrelevant when asking the respondent whether he likes or dislikes something. Specifically, how does one instruct a person to "rate price"? Which of these questions seems more reasonable, and what are the implications of each?

1. How much would YOU pay for this package of fish?
2. What is the FAIR PRICE for this package of fish?

The first question, How much would YOU pay for this package of fish? seems quite reasonable. After all, we are interested in price, so it's only natural to ask people how much they would pay. However, on thinking about this, we realized that a person might always opt for the lowest price. When we change the emphasis from what one wants to pay to what one thinks is the ***fair price***, we subtly change the "rules of the game." We move away from asking the respondent an opinion about his own behavior to instructing the respondent to "guess" what everyone else would say. Although we did not test the two questions, but rather went with a version of question 2, we were sensitive to the need to remove the respondent as an individual, and rather ask the respondent what he thought the merchant SHOULD do.

As we see in Figure 19.1, the question reads: What would you consider a FAIR price for a package of this type of FISH?

After selecting the price, respondents rated interest, using an anchored 1–9 scale, just as we have done throughout this book.

What the Data Tell Us from This First Study on Fish

Our analyses of these data begin precisely with the analysis we introduced at the start of the chapter. That is, we replace the selection of price by the actual price itself, and then run the model relating the regression analysis to presence/absence of the 16 elements to the ratings.

We'll start first by looking at the results from the entire panel of respondents, running one single model. Then we will look at some key subgroups of consumers. We've seen before that people don't necessarily differ from each other in their response to package features or concepts, at least when those features are benefits, and when the rating is interest or some other "overall evaluation." But what about price? We all grow up with price as an important determinant of our behavior, what we buy, and often even what we will consider. Does "fair price" for the fish packages vary at all with the type of person who is assigning the rating? We don't know, but we can answer the question in due course, and fairly easily with the data we have.

Let's first look at the results from the total panel, which appears in the first data column in Table 19.2.

As is our convention, we list the elements on the side, in order of silos. The columns comprise the attributes and key subgroups. We should keep in mind that the analysis works using multiple linear regression (ordinary least-squares) and that the data we use are recoded (for pricing, the rating points are replaced by "cents"; for interest, the data are replaced by 100 for ratings of 7–9, and by 0 for ratings of 1–6). See Table 19.2.

Table 19.2 Impact values (interest) and dollar price for each package element

	Interest	Price $	Expected Value $
Additive constant	–4	$3.44	–0.44
Name			
Three Star	3	0.01	0.18
Wild Alaskan Salmon	3	0.06	0.19
Kenai Wild Alaskan	3	0.02	0.14
Ocean Beauty Seafoods	1	0.04	0.06
Picture			
Slice	17	0.59	0.99
Whole	14	0.55	0.85
Smoked	12	0.62	0.75
Pieces	11	0.42	0.70
Benefit			
The American Heart Association recommends 2 servings of fatty fish per week as part of a balanced diet	5	0.02	0.20
High in beneficial omega-3 fatty acids	4	0.01	0.23
An important factor in reducing heart disease	3	0.03	0.21
Omega-3s may also reduce the risk of Alzheimer's disease	3	–0.01	0.18
Size			
3/4 lb. package	11	0.99	0.77
10 oz. bag	9	0.95	0.64
8 oz.	7	0.82	0.47
In a 4 oz. package	5	0.42	0.30

What is the story that emerges from Table 19.2? Again, it is easiest to list the results, rather than trying to make a coherent story out of what are really disconnected findings. Parenthetically, at this point, it is important to note that the approach we have taken is not the conventional approach of "weaving a story." There is no single story or a simple thread running through these experiments. We're not confirming or denying a hypothesis, quite the contrary. We are mapping out a territory, seeing what nature has in store for us as we deal with the topic of pricing in our experiments. The story emerges from the discrete points, from the results, and not the other way around. The story does not begin, and the facts do not fit gently into the story, as an artist would paint a picture. Our story is rather that of the picture painted by the pointillist and becoming clear only as we step back from the picture. See Table 19.3.

Here are eight findings:

1. When it comes to driving interest in purchasing the fish, without elements, the interest is close to 0. In fact, the additive constant shows the interest value to be –4. This makes sense. If we have nothing to show but a blank package, there should be no interest in purchasing because there is simply nothing to purchase. The ingoing product is "fish," but consumers

Table 19.3 Impact values (interest) and emotion for each package element

	Interest	Sad	Irritated	Calm	Neutral	Joyful	Relaxed	Energized
Price	6							
$4.99		2	7	1	13	3	1	1
$5.49	2	2	9	4	10	3	–1	0
$5.99	1	2	9	4	11	2	0	1
$6.49	0	2	9	3	12	1	0	0
Picture	–1							
Sliced		–4	–11	6	13	9	9	6
Whole	21	–4	–9	5	15	10	6	5
Smoked	17	–3	–10	1	17	10	7	5
Pieces	16	–2	–6	6	14	7	4	3
Benefits	12							
An important factor in reducing heart disease		2	5	5	14	1	0	1
The American Heart Association recommends 2 servings of fatty fish per week as part of a balanced diet	4	2	4	3	15	1	1	2
Omega-3s may also reduce the risk of Alzheimer's disease	4	1	5	3	15	1	1	2
High in beneficial omega-3 fatty acids	4	3	7	4	15	–1	0	1
Package Size	–1							
In a 4 oz. package		1	6	1	14	2	2	2
8 oz.	4	1	7	2	13	1	2	0
10 oz bag	3	1	5	2	13	2	3	0
3/4 lb. package	3	2	7	1	13	1	3	1

are not presented with any other basic benefits beyond the fact that this is a package of fish.

2. Only pictures and sizes drive interest in the package. All of the pictures show a positive impact, but there are differences among the pictures. Respondents pick up these differences, meaning that one can change the impact of the package by changing the nature of the product displayed on the package, or the product seen through a transparent plastic sheet covering the product. What is seen through the window makes a difference.
3. Respondents are more interested in buying the bigger-sized packages than they are in buying the smaller packages.
4. Moving to price, we see that without any other information present on the package, the respondent expects to pay about $3.44.
5. Brand names and benefits neither add much nor detract from the price that the respondent expects to pay.
6. Depending upon the particular type of picture (but not description of the fish), the respondent may pay up to an additional 62 cents.
7. The big activity occurs with respect to package size. Respondents change the amount they are willing to pay as a function of package size. For example, going from 4 oz. to 12 oz. increases the price the respondent would pay, from an extra 42 cents to an extra 99 cents.
8. Package size without associated price drives both interest and price willing to pay, as we see in Figure 19.4. As the package size increases, both the interest impact and the price one will pay increase, fairly linearly. We conclude from these relations that "size

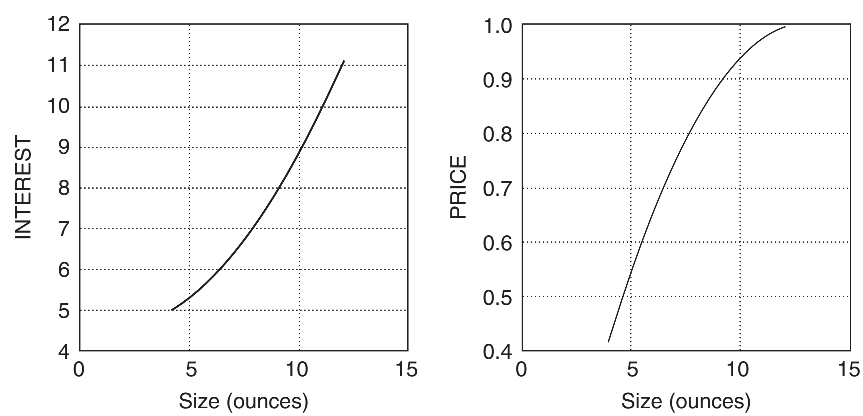


Figure 19.4 How package size in ounces "drives" both interest in the fish package, as well as price respondent is willing to pay for the package

sells," and people expect to pay more, even when there is no price information so that the respondent has to guess about the price. (In the next study, with price attached to the package, we will see that size ***may not*** sell quite as well!).

Beyond Price and Interest to the "Expected Value" of a Package Element

We have been dealing with two different response variables. One is the price that the respondent is willing to pay, expressed in U.S. dollars. The other is the interest in buying the product, expressed first in the anchored 1–9 scale, but then converted to a binary scale. The output of that second scale, interest, is a measure of the conditional probability of a person saying he would be interested in buying the package, if the element were present.

How do we combine these two responses to a single number? This number, according to economists, is called the "expected value." It is the product of the amount to be gained (viz., the amount to be paid for the package), and the probability that the amount will be paid at all (i.e., that the package will be purchased or not purchased).

Continuing this line of reasoning a bit further, we can think of the following three situations, and show how the economist computes the expected value:

Situation 1: The item sells for \$5 and the probability is 100% (i.e., 1.0) that the person will buy it. The expected value is \$ × 1.0 = \$. This makes sense. The person will buy it, and the yield from that purchase is \$5.

Situation 2: The item sells for \$10 and the probability is 50% that the person will buy. The expected value is \$10 × 0.50 or again \$5.

Situation 3: The item sells for \$20 but the probability is 0% that the person will buy. The expected value is \$20 × 0.00 or \$0.

Moving back to what we have in this study, we are working with a 9-point rating scale. We have ***defined*** ratings of 1–6 to be 0 on the probability scale (i.e., 0% probability of purchasing the item, no matter how much the person might be willing to pay for the package). And, in turn, we have further defined ratings of 7–9 to be 100% on the probability scale (i.e., interested, and therefore will purchase the package). This corresponds to 100%.

Of course, these assignments of probabilities to scale values are arbitrary, but they allow us to move beyond ratings into class membership—not interested in buying versus interested in buying. We are simply moving here beyond that membership in the class of 'not buy' versus the class of 'buy', and converting that membership into probabilities of purchase (0% when the rating is 1–6; 100% for ratings of 7–9).

Let us now move one step further, and for each test stimulus multiply the amount that a respondent would pay for the test stimulus by the probability of buying that stimulus. Keep in mind that the respondent selected a price, and that we have assigned ratings of 1–6 the probability "0," and ratings 7–9 the probability "1.0." Now we can go back to model the expected value of the items, which combines the price he is willing to pay and the probability of buying the package. The third column of results in Figure 19.3 shows the deconstruction of expected value into the contributions of the different elements.

As expected, picture and size play the greatest roles in driving expected value. Now, however, from the third data column in Figure 19.3, we discover the estimated dollar yield of each element in the package, by multiplying together purchase probability and amount he is willing to pay. That is, we know how each element will "deliver," not just how each element interests, and how each element would be valued without expected delivery taken into account.

Will I Buy at That Price? Price as an *Independent* Variable—Fish Study 2

We finish our exploration of price with perhaps the most conventional price study of all. In this second study with fish, we let the price act as an independent variable. This time we are interested in whether or not the respondent will buy the fish product at the price we offer. To make things interesting this time, we also include the selection of emotion, to discover whether price, and especially unit pricing, can drive emotional responses to the degree we seem to hear when we talk about price with friends and colleagues.

For this second study on fish, we tried to make the study as parallel to the first as we could, with a few strategic changes:

1. Three of the four silos were identical. These identical silos were picture, tagline, and size, respectively.

2. The silo of price in Study 2 replaced the silo of brand names in Study 1.
3. The four prices in the price silo were the same four prices that were used in Study 1 as responses ($4.99, $5.49, $5.99, and $6.49, respectively).
4. We first instructed the respondent to select the emotion that described how he felt when he evaluated the stimulus. Then we instructed the respondent to make the second rating, this time to rate interest in the package. This order paralleled the order in Study 1, where the respondent first selected the price, and then selected the interest rating. The interest rating was the second rating question in both cases.
5. To keep the parallelism, we used the same invitation letter, albeit with different invitees.

We see an example of the stimulus in Figure 19.5. The package looks very similar to the package that we saw in Figure 19.2. The main difference is that we now have both a price and a size as part of the test stimulus. For this particular stimulus, we show the seven emotions on the response scale, one emotion of which had to be selected for each stimulus. Thus, other than the switch in scales from price to emotion, and the insertion of price in place of tag line, we have the same type of stimulus.

What Pricing Contributes When It Is a Stimulus

The simplest analysis looks at the part-worth contribution of all of the elements. We see those results in Figure 19.4. The first data column shows the interest values. The remaining data columns show the seven emotions.

1. We calculated the interest values as we have done throughout this book, and as we did for Study 1, with ratings 1–6 coded as 0 and ratings 7–9 coded as 100. Then we ran the regression analysis, using ordinary, least-squares regression.
2. Since respondents selected one emotion from seven, we created the seven new variables for emotion, one variable for each emotion type. For each stimulus shown, the emotion selected was assigned the value 100, and the remaining six emotions were assigned the value 0. We then ran the regression analysis on the seven emotion variables. Just as in the case of coffee package and QSR experience, we did not use the additive constant when we modeled emotion by regression analysis.
3. In all cases, we related the presence/absence of the 16 elements to the dependent variable (interest, emotion).

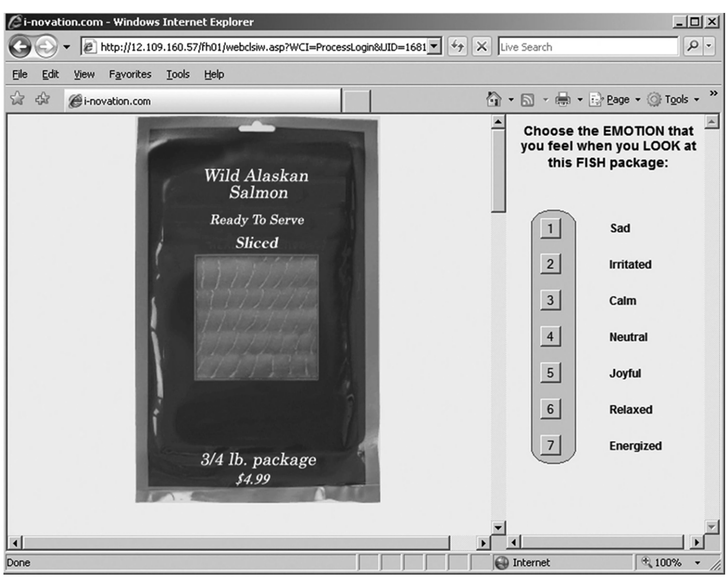


Figure 19.5 Example of the test stimulus, with price instead of tagline, and showing the emotion scale

There are five key observations to take away from Figure 19.3:

1. As the price increases, the interest decreases. However, the effect of price is small, showing only a 3-point range. Keep in mind that this change in interest occurs independently of the size of the package.
2. Pictures are still the most important drivers of interest, whereas benefits are not.
3. There is only a 2-point range for package size.
4. There are correspondingly small effects of elements on emotions, with the exception of picture, which drives the emotional reaction.
5. Price irritates the respondents.

Summing Up

The price of products is always one of those topics that bring out people's emotions. Prices are generally never right. An item is too expensive or too cheap. Yet as we see, we become accustomed to the prices we pay, and even if we complain, we often continue to buy the product, complaining all the way to the checkout counter and beyond.

What do our studies teach us about pricing? We see three major lessons from the three studies. These lessons have to do with the investigation of prices as a factor in the product, rather than with the actual prices we discover in the study.

Lesson 1: People do have a sense of the proper pricing of a food item. When you ask a person to select a FAIR price, they do so. People do not select the lowest price, or even one single price from the set. Even though critics may aver that people always want to pay as little as they can, people appear to be reasonable when it comes to selecting a fair price.

Lesson 2: Giving a person a price for an item as part of package does not overwhelm everything else. That is, price is not the dominant factor of a product, at least if the price is reasonable. We saw this in our second fish study. The change in price from \$4.99 to \$6.49 has relatively little effect on interest in the product. The picture of the fish is far more important. If the prices were extended beyond this range, to be far lower than \$4.99 or far higher than \$6.49, then perhaps we might see a greater contribution of price to interest.

Lesson 3: Increasing price may have an effect on emotion, specifically on irritation. As the price increases irritation increases as well, but again the effect is not dramatic. The connection between price and emotion certainly merits deeper investigation.

References

Bower, J.A., Saadat, M.A., and Whitten, C. (2003) Effect of liking, information and consumer characteristics on purchase intention and willingness to pay more for a fat spread with a proven health benefit. Food Quality & Preference 14, 65–74.

Carneiro, J.D.S., Minim, V.P.R., Deliza, R., and Silva, C.H.O. (2005) Labelling effects on consumer intention to purchase for soybean oil. Food Quality and Preference 16, 275–282.

Deliza, R., Rosenthal, A., Hedderley, D., MacFie, H.J.H., and Frewer, L.J. (1999) The importance of brand, product information and manufacturing process in the development of novel environmentally friendly vegetable oils. Journal of International Food & Agribusiness Marketing 10, 67–77.

Di Monaco, R., Ollila, S., and Tuorila, H. (2005) Effect of price on pleasantness ratings and use intentions for a chocolate bar in the presence and absence of a health claim. Journal of Sensory Studies 20, 1–16.

Guinard, J.-X., Uotani, B., and Schlich, P. (2001) Internal and external mapping of preferences for commercial lager beers: comparison of hedonic ratings by consumers blind versus with knowledge of brand and price. Food Quality and Preference 12, 243–255.

Jacoby, J., Olson, J.C., and Haddock, R.A. (1971) Price, brand name, and product composition characteristics as determinants of perceived quality. Journal of Economic Behavior and Organization.

Moskowitz, H. (1995) Consumer Testing and Evaluation of Personal Care Products. Marcel Dekker, Inc., New York, NY.

Ritson, C. and Hutchinis, R. (1995) Food choice and demand for food. In: Marshall, D.W. (Ed.) Food choice and the consumer. Blackie Academic and Professional, Glasgow, UK.

Chapter 20

Which Should I Choose?—Packages on the Shelf

Introduction

Much of what we have presented in this book deals with the early stage knowledge building and development of packages and their designs. We began to look at the issues of finished packages and selections in the chapter on eye tracking (Jacob and Karn, 2003). We tried to merge eye tracking with experimental design of packages. Yet, there is a whole other world to be explored. Many researchers use eye tracking to understand how packages "perform" on the shelf, at least in terms of attracting the customer's gaze. And it is to the shelf where we next turn, again using designed experiments to learn more about the dynamics of packages, this time on the shelf. We will deal with eye tracking, but for package design, in a later chapter (Chapter 22).

By this time we assume that the package has been researched, that the designer understands how the different features of the package work together, and that the key decisions have been made about pricing and the like.

There are a number of commercial research firms that do shelf evaluation, asking the respondent to find a product on the shelf, or even presenting the shelf and asking the respondent whether he saw the package, or what package he saw (Meyers and Lubliner, 1998: Pieters and Warlop, 1999). These efforts provide the marketer and merchandiser with a measure of package effectiveness. The adept merchandiser has "rules of thumb," regarding what types of packages do well, where the package will be most easily seen, and the like. Indeed, there are even technologies that allow the consumer to go through a virtual shopping exercise, entering the store, walking down the aisle, looking at a shelf, zooming in, picking up a package, looking at the package, and then selecting it or putting the package back on the shelf and moving on (Rhall, 2005).

The question we pose in this chapter is "what package will a person select?" This seems, at first, to be a relatively simple question. Why not simply put two or three packages on the shelf, and measure the number of people who select one package versus another, or even don't bother selecting any of them? This simplistic approach is perfectly valid. It's a beauty contest—put the contestant out on display and see whether the contestant is selected. Even better, put the contestant out in the presence of other contestants, and measure how many times the contestant wins.

An example of this type of beauty contest appears in Figure 20.1. The objective of the research is to find out how consumers respond to the products when they are on the shelf. As in most of commercial research, the emphasis in Figure 20.1 is on "testing" rather than on development. That is, Figure 20.1 shows the type of stimulus that a manufacturer would use in order to understand how the product performs on the shelf, which answers the immediate marketing and merchandising question. Research of the type suggested by Figure 20.1 does not, however, provide a body of knowledge about what drives selection, nor is the research meant to do so. Rather, the efforts go into measuring, rather than understanding, with understanding simply being a by-product of the research process.

Our focus in this chapter differs from conventional measurement approaches, just as our studies in previous chapters differed from the standard approaches of test and report. We are interested here in the earlier stages of design and selection from the shelf. It is one thing to understand which product will be selected. However, here we are interested in developing rules from the ground up. We will systematically vary the shelf by experimental design, measure responses (which product is selected), measure emotional reactions, and try to determine some rules or patterns. In this way, we

Packaging Research in Food Product Design and Development Howard R. Moskowitz, Michele Reisner, John Ben Lawlor and Rosires Deliza
 ISBN: 978-0-813-81222-9

Figure 20.1 Example of the shelf

continue the thrust of the book, focusing on the earlier stage aspects, where design is important, and where we want to discover patterns and, perhaps, rules.

The Allure of "Choice" in the World of Knowledge

Choice is deceptively simpler than simple ratings, yet at the same time profoundly more difficult. Think for a moment about presenting an argument for packaging to a corporate executive, say the marketing director, for a particular brand of tea. It's a lot easier to say that "our tea beats the competition, hands down" or "our tea is chosen more often than competition." It is a lot harder to say "our tea gets a rating of 8.2 whereas our competition gets a rating of 6.4." When it comes to choice, every business person intuitively understands that it's better to be chosen than to be rejected. That's a lesson from daily life that immediately transfers to business.

When it comes to ratings, we face an entirely different matter. Ratings by their very nature are "report cards." They are scores of a product or components of a product or package. After one hears about one's own rating, the natural question to ask is "what does this rating mean, is it good or bad, weak or strong?" One needs norms to interpret ratings. One does not need norms to interpret choice.

When we deal with the issues of choice in this chapter, we want to develop rules for choice, just as we developed rules for ratings. Thus, a lot of the thinking will be the same as we have used in previous chapters—systematic variation of the choice "set," instructions to the respondent to do a specific action (choose, rather than rate), and the deconstruction of the data to find relations between the stimulus and what was chosen.

Approaching the Problem Through Experimental Design

We wanted to approach the problem of choosing from a shelf in a fairly simple, direct way, without involving a lot of theory and without forcing too much theory onto what should be a simple study. There are many experiments that one can run dealing with packaging, such as where the packages are placed, how many facings a product has, the nature of the design of the package, and how it captures the consumer's attention, and so forth (Bernstein and Moskowitz, 2003; Dettman, 2001; Stinson, Jr., 1996). Indeed, the interaction of package design and the shelf can be a lifetime of work, as one seeks to understand the dynamics of what makes a good package when that package is head to head against other packages of the same type.

To simplify our task, we chose a relatively simple problem. The problem was the selection of a tea product as a function of where it is on the shelf in the competitive array, along with the flavor of the product and the price of the product. The objective of the study was to determine how each variable, location, flavor, and price drives the probability that a package will be selected.

The Experiment

Since we were interested in how the position in an array "drove" the selection of a tea product, we worked with just five variables, each having three options (see Table 20.1). The practical aspects about the choice of variables are that it is easy, that the responses generate "rules" or at least patterns, and that the experiment gives information beyond package design. The research deals with both flavor and price, beyond the product package. The experiment is not just a study that creates a piece of information that so often merely "reports but does not educate."

We created an artificial shelf comprising three locations. Each of the three locations could either be empty or contain one of the three tea brands. That is, a specific tea, such as Twinings™, could appear in the left, middle, or right position, or not appear at all.

Table 20.1 The five variables and the options for each one

Flavor 1	A1	Lemon
Flavor 2	A2	Lemon-Green
Flavor 3	A3	Orange
Tea 1	B1	Celestial Seasonings™—Left position
Tea 1	B2	Celestial Seasonings™—Middle position
Tea 1	B3	Celestial Seasonings™—Right position
Tea 2	C1	Lipton—Left position
Tea 2	C2	Lipton—Middle position
Tea 2	C3	Lipton—Right position
Tea 3	D1	Twinings™—Left position
Tea 3	D2	Twinings™—Middle position
Tea 3	D3	Twinings™—Right position
Price 1	E1	$2.39 for all three teas
Price 2	E2	$2.59 for all three teas
Price 3	E3	$2.99 for all three teas

To make things more interesting than merely the same product in different positions, we also varied prices and flavors as two independent factors. There were three prices, associated with all of the teas on the shelf, rather than associated with one particular tea. The third variable was the general flavor of the tea. With today's proliferation of teas, flavor has become an important factor in the marketing and merchandising. Each week there seems to be a new set of tea flavors, whether herbal or regular tea in exotic flavors, or increasingly frequently, mixes of flavors.

The actual experiment itself followed the same choreography of interviews that we have presented throughout the book. The respondent was invited to participate by means of an e-mail invitation. Those respondents who clicked on the embedded link were led to the orientation page, shown in Figure 20.2.

What the Respondents Saw

Since the objective of this study is to understand "choice" rather than ratings of a simple, single product, we had to create a simulated shelf, and at the same time ensure that there was sufficient variation among the stimuli on this simulated shelf so that we could develop some "rules" of choice.

Welcome to the **Tea** survey.

We need your opinions about different packages for tea.

On the following screens, you will see some ideas for packages of tea. You'll see them on a shelf like you would see them in a store where you typically purchase tea.

Please look at the entire screen before rating them. Some screens may seem similar. However, each one is Unique.

Please read each screen carefully and rate it on the following 2 questions:

If you were shopping for tea right now, which of these tea packages would you choose?
1 = Twinings, 2 = Celestial Seasonings, 3 = Lipton, 4 = Not Sure, 5 = Choose None

Choose the emotion that best describes how you feel when looking at this tea shelf?
1 = Sad, 2 = Irritated, 3 = Calm, 4 = Neutral, 5 = Joyful, 6 = Relaxed, 7 = Energized

At the end, we will also ask you a few questions to help us understand your needs better.

Please click the continue button to proceed to the study.

Continue

Figure 20.2 Orientation page

When creating a shelf, one's imagination is the only limitation. Yet, at the same time, since we were exploring choice in a systematic way, we thought it best to keep the shelf simple. Figures 20.3 and 20.4 present two simulated shelves, respectively, that were tested in the study. These were originally presented in natural color on the computer screen. That is, the products were scanned from the original packages so that the artwork was vertical.

In both Figures 20.3 and 20.4, all three teas appear. In a number of other test stimuli, one of the three teas was absent. This absence was deliberate, to allow us to understand the absolute contribution of each tea in each of the three positions. We have called this the "zero condition." It is necessary for the regression analysis to understand how each element drives the response. In Figure 20.3, we see an example of the three teas with the flavors listed, but not the prices. In Figure 20.4, we see

Figure 20.3 Example of a "shelf" comprising three products, all of lemon green tea flavor, but with no prices

Figure 20.4 Example of a "shelf" with three teas, no flavor specified, all priced at $2.39 for a box

an example of a different order of the three teas on the shelf, with the prices, but without the flavor on the top.

To reiterate, each test screen had a different order of teas, with occasionally one of the three teas missing, as well as having a single price for all teas shown, and a single flavor for all teas shown.

Setting Up the Data for Analysis

Look at Table 20.1, where we specify the independent variables. We have five silos, our independent variables, each with three elements. We recode these five silos and three elements into 15 different independent variables, much like what appears in Table 20.2.

The recoding is straightforward, following these simple guidelines:

1. We have one silo comprising three flavors. This silo generates three independent variables, A1–A3. For any test stimulus, there will be either one flavor for all three teas, or else the flavor will be missing. When a specific flavor is present, the independent variable corresponding to this flavor will be given the value "1," and of course the remaining two flavors will be given the value "0." For those vignettes having no flavor at all, all three flavors will be assigned the value "0."
2. We have one silo of three prices. This silo also generates three independent variables, with the same coding rules (1 = the particular price present in the stimulus, 0 = the particular priced absent in the test stimulus).
3. There are three teas, Celestial Seasonings, Lipton, and Twinings™, respectively. There are also three positions on the shelf (left, middle, right), as well as the product being absent from the shelf entirely. Thus, each tea generates three variables, corresponding to that tea present in the left position, in the middle position, or in the right position. For any specific shelf stimulus, only one of those three variables can take on the value "1" for the simple reason that a particular tea (i.e., Celestial Seasonings™) can appear at most in one position. The tea can either be at the left, in the middle, at the right, or absent entirely from that shelf. For example, in Shelf #3, the experimental design dictates that Twinings™ be entirely absent from the shelf.
4. Looking at the first stimulus (i.e., column labeled "Shelf #1"), we see that the shelf featured orange-flavored tea, priced at $2.99. On the shelf were Lipton on the left, Twinings™ in the middle, and Celestial Seasonings™ on the right.

Table 20.2 Example of the recoded design for four shelves, and the data from one respondent who was presented with each of the four shelves

	Shelf #1	Shelf #2	Shelf #3	Shelf #4
Independent variables under experimenter control				
Features of the shelf				
Flavor—Lemon	0	1	0	0
Flavor—Green Tea	0	0	0	1
Flavor—Orange	1	0	1	0
Price—$2.39	0	0	1	1
Price—$2.59	0	0	0	0
Price—$2.99	1	1	0	0
Left—Celestial Seasonings™	0	0	1	1
Middle—Celestial Seasonings™	0	0	0	0
Right—Celestial Seasonings™	1	1	0	0
Left—Lipton	1	0	0	0
Middle—Lipton	0	1	1	0
Right—Lipton	0	0	0	1
Left—Twinings™	0	1	0	0
Middle—Twinings™	1	0	0	1
Right—Twinings™	0	0	0	0
Dependent measures—respondent behavior				
Product Selection	4	3	2	1
Choose Twinings™	0	100	0	0
Choose Celestial Seasonings™	0	0	100	0
Choose Lipton	0	0	0	100
Choose None	100	0	0	0
Emotion Selected				
Sad	0	0	0	100
Irritated	0	0	100	0
Calm	0	100	0	0
Neutral	100	0	0	0
Joyful	0	0	0	0
Relaxed	0	0	0	0
Energized	0	0	0	0

Moving on to the data, and looking at the first four stimuli, we see the following information gathered from the study, for this particular respondent.

1. The respondent was instructed to make one of four choices (select a tea, or select none). For this first shelf, the respondent selected "none." For the second shelf (column marked "2"), the respondent selected Twinings™, and so forth. Allowing the respondent the "no choice" is meaningful because when one shops, there's no need to select a product.
2. The respondent was instructed to select one of seven emotion statements to describe the feeling (Laros and Steenkamp, 2005). There is no additional information to guide the respondent about what criteria to use when selecting an emotion. The respondent selected "neutral" as the appropriate emotion for the first shelf, and calm as the appropriate emotion for the second shelf, etc.

Analyzing the Results—What Teas People Choose and the (non) Effect of Position

The easiest analysis of the choice data measures how frequently the respondent chooses each of the three teas, or decides not to choose any (Buisson, 1995; Hoyer and Brown, 1990). Of course, the choice of the tea depends upon what teas are featured, how much each person likes the tea, what flavors are featured, and at what price the tea is offered.

Our first analysis will look at the selection of each tea, as a function of the variables under our control. In fact, we have 15 such variables under our control. The experiment was set up so that each tea appeared only once or not at all, in one of the three positions. This was a necessary "constraint," to ensure that there could be only one tea in any of the three positions, that two teas or three teas always appeared, and finally, that when a tea was missing, it was missing from all of the locations, not just from one end or another.

With this in mind, let's move into the analysis. We first discuss how to analyze this type of data, and then discuss what we found. First, keep in mind that the respondent could make only one of four selections: choose one of the teas or choose none of them. We had no gradations of choice. Rather, choice was a so-called "all or none" response. One either chose the tea or did not.

Look now at Figure 20.3. We see four columns of numbers. Each column corresponds to a specific choice of one of the three teas, or the option to choose no tea. Recall that after the respondent inspected the "test shelf," the respondent could select any one of the three teas, or decide to select none of the teas. This is the typical way one does a choice test, although to be fair, one might allow the respondent to select any set of the teas, and in any quantity. That "fairness" would simply complicate matters, forcing us to think of many more options, and making the results more cloudy, less clear.

In the left-most column in Table 20.3, we see the stimuli that were presented to the respondent on the shelf. These were different flavor names, different prices, and, of course, the different teas. Furthermore, the teas were in different positions, so the same tea, such as Celestial Seasonings™, could be at the left side, in the middle, or at the right side, or not present at all.

To make things easy, we will focus on the first data column, labeled ***Choose Celestial Seasonings™***. We

Table 20.3 How the variables under the researcher's control (flavor, price, tea) drive choice

Stimulus present on the shelf	Choose Celestial Seasonings™	Choose Lipton	Choose Twinings™	Choose None
Constant	23	22	21	34
Flavor				
Flavor—Lemon	–4	1	–4	7
Flavor—Green Tea	–3	–4	–2	9
Flavor—Orange	–2	–9	–3	14
Price				
$2.39	1	1	1	–3
$2.59	–1	2	1	–2
$2.99	–1	0	–1	1
Celestial Seasonings™				
Left	18	–7	–8	–3
Middle	19	–7	–9	–3
Right	16	–3	–10	–3
Lipton				
Left	–8	28	–6	–14
Middle	–7	28	–8	–13
Right	–10	31	–10	–12
Twinings™				
Left	–6	–6	13	–1
Middle	–7	–4	12	–2
Right	–7	–2	11	–2

want to see what factors on the shelf drive a respondent to select the Celestial Seasonings™ package. We will also comment on some of the other teas as well, as we move through the data.

1. The additive constant is the conditional probability or the percent of the respondents who would select Celestial Seasonings™ without any additional information. The number is 23, meaning 23% would select Celestial Seasonings™ if they knew nothing else.
2. The additive constant for all four options should add to 100, which they do. The reason is simple. These are the four options that are available. So, knowing nothing else, the total percent of respondents who select one of the four options MUST BE 100. There are no other options.
3. Let's now turn to flavor. All three flavors drive consumers away from selecting Celestial Seasonings™. The effect of flavor is quite small. Only when we get to orange do we see a large effect of flavor, with the effect being to "drive to no choice." Orange flavor is clearly a negative factor, especially when it is associated with Lipton. Now for the interpretation, which is not part of the science but a value added from the researcher. Perhaps this strong, negative interaction, between Lipton and orange flavor is due to the fact that Lipton is not usually paired with orange flavor.
4. We now turn to price, which again plays little role. The prices were all the same, so perhaps the real role of prices emerges when the price differs (comparison shopping), or when the price is so high that the consumer decides not to buy at all. For our tea study, here we see that in the range of \$2.29 to \$2.99 as the price moves to \$2.99, we move toward choosing none. That is, when the price is \$2.39 we have 31% choosing none (additive constant of 34 and impact of −3, or 31. When we move toward \$2.00 price per unit, we move toward 35% choosing none (additive constant of 34 and impact of +1, or 35).
5. We move now to the three teas. Our first question is whether the location on the shelf makes a difference. It makes no difference at all where the tea is located. For each tea we see no change in the probability of choosing the tea whether the tea is in the left, middle, or right position.
6. Among the three teas, Lipton is the strongest, Celestial Seasonings™ is weaker, and Twinings™ is the weakest.

Emotions and Selection

The second rating required the respondent to select one of seven emotions (Wood, 2007). Looking at Figure 20.5, we see that the most frequent emotions are calm and neutral. We should expect that the key emotion to turn up with the selection experience is neutral or calm. Thus, we should be looking for exceptions to this pattern.

Our basic analysis of emotions follows the same pattern of analyses that we have done for choice. As we did previously for coffee (see Chapter 17), we recode the selection of emotion into seven different scales, one for each rating. We then look at each of the 6,840 different stimuli, to determine which emotion was selected for that particular stimulus. The emotion that was selected is then assigned a value of 100 for that stimulus. The remaining six emotions were, by definition, not selected, and are assigned values of 0.

With this recoding, we are now in a position to determine whether there is any effect of flavor, price, or tea on the emotions that a person experiences when making a selection, or at least a virtual selection. The analysis follows what we have done for the other emotion studies. We run the standard, ordinary least-squares regression, with no additive constant. We have 15 independent variables (flavor, price, tea brand by shelf location), and 7 dependent variables (the 7 emotions).

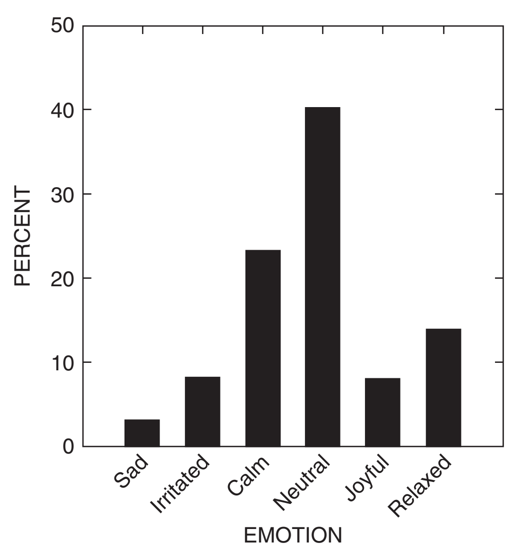


Figure 20.5 Distribution of emotion "ratings" across the full set of 6,840 test stimuli

Table 20.4 How the shelf elements (flavor, price, brand) drive emotions when the respondent has to select a tea from the shelf

	Sad	Irritated	Calm	Neutral	Joyful	Relaxed	Energized
Lemon	1	6	1	11	1	1	0
Green	2	6	1	11	0	1	1
Orange	2	7	0	12	0	1	0
Price $2.29	1	3	5	10	1	1	0
Price $2.59	1	5	4	9	0	1	0
Price $2.99	2	7	2	9	1	0	0
Celestial Seasonings™ Left	2	2	5	6	0	4	1
Celestial Seasonings™ Mid	2	1	6	6	0	6	0
Celestial Seasonings™ Right	1	3	7	5	1	4	0
Lipton Left	–1	–6	8	6	5	6	3
Lipton Mid	–1	–6	8	7	5	8	2
Lipton Right	–1	–5	11	4	5	6	2
Twinings™ Left	0	0	4	12	2	1	2
Twinings™ Mid	0	0	4	12	2	3	1
Twinings™ Right	0	1	8	10	1	0	2

We can summarize the emotion results by a single table, Table 20.4.

The data suggest some very simple patterns:

1. There is very little effect from flavor. Flavor simply doesn't drive emotional reaction, except for irritation.
2. Price also does not seem to have any effect. Even though we have seen previously that price does generate an emotion (at least for the dining experience, see Chapter 7), it is clear from these data that there is no effect of price on any emotion. It may well be the case that we did not extend the price high enough or low enough to show an effect.
3. Lipton tea is the only tea that generates emotions in the choice task. These emotions are calm and relaxed.
4. However, the shelf location of the tea (left, middle, right) does not have any effect on emotions, or in the case of Lipton, has equal effect on the emotions. Thus, again we see that location as we have operationally defined it here (eye level, direct gaze in front of respondent) makes no difference when the shelf is of limited size, with one facing of each product.
5. None of the teas drive the emotion sadness nor energized.

Summing Up—What Does Choice Research Begin to Tell Us?

When we deal with choice, we deal with a different world (Davies, 2001). Choosing is not the same as rating. Choosing means a yes/no response—either choose the product or not choose. There are no gradations in choice.

That being said, we learn from this first and rather basic experiment that respondents can make choices, and that the choices appear to be reasonably consistent, that price and flavor do not influence choice when they apply equally to all of the teas. We also learn that the brand of tea is the most powerful determinant of choice as well as emotion. Finally, we learn that position makes no difference, at least in the fashion operationally defined in this project (frontal view, equal height, one facing, small shelf).

Of course, this simplistic experiment is only the start of what itself could be an entirely separate, self-contained array of studies using choice as a measure and using packaging as the stimulus. Researchers are only at the start of the question. There is much more to come and far more exciting results to be had. It is just a matter of resources, vision on the part of researchers, and the recognition that experimental design can be applied to the shelf as directly as it is applied to the single package itself.

References

Bernstein, R. and Moskowitz, H.R. (2003). The Marriage Of Graphic Design & Research—Experimentally Designed Packages Offer New Vistas & Opportunities. In: Persuasive Imagery: A Consumer Response Perspective. Scott, L.M. and Batra, B. (Eds.) Lawrence Erlbaum Associates; Mahwah, NJ, 337–348.

Buisson, P.D. (1995) Developing New Products for the Consumer. In: Marshall, D. (Ed.) Food Choice and the Consumer. Blackie Academic and Professional, London, UK, 182–216.

Davies, S. (2001). Food Choices in Europe—The Consumer Perspective. In: L.J. Frewer, E. Risvik, and H. Schifferstein (Eds.) Food, People and Society—A European Perspective of Consumers' Food Choices. Springer-Verlag: Berlin, Germany, 365–380.

Dettman, D. (2001). The whole package—When it comes to boxing up your product, don't cut any corners. Available online at http://www.entrepreneur.com. Accessed 12/05/08.

Hoyer, W. and Brown, S. (1990) Effects of brand awareness on choice for a common repeat purchase product. Journal of Consumer Research 17(2), 141–148.

Jacob, R.J.K. and Karn, K.S. (2003) Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises. In: Radach, R., Hyona, J., and Deubel, H. (Eds.) The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research North-Holland/Elsevier, Boston, MA, 573–605).

Laros, F.J.M. and Steenkamp, J-B.E.M. (2005) Emotions in consumer behavior: A hierarchical approach. Journal of Business Research 58, 1437–1445.

Meyers, H.M. and Lubliner, M.J. (1998) The Marketer's Guide to Successful Package Design. McGraw-Hill; New York, NY.

Pieters, R. and Warlop, L. (1999) Visual attention during brand choice: The impact of time pressure and task motivation. International Journal of Research in Marketing, 16(1), 1–16.

Rhall, T. (2005) The Effectiveness of Online Advertising. Available online from http://info.hktdc.com/acninsights/104_01.htm. Accessed 12/05/08.

Stinson Jr., W.S. (1996) Consumer packaged goods (branded food goods). In: Product, new product development, product development. Rosenau, Jr., M.D., Griffin, A., Castellion, G.A., and Anschuetz, N.F. (Eds.) John Wiley & Sons, New York, NY, 297–312.

Wood, O. (2007) Using Faces Measuring Emotional Engagement for Early Stage Creative. Available online from http://www.brainjuicer.com. Accessed 12/05/08.

Part V

Temptations of Technology

Packaging Research in Food Product Design and Development Howard R. Moskowitz, Michele Reisner, John Ben Lawlor and Rosires Deliza
 ISBN: 978-0-813-81222-9

Chapter 21

Response Time

Introduction

Psychologists studying how quickly people respond to stimuli believe that they are dealing with one of the most important variables that one can measure in the search to learn what's going on "in the respondent's head." Indeed, it is probably not an exaggeration to say that the psychological construct of response time, developed by F.C. Donders in 1868, is to experimental psychology what blood pressure is to medicine—a key indicator, filled with meaning, because response time is signaling that the brain is working, doing things, processing information, making decisions. Of course, there is nothing intrinsic to the response time measure itself. We are only measuring the length of time between stimulus and response. Rather, it is what this measure may tell us, what other behaviors and attitudes correlate with it, and what we might hypothesize is going on.

There is a wonderful history and stream of current research involving response time. Let's dip into the stream a bit, get wet, and look around. To help us, let's first use the organizing principles that some modern researchers have created. That will help us in the journey. Tourangeau (2000) described response time information processing in four stages, two of which are interesting in general and set the stage for us: interpretation and information retrieval. This is the theory, the bare bones of what should be an interesting area when it comes to package research.

Let's move on to some of the thinking that has appeared in the published scientific literature. Since being recognized as an important "marker" variable, response time has been the subject of countless experiments. Simplicity and relative ease of measurement made response time attractive to researchers in many disciplines, including psychology, marketing (particularly in attitude research), and political science (Tyebjee (1979a, b; Aaker et al., 1980; Bassili and Fletcher, 1991; Ratcliff and Rouder, 1998; Mulligan et al., 2003). For instance, Tyebjee (1979b) used response time to measure conflict and brand choice for different beer brands, Mulligan et al. (2003) recommended the use of response time in public opinion research to demonstrate the influence of attitude accessibility on political attitudes and behavior. And the list could go on and on and on.

The links between response time, attitudes, and behavior have been attributed to the fact that attitudes that are expressed quickly are more probably accessible, more likely to be elicited automatically, and more likely to guide choice behavior (Bassili, 1995; Fazio, 1994; Fazio and Williams, 1986). But these behaviors come from somewhere. They must be guided. Presumably the response time prior to behavior is an indication that mental processes are occurring that will guide the behavior just afterward.

The attempt to deconstruct response time into the time to process the components of a complex stimulus is, of course, a natural outcome of research efforts. If we can measure interest in a complex vignette and deconstruct that interest rating to the part-worth contribution of components, why not perform the same deconstruction with response time in place of a rating or in addition to a rating? The analysis shows how much of the response time cannot be allocated to specific stimuli, how much can be allocated, the impact of interesting elements versus boring ones, the impact of stimulus length (number of letters), and so forth (Moskowitz et al., 2000; 2001).

Just from these citations alone, we see that researchers in the more recent past have taken this simple measure of latency between stimulus and response, using it to hypothesize all sorts of underlying psychological mecha-

Packaging Research in Food Product Design and Development Howard R. Moskowitz, Michele Reisner, John Ben Lawlor and Rosires Deliza

 ISBN: 978-0-813-81222-9

nisms. So there is every reason to use response time for packaging research, which has been done already, of course, and merge response time measures with experimental design, which hasn't been done frequently at all.

Pragmatics and Thoughts About Experiments

In the studies we cite in this book, we measure response time in the simplest possible way, using the computer and the Internet. It was part of the vision we had, to introduce research ideas that would be scalable, easy, doable worldwide.

Given the nature of today's technology, respondents differ in the bandwidth available to them on the Internet. So, in light of those massive differences, we had to develop a method that was relatively independent of bandwidth. We developed this method several years ago, at the start of our projects on the Internet, really in order to speed up the interview process, but as it turns out, in a way that lets us measure response time effectively.

The process is relatively straightforward, but produces a wealth of data that can be productively analyzed. At the start of the interview, as we introduce the topic to the respondent, we download the specific elements that we later combine into package concepts. The download is done in the "background," in order not to interfere with the respondent's task. These elements are then reassembled quite quickly "on the spot," by the respondent's own computer.

One consequence of this technology is that, with fair accuracy, we measure the length of interval between the completion of this "local assembling of the concept" and the first key press, which signals a rating. We measure this interval in tenths of a second.

With this measure of response time, that is easy to acquire and independent of Internet bandwidth, we can now address questions such as the length of time needed to respond to a concept, the deconstruction of the response time into the contributions from the different elements, and, finally, differences among people in terms of the way that they process information.

What We Knew *Before* These Studies and How That Knowledge Informed Our Efforts

Before we began serious work with consumers using the Internet, we had already worked with computer-aided presentation of stimuli. Our initial forays into response time took place in the mid-1990s, when we worked with the larger version of our approach, called IdeaMap®.

Some of our earlier work involved adults and children, responding to concepts about toys. This work has been already written up as a chapter in a book about concepts (Moskowitz, Poretta, and Silcher, 2005), in this Blackwell series of books about products, concepts, and, now, packages. In that study on toys, as well as in other studies on coffee, we discovered, as one might expect, that longer concepts with more text resulted in longer response times, because respondents needed time to read the text and assign their ratings. We also discovered that there were differences among the patterns of response times generated by adult women versus children, suggesting different styles of assimilating information.

With this sketchy initial introduction, let us now go further into an analysis of response time for two studies of fish packaging that we ran when preparing this book. The analysis comes from the two fish studies that we ran to investigate pricing. See Chapter 19.

For this chapter we will ask three sets of questions and answer them with data:

1. How long does a person look at the stimulus before he responds? This period of time is called "latency of response." Is a person consistent in his response times, or does it change?
2. What elements "hold the eye" and drive response time?
3. Do segments of respondents exist who show different patterns in the amount of time they spend looking?

How Long Does a Person Look Before He Responds?

Let's begin our analysis by looking at the two response times. Recall that in the two fish studies, the respondents evaluated packages comprising elements chosen from four silos. Each respondent evaluated 36 different packages, first selecting the appropriate price (Study 1) or the appropriate emotion (Study 2). Then the respondent rated interest as the second question. We also measured response time for each study; and we introduce the analysis in this chapter.

Does an Individual Demonstrate a Consistent Style that Can Be Evidenced Through Response Time?

Our first question has to do with the individual styles, at a very general level. Specifically, does a specific person use a characteristic style, so that the person takes either a long time or a short time to read and rate the test stimulus?

We are going to answer this question at the level of the individual respondent by comparing the median response time for the first rating question (select the price or select the emotion) to the median response time for the second question (rate interest).

Let's look at the pattern shown in Figure 21.1. Each filled circle in Figure 21.1 corresponds to the median of 36 response times for each question, from a single respondent. Recall that in that study the respondent evaluated 36 different test stimuli, allowing us to measure 36 pairs of response times. For each test stimulus we defined the two response times, as follows:

Response Time 1 was defined as the time between the full presentation of the test stimulus and the respondent's first rating, whether that was the selection of the appropriate price (Study 1) or selection of the appropriate emotion (Study 2).

Response Time 2 was defined as the time between the first response and the second response (interest).

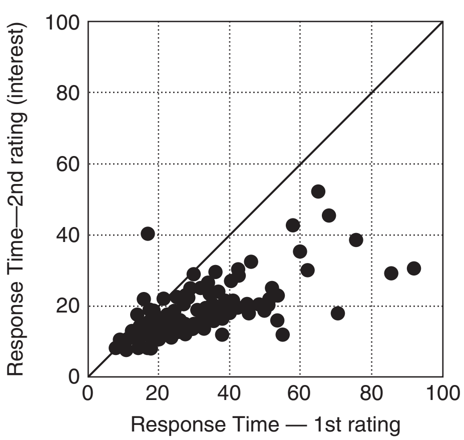


Figure 21.1 Comparison of individual median response times for the second question versus the median for the same person from the first question. Each filled circle is from an individual respondent, who evaluated 36 stimuli. The data come from two studies on fish, where the respondent rated 36 stimuli on two attributes. The 45-degree line shows equal response times. The two studies show similar patterns. The second response time (evaluate interest) is almost always faster than the first response time (select price or select emotion, depending on the particular study).

The 45-degree line in Figure 21.1 shows what to expect when the two response times are equal. *In general, we can say that a person who takes a long time to rate the first attribute will take a long time to rate the second attribute. However, the second rating is faster than the first rating, which makes sense since the respondent has already read, intellectually processed, and then evaluated the test stimulus.*

Do People "Learn" to Respond Faster with Practice?

When a respondent begins to evaluate stimuli on the computer, the most natural behavior is to sit and read the stimulus thoroughly. Over time, however, the respondents become more test-wise and learn how to take in the information in a more efficient way.

Do we see any evidence of this learning behavior for our study? That is, over the course of the 36 test stimuli, do respondents become faster?

One way to answer the foregoing question divides the response times into those for the first set of stimuli (i.e., the first half of the packages, stimuli 1–18) and those for the second half of the packages (i.e., stimuli 19–36). We then compare the distribution patterns of response times for these two sets of stimuli. When the distributions look similar, the odds are that there is no "learning to judge." The respondent does not seem to become more efficient. On the other hand, if the distribution of the later-evaluated stimuli is narrower, with a lower mean, we conclude that the respondents learn and become more efficient.

We see evidence for this learning in Figure 21.2. In the first half of the interview, the response times distribute differently than they do during the second half of the interview. In the second half we see far more "short" response times (See page 224).

What Holds the Eye—How Elements Drive Response Time

We know from our foregoing analyses that people differ from each other in the patterns of their response times. We also know that stimuli differ. Up to now we have seen that people are somewhat consistent, that there is

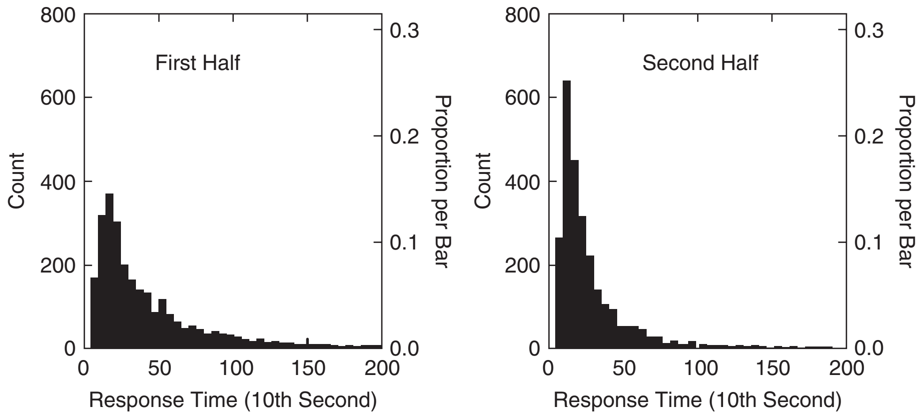


Figure 21.2 Distribution of response times for first and second halves of the interview. The distribution suggests that on the average, the respondents answer more quickly later in the interview. These data come from the first rating (select the appropriate emotion).

learning over the course of the interview with that learning affecting response time, and of course we know that there is a great deal of variability in response times.

Moving forward, the next question to address is "What drives response time?" Can we allocate the number of tenths of seconds that a person spends evaluating a stimulus to the different elements of that stimulus? Can we identify the number of tenths of seconds that each stimulus element might demand?

Throughout this book runs the thread of experimental design, which states that perhaps the best way to truly "understand" what drives a perception or a response is to discover the quantitative relation between that perception or response and factors under the experimenter's control. This point of view, an offshoot of the philosophical tradition known as *operationism* (Bridgman, 1938), applies to understanding response time.

Let's now answer the question "What drives response time?" We can allocate the response times we observe to elements by using regression analysis, in the way that we have done throughout this book. There are two minor departures from previous analyses, however, that are necessary:

1. The dependent variable is the number of tenths of a second for each test stimulus. Furthermore, since in theory the respondent could have waited a minute or two and done something else, we had to truncate the response time to a meaningful range. Using judgment, we eliminated from the analysis those response times greater than 300 (i.e., 30 seconds). We assumed that if the respondent took more than 30 seconds to rate a stimulus package, more than likely the respondent was multitasking, doing something else, and not paying attention to the particular stimulus being evaluated.
2. We used a form of ordinary least squares with *no additive constant*. This form of analysis assigns each element a specific number of tenths of seconds.

Armed with this way of thinking about the contributors to response time, let us now look at the number of tenths of seconds for each study. We will put all of the data in Table 21.1 so that we can compare across the two studies as well as within a study across the two rating attributes. (See page 225).

Looking at Table 21.1, the first thing that should strike us is that most of the response times for the first question (select price, selection emotion) are longer than the corresponding responses times for the second question (rating of interest). We saw that difference in Figure 21.1, where the response times for the second rating question are universally lower than the rating times for the first question.

That's the simple, general result. What about particulars? Looking closely at Table 21.1, we may conclude the following as a first approximation of what's going on:

1. Response times are similar with two different task "demands": When the same elements (pictures, benefits) are tested in two different studies, for the most

Table 21.1 How the response time breaks down into the contributions of the four elements. Numbers in the body of the table are tenths of seconds.

	First	First	Second	Second
	Select Price	Select Emotion	Rate Interest	Rate Interest
Brand				
Wild Alaskan Salmon	13	NA	8	NA
Kenai Wild Alaskan	15	NA	8	NA
Ocean Beauty Seafoods	15	NA	7	NA
Three Star	14	NA	7	NA
Picture				
Sliced	7	6	5	5
Smoked	6	5	4	5
Pieces	7	6	6	4
Whole	9	7	6	5
Benefit				
High in beneficial omega-3 fatty acids	13	10	5	6
An important factor in reducing heart disease	14	11	5	6
The American Heart Association recommends 2 servings of fatty fish per week as part of a balanced diet	18	10	5	5
Omega-3s may also reduce the risk of Alzheimer's disease	13	11	5	6
Pack Size				
In a 4 oz. package	6	13	7	7
8 oz.	8	16	7	7
10 oz. Bag	8	14	8	6
3/4 lb. package	10	16	7	8
Price				
$4.99	NA	9	NA	5
$5.49	NA	9	NA	5
$5.99	NA	11	NA	6
$6.49	NA	10	NA	5

part, they show the same response times. This similarity in the pattern of response times means that response time is a function of "processing" the information. Furthermore, the same information is probably processed the same way in the two studies even though the first study instructed respondents to select a price, and the second study instructed respondents to select an emotion.

2. Brands take up time: Surprisingly, the brand accounts for a lot of the response time.
3. Pictures are processed more quickly: People allot only about 50% to 60% as much time to pictures as they allot to brand when they have to select a price. Brand is the stronger driver when a person has to select a price.
4. Words take time to process: The benefits, complex text messages, also take up a lot of time, and in fact as much time as the brand even though the brand has few words and the benefits have many.
5. Economic issues are processed more rapidly than emotion questions: Pack sizes take up an intermediate amount of the response time when the respondent's task is to select the price. Pack sizes take up much more response time when an emotion is chosen. This is worth remarking upon a bit more. *It looks like the "economic equation" of price and pack size is processed very quickly.* People don't appear to think deeply about the price and price/value of a particular pack size; that processing is done very quickly. The same stimulus, pack size, has to be processed "more intellectually" when it comes to selecting an emotion, as if that task (selecting the emotion) is less "natural."
6. Simple questions such as "interest" are processed similarly across elements: When it comes to the relatively simpler question of overall interest, there are few differences across studies or across elements. About 0.5–0.8 seconds are generally allocated for each element. This corresponds to 5–8 "tenths of a second."

Individual Inspection Styles—Are There Segments?

Throughout this book we have emphasized that the consumers do not comprise a single, monolithic group of individuals. Nor, in fact, can consumers be easily understood by dividing them by conventional geo-demographic variables such as age, market, income, or gender. And, even the standard psychographic segmentations may not be sufficiently granular to add much value. Knowing that a person belongs to a segment interested in "health" does not really tell us which messages or pictures among a wide number will attract the respondent.

Following the worldview that segmentation is critical, we now again come to the issue of how we should segment people in order to understand person-to-person differences. Let's segment people on the basis of what catches their eyes, rather than on the basis of what interests them. The segmentation approach follows the pattern of segmentation that we used to develop the model for the entire panel. This time, however, the modeling is done at an individual-by-individual level.

Each respondent evaluated 36 different combinations. For each combination, the computer measured the first and second response times in tenths of seconds, with the first response time being that needed to make a selection decision (price, emotion) and the second response time being that needed to make a rating decision (degree of interest).

Then, on an individual-by-individual basis, we modeled the relation between response time in tenths of seconds and the presence/absence of the 16 elements.

We did this analysis separately, for both fish studies. Table 21.1 shows how we did this analysis for the total panel. We merely had to do the analysis separately for each respondent.

Finally, we segmented the individual models, dividing our consumers into three segments or clusters.

Let's see what happens when we create clusters of respondents based on the pattern of what holds the eye. Keep in mind that we are not focusing on interest or emotion. The only information that interests us right now is what elements hold the respondent's eye. That is, we empirically measure the response time. The data of interest are what give rise to that measured response time.

We analyzed the data from each experiment, to create an individual-level model of response times. For each respondent we "knew" how the 16 elements generated the response time for the first attribute rating (selection of price or emotion, respectively), and how the same 16 elements generated the response time for the second attribute rating, interest. We discovered that the major "drivers" for the segmentation came from the first rating—selection of price or emotion, respectively. *There was no segmentation based on the second rating of interest, something we did not expect, but which simplified our analysis.*

Look at the results in the *top half* of Table 21.2 for Fish Study 1, where the respondent first selected the "right price" for the package of fish, and then rated interest. We see evidence for three segments.

1. The first segment pays attention to health messages and then to the pictures. This first segment doesn't even pause to look closely at package size.
2. The second segment pays much more attention to size and then almost as much attention to health messages.
3. The third segment is "all over the place." It's not clear what this segment looks at. Nothing really holds their eye. They look at brand, pictures, and size. They do not pay attention to the health messages. Those might be important, but the respondents don't look like they stop to read these phrases on the package.

Now let's move to the results in Fish Study 2, where the respondent first selected the "right emotion" he felt when looking at the package stimulus, and then rated interest. These results appear at the *bottom half* of Table 21.2. In this second study, we replaced the four brand names with four prices. We extracted three new segments for this second study.

1. The first segment pays attention to price and to size. These respondents are the typical price shoppers.
2. The second segment primarily pays attention to price.
3. The third segment pays attention to size and then to health messages.

It is clear from these two studies that when present on the package, price plays a key role. When price is put onto the package it immediately becomes a key focal point for more than half of the respondents (Study 2). When price is not the focal point, then size becomes important.

Summing Up

Experimental psychologists have long considered response time (or in their terms, reaction time) to signal that underlying mental processes are occurring in people. One research strategy measures response times to a variety of stimuli (i.e., products or packages on a shelf), during the course of performing a task. The report is the "response time."

Table 21.2 Elements that "hold the eye" for segments of respondents defined by pattern of response times. The numbers in the body of the table are the numbers of *tenths of seconds* of response time that can be attributed to the particular element. The table shows only those elements that hold the eye more than 2 seconds (20 tenths of seconds).

	Seg1	Seg2	Seg3
Fish Study 1: Task was to first choose appropriate price, and second to rate interest in the fish package	N = 50	N = 60	N = 25
Segment 1: Health then pictures			
The American Heart Association recommends 2 servings of fatty fish per week as part of a balanced diet	31	16	–47
An important factor in reducing heart disease	29	17	–49
High in beneficial omega-3 fatty acids	25	15	–42
Picture: Smoked	24	–6	44
Picture: Whole	21	–3	–26
Omega-3s may also reduce the risk of Alzheimer's disease	19	17	8
Picture Pieces	16	–4	42
Segment 2: Size then health			
8 oz.	–10	20	–4
10 oz. Bag	–4	18	1
3/4 lb. Package	–10	18	23
An important factor in reducing heart disease	29	17	–49
Omega-3s may also reduce the risk of Alzheimer's disease	19	17	8
The American Heart Association recommends 2 servings of fatty fish per week as part of a balanced diet	31	16	–47
In a 4 oz. Package	–10	16	23
Segment 3: Brand Pictures Size			
Ocean Beauty Seafoods	10	15	63
Kenai Wild Alaskan Salmon	13	14	55
Wild Alaskan Salmon	12	13	47
Picture: Smoked	24	–6	44
Picture: Pieces	16	–4	42
Three Star	10	12	40
3/4 lb. Package	–10	18	23
In a 4 oz. Package	–10	16	23

	Seg1	Seg2	Seg3
Fish Study 2: Task was to first choose appropriate emotion, and second to rate interest in the fish package	N = 49	N = 38	N = 60
Segment 1: Price and size			
$6.49	23	16	–5
8 oz.	23	1	21
10 oz. Bag	23	–4	20
$5.99	21	15	–4
In a 4 oz. Package	19	–10	19
3/4 lb. Package	19	5	22
$5.49	17	19	–4
$4.99	16	18	–3
Segment 2: Price			
$5.49	17	19	–4
$4.99	16	18	–3
$6.49	23	16	–5
$5.99	21	15	–4
Segment 3: Size then health			
3/4 lb. Package	19	5	22
8 oz.	23	1	21
10 oz. Bag	23	–4	20
In a 4 oz. Package	19	–10	19
An important factor in reducing heart disease	6	12	18
Omega-3s may also reduce the risk of Alzheimer's disease	4	12	18
The American Heart Association recommends 2 servings of fatty fish per week as part of a balanced diet	6	13	16
High in beneficial omega-3 fatty acids	2	13	16

However, it is not response time alone that is important. Without an underlying structure, response time is simply a measure of behavior, similar to a rating scale, albeit with perhaps some greater intuitive appeal because it appears to be "objective."

In this study, as in other factorial studies with underlying statistical design, response time takes on a new meaning. We know that response time can be interpreted as the time needed for information processing. What we typically don't know is what is going on in the consumer's mind. Indeed, we may never know. However, being able to trace response time to the presence/absence of elements of known type (brand, price, size, message, picture, etc.) gives one a sense that there is an underlying pattern to be discovered. Furthermore, the ability to run models relating response time to elements, and then to segment people on the basis of these models, taps in to different patterns in the way people process package information. Such a sense allows for true insights, based on repeatable, expandable, modifiable experiments. These data suggest just one small step in a world of insights that are waiting to be uncovered by the researcher.

Postscript—Response Times, Network Effects, and the Value of Designed Experiments

This chapter is a good place to expand a bit on the value of "networks." We know that the so-called "network effect" acts as a multiplier of value. If one person is connected to a network, there is little value. There is no relation between that one person and of course, himself. Now let's look at the same network, this time with two individuals or points on the network. We have the connection between people. They can share things, ideas, etc. They can be compared to each other.

Moving onward, we can look at networks of 3, 4, 5 people, etc. The more individuals or points we have in the network, the more possible connections there are. In fact, the "value of the network" is often stated to be proportional to the square of the number of points or people in the network. (Technically, this squared value comes about because for N points in the network there are ((N)*(N-1))/2 connections).

What does this have to do with our experiment dealing with response times? The answer is straightforward. Look at Figures 21.1 and 21.2. These figures provide merely "point measures" of single stimuli. We really don't have the structure that allows us to make comparisons, which in turn, make us feel we've "learned something." These two figures *describe* what we measure. They are networks with one point or two points.

Now let's look at the models, shown in Tables 21.1 and 21.2. We feel that we have learned something about nature's rules. Why? Simply, we have a network effect operating. We have 16 points (really 16 stimuli that have been mixed and matched). We can compare these points to each other in terms of impacts. The points or elements have some cognitive meaning. From what these elements or points mean, we can report findings, patterns, and even rules that would otherwise elude us, as they elude us in Figures 22.1 and 22.2. And that power to discover results and learn rules, expressed in terms of today's new age of computation, is the "why and rationale" of this book that we are presenting to you.

References

Aaker, D.A., Bagozzi, R.P., Carman, J.M., and MacLachlan, J.M. (1980) On using response latency to measure preference. Journal of Marketing Research 17, 237–244.

Bassili, J.N. (1995) Response latency and the accessibility of voting intentions: What contributes to accessibility and how it affects vote choice. Personality and Social Psychology Bulletin 21, 686–695.

Bassili, J.N. and Fletcher, J.F. (1991) Response-time measurement in survey research. Public Opinion Quarterly 55, 331–346.

Bridgman, P.W. (1938) Operational Analysis. Philosophy of Science 5, 114.

Donders, F.C. (1868). Over de snelheid van psychische processen. Onderzoekingen gedaan in het Physiologish Laboratorium der Utrechsche Hoogeschool. Tweede reeks, II, 92–120. Translated by W.G. Koster in Acta Psychologica 30, 412–431, 1969.

Fazio, R.H. (1994) Attitudes as object-evaluation associations: Determinants, consequences, and correlates of attitude accessibility. In: R.E. Petty and J.A. Krosnick (Eds.) Attitude Strength: Antecedents and Consequences, Erlbaum, Hillsdale, NJ.

Fazio, R.H. and Williams, C.J. (1986) Attitude accessibility as a moderator of the attitude-perception and attitude-behavior relations: An investigation of the 1984 presidential election. Journal of Personality and Social Psychology 51, 505–514.

Moskowitz, H., Cohen, D., Krieger, B., and Rabino, S. (2001) Interest and reaction time analysis of credit card offers: Managerial implications of high level research procedures. Journal of Financial Services Marketing 6, 172–189.

Moskowitz, H., Hjelleset, T., and Rabino, S. (2000) Children versus adults: On the structure of their responses to systematically varied concepts. Canadian Journal of Marketing Research 19, 77–88.

Moskowitz, H.R., Poretta, S., and Silcher, M. (2005) Concept Research in Food Product Design and Development. Blackwell Publishing Professional, Ames, IA.

Mulligan, K., Grant, J.T., Mockabee, S.T., and Monson, J.Q. (2003) Response latency methodology for survey research: Measurement and modelling strategies. Political Analysis 11, 289–301.

Ratcliff, R. and Rouder, J.N. (1998) Modeling response times for two-choice decisions. Psychological Science 9, 347–356.

Tourangeau, R., Rips, L.J., and Rasinski, K. (2000) The Psychology of Survey Response. Cambridge University Press, Cambridge, UK.

Tyebjee, T.T. (1979a) Response latency: A new measure for scaling brand preference. Journal of Marketing Research 16, 96–101.

Tyebjee, T.T. (1979b) Response time, conflict, and involvement in brand choice. Journal of Consumer Research 6, 295–304.

Chapter 22

Combining Eye Tracking with Experimental Design

In the last decades, innovation and advances in technology have radically altered the packaging industry revolutionizing the four main functions of packaging: containment, protection, convenience, and communication (Robertson, 2005). Many of these technological marvels were then "beautified" by top designers, evolving these advances from technology into art. On the other hand, this revolution all too frequently led to creations of art on the shelves of supermarkets without regard to consumer needs and tastes, sustainability, and the like.

Enough of art criticism, however! Now let's look at how technology can help us understand how we respond to package design. In the previous chapters of this book, we looked at the contribution of science and statistics to help "design" the package, and then analyze the results from the perspective of what the consumer felt about the packages, whether in terms of overall judgment, emotions, image, etc.

We now move to an emerging field—technology to measure the behavior of a person's "scan" when the person looks at a package. We are thus evolving from science and technology as aids for design and subjective measurement to technology as a measure of what we do, how we behave, albeit unconsciously. In some sense, this "evolution" from science measuring feelings to science measuring unconsciousness, is not new. Psychologists who worked with subjects in experiments in the 1920s were confronted by the same innovations in technology, this time in the form of the Galvanic Skin Response (GSR) meter, which measured electrical conductance on the skin, and which later was to take its place as a standard physiological measure in the world of psychology. A dozen or so years later after the GSR technology had emerged, researchers began to measure the electrical properties of the skin, especially those of muscles, and then later take complete electroencephalograms (EEGs) of the head. Another well-known GSR application was the lie detector or so called polygraph. All of these efforts were to supplement current ways of measuring subjective responses and perhaps opening up new paths. The newest of these technologies is brain scans (Coon, 2005).

The Role of Eye Tracking in Package Design

With the foregoing brief introduction to "new" methods of measurement, we turn now to eye tracking as a method that may have unique ability to contribute to the understanding of package design. Up to now, we have been using the respondent to "judge" the features of a package. Our tools have been questions and answers. Of course, we were able to modify the questions, and by so doing enter more profoundly into the mind of the person inspecting a package. Beyond such relatively crude methods of measurement, fraught with errors and with the inevitable subjectivity of the "question as tool," we have the person-to-person variability. The list goes on. No wonder we are interested in other measures; perhaps their so-called "objective nature" may help us.

During the past decade, the notion has been developing that eye tracking or analyzing one's gaze could provide additional information. Eye tracking "follows" the eye of the user, using technology to track where the eye moves when looking at an object, in our case, inspecting a package. Thus, at a more cognitively meaningful level, eye tracking provides a way for the researcher to know what the person is looking at, even if the person cannot articulate what he sees (Jacob & Karn, 2003).

There's another facet to eye tracking. Whereas a person rates a package on at most a few dozen attributes, for that's all a person can really report, eye tracking gathers a lot more data. When the package is divided into so-called areas of interest (AOI), eye tracking can

Packaging Research in Food Product Design and Development Howard R. Moskowitz, Michele Reisner, John Ben Lawlor and Rosires Deliza
 ISBN: 978-0-813-81222-9

identify which AOI is looked at say, every tenth of a second. The technology to do so is already available.

The typical outcome of such eye tracking is a "heat map," showing where most of the looking takes place. By dividing the package into areas of interest, and by measuring the location of one's focus several times, eye tracking can tell the designer the package areas gathering the most gazes, the most attention. This information is not the same as what is important. Rather the information tells what "attracts the eye."

Eye tracking as a measurement method has become popular among qualitative researchers, as a way to identify where a person looks. Once the researcher discovers the key locations, it becomes easy to probe the consumer, who may not be aware of where they looked, but can certainly offer opinions about the different package AOIs that are being measured by eye tracking (Pieters and Wedel, 2007).

More About the "Innards" of Eye Tracking

Formally defined, "eye tracking" is a general term for techniques for measuring the point of gaze, where a person looks. Since human behavior and thinking are linked to where people look, the ability to measure eye gaze adds value to behavioral research and analysis.

Eye tracking originated in the late 1800s when it was first used by experimental psychologists and physiologists. Their equipment was intimidating, probably because most researchers at that time had to craft their own equipment with materials at hand. There were devices attached to the eyes of people, in what probably was a fairly painful or at least annoying way.

Equipment advances with technology breakthroughs. Eye tracking is no different. Over time the barriers have fallen, reducing the intrusiveness of equipment (always a discouraging factor), increasing robustness, and improving means to compute results rapidly and automatically. Finally, and probably most important, the price for the equipment has dropped, making the technology in reach of those who are interested in the problems, not simply those who are fortunate to have a skilled equipment-maker, and the financial resources to make things happen (Dongheng, Babcock, and Parkhurst, 2006).

The first of these more modern, more recognizable technologies for eye tracking appeared on the scene about 70 years ago in the 1930s. The trackers used beams of light that were shined, reflected by the eye, and recorded on film. Today there are modern eye trackers that do not affect test subjects or users and that do not require extensive technical expertise. Non-intrusiveness and ease of use have been keys for taking eye tracking out of the research labs into broader use (Tobii, 2008).

Eye Tracking Comes of Age in the World of Advertising and Package Research

Eye tracking is now used in virtually every kind of marketing research to measure what works and what doesn't, or at least where the eye looks in a stimulus that has many different AOIs. The world of eye tracking encompasses such diverse entities as TV ads, billboards, product packaging, and websites.

With the increasing technological prowess comes increasing sophistication in the way the data are analyzed. For example, packages and other objects are scanned differently by each visitor based on individual perception, interest, need, age, education level, computer monitor, browser settings, and other variables that can be tracked in empirical, eye-tracking studies (Tanner, 2007). Although cognitive processes cannot be observed directly, they are reflected in the pattern of gaze behavior. People do not explore an image or package randomly while looking on it. For example, while viewing an image, the items in the foreground get more attention than the items in the background (Babcock and Pelz, 2004). People usually pay more attention to certain distinct features such as the edges of an object, colors, or asymmetries processing a significant part of the visual information on a pre-attention level.

Researchers working with eye tracking have developed some normative data about how to measure the different aspects of eye, which in turn yields some fascinating results. For example, *Gaze path* is a sequential combination of fixations and saccades produced by the eyes. *Fixation* is a relative stability of the eyes for a brief period on a specific location. On average, fixation lasts for 200–300 milliseconds. In general, more than 150,000 eye movements occur each day for one person. *Saccades* usually last between 50 and 150 milliseconds and occur 2–3 times per second. It is curious that people have clear vision during the fixations, but not during the saccades (Pannasch, Dornhoefer, Unema, and Velichkovsky, 2001).

The issue that we deal with in this book focuses specifically on eye tracking and experimental design of packages. We know that the technology can measure what a person looks at, as well as how long. Let us now merge that information with experimental design, where the researcher and the designer can control the stimuli, presenting known combinations to the respondent. What type of link exists between what the researcher can do to the stimulus by such systematic design, how the eye tracks these changes, and what type of response the participant in a study might make (i.e., interest, statement of emotion)? We will explore this new interlinked approach for the rest of the chapter, working with a popular new product, wine in a box.

Boxed Wine Case Study

It is difficult to overstate the role of correctly choosing the right visual parameters for packaging. Recent studies showed that even when shoppers are open-minded and directly considering a category (as opposed to picking up their usual brand), over one-third of the brands displayed are completely ignored. However, a unique and striking appearance consistently helps to attract the shopper's gaze (Young, 2008; Jarman, 1999; Mervis and Janiszewski, 2002), and perhaps may in the happier of circumstances generate purchase.

There is no lack of experts and readily available best practices that guide many producers in the work of package creation, especially when the product moves beyond simple function to image. Sometimes the packages become the unintended victims of these conflicting approaches separately favored by marketers, designers, product developers, brand managers, etc. Focus groups and other forms of direct questioning, although still popular, do not usually resolve the problem. The only reliable way to satisfy the consumers is by involving them in the process of package creation (Thomas, 2008). We have seen this message of consumer involvement weaving throughout the chapters of this book. Let's now see the same consumer involvement explored, but this time "married to" the technology of eye tracking.

Environmental sentiments coupled with the ever-operative business sense have driven marketers and thus designers, to look at wine packaging as a new opportunity for environmental and personal "sustainability." One of the interesting areas, an offshoot of the dairy industry, is wine in a box, which would enjoy long shelf life. The box, or similar package, is easy to manufacture, more efficient in terms of the demands on transportation, is easy to store, and maintains the wine quality. Finally, and most important for this book, is that wine in the box provides a unique opportunity for new packaging design, to complement the new storage container.

Let's now dive again into the world of systematic design as we have done so many times in this book. This time we will couple eye-tracking technology to the design and to subjective responses by the consumers. In this way, the chapter provides a true "first"—experimental design of package, coupled with high-level objective measures of eye tracking, and finished off with consumer ratings of both interest and emotion, respectively.

We begin with the test stimuli. For this particular project, we "invented" a package for boxed wine, comprising four silos, each with three elements. The structure of the wine package appears in Figure 22.1. The very-simplified package comprises a label, a background medallion for the label, a center picture, and a tag line, respectively. As we have done in the other designs, whenever we have a silo missing there is always a "background" behind the stimulus so that the eye doesn't sense a discontinuity. That is, the silo may be missing, but otherwise the package looks reasonable.

Silo 1 (Label Fonts)

Working with wine labels can be fun, because the wine category lends itself to different fonts as well as messages. We didn't know what to expect about fonts, although we realized that fonts are often used to indicate high quality, especially when the script is fancy and provides a sense of "old luxury." Following this notion, we looked at a single label text (*Château du Vin*) executed in three different ways: with a fancy font, a regular large font, and a regular small font, respectively.

Silo 2 (Medallions)

The design comprised three medallions: a banner, a hexagon, and an oval.

Silo 3 (Picture)

The center picture placeholder also has three options: two color variations of the bottle and a stylized picture of grapes.

Silo 4 (Tagline)

The tagline has the same text for all three options (*100% Organic Wine*) rendered a normal font, large script, and small script.

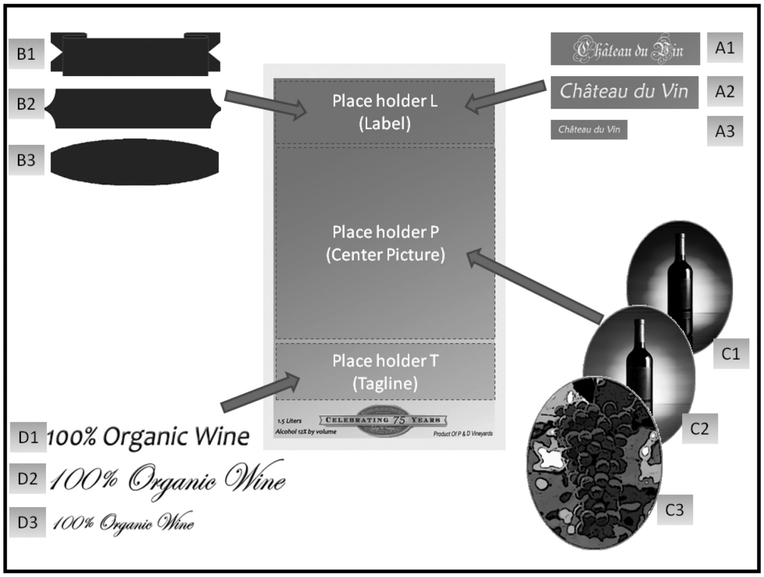


Figure 22.1 Template of the Wine Package Design Project with the tested elements (not to scale)

With four silos, each comprising three options, the experimental design calls for 27 unique combinations. The experimental design ensures that each respondent tests a different set of combinations, followed by a classification questionnaire. So far, this is standard operating procedure. One further difference is worth mentioning. The eye-tracking monitor must be installed at a test facility. Consequently, the respondents were invited into the facility to participate, and of course, they were paid for their efforts. In many of the studies we present in this book, especially the design studies, the respondents evaluate the test stimuli at home, with the stimuli presented over the Internet.

Running the Interview

We began the project with the desire to answer two questions, both of which we have dealt with earlier in this book. The first question was the traditional evaluation of the package, "overall" purchase intent. This is a fairly simple question, requiring the respondent to assign one number to the entire wine package. We used the purchase intent rating, rather than liking, although if truth be known, for most products that are purchased for sensory pleasure, like wine, liking and purchase intent ratings show the same pattern across stimuli. It's only when we add a variable such as price or a nonsensory health benefit such as reduced calories that we discover liking and purchase intent to diverge.

Our first rating scale was the nine-point purchase intent question: "*How LIKELY are you to PURCHASE this wine? 1 = Not at All Likely … 9 = Very Likely*."

Our second focus for the subject reactions was the emotional reaction to the package. We introduced the idea of emotion in previous chapters dealing with emotions and package design and emotion of experience. In those chapters, we found that the respondent appeared to have no problem choosing the emotion that best fit a test package or concept. Thus, we thought that emotion might be interesting here as well, since there might be a relation between eye movement and emotion, although we did not know what to expect. As in the previous cases, we asked the respondent to select the single emotion that best fit the concept. The respondent had a choice among seven alternatives, including one "nonemotion" response (neutral). These seven were presented for each concept. The respondents were instructed to select the one emotion most appropriate for the wine package, from this group: *Sad, Irritated, Calm, Neutral, Joyful, Relaxed, and Energized*.

From One Objective Measure to Two

Figure 22.2 The Tobii eye-tracking device used in the project

For the objective measure, we began first with the eye-tracking data. We partnered with a Swedish company, Tobii Technology. Tobii is one of the world's leaders in hardware and software solutions for eye tracking. Tobii's expertise comes from their ability to design both for the scientific community and for helping the disabled communicate by eye movement.

Important for the research was Tobii's ability to create a product that was not intrusive. That is, respondents did not see any indications that their eye movements were being monitored and recorded. In fact, the only way the respondents knew about their eye movement being tracked is due to information provided during their recruitment and orientation. The Tobii devices are very similar to traditional LCD displays and do not bias respondents more than other computer-aided technologies for interviewing people.

We also explored a second "objective measure"—"response time." There is always some time between the presentation of a stimulus package on the computer screen and the first response that a person makes (i.e., rating). This intervening time is presumed to be the time it takes for a respondent to process the information and make a judgment. Experimental psychologists call this "reaction time." More than a century ago, psychologists began to measure reaction time to stimuli. Over the years, this measure, like blood pressure for a doctor, started to yield its secrets. Response or reaction time wasn't simply something fixed. It varied with the stimulus, suggesting that this "dead time" might actually correspond to interior "mental processing."

Our research tool measured the interval between the time when the rendering of the package is complete on the respondent's screen and the rating assigned by the respondent. It's important to keep in mind that computers differ in the speed at which they are able to download the particular "image" of a package and display that image. We designed the response time program to pick up that time between the completion of rendering and the respondent's rating, because that definition of response time cannot be influenced by the speed of the computer.

Overall, the data generated by this first study using experimental design and eye tracking provide a very rich source of information. In addition to the experimental design, we have interest and emotion, coupled with eye movements (at least for the first 5 seconds), and response time.

We executed the project in cooperation with Tobii Technology and its associate Realeyes Data Services, Ltd., in August of 2008 with 50 prerecruited respondents in a central location in the United Kingdom using Tobii eye-tracking devices similar to one shown in Figure 22.2. Since this project was the first known case of integrating the three previously unrelated methods, we focus the rest of this chapter both on what we found, as well as our reactions as researchers to the process.

Looking at the Results—What RDE Tells Us About What "Works" and How People Feel

Whenever a researcher begins exploring nature with a new method, it always helps to start from what is known and from what feels comfortable. Our study with boxed wine is no different. Let's begin by deconstructing responses to the packages using the interest measure as the dependent variable. Recall that we relate the presence/absence of the package features, our independent variables, to the rating of interest, defined as 0 (for original purchase ratings 1–6), or defined as 100 (for original purchase ratings 7–9). This deconstruction tells us "what wins."

One can break the data many different ways, as we see from Figure 22.3. As in the previous chapters, it's best not to look at the data, one number at a time. That strategy soon becomes mind-numbing and rather

			Gender		Age		Income		Consumption		Segments	
	Element	**Total Sample**	**Male**	**Female**	**18-29**	**30+**	**Less than £40K**	**Greater than £40K**	**Once a week or more**	**Less than once a week**	**Seg. 1 – visually oriented**	**Seg. 2 – no nonsense**
Base size		**50**	**27**	**23**	**32**	**18**	**26**	**24**	**35**	**15**	**38**	**12**
Additive constant		*-8*	*1*	*-18*	*-10*	*-4*	*-13*	*-2*	*-2*	*-21*	*-14*	*12*
Silo 1 – Font												
A1	Fancy font label	11	6	17	14	7	12	11	8	19	11	12
A2	Large normal font label	10	4	17	9	10	9	10	6	18	13	0
A3	Small normal font label	4	0	9	4	4	0	8	4	4	6	-3
Silo 2 – Medallion												
B3	Oval	3	3	4	6	-2	1	5	3	4	2	7
B1	Ribbon	2	4	-1	3	0	4	-1	-3	14	2	-1
B2	Fancy hexagon	0	3	-2	2	-3	-1	2	0	2	1	0
Silo 3 – Picture												
C1	Purple bottle	11	11	12	11	12	16	6	11	12	17	-7
C2	Green bottle	8	5	11	9	6	16	-1	5	14	14	-12
C3	Grapes	7	-1	15	6	8	10	4	3	16	15	-18
Silo 4 – Tagline												
D2	Large script font tagline	4	0	9	5	3	7	1	3	7	4	5
D3	Small script font tagline	3	-4	10	4	-1	7	-2	0	8	1	7
D1	Normal font tagline	2	-2	7	3	0	5	-1	2	3	2	3

Figure 22.3 RDE—How features of boxed wine labels drive purchase.

counter-productive. Rather, the more productive way is to get a sense of what the experiment is trying to tell us. This holistic approach is even more important when we realize that we are dealing with elements of visual style that do not have a more profound meaning, as we might encounter were we to do the same experiment, but with words. Essentially, when it comes to package design, "What you see is what you get."

In package designs there are typically modestly performing elements, rather than very strong or very poor performers. Boxed wine is no different. Look at Figure 22.3, which shows the impact values for the 12 elements. Let's first look at the total panel:

1. The additive constant is below 0, which makes sense. If there are no elements, there's nothing to look at. The baseline should be around 0. Unlike the concept of a boxed wine, we're really looking at the package itself. Without content, the package is meaningless.
2. Averaging the data across all of the respondents reveals a few good (but not great) elements that drive up their purchase intent. The consumers are generally neutral or slightly negative to the idea of the boxed wine, but it could be influenced by selecting "right" elements for the package: the fancy font label (A1) or the large regular font label (A2). Another winning element is the purple bottle (C1). There are no negative elements for the Total, although half of the elements are neutral.
3. The font plays a role. In fact, the key to success is to have big fonts, not small fonts.
4. Medallions play a much smaller role. It really doesn't matter what type of medallion is used. However, when it comes to choosing a medallion, the visually simple oval is best.
5. The pictures make a difference. All of the pictures do well. The color of the picture is important, with a purple bottle (C1) doing better than a green bottle (C2). But it's not just color—it's also topic. The stylized grapes do well (C3), but again not as well as the purple bottle. There's really no functional reason for the pictures—it's a matter of artistic taste.
6. The taglines are all the same—slightly positive to neutral.

We can look at the same data, but break out the results by gender, age, income, consumption frequency, and lastly by segments defined in terms of their responses to the package elements. We'll just look at big exceptions to the patterns defined by the total panel.

1. When it comes to gender, females are far more responsive to the visual elements than males are. Females start with a much lower baseline (−18), so it is the visual elements that do the work. Males, in contrast, start with a higher baseline (1), but most of the visual elements don't perform particularly well. The message here is clear—females are more responsive than males. Females are quite negative to the general idea of the package but selecting the fancy or large font label with the picture of the grapes dramatically increases their purchase intent (A1 or A2 and C3). In fact, the three above mentioned elements would convince an additional 42% of the female consumers to buy the wine. Males are different. They are completely indifferent to the picture of grapes and to the messaging about organic content, whereas females love both.
2. When it comes to age, the two ages show random differences. One interesting "factoid" is that the older respondents like the elaborate font more than the younger respondents do. If that is an instance of a general rule, then the font may be very important as a covert attractor of the appropriate age group.
3. Income makes a difference, with the higher income respondents beginning with a much lower base (additive constant = −13), but with the elements doing all of the work.
4. Moving on to consumption patterns, we see the patterns that emerged previously. Those who consume wine frequently start out with higher baselines, but the elements are not particularly powerful. In contrast, those who consume wine less frequently start out with low baselines, but respond more strongly to some of the elements, especially the pictures.
5. We can segment the respondents based on the patterns of their responses. The segmentation shows the most profound differences among "complementary" groups of respondents. The two segments that emerge differ dramatically in the pattern of what they respond to.
   a. Segment 1, "Visual," comprises about 75% of the respondents. The "Visual" is negative to the general idea of boxed wine. However, merely featuring a simple large font label (A2) along with the purple bottle (C1) could sway an additional 30% of the respondents to say that they would buy the packaged wine.
   b. Segment 2, "No Nonsense," comprises about 25% of the respondents. The "No Nonsense" is moderately positive to the general idea of boxed wine. About 12% would be interested in buying the package without any elements shown to them. However, things go downhill from here; there are not many positives that influence these consumers. With the exception of the fancy font label (A1) on the oval banner (B3) and small font tagline (D3, a modest impact), nothing sways them. Quite the opposite; any picture reduces their interest in purchase.
6. Our excursion into the innards of the experimental design suggests to us that the pictures make a lot of difference, as do the fonts. The medallions make less of a difference, and the taglines hardly contribute.

Beyond the What the Respondent Says to What the "Eyes Do"

Having gone through our "standard analysis," we know something about how the respondents feel. Now, for the same stimuli, let's see what the respondents do. We are NOT looking for a correspondence between, for example, purchase intent and what the eye spends most of its time on, although to find that would be "nice." Our goal is different. We want to approach the problem of package through a different path. We want to see what people actually look at.

Although we might think that we evaluate packages as whole, the reality is quite different. We saw that through experimental design. Now we move to eye tracking, where we can take precise readings many times a second to see where the eye is gazing. The eye-tracking technology records the so-called "gaze location" about 50 times a second! For the purposes of our project, that amount of information is simply too fine-grained. We are going to look at the "gaze location" only 10 times a second.

Let's first look at what type of information comes from eye tracking. Consider a respondent who is presented with a package. Tobii's eye-tracking technology will record the "gaze location." Let's superimpose that gaze location on the package so we can get a sense of the pattern of gazes our respondent made. Figure 22.4 shows an example of the gaze path of a respondent. The gaze path starts from the middle of the package, which is the most typical starting point. The gaze then moves from location 1 to location 2, and to the label located at

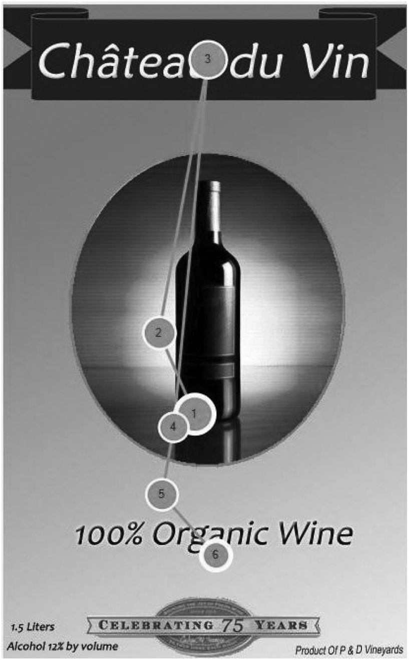


Figure 22.4 Gaze path of a single package of a single respondent

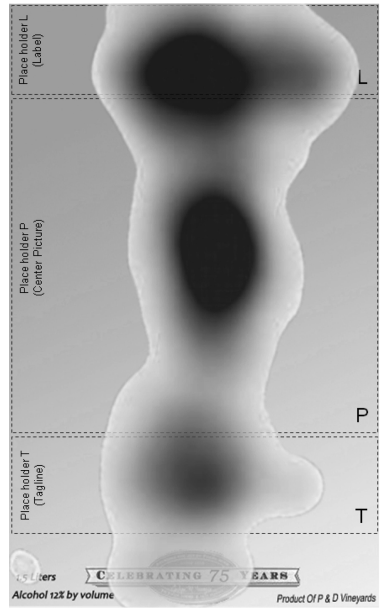


Figure 22.5 "Heat map" (intensity of the combined gazes) for the total sample overlaid on the template (in order to show the locations). On the original, the red color indicates more intensive accumulated gaze, yellow is medium intensity, and green is a light one. On the B/W version, red and yellow merged into black and dark grey while green is light gray

location 3. After stopping at location 3, the gaze moves through the middle area again (locations 4 and 5), and to the tagline (6). People differ, however, so the pattern from this one respondent will differ from the pattern of the next respondent. Each person has a unique pattern of gazes. Eye tracking opens this hidden information to the researchers and designers.

A very effective, graphical way to present eye-tracking data uses group data displayed in the so-called "heat map." Heat maps provide an overall view of eye activity of the respondents evaluating the packages. Heat maps are developed by combining the data from all eye gazes on the packages. The more people look at a location, the more "heat" is assigned to that particular location. Heat maps are color-coded, to display the information graphically, and to make the insights immediately accessible. As one might expect, the color red corresponds to the most gazed at areas, yellow corresponds to areas that are looked at, and finally, green corresponds to the least looked at. We see an example of a heat map in black and white in Figure 22.5.

Most gazes concentrate on the middle of the center picture area and the label, with less time spent on the tagline. The fixed area at the footer of the package that did not vary showed different patterns of viewing. For example, males looked more at the area with the size of the package and females more at the brand name. The full analysis of the heat maps is beyond the scope of this chapter.

Dissecting the Patterns of the "Look"

As we have mentioned above, the technology of eye tracking reveals where the consumers look, for how long, in what sequence, etc. We might think that the respondent spends most or at least a lot of his time evaluating the stimuli (pictures). Let's see whether or not our hypothesis is correct by looking at the data from this study on boxed wine.

We sampled the gaze location during the first 5 seconds starting from the moment the package was completely rendered in the browser. In most cases, the package came on almost immediately (all of the parts were preloaded at the start of the interview). Furthermore, we realized that most people make their decisions within a few seconds. It seemed "overkill" to measure the gaze for longer than a few seconds. Following this idea, we recorded the gaze for exactly 5 seconds. We did not record the gaze location after 5 seconds, even if the respondent had not yet made a judgment.

So, what did we find? Here are some of the highlights, such as how long people look, where they look, etc.:

1. Respondents don't look at the package all of the time. In fact, on the average, respondents spend only about 22% of the first 5 seconds (about 1 second in total) looking at the package itself. The rest of the time (78% or almost 4 seconds), they either read and answer the questions or simply "wander" around, outside the package.
2. In the full sequence of the 5 seconds, just 8% of the time do the consumers look at the label, area L; only 11% of the time do they look at picture, area P; and a measly 3% of the time do they look at the tagline, area T.
3. These numbers are based on the short initial period (the 5 seconds only) of each screen. When we tabulate the time between the screen rendering and the actual rating time (the moment the respondent presses the rating button), we discover that the proportion of time spent evaluating the actual package is lower.

So, the respondents do not spend a lot of time evaluating the packages. Maybe they are cheating or just bored during the long exercise?

Do respondents cheat? Let's first see whether the short evaluation time suggests "cheating." The truth of the matter is that our observed, perhaps disturbingly short, time of gaze is not necessarily a bad thing. The respondents should not overintellectualize the exercise—they should provide their first reaction, their "gut" feeling. People form their general opinion about visual objects very fast, but that opinion is usually strong. Because the survey is conducted via the Internet, we should be concerned with the time needed for the web users to form their opinion about a web page. A recent study by researchers in Canada showed that the snap decisions Internet users make about the quality of a web page have a lasting impact on their opinions. They also reported that impressions were made in the first *50 milliseconds of viewing* (Lindgaard et al., 2006). The findings suggest that it is mostly the main features and the general appearance of the objects that make a difference, not those small details that require more time to evaluate.

Do respondents change the pattern of inspecting packages with experience? We use experimental design in order to generate a set of systematically varied prototypes to be rated by the respondents. Depending on the number of variables, a respondent might have to evaluate dozens of similar-looking screens (in our case, 27 although in some designs, the number is much higher). This systematic creation of a relatively large array of packages, and a possibly onerous task, actually help us to discover whether or not the respondent changes his personal strategy of gazing, as this 27-screen interview moves along.

Let's now do this analysis, first by dividing the 27 *unique* screens evaluated by each respondent into thirds—the first (screens 1–9), the second (screens 10–18), and the third (screens 19–27). Let's then look at the FIRST location the respondent looks at. This is the start of the pattern. People have individual ways of acquiring information. On average, the person presented with the same type of stimulus should land at the same place, and explore in more or less the same way. When respondents start changing their place of landing when they begin the interview (screens 1–9) versus when they end the interview (screens 19–27), we have evidence that they are less attentive and more random in their visual search.

Let's look at Figure 22.6. We see that regardless of the interview part, whether at its beginning, in the middle, or at the end, about three-quarters of the consumers start their first gaze at the same place—the center of the package (location P, main picture). By the end of the study, this number falls just slightly.

One possible explanation of the findings is that they might be related to the habitual way a person looks at a package, and thus represent "automatic behavior" that is not under the influence of boredom. In the majority of cases, the gazes are center-loaded. The number of first gazes at the center slightly drops to the end of the interview (from 77% to 71%) whereas the label area gains in returns of gazes (20% to 28%). The tagline loses some (from 3% to 1%) gazes at the end of the surveys. The latter (both the absolute values and the trend of reducing interest) hints that the tagline (location T) should not be considered as the most preferred place for important messages.

Let's look at the percent of the time that a respondent gazes at each of the three main areas of interest: label,

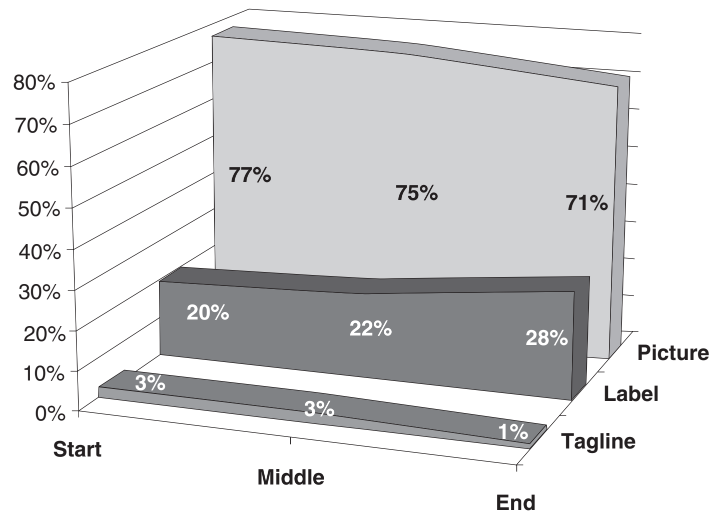


Figure 22.6 Change in the location of the *first gaze* as the study progresses: start of the interviews (screens 1 to 9); middle of the interviews (screens 10 to 18); and the end (screens 19 to 27). The numbers represent the averages of the respondents with the first gaze at the corresponding locations.

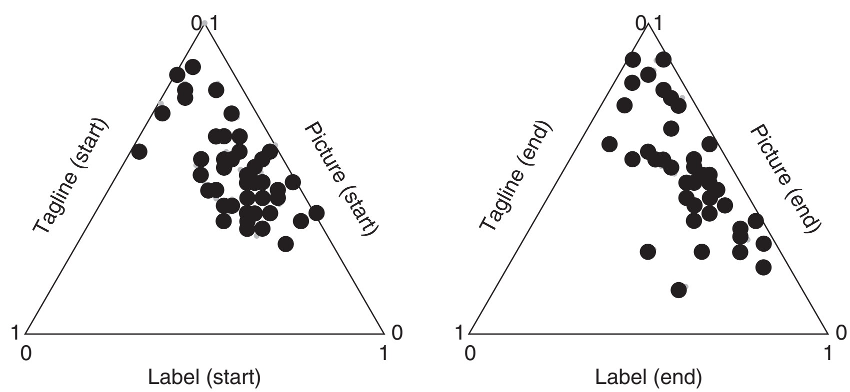


Figure 22.7 Even with continued practice, people do not change the way they look at the package. The figures show the percent of gazes at the three locations, for the first nine versus the last nine stimuli. The pattern of people is fairly similar.

picture, and tagline. The totals add up to 100%. That is, the respondent who looks at the package must be looking at one of these three areas. We compute the percents *for each respondent, separately* considering the first third of the packages that person evaluated (stimuli 1–9) and the last third of the packages (stimuli 19–27). Each of those nine packages generates its own set of percents for the respondent. Now look at Figure 22.7 where we plot the 50 respondents in the triangle plot. The pattern is almost the same, meaning that the general distribution of gazes across 50 people looks pretty much the same at the start of the evaluation and toward the end of the evaluation. *People do not change their gaze patterns, even with practice*.

All in all, we take the results to suggest that *people don't seem to change their gaze patterns with repeated exposure. Boredom may set in, but people move their eyes in the same general pattern.*

What Happens After That First Gaze? What Eye Tracking "Really Teaches Us"

We're all familiar with the newsstand, the book, and of course, the beach. Sometimes our gaze lands on an object, and we feel immediately compelled to look more closely, to explore more, and to learn about what we just saw. Far more frequently, however, the first object of the gaze path reduces our interest and we pass on the surroundings.

Let's see how the location of the first gaze of the respondents in the wine package project influenced other aspects of their behavior. In our study the respondents actively evaluated the package. What do their eyes tell us that they cannot?

1. The location of that first gaze correlates with the total amount of time this individual will spend exploring the entire package.
2. Going more deeply into this analysis, we discovered that when a person's first gaze landed on the main picture (P), as was the case for most of the package evaluations, then most of the time the eyes would spend about half as much time looking at the label (L) as they had looked at the picture.
3. Now let's look at what happens when the person first looked at the label. The same type of pattern occurs. Let's call the time the person would spend looking at the label LT (label time). The person would then move away from the label, and spend a total of approximately 0.65 LT devoted to *the entire rest of the package.*
4. Despite the emerging regularities of the eye movement, the location of the first gaze neither correlates with the emotion selected, nor correlates with the time that the respondent needs to make a decision. Recall that this response time was defined as the time between the completion of the rendering of the package on the screen and a reaction of the respondent. We conclude from this that the location of the first gaze defines the information a person "takes in," but does not tell us about any emotional response.
5. Where the eye lands can influence the purchase rating. There seems to be also somewhat of a relation between the location of the first gaze and purchase intent. To make this discovery we divided the respondents into groups, defined by the location where their eyes landed. We did this at package-by-package level. We divided the 1,350 screens (27 from each of 50 people) by the location where they first landed. The most interesting discovery, probably not leading to a rule, however, is for the Fancy Front Label (element A1). When the first gaze landed in the label L, and it happened to be this particular element, a full 23% more respondents would buy this package, rather than just 8% if the first gaze landed in the picture area. Getting the first landing right, on the label, and getting the first gaze there, may have some unexpected impact.

What "Sticks" the Viewer's Eyes to the Package?

Some parts of the package grab the attention of the respondents and keep them for a while. Others might be as striking and catch the gazes but not for long. The eye moves away after looking for a moment. In the best of worlds, the designer and marketer want the consumer to look at the different parts of the package, not just focus on one part and ignore the other. The various parts of the package convey information—whether brand, flavor, price, nutrition, etc.

Of course, in the world before eye tracking and experimental design, these same issues were faced, discussed, and designed for. The benefit of eye tracking is that the movements of the eye can be quantified to see whether, in fact, one part of the package design is "hogging" all of the attention. The benefit of Rule Developing Experimentation (RDE) (Moskowitz and Gofman, 2007) for systematic variation is that one can discover the combinations of package features that either generate this "hogging" or reduce it.

Like so many other aspects of science, one of the key problems that investigators face is clearly defining their terms. What does it mean to say that a part of the package is "sticky," or that a section of the package "hogs" the attention? How do we operationalize those terms, create a series of metrics for them, and then use the eye-tracking technology to give numbers to those metrics? How do we use experimental design to engineer those metrics so that they have the desired values based on marketing objectives?

We measured the proportion of time for each test package that the eye spent on the different areas of interest. If we want the eye to move around, then we need to develop a number that shows us the "variation" in the time spent on the three major areas of interest. (Recall that the label and the medallion were located together, so we deal here with the label/medallion, the picture, and the tagline.)

One easy way to do "stickiness" analysis computes a measure of variation in gaze location, for each individual package. For instance, we empirically measure the *percent of time* that the eye gazes on the three different areas. If the person spends equal amounts of time, regardless of how much, gazing at the three different locations, then we should have a standard deviation for this stimulus for this person equal to "0." The reason is simple. There is no variation in the percent of time the respondent spends in one location versus another. They are all equal, and the standard deviation of the equal numbers is "0." If, however, the person spends a lot of gaze time looking at the picture and little gaze time looking at the label or the tagline, then the standard deviation should be high.

A little-worked example will show the logic. Let us say that we have five scenarios, listed below, in scenarios 1–5 shown below. Each scenario describes a different pattern of gaze times, defined as percent of the total time a person looks at the test package. We are not really interested in the length of time because one person may look at the package for a long time but show the same percentage distribution of gazes as another person who looks at the package for just a very short time.

1. Scenario #1: 33% looking at the label, 33% looking at the picture, 33% looking at the tagline, standard deviation = 0.00. This person shows no preference. All three locations on the package are equally sticky.
2. Scenario #2: 45% looking at the label, 45% looking at the picture, 10% looking at the tagline, standard deviation = 0.20. This person shows no preference for label versus picture, but does not pay attention to the tagline.
3. Scenario #3: 60% looking at the label, 30% looking at the picture, 10% looking at the tagline, standard deviation = 0.25. This person focuses primarily on the label, so we might say that the label "hogs the gaze."
4. Scenario #4: 66% looking at the label, 33% looking at the picture, 0% looking at the tagline, standard deviation = 0.33. In this scenario, we see an extreme example, with the person looking 2/3 of the time at the label, 1/3 of the time at the picture, and never at the tagline.
5. Scenario #5: 100% looking at the label, 0% looking at the picture, 0% looking at the tagline, standard deviation = 0.58. This is the most extreme example.

With this in mind, let's look now at the distribution of all of the pictures, to see how the different elements drive the gaze.

With 50 respondents and with 27 test packages evaluated by each respondent, we generate a total of 1,350 "different" stimuli. Let's look at each of the 1,350 stimuli separately, and create a standard deviation for that stimulus showing the percent of the time that the respondent gazes on each of the three different areas of interest (label, picture, and tagline). We can get a sense of the eye movements by looking at the distribution of the standard deviations, keeping in mind that a standard deviation of 0 means that the eye gazes at all three areas equally in terms of time, whereas a standard deviation of 0.58 means that one location monopolizes the view. Figure 22.8 shows the distribution of these 1,350 standard deviations, 1 standard deviation per stimulus package. The abscissa runs from a low of 0 corresponding to equal view of all areas on the package, to a high of 0.58 where only one area out of the three is looked at for the entire time.

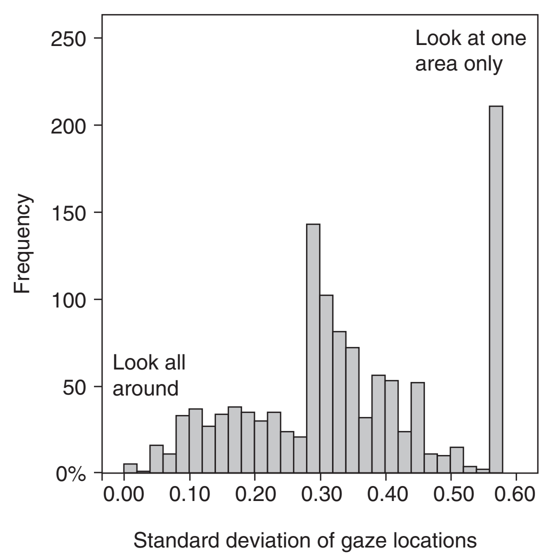


Figure 22.8 How people distribute their gazes on a package, for 1,350 packages. The figure shows the distribution of standard deviations for gazing at three locations. A standard deviation of 0 means that the gaze for a single person, single stimulus is distributed equally across label, picture, and tagline. A standard deviation of 0.58 means that the same gaze focuses on only one of the three locations for the entire evaluation.

In about 16% of the cases (the large bar on the right end of the histogram), the respondents concentrated on only one area of the package ignoring the rest. In just a very few cases (the left end of the chart), the respondent allocated an even amount of time for every area of the package. The bulk of the gaze patterns fall in the middle of the spectrum where the respondent looked at the package all over with reasonably varied amount of time spent on each location.

Continuing this analysis of gazing, let us now look at the contribution of each of our 12 elements to this "stickiness." Keep in mind that stickiness is defined as a high standard deviation, with one area attracting more gaze time than 33%.

Since the experimental design systematically varies the different elements, we can relate the size of the standard deviation to the presence of each of the 12 elements. The higher the contribution of a particular element, the more that element forces the respondent to focus on one location and ignore the others. In contrast, the lower the contribution of an element, the more that element forces the eye to distribute its gaze equally. See Figure 22.9.

Armed with that, consider the results in Figure 22.9. In an ideal case, the designer wants the consumer to look around the package, rather than concentrating on only one part of the package, ignoring the rest. It's important for foods to convey both brand and functional delivery, such as nutrition. So, what are the features that keep the consumer's eye wandering? Or, in the terminology of the standard deviation that we are using here, what are the package elements that reduce the size of the standard deviation?

Figure 22.9 shows how the 12 elements "drive" the gazing. Let's now list some of the patterns that appear to emerge from this figure.

1. The additive constant is the estimated standard deviation in the absence of elements. We might consider this constant as the "propensity to shift one's gaze." Recall that larger numbers mean that the gaze does not shift as much, whereas smaller numbers mean that the gaze shifts around more. Males show higher constants for the model (0.45), and females show lower constants (0.37). Thus, we first conclude that females vary their gaze more than males do.
2. The printed information (conveyed by fonts) drives people to move around more. The label medallion and the picture drive people to move around less and concentrate more. People do not stop and gaze at text; they read and move on. People do stop and gaze at pictures and at medallions.
3. The impact of text differentiates males from females. Males are more likely to move their gaze around

		Total	Males	Females
	Additive Constant	*0.42*	*0.45*	*0.37*
	Elements keep gaze moving			
A3	Small normal font label	–0.07	–0.09	–0.04
D2	Large script font tagline	–0.07	–0.07	–0.07
D3	Small script font tagline	–0.07	–0.08	–0.06
A1	Fancy font label	–0.06	–0.08	–0.04
A2	Large normal font label	–0.05	–0.07	–0.03
D1	Normal font tagline	–0.05	–0.05	–0.04
C3	Grapes	–0.03	–0.02	–0.05
C1	Purple bottle	–0.02	–0.02	–0.02
C2	Green bottle	–0.02	–0.01	–0.03
B3	Oval medallion	0.01	0.01	0.00
B1	Ribbon medallion	0.02	0.02	0.02
B2	Fancy hexagon medallion	0.02	0.02	0.01
	Elements keep gaze fixated			

Figure 22.9 How each of the elements of the boxed wine package drive the eye to focus on a single area of interest (high numbers) or move around the package (low numbers). The dependent variable for the model is the standard deviation of the percent of time spent on label versus picture versus tagline.

when confronted with text. Females are somewhat less likely to do so and probably read the text.

4. Males and females are more similar in their reaction to the visual stimuli of medallion and picture.

Keep in mind that there is no "good" versus "poor" performance here. Knowing what package design features keep the gaze moving provides information to the designer about how to engineer the package. One could envision application of such findings to the packages with multiple messages/images (i.e., health, taste, authenticity, etc.). The marketer would want to find a combination that keeps the eyes on the package longer but not at one location. In some situations, it might be preferable to distribute the time of the gazes more or less evenly between the elements of the package.

Linking Emotions

Emotions drive everything that people do and often guide our everyday choices. Without emotions, there will be no action. Of course, the emotions steer our purchase decisions. Sometimes a small nuance in a package feature, its color or size, can significantly influence the emotions. The "right" emotion could produce the desired action. Some marketers want their products to cause people to feel joyful— others, relaxed—yet others, energized. Can we "engineer" the emotions that the consumer might experience looking at the specific package and thus help the marketers and designers to create an emotional link or context for a more successful product?

Emotions can be measured either through self-reporting, where the consumers indicate their feelings, or through psycho-physiological measurements where the emotions are indirectly measured through physiological parameters. The former method of direct self-reporting is much more practical and has been used in our case study here.

As part of our efforts in this study, we wanted to see whether there was any correlation between eye movements and emotions. The method of having a respondent select the emotion that most characterized a particular package seemed appropriate for this study as well. It's easy for respondents to select the emotion, although such an approach does not get at the rich set of multiple emotions that might underlie a specific package. There were no expectations about what might be the relation, if any. In a sense, we were again breaking new ground here.

The second rating question of our experiment instructed the respondent to select the ONE emotion he felt when he inspected the particular package for boxed wine. We used the same set of emotions that we had used previously in the other work on emotions: sad, irritated, calm, neutral, joyful, relaxed, or energized, respectively.

The information collected based on the systematically varied packages gave us a large database of emotions and their associations. Figure 22.10 shows the distribution of the emotion ratings of the consumers in the packaged wine project. This distribution is done in the same way as the distribution of response times. That is, we learn what happens in general, from looking at the data. We do not yet know what package design features covary with the selection of emotions. However, we do know that there are a disproportionately high number of neutral and calm ratings.

Aside from "neutral," the most frequently selected emotions were "calm," "irritated," and "relaxed." This could be explained by some polarized opinions about alcohol consumption and different mind-sets of the respondents. It is possible that the consumers that see wine as a special social occasion would not possibly like

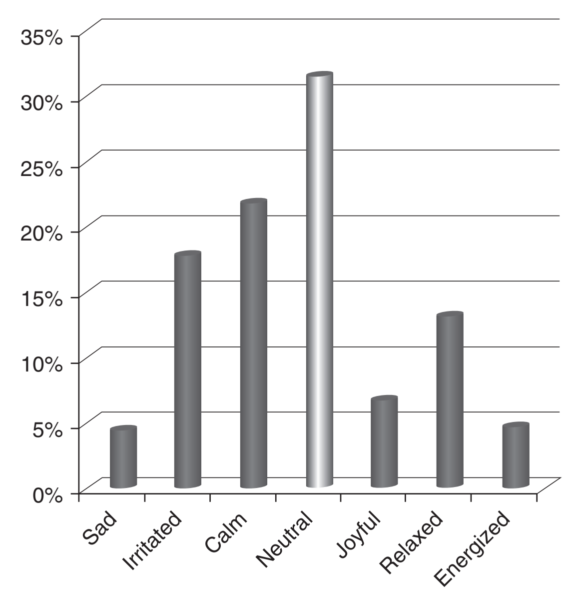


Figure 22.10 Distribution of the emotions selected across the 1,350 experimentally designed boxed wine packages. Each package had to have one emotion attached to it.

the package preferring instead a classical bottle. Others associate wine with free and relaxing time. These are hypotheses that can be checked out by other types of research.

How Do Package Features Drive Emotions?

We have already seen how the individual features drive interest in the entire package. The statistical method of regression analysis generates these individual impact values for each package element. The same approach works for emotions, but this time we have to deconstruct the seven different emotions from one scale into seven different emotion scales.

When we create the seven new emotion scales, we create seven new dependent variables. Let's just take one emotion, such as irritated. We have 1,350 different packages that we showed to our 50 respondents. The default value for "irritated" is "0," which means that the package was not selected as being irritating. Now, let us go through all 1,350 packages and see what emotion was actually selected for each package. If we look at Figure 22.10, we see that about 17.5% of the packages were associated with the respondent saying he was "irritated." For each of those packages, we replace the default value "0" by the value "100." Now we have 1,350 rows of data, wherein we know what elements were present, as well as knowing whether the respondent selected "irritated" as an emotion (those rows with 100 in the irritated column), or whether the respondent selected some other emotion (those rows with 0 in the irritated column). This matrix of 1,350 rows, with package elements, and the two values, 0 or 100 for "irritated," is easily analyzed by regression. See Figure 22.11.

		Sad	Irritated	Calm	Neutral	Joyful	Relaxed	Energized
Additive constant		14	37	14	50	–8	–6	–2
Silo 1—font								
A1	Fancy font label	–4	–10	2	–3	4	7	4
A2	Large normal font label	–4	–18	7	1	4	8	2
A3	Small normal font label	–3	–19	13	0	3	5	1
Silo 2—medallion								
B1	Ribbon	1	1	4	–7	2	2	–3
B2	Fancy sextagon	1	3	3	–6	1	0	–3
B3	Oval	3	5	1	–8	4	–2	–3
Silo 3–picture								
C1	Purple bottle	–9	–10	4	–11	5	15	5
C2	Green bottle	–10	–7	3	–7	4	8	9
C3	Grapes	–4	–2	0	–14	13	5	3
Silo 4—tagline								
D1	Normal font tagline	–1	–3	–5	0	4	3	2
D2	Large script font tagline	–1	–3	–4	–4	3	7	2
D3	Small script font tagline	–1	–2	0	–3	1	3	3

Figure 22.11 Individual contributions of the elements of the wine box to different emotions as selected by the respondent.

Let's see how the elements drive the emotions, if they do at all. The data are summarized in Figure 22.11, which shows how each element drives each emotion. Of course we can always run the statistics. The real question is what do the results suggest, and more important, how can the designer use these results in practice?

1. *Create a package that makes the respondent feel "relaxed."* The designer should consider using the large normal font label (A2) on the ribbon banner (B1), the picture of the purple bottle (C1), and a large script font tagline (D2). This combination maximizes the sum of the utilities for the column "relaxed." When we replace the large normal font, A2, by the small normal font, A3, keeping the rest of the elements unchanged, this slightly revised combination minimizes the feeling of irritating.
2. *Create a package that makes the respondent feel "energized."* The design should consider a package with the fancy font logo (A1) on either of the banners (B1–B3), the picture of the green bottle (C2), and the small script font tagline (D3).

Summing Up

In this chapter, we have begun to explore how tracking eye movements can provide additional information about package design. Eye tracking, as well as other psychophysiological measures, provide some information about what the body does. This information can, of course, be linked to the stimulus and reports generated. For the most part, eye tracking and other such measures have been used as simple dependent variables. The researcher makes a change in the stimulus, and measures the eye-tracking behavior.

A much richer world can be created, promising greater understanding, if we systematically vary the stimuli, and look for generalities, for rules, that transcend simple "point measures" such as "Package X generates more heat around the label than does Package Y."

We have begun to create this world in two ways. First, we have looked at eye tracking as a psycho-physiological measure, and both interest and emotion as two different classes of subjective measures. Second, we have used experimental design of the stimuli, rather than simply presenting two stimuli and looking for an "effect."

From this approach and from the empirical results, we may conclude the following, which stand as the start of research efforts rather than as hard and fast rules:

1. Experimental design enables the design to consider different options for a package, doing so in a knowledge-based way. Experimental design provides improved understanding of the general consumer mind, highlighting the differences between the demographic groups and empowering the designer with actionable attitudinal segmentation. It gives input for the rough first iteration in the package design based on the general directions of understanding of the consumer mind.
2. Eye-tracking data of the systematically varied packages generate insights into consumers' gaze patterns. The information allows fine-tuning the packages, helps select the right locations of the features as well as adjusts font size, colors, etc., to achieve marketing goals.
3. The analysis of emotions connected with package designs give a sense of how the consumer feels about the viewing experience. The emotion data might help the astute designer looking to create an emotional link between the package and consumer or create a proper emotional context stimulating purchase decision.
4. Used together, experimental design, eye tracking, and multiple ratings (evaluative, emotion selection) could generate an actionable database for particular projects. The approach we suggest does not provide "general rules." Rather, the rules or patterns are appropriate for each particular study. From the results of many of these studies, there should emerge an even more powerful understanding of how people react to packages.

Acknowledgments

This chapter was written by Alex Gofman, Vice President, Moskowitz Jacobs Inc.

Sincere thanks to David Moskowitz and Paolo Gentile of Moskowitz Jacobs Inc., Johanna Fyrbjork of Tobii, Niall Claudio Bellabarba and Mihkel Jäätma of Realeyes Data Services, Ltd., for their hard work making this project happens. We are also grateful to Nick Harmsen, formerly of Tobii, who suggested collaboration in several fruitful, early meetings that acquainted us with the eye-tracking technology.

References

Babcock, J. and Pelz, J. (2004) Building Lightweight Eye Tracking Gear. In: Eye-Tracking Research and Applications Symposium, 109–114.

Coon, D. (2005) Psychology: A Modular Approach to Mind and Behavior. Thomson Wadsworth.

Dongheng, L., Babcock, J., Parkhurst, D. (2006) OpenEyes: A low-cost head-mounted eye-tracking solution. In: Proceedings of the 2006 Symposium on Eye Tracking Research & Applications, San Diego, California, pp. 95–100.

Jacob, R.J.K. and Karn, K.S. (2003) Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises. In: R. Radach, J. Hyona, and H. Deubel (Eds.). The mind's eye: cognitive and applied aspects of eye movement research. North-Holland/Elsevier, Boston, pp. 573–605.

Jarman Jr., J.B. (1999) How To Take Your Package Design to the Next Level. Frozen Food Age, Nov. 99, 48(4), 22.

Lindgaard, G., Fernandes, G.J., Dudek, C., and Brown J. (2006) Attention web designers: You have 50 milliseconds to make a good first impression! Behaviour & Information technology 25(2): 115–126.

Mervis, T. and Janiszewski, C. (2002) Consumers' Beliefs about Product Benefits: The Effect of Obviously Irrelevant Product Information. The Journal of Consumer Research 28(4):618–635.

Moskowitz, H. and Gofman, A. (2007) Selling Blue Elephants: How to make great products that people want BEFORE they even know they want them. Wharton School Publishing, Upper Saddle River, NJ.

Pannasch, S., Dornhoefer, S.M., Unema, P.J.A. and Velichkovsky, B.M. (2001) The omnipresent prolongation of visual fixations: Saccades are inhibited by changes in situation and in subject's activity. Vision Research 41(25–26), 3345–3351.

Pieters, R. and Wedel, M. (2007) Goal Control of Visual Attention to Advertising: The Yarbus Implication. Journal of Consumer Research 34(August), 224–233.

Robertson, G.L. (2005) Food Packaging: Principles and Practice, Second Edition (Food Science and Technology). CRC/Taylor & Francis, Boca Raton, FL.

Tanner, J. (2007) Eye Tracking, Statistical Analysis and Site Success. SiteProNews: http://www.sitepronews.com/archives/2007/sep/17prt.html

Thomas, J. (2008). Begin at the Beginning: Research should be Involved at the Outset of any (re)Packaging Process. QUIRK'S Marketing Research Review XXII(1):52–56.

Tobii. (2008) What is eye tracking? http://www.tobii.com/corporate/eye_tracking/what_is_eye_tracking.aspx

Young, S. (2008) New and Improved Indeed: Documenting the Business Value of Packaging Innovations. QUIRK'S Marketing Research Review XXII(1):46–50.

Chapter 23

Taking Stock and Summing Up

Our trip through the world of packaging is approaching the end, as we finish our book. If we were to stand back and take stock of what we have learned during the course of the writing, what sentences would we use to close? What are the lessons we have learned, what wisdom have we gained, and what might we say to those who read these chapters, and want to go on?

Summing up is always difficult. Bidding goodbye to one's efforts, to a world just beginning to open up, has become increasingly hard for us. Each of the chapters opens up new ideas, creates new worlds to explore, new dimensions to understand. And, as we move from chapter to chapter, we realize how little we have done, and how much remains to do, both in theory and in practical application.

But, enough of these platitudes and *mea culpas*! Let's synthesize our learning into some meta-principles, some generalities that we believe to be our starting contribution to a field that is just now entering the world of science, while keeping its artistic heritage in the world of design.

Observation 1. In a World of Ever Faster, the Package Will Become an Even More Important Medium for Communication

When we look inside the corporation, sitting in on meetings held by packagers, by brand marketers, and by advertising agencies, we notice something very unusual. When it comes to discussing anything about the product, about the formulation, about packaging and the like, we see what appears to be interminable discussion.

Let's move our focus a bit from the inside of the company, outward to a different venue. This venue is the focus group facility, the hallowed room with the one-way mirror. On one side of the mirror, we see poised in front of us a smiling moderator to lead the discussion, and a group of current or potential consumers of the product, or perhaps a group of disaffected users, the so-called trier-rejecter. On the other side of the mirror sit the cadre of packaging and product people, the marketers, a researcher or two, and others from the focus group facility.

What do we see? If we were to focus on the dialogue between the moderator and the consumer(s), we would see the consumer examining the package. One consumer after another would opine, sometimes for a very long time, sometimes holding forth in what seems to be a forever analysis, detailing every aspect of the package. We watch and hear people focusing on packaging in a way that seems to us both so terribly analytical, and in some ways, absurd in the detailed activities that we see happening in front of us.

Let's now visit a store, as ordinary consumers do often, and as marketers and designers do on occasion. Let's watch what happens, as the shopper enters the store. We don't need to work with elaborate technology to see what is happening, nor measure with a stopwatch or other equipment how much time a consumer spends looking for a product on the shelf. There is a lot of that research around, often very good (i.e., Mariampolski, 2001).

What we see is shoppers rapidly scanning the shelf, at least for the most part, often followed by either a deliberate selection ("I was looking for this specific product … and here it is."), or selection on the spur of the moment ("I just wanted to have it."). The strategies for searching the shelf, the momentary reasons for selecting products, can take up books by themselves. What concerns us here is the speed of the decision.

People make decisions selecting a product in a way that belies the detailed analysis of verbatims in the focus group, or the well thought-out set of ratings in a question-

ISBN: 978-0-813-81222-9

naire. For the most part, product selection is quick, occasionally automatic, often following a mental script ("I have to replenish my pantry and this is my brand.").

With this in mind, where does package research come in? Is the research to understand why the person selects a package, or just to find out what package is selected, and how to engineer the selection of that package once again? With the speed of selection, it seems to us, the authors, that there is an even greater need to understand the factors that drive one to select a package. It's not just a question of deciding which package is best—the so-called beauty contest to which we have referred again and again. Rather, we believe that there needs to be a "science" of packaging, telling the designers, the marketer, the trade merchandiser, what "works," what doesn't, how consumers react, and what to feature on the shelf to drive the proper response.

Speed of selection now becomes a friend to the researcher. Rather than relegating packaging to simply a creative art form, something that is necessary for the product, packaging is becoming a science and must continue to evolve in that direction. As the opportunities to select products increases, it's no longer a question of creating packages that may work, but rather an opportunity to use the science of consumer packaging to engineer a package that *will* work.

Observation 2. Package Design is Becoming a Key Focus of Today's World of Packaged Goods, and in So Doing Package Design Will Become Both an Art and a Science

Anyone in consumer-related business today cannot help but be aware of trends. Today's business world is interconnected, thanks to the Internet, thanks to virtual meetings that are held with groups dispersed worldwide, and thanks to the outsourcing.

All of these interconnections end up leading to continuing education, training, magazines, short courses, conferences, and the like. We can often tell about the progress of a field from reading its publications, seeing its want ads in trade magazines, and scanning the types of business conferences that are devoted to the topic.

When we look at package design, comparing today (early 2009) to a decade ago, we see some interesting patterns that tell us just how important design is becoming. First we see recognition of design by academics in books. In the early 1980s, one book, edited by Walter Stern, appeared that dealt with design (Stern, 1981). A decade and a half later another book edited by Ayn Gelinas appeared, courtesy of Committee E-18, Sensory Analysis, of the American Society for Testing & Materials. This book was called *Creative Applications of Sensory Techniques Used in Conducting Package Research with Consumers* (Gelinas, 1996). Finally, in 2003, Scott and Batra edited a book on "Persuasive Imagery: A consumer response perspective."

Books alone don't provide the necessary markers. There are the articles that appear in trade journals. Articles on packaging appear more frequently these days in trade magazines, but more tellingly in short e-mail newsletters that can be had by signing up for a subscription (i.e., Daily Briefing from the *Food Production Daily* website.) These articles, and especially the daily newsletters, spend an increasing amount of their space on packaging topics, for two reasons. One reason is that they cover nutrition issues. Nutrition labeling is a hot area worldwide, and in being such a focus of attention, the issue of packaging also is drawn into the fray. The other reason is the increasing cost of food commodities, which drives up the prices. One way for companies to cope with the rising price is to reduce the size of the package. Such size and of course weight reductions are the grist for many newspaper and newsletter articles, whether the article is tilted as an "expose" (more for mainstream media), or is tilted toward factual reporting, albeit wordsmithed by the PR department of the company releasing the story (an increasing hallmark of the Internet briefings and newsletters).

These two types of stories, nutrition labeling and size reductions, represent only one side of the world of packaging, the side having to do with the more "hygienic" part of design and packaging. They deal with issues that are arising that pertain to necessary packaging information and product economics.

The "flip" side of these magazines and newsletters is creative use of packages to signal new benefits, to gain attention from consumers, to grab attention in a shelf. Stories dealing with these issues don't often make the magazines and newsletters, except when the story topic is packaging in general. There are examples of such stories in magazines such as Food Product Design. At the same time, there is a world of short courses growing, such as those offered by the Institute of International Research in New York and Europe. IIR presents a variety of conferences. For almost a decade now, IIR has presented a two-day conference dealing with product packaging. At this conference, suppliers and business clients alike talk about what worked, and more interestingly, what didn't work. These conferences give the audience a sense of the field, who is doing what, what is the progress. And, looking at the history of the conferences, the speakers, the topics, the nature of the attendance, how many have attended, how many reminder notices must be sent, all tell us about the state of the field. From the attendance at these IIR conferences, and factoring in the economic situation for the year of the conference, something that always has to be done in business to account for cycles, we can pretty well conclude that packaging is becoming increasing visible, increasingly relevant to the business community, And well it should be for business students in marketing around the world are taught the Five Ps—product, price, PACKAGING, positioning, and promotion. The history of business is witness to the acceptance that packaging is truly important for the product.

Observation 3. Economic Factors Will Advance Package Design in Two Ways—New Designs (Art/Technology) and Creative Business Applications (i.e., Size Reduction)

When we began this trip through packaging, we recognized that a lot of the topics that we would cover would deal with the creative aspect of packaging. In a way, that's where the fun is. Anyone who has ever worked at an advertising or design agency knows that the soul of the creative person is often happiest when the problem is new, when there are many alternatives that one can execute, and when one is leaping and bounding through new fields, hitherto unexplored, or at least not particularly well worked over.

In our travels we discovered that there was a recurring theme. When we told people about the book, our experiments, what we were learning, what we were confused about, and so forth, the notion of "nutrition labeling" and "size reduction" kept emerging. We would talk about designs for products that "walked off the shelf," grabbing the customer's attention, and in return we'd hear about "what works with nutrition, what should we say to make the legal folks happy, what is the effect of size on perception of our product, and what will we do when we have to reduce the amount."

Suffice it to say, we have dealt in some ways with those continually reemerging issues. We probably have not given those two key topics sufficient treatment, especially with today's focus on nutrition and cost of goods. Everyday one reads in the newsletters and newspapers, in the trade press, and the like, about new regulations worldwide to better communicate nutrition. One cannot escape the distressing news that the product sizes have to retreat from supersizing to subsizing, as company after company recognizes it cannot survive by giving away what it should be selling.

And so this book compromises a bit, as any book must. To treat labeling adequately requires an entire book itself, and perhaps to our joy, some book will appear that takes our approach and expands it to different foods, different countries, and by so doing creates the integrated body of knowledge that we have foreseen (Moskowitz and Gofman, 2007; Moskowitz, Reisner, German and Saguy, 2005). We haven't seen it yet, but as the poet Percy Bysshe Shelley says in "Ode to the West Wind" about spring, "if winter comes, can spring be far behind?"

Observation 4. Measurement Is Here to Stay

We began this book with a peek at how experimental psychologists, researchers by nature and vocation, might look at packaging. Our little foray began in the psychophysics laboratory of Harvard University, where researchers had spent a number of years trying to understand the dos and don'ts of subjective measurement. Their pioneering work along with that of other psychology laboratories, then government, and finally industry laboratories established the reality that people could act as measuring instruments.

What's the key learning that we ought to take away from the widespread use of subjective measures of perception, and especially their increasing use in package research? One hint is that it's not the specific scales that people use to measure response to packages. There don't seem to be general scales of perception or of attitude, no matter how strongly one wishes to argue to the contrary. Certainly some scales are more sensitive than others, and some make more sense. But that's not the real news here.

The real news is the headline of this observation. Subjective measurement is here to stay. This may sound like a truism, but it's not. Throughout the history of psychology, and now through the history of consumer research, there have been attempts, again and again, to dethrone subjective measurement, and instead find some type of objective instrument that could scan the stimulus and come up with a number to reflect how a person might respond to the stimulus.

In our journey through packaging research we looked at eye movements as one of these methods. There are others, such as brain waves, and a host of machines that will measure reaction time, and the like. All of these technologies are worth investigating further, except perhaps objective measures of one's skull, such as that used by phrenologists. But, all jesting aside, these objective measures are worth pursuing. On the other hand, they are just that—objective measures. They are not perceptions; they do not equate to the person, nor do they capture the richness of experience that might be captured in a set of scales, or even more, by a qualitative interview. Subjective measurement, the kind by people reporting about objectives, is here to stay, at least for the foreseeable future. The technologies may be aids, but the ultimate will probably come from the human consumer, responding by numbers or in text-based discourse. It is the person, the private, subjective world of the consumer, that is bound, more than anything else, to yield insights about packaging and package design.

Observation 5. Experimentation and Experimental Design Are Increasingly Accepted

Anyone who has worked in an industrial setting, particularly in a marketing-driven environment, soon senses that there is a wide gulf between the artist/designer and the researcher. Of course, as we just stated in Observation 4, a lot of that feeling comes from the distaste in being measured.

There is another gulf, one that is also being bridged, shortened, and may eventually disappear altogether. This is the dynamic tension between art and experimentation. Of course, in departments that teach design, as well as in schools that teach advertising, the gulf between creativity and experimentation is usually glossed over. It's not politically correct to aver, at least in a school and certainly among more powerful colleagues, that one dislikes experimentation. For heaven's sake, the politico might say that everyone experiments, all the time. For what other way is there to create designs that work?

The truth of the matter is somewhere between the isolated genius designers who create masterpieces that fly off the shelf and the far less-inspired drone, who somehow manages to test thousands of packages, in sort of a routine, mind-numbing way.

Systematic experimentation is part of the future. During the past few years, the authors have seen the beginnings of acceptance by designers. A little history will put this into perspective. Ten years ago, as we headed toward the year 2000, there was very little acceptance of experimental design among designers. Most designers, if they would talk to us at all, spoke about creativity, inspiration, and learning from the consumer in a sort of mysterious and unstructured way. No one was brash enough to say that the consumer was irrelevant. Almost all designers we encountered, however, were sufficiently bold to say that they simply do not need this newfangled thing called "systematic experimentation," that the approach might work for chemistry but certainly would not work in the world of art where there are so many options that one dare not experiment so much. It would break the budget and, besides, who could be sure that the experiments were dealing with the proper stimuli? Better to rely on knowledge of the consumer through the talented eyes of the designer.

Ten years later, we are facing a new world. There is an increasing acceptance of the notion that one can learn from experimentation. Such acceptance does not mean, however, whole-hearted commitment to experimentation. No, such commitment from the artistically oriented designer cannot happen yet. It's too early. Experimentation is messy, sloppy, filled with numbers, filled with outcomes, structured in a way that seems to crimp art, or so the feeling goes.

So, why is experimentation being accepted, if not for its intrinsic value? The answer is the eternal "WIIFM," the old radio station "What's in it for me?" Designers, like everyone else in business, are facing an increasingly complex world. Designers want all the information that they can get to create better packages that will sell. Ultimately, they realize that they are not artists, that they cannot wait two or three decades to be discovered, and that they have to put food on the table. These real needs make the designer sit and listen, respond to what might be an opportunity to win in the marketplace.

And as for the role of experimentation, where does that fit? For most designers right now, and that also includes other creative artists in business such as perfumers, flavorists, braumeisters for beer, and the like, experimentation is a necessary evil, something that probably works, and that should be tolerated. Experimentation is not close to their hearts, not by any means. Experimentation is simply the vehicle by which some of the knowledge of consumers can be obtained. Designers don't want to be bothered with the methodological niceties. They simply want to know what works, what doesn't, and how to move forward. The fact that the information comes from experiments is of little interest to them, at least TODAY. For tomorrow, perhaps they will come to embrace experimentation in design, rather than tolerate it, which is certainly better than the fearful disdain they had for experimentation a mere decade ago.

Observation 6. Expect More Technology for Package Research

As we have noted, people like technology, or at least people in business like technology. There's something to be said for technology, which gives the sense of objectivity. Just look back at the history of product testing, not to 10 or 20 years ago, but to 80 years ago and longer. The beginning and the middle of the twentieth century witnessed the growth of interest in the sensory characteristics of products. As foods became increasingly stable because of processing and packaging, people evolved out of a survival mentality (I'll eat it if it's not rotten and tastes reasonably good), and into a choice mentality (I'll choose what I like).

This evolution of course focused on the sensory characteristics of food, such as appearance, texture, aroma, and flavor. What were the first efforts in these fields? They were not studies of the perceptual characteristics. No. Rather, these first forays were attempts to build instruments that could parallel or replace the eyes, the nose, and the like. It was only in the middle 1960s and onward that practitioners in the food industry realized that they could not really replace the nose and mouth with instruments, although they keep trying, with state of the art devices that analyze the chemical and physical properties of food.

What does this mean for the world of packaging? It means that the predilection of research is toward creating instruments. Perhaps our scientific world, the world of technology, gives people the false sense that an instrumental measurement of a package is somehow better than a subjective one. This is not the time, nor the place, to argue such a thesis, to defeat it or to defend it. Rather, it should be recognized, acknowledged, not forgotten.

Recognizing that people like instruments because it makes them feel objective means that in the world of package research we should expect to see more technology. The technology will be of two types: presentation of stimuli and measurement of response.

Presentation of Stimuli

As computers matured, they were able to do graphics better, more realistically. Early shopping experiences through the store, done on the computer, programmed painfully by visionary pioneers such as Mike Gadd in Canada and Ray Burke at Harvard and Indiana, with associated commercial interests, have given way to virtual package and shelf testing. We should expect to see lots more of this work, some motivated by theory, other work motivated by the sheer applicability of graphics technology to package research, simply because money can be made. But not to naysay it all—such presentation methods may provide new learning. The jury remains out, but early reports suggest that in the mountains of data that such methods provide, there are nuggets of gold to be discovered and mined.

Measurement of Response

Just as people love instruments to measure the "sensory properties of products" so they love instruments to measure what they often believe to be the "true, underlying" perception. It's not sufficient to ask people about the package. Someone has to measure responses, whether eye tracking and heat maps, or fMRI (functional magnetic resonance imaging), which measures blood flow in the brain. The reports of these studies make for fun reading, for exciting if not exactly profound presentations at conferences, and for nurturing the hopes of yet another generation that this technology would truly "discover" how the brain responds to packaging. From the point of view of the authors, "*it ain't necessarily so*." We believe that there may be things to learn with instrumental measurement of responses to stimuli, but it will be hard to know what is real learning and what is the inevitable correlation with responses that were more easily obtained simply by asking the respondent. Alas, it is probably impossible to cure this addiction to "objective measurement of responses" and probably a thankless task anyway.

Observation 7. We Still Don't Know What Works

Pervasive uncertainty with some glimmers of insights always accompanies new approaches. There's no way to avoid uncertainty and the unpleasantness that it brings. People, designers, marketers, storeowners, and of course, consumers like to be certain. They like to be sure about what they buy, that they are eating nutritionally adequate food, prepared under safe conditions, at a fair price.

Designers know that to get their products selected from the crowded shelf they have to communicate the right messages—visually and textually—while appealing both to the rational mind and to the less rational, emotional mind. Furthermore, people differ from each other. We have seen this segmentation again and again. It's most pronounced, of course, when we deal with concepts or ideas, where people can agree with the idea or disagree with it, where people can want what the idea proposes, or not even care at all.

Matters are a bit different with design. If we move beyond the messaging that conveys strictly positive or negative characteristics (i.e., nutrition), the situation becomes a bit more clouded. The designer doesn't really know what works, or it if does work, doesn't know why. Design, for the most part, is not verbal but visual. We see from designs that there are combinations that are liked and combinations that are not liked. Yet, there are no simple rules. When we move beyond factual information to supposed emotion-driven stimuli on a package, we move beyond simple good-bad to far more subtle responses. We can measure, because people know what they like and don't like. But, as often as not, we cannot isolate why they like what they like. We can move a bit closer to the "why" through experimentation, but even so, the experimentation does not show the strong positives and negatives with design variables that it shows with text and messaging.

And, of course, in that lack of knowledge of truly what works lies waiting an entire lifetime of research opportunities.

References

Daily Briefing, Food Production Daily website http://www.foodproductiondaily.com.

Gelinas, A. (Ed.) (1996) Creative applications: Sensory techniques used in conducting package research with consumers. American Society for Testing & Materials, West Conshohocken, PA.

Mariampolski, H. (2001) Qualitative Market Research: A Comprehensive Guide. SAGE Publications, Inc., Thousand Oaks, CA.

Moskowitz, H.R. and Gofman, A. (2007) Selling Blue Elephants: How to make great products that people want before they even know they want them. Wharton School Publishing, Upper Saddle River, NJ.

Moskowitz, H.R., Reisner, M.R., German, J.B., and Saguy, I.S. (2005) Unveiling health attitudes and creating good-for-you foods: The genomics metaphor, consumer innovative web based technologies. CRC Critical Reviews in Food Science and Nutrition 45, 165–191.

Scott, L.M. and Batra, R. (Ed.) (2003) Persuasive imagery: A consumer response perspective. Erlbaum, Mahwah, NJ.

Stern, W. (Ed.) (1981) Handbook of Package Design. John Wiley and Son, New York.

Index

 ISBN: 978-0-813-81222-9

 ISBN: 978-0-813-81222-9