GENERALIZING FROM PURCHASE OUTCOMES

By

WOUTER VANHOUCHE

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2005
I dedicate this dissertation to my wife and the little miracle she is carrying with her.
ACKNOWLEDGMENTS

Initiation, development and completion of this dissertation would not have been possible without the contribution of a number of people. I thank Joe Alba, my dissertation chair, for making me a better researcher and for his dedicated guidance throughout the dissertation process. I have also benefited from interactions with Chris Janiszewski and thank my other committee members--Lyle Brenner, Rich Lutz and Gary McGill--for their valuable comments on the dissertation.

I am grateful to Jef Nuttin, Hans Vertommen, and Frank Baeyens. Each in his own way has played a significant role in my development from a student to a researcher. Special thanks go to Luk Warlop, without whom I might never have pursued a Ph.D. abroad.

My parents have provided invaluable emotional support, and the same is true for my wife, Maddy, who has made our time in Gainesville so much fun. I thank her for enabling me to initiate this Ph.D. and for her support in completing it. I look forward to the next stage in our life as a family of three instead of two.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>viii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>ix</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2 LITERATURE REVIEW AND THEORY DEVELOPMENT</td>
<td>3</td>
</tr>
<tr>
<td>Generalization and Induction Research</td>
<td>3</td>
</tr>
<tr>
<td>Causal Lay Theories as Alternative Framework</td>
<td>7</td>
</tr>
<tr>
<td>Do People Hold Causal Lay Theories about Events that Have Occurred Once or Twice?</td>
<td>9</td>
</tr>
<tr>
<td>Is the human cognitive system designed to draw causal inferences based on very small samples?</td>
<td>9</td>
</tr>
<tr>
<td>How many theories do people typically generate?</td>
<td>10</td>
</tr>
<tr>
<td>Are Causal Lay Theories Used as Input into Generalization Judgments?</td>
<td>12</td>
</tr>
<tr>
<td>3 RESEARCH QUESTIONS AND INITIAL EVIDENCE IN MARKETING</td>
<td>16</td>
</tr>
<tr>
<td>Research Questions and Hypotheses</td>
<td>16</td>
</tr>
<tr>
<td>Initial Evidence in Marketing Context</td>
<td>19</td>
</tr>
<tr>
<td>4 TEST OF EXPLANATORY POWER OF COMPETING ACCOUNTS</td>
<td>22</td>
</tr>
<tr>
<td>Experiment 1</td>
<td>22</td>
</tr>
<tr>
<td>Method</td>
<td>23</td>
</tr>
<tr>
<td>Predictions</td>
<td>23</td>
</tr>
<tr>
<td>Results</td>
<td>24</td>
</tr>
<tr>
<td>Discussion</td>
<td>25</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>27</td>
</tr>
<tr>
<td>Method</td>
<td>28</td>
</tr>
<tr>
<td>Predictions</td>
<td>29</td>
</tr>
<tr>
<td>Results</td>
<td>30</td>
</tr>
<tr>
<td>Discussion</td>
<td>32</td>
</tr>
</tbody>
</table>
APPENDIX

A  STIMULI FOR EXPERIMENT 1 .................................................................67
B  STIMULI FOR EXPERIMENT 2 .................................................................69
C  STIMULI FOR EXPERIMENT 3 .................................................................70
D  STIMULI FOR EXPERIMENT 4 .................................................................71
E  STIMULI FOR EXPERIMENT 5 .................................................................72
F  STIMULI FOR EXPERIMENT 6 .................................................................73
G  STIMULI FOR EXPERIMENT 7 .................................................................75
H  STIMULI FOR EXPERIMENT 8 .................................................................78
I  STIMULI FOR EXPERIMENT 9 .................................................................79
J  STIMULI FOR EXPERIMENT 10 ...............................................................80

LIST OF REFERENCES ..................................................................................81

BIOGRAPHICAL SKETCH ............................................................................88
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-1. Patterns of results as anticipated by various accounts (A, B and C) and as observed in Experiment 1 (D)</td>
<td>24</td>
</tr>
<tr>
<td>4-2. Possible outcomes in Experiment 2</td>
<td>29</td>
</tr>
<tr>
<td>4-3. The observed pattern of results in Experiment 2</td>
<td>31</td>
</tr>
<tr>
<td>4-4. Pattern of results as anticipated by several accounts (A, B and C) and as observed in Experiment 3</td>
<td>36</td>
</tr>
<tr>
<td>5-1. Classification of causes as either stable or unstable in Experiment 4</td>
<td>41</td>
</tr>
<tr>
<td>5-2. Average generalization as a function of number of experiences and purchase context (product service) in Experiment 5</td>
<td>43</td>
</tr>
<tr>
<td>5-3. Mean generalization per replicate in the product and in the service condition in Experiment 6</td>
<td>45</td>
</tr>
<tr>
<td>5-4. Mean generalization as a function of replicate and dimension in Experiment 7</td>
<td>48</td>
</tr>
<tr>
<td>5-5. Mean generalization as a function of purchase context and behavior in Experiment 8</td>
<td>51</td>
</tr>
<tr>
<td>6-1. Mean generalization as a function of replicate and number of experiences in Experiment 9</td>
<td>54</td>
</tr>
<tr>
<td>6-2. Mean generalization per condition in Experiment 10</td>
<td>56</td>
</tr>
</tbody>
</table>
Abstract of Dissertation Presented to the Graduate School of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

GENERALIZING FROM PURCHASE OUTCOMES

By

Wouter Vanhoucke

August 2005

Chair: Joseph W. Alba
Major Department: Marketing

Induction is a ubiquitous but rarely investigated process in marketing contexts. Consumers frequently interact with a vendor, usually through purchase, and then must assess the likelihood that future interactions will produce the same outcome. Research in decision science suggests that generalizing from small samples is common but ill-advised. Research in cognitive psychology suggests that generalization varies as a function of the perceived typicality of the episode or exemplar. In line with a hypothesis about conceptual coherence in the categorization literature, I argue that induction will be driven primarily by the theories consumers have regarding the reason for an outcome. I refer to this perspective as the “causal lay theory” view of induction. Moreover, I hypothesize that generalization will be greater when the driving mechanism is perceived as stable rather than unstable.

I find systematic support for this hypothesis in a series of experiments. Specifically, consumers generalize more quickly from a positive outcome than from a negative one, and more quickly in a product context than in a service context. Surprisingly, a strong
typicality manipulation failed to impact generalization. With multiple inconsistent outcomes, generalization is determined by the valenced order of the outcomes and the time lag between those outcomes, which is consistent with the lay theory perspective. With multiple consistent outcomes, generalization quickly reaches asymptote, even when an unstable mechanism had initially been assumed.

Results suggest that some major findings from the induction literature are not transferable to the non-taxonomic stimuli encountered in consumer contexts. The Law of Small Numbers, too, cannot easily be generalized to a consumer context. Instead, the results are more in line with a hypothesis in the category-formation literature and with findings in the social realm.
CHAPTER 1
INTRODUCTION

The question of how people generalize from a limited sample of observations to a larger population of observations has received considerable attention in decision science and cognitive psychology. Philosophy, too, has recognized the fundamental importance of generalization, as well as its ubiquitous nature. Reichenbach (1951) contended that generalization is the essence of knowledge, and long before him Aristotle (1963) maintained that inductive reasoning, rather than divine revelation, is the prime source of human beliefs. Psychologists have added that generalization is perhaps the simplest and most pervasive of everyday inductive tasks through which people come to know their physical and social world (Krueger and Clement 1996; Nisbett et al. 1983).

Generalization from instances is surely ubiquitous and important in a marketing context and has obvious implications for repeat purchase and customer loyalty. Surprisingly though, behavioral consumer researchers have raised hardly any explicit questions on the topic. Is one purchase experience with a vendor enough to induce strong expectations about future quality with that vendor? If not, does a typical purchase experience induce stronger generalization? Does a negative experience lead to stronger generalization than a positive one? Does the nature of the purchase determine the level of generalization, that is, does a product experience lead to different generalization than a service experience? Is the objective base rate of the occurrence of a certain outcome the sole driver of generalization? Although insights from a variety of literatures are
suggestive, systematic behavioral research is missing in marketing (Folkes and Patrick 2003 for an interesting exception).

The goal of this dissertation is to directly address these unanswered questions by systematically testing the validity of pre-existing generalization frameworks in a consumer context and ultimately proposing a framework that has not been used systematically in the generalization or induction literature. Results suggest that pre-existing frameworks do not hold up as well as might have been anticipated. The newly proposed framework fits the overall data pattern better.

The focus is on generalization from one or two purchase experiences with a vendor. The dependent measure probes the degree to which consumers find this small sample representative for a large population of experiences, often expressed in percentages. For instance, in a restaurant setting, the question pertains to the percentage of all meals on the menu that respondents believe will have the same quality as the meal chosen.

Presentation and discussion of empirical research make up the bigger part of this dissertation (Chapters 4-6), but a theoretical overview is outlined first (Chapter 2). Three pre-existing frameworks are introduced before an alternative approach is outlined, one that is a virtual stranger to classical generalization or induction research. Chapter 3 explicates the research questions and evaluates initial evidence in the marketing domain.
CHAPTER 2
LITERATURE REVIEW AND THEORY DEVELOPMENT

Generalization and Induction Research

Three distinct frameworks emerge from research closely and directly related to the question of how and when consumers generalize: (1) the Law of Small Numbers, (2) the heterogeneity account, and (3) the typicality account. A seminal marketing paper relies on the first of these accounts and is the starting point of this literature review.

Hoch and Deighton (1989) propose that generalization will occur even when only very small samples are available as input to the estimate. Although not focusing on generalization, they suggest that consumers do not generate many hypotheses when confronted with a certain purchase outcome; for instance, a bad meal in a restaurant. Instead, consumers jump to conclusions fairly quickly and generalize, even when statistical criteria do not support such conclusions.

Hoch and Deighton (1989) based their view on two streams of literature, the first being concerned with the Law of Small Numbers, which states that people generalize quickly because they believe that even small samples are representative (Tversky and Kahneman 1971). An example in a marketing context is the consumer who labels a restaurant as bad on the basis of one negative experience (e.g., a bad meal). The second, arguably less well-known literature on hypothesis generation seems to confirm that people often generate no more than one hypothesis (Gettys and Fisher 1979; Gettys, Mehle, and Fisher 1986). Moreover, hypothesis formation may happen in a rather passive way so that salient aspects of the problem drive the content of the actual hypothesis.
(Hoch and Deighton 1989, p. 4). Thus one bad meal may easily be taken as sufficient evidence of a bad restaurant. The Law of Small Numbers and the work by Hoch and Deighton are clearly in line with the claim that people perceive less variability than is actually present (Kareev, Arnon, and Horwitz-Zeliger 2002).

Other researchers have argued that people may not always generalize from small samples. In a reply to Tversky and Kahneman (1971), Nisbett et al. (1983) show that heterogeneity is perceived, in some situations at least, and thus generalization is low. This view will henceforth be referred to as the “heterogeneity account.” Empirical evidence is provided in a scenario requiring participants to imagine they arrive at a previously undiscovered island. They have to estimate the percentage of tribe members, Barratos, who are obese. The only input available for the estimate is the knowledge that a given sample is obese. The size of the sample is varied: 1, 3 or 20 obese tribe members. Clearly, generalization is low when the sample consists of 1 member (~38%), higher with a sample of 3 (~56%), and highest with a sample of 20 (~75%).

In stark contrast, generalization about an object that is found on this imaginary island, floridium, which is said to burn with a green flame when heated, is extremely high, even with an n=1 sample. Nisbett et al. (1983) argue that obesity is perceived as containing more heterogeneity in a tribe population than does a population of floridium elements. This difference in perceived heterogeneity allows Nisbett et al. to make their point that, at least in some situations, people perceive heterogeneity even though only a small and arguably homogenous sample is overtly available. In those situations, people are notably less likely to generalize and will instead reason statistically. This “generalization” literature, concerned with the Law of Small Numbers and perceived
heterogeneity, focuses on statistical reasoning and whether or not people engage in it. However, at least two important questions remain largely unanswered. When are people likely to perceive heterogeneity and when are they not likely to do so? Also, and in more general terms, what drives perceived heterogeneity?

One seemingly obvious factor that may affect perceived heterogeneity is the typicality of the sample. When a given sample is atypical for its population, perceived heterogeneity may be high and thus generalization low. Conversely, when the sample is typical for its population, perceived heterogeneity may be low and generalization high (Rothbart and Lewis 1988). This pattern of results is anticipated by what is usually referred to as the “induction literature” in cognitive psychology, which will be referred to here as the “typicality account” (Heit 2000; Osherson et al. 1990; Rips 1975; Sloman 1993).

The typicality effect is perhaps the single most-demonstrated effect in the induction literature, and typicality as a factor may account for more explained variance than any other factor. A priori, the induction literature may be considered as the single most-important source that can help answer questions raised in this dissertation because of its rich history, its significant volume, and the force of one of its main arguments, that typicality drives generalization.

Although the typicality effect is compatible with the notion of perceived heterogeneity and the shared focus on generalization in the induction and generalization literature, the induction literature does not build on the generalization literature. Reasons for this lack of cross-referencing are unclear, but focus in the two streams of research differs slightly. Induction research focuses on “how people project information from
known cases to the unknown” (Heit 2000, p. 569), rather than on the (non)normativeness of any such generalization, which concerned Tversky and Kahneman (1971). Also, the induction literature relies almost exclusively on natural or biological categories (e.g., animal categories) for its stimuli, while Tversky and Kahneman and Nisbett et al. (1983) rely more on “social” stimuli. And although there are a variety of induction models, most assume some kind of typicality calculation underlying the induction judgment. The output of the typicality judgment almost automatically leads to the induction judgment. That is, very basic processes, sometimes called “bottom-up” processes, drive generalization. In contrast, the framework developed in Chapter 3 assumes that higher level processes--"top down” rather than “bottom up”--drive generalization.

Induction research focuses on the typicality of the sample in relation to the population, but explicitly ignores the dimension or characteristic on which the sample is considered (Heit and Rubinstein 1994, for an exception). Like Nisbett et al., the framework developed here focuses less on the typicality of the sample than on the importance of the dimension or characteristic. Finally, generalization in induction research is usually based on category knowledge rather than object knowledge: an entire category such as all sparrows rather than just one or a few sparrows. Despite the differences between the two lines of research, a 25-year review of induction research by Heit (2000) concluded that most induction models apply widely. Typical samples lead to stronger generalization than do atypical samples (e.g., a “sparrow” sample leads to stronger inferences concerning “all birds” than does a “penguin” sample). An equivalent example in a marketing context would be generalization to all meals on a steakhouse menu from a steak rather than a pasta dish. In sum, given the dominant status of the
induction literature and the extensively documented typicality effect, typicality may well be the most important factor that drives perceived heterogeneity and thus generalization.

Available insights in the generalization and induction literature suggest that consumers are highly likely to generalize, even with very small samples. If they do not, the perception of heterogeneity of the characteristic in its population is the inhibiting force. Despite the lack of cross-referencing between the induction and generalization literature, typicality of the sample appears to be an extremely important factor in the perception of heterogeneity and thus generalization. However, a fourth account is introduced to explicitly consider consumer contexts.

**Causal Lay Theories as an Alternative Framework**

Although insights from the literature review seem compellingly documented and comfortably intuitive, easy generalization to a consumer context is not guaranteed. Do consumers always generalize from small samples, as the Law of Small Numbers suggests? The induction literature, as well as research by Nisbett et al. (1983), suggests they may not. If they do not, will typicality be a major determining factor? Or will other factors also drive perceptions of heterogeneity and thus generalization?

In an effort to understand generalization in a consumer context, I introduce the concept of causal lay theories as a determinant of generalization, an alternative option to the frameworks already reviewed. Unlike the Law of Small Numbers, the causal lay theory framework assumes that consumers will not always generalize. It is also broader than the typicality account in allowing for low generalization when the sample is typical and for high generalization when the sample is atypical. It is more specific than the perceived heterogeneity account by assuming a definite source for perceived
heterogeneity: the causal lay theory that is invoked to understand the occurrence of a particular outcome.

Lay theories are causal knowledge structures about how the world functions or is organized. Depending on the specific causal lay theory generated, generalization will vary. When the mechanism underlying the theory assumes stability or systematicity, generalization tends to be high, even when based on very small samples. When the assumed mechanism reflects instability, generalization is low. An example of a stable mechanism may be the expertise of the chef in a restaurant scenario, whereas an example of an unstable mechanism may be an “off day” for the chef. If the chef scores low on expertise today, he probably did so yesterday and is likely to continue to do so tomorrow. However, a chef having a bad day may have had a good day yesterday and may be expected to have another good one tomorrow. Depending on which mechanism is invoked, generalization from a single experience will vary.

Note that lay theories--the central concept of the framework--often do not reach the rigor and consistency expected of a scientific theory (Nisbett and Ross 1980; Tversky and Kahneman 1980). While lay theories have been shown to improve judgment in some situations in the social realm (Wright and Murphy 1984), they may hurt judgment in other situations (Chapman and Chapman 1969). However interesting, the appropriateness of the impact of those lay theories is not the focus of this dissertation. Instead, the emphasis is on whether consumers hold such causal lay theories and if so, whether they drive generalization. The induction literature implicitly suggests that they do not and that typicality calculations drive generalization. A major goal of this dissertation is to address these questions explicitly, empirically (Chapters 4-6) and theoretically.
The remainder of this chapter reviews a variety of literatures that may be theoretically relevant but that has not been relied on by induction or by generalization researchers. Insights from the literature on causal reasoning and that on hypothesis generation address the question of whether people hold causal lay theories. Evidence from the social psychology literature on stereotyping, the correspondence bias, the negativity effect, and person versus group perception is discussed to provide initial evidence for the second question of whether causal lay theories drive generalization. The suggestion is that people can and do hold causal lay theories, even about events that have occurred only once, and that such theories impact generalization, at least in the social realm.

**Do People Hold Causal Lay Theories about Events that Have Occurred Once or Twice?**

This question implies two more subquestions. First, is the human cognitive system able to draw causal inferences from very small samples? Second, if so, how many of those inferences (theories) are typically made per observation?

**Is the human cognitive system designed to draw causal inferences from very small samples?**

If people have a causal lay theory--if they draw a causal inference--about a one-time event, then they must feel they know why something happened after having experienced it only once. Such a claim may run counter to common wisdom as well as to insights from some of the most influential thinkers in psychology (Heider 1958, Kelley 1967) and philosophy (Mill 1973). For instance, at the heart of Kelley's ANOVA model lies the necessity to observe covariation between presumed cause and effect before a causal inference can be made. According to the model, causal inferences will be drawn to the extent that cause and effect coincide in a series of observations. One observation is,
by definition, not a series; so according to this account, it seems unlikely for people to
draw causal inferences based on only one observation.

When two observations are presented in this dissertation, only outcome
information without information about a cause was provided. As such, covaration
information is again not overtly available and causal inferences seem unlikely. The
ANOVA and similar accounts may make one wonder whether people actually draw
causal inferences in the type of situations studied in this dissertation. However, it has
been recognized (Hilton and Slugoski 1986) that causal attributions are not always made
as prescribed by the normative ANOVA or any other covariational model (Cheng 1997;
White 2002).

Different approaches have been proposed (Ahn et al. 1995; Ahn and Kalish 2000;
White 1990). The approach advanced by Ahn and her colleagues may reveal relevant
insights. The basic argument is that people seek out causal mechanisms in their
knowledge base to develop an explanation for a specific event, rather than relying solely
on covariation information to identify a causal relationship between sometimes arbitrary
factors. Somewhat simplified, this means that people confronted with a certain event will
search for an explanatory mechanism or theory. Clearly, the suggestion is that it is not
impossible for the human cognitive system to draw causal inferences from extremely
small samples. A remaining question then is how many reasons people typically generate.

**How many theories do people typically generate?**

As previously suggested, people seem able to draw causal inferences based on
only one observation, at least in some situations. A following question pertains to the
number of hypotheses generated for a single event or observation. Granted that any given
outcome is likely to be caused by multiple factors, it is not unreasonable to expect that people may generate multiple reasons or theories. If so, uncertainty about exactly which theory holds may increase and the degree of generalization may consequently decrease. The literature on hypothesis generation may reveal valuable insights inasmuch as it deals with the number of hypotheses (theories) people tend to generate for a particular problem (event or instance in our analysis). (The terms "hypotheses" and "theories" are used interchangeably in this section.)

The possible number of hypotheses or theories one can consider for a given problem can theoretically be large, and people are clearly able to take into account or generate multiple theories for a given problem (Alba, van Osselaer and Vanhouche 2003; Arocha, Patel and Patel 1993; Bockenholt and Weber 1993; Fisher et al. 1983; Gettys and Fisher 1979; Koehler 1994; Kruglanski 1990; Liberman et al. 2001; McClure, Jaspars and Lalljee 1993; Mehle 1982; Trope and Liberman 1996; Weber et al. 1993). For instance, it is likely that multiple theories will be generated by a physician faced with a patient’s problem or by a detective dealing with a crime.

However, an extensive review on hypothesis (theory) generation and testing (Sanbomatsu et al. 1998) concludes that hypothesis testing is often a first-come, first-confirmed process. The first viable theory generated has an enormous comparative advantage (p.202). This research suggests that people often generate only one explanation for a given problem. This is consistent with work by Simon (1956, 1982): People search for a satisfying but not necessarily optimal solution (Garst et al. 2002). This tendency is consistent with research that suggests that people tend to pursue confirmatory strategies (Arocha et al. 1993; Hoch 1984; Hodgins and Zuckerman 1993; Klayman 1995; Schaklee
and Fischoff 1982; Tschirgi 1980; Zuckerman et al. 1995): Most of us look for the presence of what we expect, not for what we would not expect. In addition, it has been suggested that once a hypothesis becomes focal (i.e., has been generated) its strength is overestimated (Sanbomatsu, Akimoto and Biggs 1993). This renders a causal inference more plausible and makes the “hypothesizer” more confident (Gettys, Mehle and Fisher 1986; Koehler 1994).

If the foregoing analysis of the literature on causal reasoning literature and hypothesis generation is valid, it makes sense to expect that people draw causal inferences, even from samples with only one observation. This is exactly what Read (1983, 1984) has empirically demonstrated. In line with the work of Ahn et al. (1995) he has shown that causal inferences based on one observation are more likely when a plausible theory (an analogy) is available (Read 1984; Weber et al. 1993).

It seems fair to conclude that people can and do draw causal inferences based on samples with as few as one or two observations. We tend to do this when we have a lay theory that explains the datum in a satisfactory way. As soon as one lay theory has been activated, the search ceases, and uncertainty originating from the plausibility of multiple theories is not experienced. A subsequent question is whether a person’s causal theories are used as input in generalization judgments. I address this issue empirically in a marketing context (Chapters 4-6), but first outline evidence from social psychology.

Are Causal Lay Theories Used as Input in Generalization Judgments?

Evidence supporting the impact of lay theories comes from research streams that are not necessarily focusing on generalization or induction per se, but instead deal with lay theories in one way or another. Four significant bodies of research, all from the social psychology literature, are briefly discussed.
Research on person versus group perception indicates that people perceive more heterogeneity when a certain behavior is performed by a group rather than by an individual (Hamilton and Sherman 1996). Consequently, generalization is stronger regarding a person than a group. That is to say, people seem to factor in the origin of a certain behavior and thus they seem to have a certain theory about its causes.

Research on stereotyping is rather explicit about the impact of lay theories. It no longer questions whether people’s lay theories--in this case called “stereotypes”--affect judgment, but instead accepts that they do and asks when they are activated and applied and when they are not (Kunda and Spencer 2003). For instance, if people hold the causal lay theory that advanced age slows down certain responses, the question in this literature is no longer whether the causal belief affects future judgment about speed-related behavior but when it does. The impact of lay theories is found to be so pervasive that research is even focusing on situations in which judgment is not impacted. For the purpose of this dissertation, the message is that research other than the core induction or generalization work suggests that lay theories impact judgment.

Research on the correspondence bias also seems relevant (Gilbert and Malone 1995). Participants in a representative experiment draw conclusions about the personality characteristics of an actor, even on the basis of one isolated observation of behavior that is explicitly said to be caused by situational factors rather than personality traits (e.g., Jones and Harris 1967). The social psychology research is interested in the fact that people perceive personality traits to be stable predictors of behavior, whereas situational characteristics often are (set up to be) the actual predictors. The point of interest here is that people seem to bring a specific causal lay theory to the scene, one that attributes
behavior to personality traits, not to certain cues in the environment. It is this stable lay theory that determines the level of generalization: People generalize extensively regarding the future behavior of the actor, even on the basis of a single observation. The implication is that they would have generalized much less—or at least much differently—if the lay theory concerned situational cues rather than personality traits. That is, not only is a lay theory driving generalization, but impact of the stability of the assumed mechanism is implicitly recognized as well.

A fourth significant body of research finds that negative information receives more weight than positive information (Folkes and Kamins 1999; Herr, Kardes and Kim 1991; Peeters and Czapinski 1990; Reeder and Brewer 1979; Skowronski and Carlston 1987). In a social context, Ybarra (2002) translates this observation into the claim that people believe that negative behavior is caused by stable personality traits while positive behavior is caused by variable situational cues. The implication is that stronger generalization occurs based on negative rather than positive behavior, which suggests that consumers will generalize more strongly from a negative rather than from a positive purchase experience. Further, theories featuring a “personality” mechanism lead to strong generalization while those underlying a “situational” mechanism lead to weak generalization. Ybarra goes even as far as claiming that the perception of social behaviors is driven by the goal of inferring underlying causes in the person. That is, not only do people apply lay theories when available, they actively search for them, at least in the social realm.

To summarize, research using social stimuli tends to support the notion that lay theories impact generalization judgment and that hypothesized stable mechanisms lead to
high generalization while unstable mechanisms lead to comparatively low generalization. In addition, it is argued that people actively search for causal lay theories. Before considering whether lay theories impact generalization in a *marketing* context, I compare the four theoretical frameworks, advance hypotheses to test them, and review the scarce evidence available in the marketing literature in light of the four accounts.
CHAPTER 3
RESEARCH QUESTIONS AND INITIAL EVIDENCE IN MARKETING

Research Questions and Hypotheses

When and how do consumers generalize? This general but central question is approached by testing the validity of the three frameworks that have been extensively investigated in the generalization or induction literature and by examining a fourth approach that has remained virtually untested as a generalization framework. This chapter explicates how each of the accounts differ and advances hypotheses to test their validity.

One stunningly simple yet powerful answer to our general question comes from the Law of Small Numbers (Tversky and Kahneman 1971) in the generalization literature: Consumers always generalize because they find even the smallest sample representative. Several experiments allow testing of this hypothesis. The first includes two manipulations--typicality and outcome valence--that can be expected to induce lower generalization.

The perceived heterogeneity view is non-compatible and posits that consumers will not always generalize. Generalization will be low when perceived heterogeneity is high (Nisbett et al. 1983). Confirmation of this hypothesis implies disconfirmation of the previous one. Experiment 1 is again applicable in this regard, and Experiment 2 tests whether generalization is driven only by objective heterogeneity (in the sample and/or the population).

The typicality account (Heit 2000) is also incompatible with the Law of Small Numbers but is more specific than the heterogeneity framework in explicitly anticipating
lower levels of generalization when the sample is atypical rather than typical. It further
distances itself from other accounts by suggesting that generalization occurs virtually
automatically as a result of low-level typicality calculations. If valid, the implication is
that generalization necessarily covaries with typicality. This hypothesis is also tested in
Experiment 1.

The fourth account originates from research other than strict generalization and
induction work. Like the typicality and heterogeneity accounts, the lay theory framework
is incompatible with the Law of Small Numbers in that it assumes low generalization in
some situations. It is broader than the typicality account in allowing for high
generalization when typicality is low, and low generalization when typicality is high, and
it is more specific than the heterogeneity account in assuming a specific source of
perceived heterogeneity: the lay theory invoked to account for the outcome of a specific
purchase situation. Although the exact content of lay theories can vary greatly, the
stability of the underlying mechanism is one-dimensional. This stability supposedly
drives generalization. The lay theory framework, contrary to the other three accounts, has
remained virtually untested as an explicit generalization account. To make up for this
lapse, this research project started testing specific marketing-relevant hypotheses, as
outlined below. Most are derived from the extant literature.

If the causal lay theory perspective is to be a valid generalization account, it must
(1) show that consumers hold causal lay theories and (2) that the underlying stability of
these theories drives generalization. For this approach to differentiate itself from the
typicality account, it is further desirable (3) that it shows stability and instability in
contexts independent of typicality. Experiment 4 explicitly addresses the first and third
requirements, testing whether negative outcomes elicit stable theories while positive outcomes elicit unstable mechanisms (Ybarra 2002) or whether the opposite occurs, as claimed by Folkes and Patrick (2003).

Experiments 3, 5 and 6 address the second requirement. The question is whether consumers’ lay theories about product versus service experiences affect generalization, and whether their theories are similar to those held by academics. For instance, Zeithaml et al. (1985) recognized that heterogeneity is higher in a service than in a product context. If consumers’ lay theories conform, generalization is expected to be higher in a product than in a service context because more stability is assumed in the former.

Causal lay theories about single outcomes are not the only theories investigated. Experiment 2 tests the presence of lay theories about a sequence of two inconsistent outcomes. This allows for a test of the widely held belief that “the first impression matters most” (Nisbett and Ross 1980) and thus that more stability is inferred from a “first” rather than a “recent” experience.

If a lay theory perspective turns out to be defendable, another question pertains to the speed of belief (theory) updating. For instance, consumers may hold a positive belief about a certain company but encounter a negative experience that is attributed to an unstable cause. In such a situation they may generalize only to a limited degree. It is imperative to investigate how many more negative experiences are needed before the negative outcomes are perceived to be caused by a stable mechanism. Experiments 3, 4, 9 and 10 test the hypothesis that consumers update beliefs slowly (Boulding, Kara and Staelin 1999) but also consider the possibility that consumers update quickly. Fast
updating would not be inconsistent with a lay theory perspective that assumes an unstable mechanism after one experience but a stable mechanism after two experiences.

**Initial Evidence in Marketing Context**

In evaluating previous research in marketing, I identified three articles that deal explicitly with the generalization question. These give supportive evidence for the typicality and the perceived heterogeneity accounts. An unmodified Law of Small Numbers does not seem to fit the scant available data and evidence for the lay theory framework is mixed.

Using a paradigm similar to that used in standard induction research, Joiner and Loken (1998) studied the typicality of the “population” rather than of the sample. For instance, employing the traditional induction paradigm whereby participants are exposed to the following format, they found that (2) is rated as a stronger argument than (1).

1. Sony TVs have attribute X; therefore, Sony bicycles have attribute “X”.
2. Sony TVs have attribute X; therefore, Sony cameras have attribute “X”

In other words, generalization from one specific category (e.g., TVs) to another is stronger when the latter is more typical (e.g., camera rather than bicycle). Although this is not the “typical” typicality effect, the results clearly show an effect of typicality on generalization. It is important to note, however, that as in most induction experiments, typicality information is made salient by explicit comparison of the typical and atypical information.

Boulding et al. (1999) do not investigate the impact of typicality but advance suggestions about the pace with which evaluative beliefs about a service experience are updated as a function of the number of experiences with the service. Specifically, the authors investigated generalization from one hotel experience (sample) to the overall
population of experiences with the hotel. They conclude that "any time a firm wants to change consumers' beliefs of its perceptual positioning through the delivery of goods or services, managers must recognize that these beliefs may be slow to adjust" (p. 481). It may be informative to note that some of the data reported by Nisbett et al. (1983) seem in line with this claim. Even with a sample of 20 observations generalization remains short of asymptote in estimates of tribal obesity. Boulding et al. argue that multiple observations are needed to update belief. This seems counter to a lay theory perspective, which could conceivably allow for fast updating as a function of an increasing number of observations. For instance, while one negative experience in a restaurant may be perceived as being caused by an unstable mechanism, a second negative experience may induce the perception of a stable cause and thus high levels of generalization.

The third and most recent study (Folkes and Patrick 2003) is less pessimistic about the impact of lay theories on generalization. Although these authors did not manipulate typicality either, they observed differential generalization as a function of valence of the outcome. Generalization to the population of colleague service providers was more pronounced when the sample consisted of a friendly, rather than unfriendly, service employee. That is, a positivity effect is observed in a service context. The mere observation that generalization is rather low, based on negative outcomes, may point to the limited generalizability of the Law of Small Numbers in a marketing context. The combination of low generalization in the negative outcome condition with high generalization in the positive outcome condition is taken by Folkes and Patrick as evidence for the existence of lay theories. Such lay theories would imply that a good outcome is perceived as caused by a stable mechanism (e.g., the policy of the firm),
whereas a negative outcome is perceived as caused by an unstable mechanism (e.g., the personality of one person who decided not to follow the firm’s policy). Direct evidence for these theories is not provided, however.

To summarize, the Law of Small Numbers is not always consistent with the available evidence. For example, generalization is lower in the negative outcome condition in Folkes and Patrick (2003). Both the typicality (Joiner and Loken 1998) and lay theory accounts (Folkes and Patrick 2003), and thus also the perceived heterogeneity perspective, seem to receive at least some support. However, Boulding et al. (1999) anticipate slow updating of beliefs, which is not necessarily anticipated by a lay theory perspective.
CHAPTER 4
TEST OF EXPLANATORY POWER OF COMPETING ACCOUNTS

The empirical section is organized around three sets of experiments, each presented in a separate chapter. Chapter 4 explores the explanatory power of the four theoretical accounts. Chapter 5 targets the lay theory account and examines the characteristics and impact of specific lay theories on generalization, especially when the sample is n=1. Chapter 6 focuses on those situations when two observations are available as input, the first of which is perceived as caused by an unstable mechanism.

Experiment 1

The main goal of Experiment 1 is to investigate the effect of typicality on generalization in a marketing context. At the same time the valence of the purchase outcome is manipulated to investigate whether the negativity effect widely observed in the psychology literature can be replicated.

The typicality effect has been demonstrated so extensively in the induction literature that it may seem redundant to aim for replication. However, as noted before, a specific set of stimuli and a specific paradigm have been used in the induction literature, and both may have favored the impact of typicality in induction research in an artificial way.

The induction literature does not make differential predictions regarding generalization about positive versus negative purchase outcomes. However, this distinction is obviously relevant to marketers, especially since most available evidence
suggests that generalization is more extreme when based on negative rather than positive experiences. Interestingly, some consumer researchers have reported a weak negativity effect (Ahluwalia 2002) or even a positivity effect (Folkes and Patrick 2003).

**Method**

Seventy business majors at the University of Florida participated in the experiment for partial class credit. Participants imagined they went to a new restaurant with a group of friends. For at least 30 seconds, they read a menu with seven meat dishes (e.g., New York steak of black angus beef, prime rib of beef) and one pasta dish (Appendix A for the full stimuli). Half of the participants were told they had selected a beef item (typical), while the others learned they had selected the pasta (atypical item). Contrary to the standard induction experiment, typicality in this experiment needs to be inferred. Typicality of the selected item is crossed with the outcome of the meal. After having selected their meal, half of the participants learned the quality of their meal was low while the other half learned it was high. Participants were next asked (1) what percentage of all meals in this restaurant did they think would be of the same quality as the meal they had selected and (2) what was the rationale behind their generalization estimate. This experiment, as well as all the others, was conducted entirely on computer.

**Predictions**

Various predictions can be made. First, the induction literature anticipates a main effect of typicality in that generalization is more extreme in the typical than atypical condition (Figure 4-1A). Second, a large body of evidence anticipates higher generalization with a negative rather than positive outcome (Figure 4-1B). Third, the Law
of Small Numbers proposes that generalization will be high across all conditions (Figure 4-1C).

![Graph](image)

Figure 4-1. Patterns of results as anticipated by various accounts (A, B and C) and as observed in Experiment 1 (D). A) Predicted pattern by typicality account, B) Anticipated pattern of negativity effect is observed, C) Pattern anticipated by Law of Small Numbers, and D) Observed pattern in Experiment 1.

**Results**

**Manipulation check.** In an independent study, 32 undergraduate students judged the degree to which they found their item (steak or pasta) typical, given the menu. Despite the relatively low cell sizes (n=16), the difference between the typical and atypical conditions was highly significant \[t(30)=5.4; p<.001\], thereby demonstrating the strength of the typicality manipulation.
Numerical data of main experiment. Figure 4-1D clearly shows that neither of the three predicted results is actually obtained. A 2x2 (typicality x outcome valence) between-subject ANOVA revealed no main effect of typicality (F<1); instead there is an outcome effect of lower generalization in the negative condition [F(1,66)=4.3; p<.05]; no evidence for an interaction is obtained (F<1).

Cognitive responses. Two independent judges analyzed the cognitive responses that were recorded after the generalization estimate had been made. In all experiments except Experiment 2 the coding categories were “stable,” “unstable,” and “ambiguous.” “Stable” indicates a stable mechanism, such as the expertise of the chef, while “unstable” refers to an unstable mechanism such as the chef having a bad day. Responses that did not fit either category were coded “ambiguous.” The relation between the level of generalization and the assumed underlying mechanism is established by calculating the correlation between the two, with the “mechanism” variable being a dichotomous variable (unstable=0, stable=1). The goal of this measure is to provide directional evidence for the effect of lay theories, even when a specific manipulation does not reveal the anticipated impact of lay theories. In Experiment 1, the judges agreed in 91% of the cases and came to a joint conclusion in the other 9%. The resulting correlation between generalization and stability was r=.56 (p<.001).

Discussion

At least three observations are surprising in that they do not correspond to predictions of prominent literatures: (1) contrary to predictions in the induction literature, typicality failed to impact generalization; (2) a positivity instead of a negativity effect is observed since generalization is higher in the positive condition than in the negative
condition; (3) contrary to predictions of the Law of Small Numbers, one condition showed significantly lower generalization than another.

Do the results imply that typicality does not impact generalization “naturally”? A claim based on this one marketing experiment would be far-fetched, but the results suggest that generalization in the marketplace is not driven by mere typicality calculations as suggested by the induction literature. Moreover, the results may indicate that, if lay theories drive generalization, typicality may not always be the primary input for such theories. Perhaps typicality does not “jump out” as much as previously believed.

The second surprising observation refers to the non-occurrence of a negativity effect. Instead, generalization is higher in the positive than in the negative condition. This result, too, is interesting for several reasons. First, it validates the “null effect” of the typicality manipulation in that the latter cannot be attributed to negligent or inattentive participants who slid the scale without thinking about or even reading the scenario shown to them. Second, the mere observation that generalization varies between any two conditions rules out an unmodified Law of Small Numbers as a framework to account for the results. Third, whereas large bodies of literature anticipate a negativity effect, Experiment 1 shows a positivity effect.

One may wonder why a positivity effect is observed in this experiment (and in Folkes and Patrick’s 2003 study) while a negativity effect is found in so many others. Also, what drives the positivity effect? In line with the positive correlation between generalization and stability of the underlying mechanism, one could argue that lay theories drive the results and thus the positivity effect. If so, consumers believe that positive outcomes are caused by stable mechanisms, such as the expertise of the chef,
while negative outcomes are caused by unstable mechanisms, such as the chef having a bad day. This would suggest that people have different lay theories about positive and negative outcomes in the commercial world--where a positivity effect is observed--than in the social world, where a negativity effect has been reported (Ybarra 2002).

It is also possible, however, that generalization is not driven by typicality calculations or by lay theories. Instead, people may know and use the objective percentage of positive or negative experiences in restaurants. That is, base rates may drive the positivity effect and thus generalization. Although we do not know the exact base rates of positive and negative experiences in restaurants, it seems reasonable to expect more positive than negative experiences, rather than the opposite.

In sum, exactly which factors are driving the results in Experiment 1 is impossible to determine at this point. It is clear, however, that typicality does not drive generalization. Furthermore, the overall data pattern is inconsistent with the Law of Small Numbers, but not with the perceived heterogeneity framework, the lay theory account, or a base rate explanation. The plausibility of a base rate explanation is addressed more explicitly in Experiment 2.

**Experiment 2**

Experiment 2 allows further investigation of the impact of base rates on generalization by presenting participants with two experiences, one positive and one negative. The order in which they appear and the time interval between them are manipulated. Neither factor should impact generalization when objective base rates about the occurrence of good and bad outcomes drive generalization. However, manipulation of both factors may induce differential lay theories about why the specific set of outcomes
occurred. If lay theories drive generalization, differences may be expected as a function of those factors. A third goal is to investigate whether consumers reason statistically when confronted with overt heterogeneity: the positive and the negative purchase experience. Nisbett et al. (1983) argued that statistical reasoning is more likely to occur when people have reason to believe that there is heterogeneity on the considered dimension. This experiment explores whether overt heterogeneity is a sufficient source to fuel perceived heterogeneity, induce statistical reasoning and thus inhibit generalization.

Method

In anticipation of a manipulation in Experiment 3, the restaurant context used in Experiment 1 was changed to an explicit product or service context. Participants from the University of Florida (n=182) and Erasmus University (n=110) read a scenario in which a consumer experience—either a product or service experience—was depicted with either a positive or a negative outcome. Before any assessment was recorded, participants were told they experienced the same brand again—either one year or one week later, this time with the opposite outcome as result (Appendix B for the full stimuli). The dependent variable pertained to expectations about a third experience being positive, negative, or unpredictable. A 201-point scale was presented, ranging from very negative (-100) over fifty-fifty (0) to very positive (100). As in Experiment 1, participants were allowed to express the rationale behind their response in an open-ended question.

A 2x2x2 between-subject design was employed, crossing the order of outcomes (bad-good vs. good-bad) with delay between outcomes (one week/one year) and purchase context (product/service).
Predictions

If people reason statistically—that is, they treat their two outcomes in the same way they would treat two coin flip outcomes—the average estimate in all conditions should not deviate from zero (Figure 4-2A). However, it is also possible that consumers plug in base rates when generalizing. As Experiment 1 suggests, people may believe that positive purchase experiences outnumber negative experiences. If such base rates are the basis for generalization, a pattern similar to the one shown in Figure 4-2B can be expected.

Figure 4-2. These are some possible outcomes in experiment 2. A) Anticipated pattern of people reason statistically, B) Expected pattern of base rates dominate, C) Pattern of results indicating a primacy effect and D) Pattern of results indicating a negativity effect.

A third possibility is that consumers have a theory as to why their specific set of outcomes occurred. A variety of theories can be adopted, two of which are considered here. First, consumers may emphasize either the more recent or the first experience
(Figure 4-2C). This would result in either a recency or a primacy effect. Both phenomena have been reported extensively by a variety of researchers. For instance, in the social realm, Nisbett and Ross (1980) stated that “primacy effects are overwhelmingly more probably” (p.172). Research on cognitive ability (Jones et al. 1968) as well as a review paper on order effects suggest that, indeed, the first impression is often considered the most important one, at least in a short and easy task such as that used in Experiment 2 (Hogarth and Einhorn 1992). However, research in other domains suggests recent information may be more dominant (Cuccia and McGill 2000 regarding accounting, Davis 1984 regarding jury decision-making).

Alternatively, but not necessarily mutually exclusively, consumers may hold theories in which either the negative or positive experience dominates while the other is perceived as an outlier or the beginning or end of a trend. In line with the result of Experiment 1, it may be that the positive outcome will dominate. In that case, a recency effect would be observed in the bad-good and a primacy effect in the good-bad condition. Note that such a result would be very much in line with a base rate explanation. Alternatively, as suggested by a large body of other evidence, it is possible that negative information will dominate the estimation responses (Figure 4-2C).

**Results**

Brief inspection of the means in Figure 4-3 indicates that none of the predicted patterns is obtained. Instead, the means show a combined emphasis on recent and positive information. In addition, generalization is stronger when there is a one-year, as opposed to a one-week, delay between the two outcomes. This pattern is confirmed by a 2x2x2 [delay (one week/one year) x order of outcome (good-bad/bad-good) x purchase context (product-service)] ANOVA on the absolute values, with all three factors considered as
between-subject factors. Absolute values are used rather than the raw data because the raw data may fail to pick up a time lag effect even if it is present. The main effects of time lag \(F(1,284)=9.3; p<.01\) and order of outcome \(F(1,284)=18.2; p<.001\) confirm, respectively, that generalization is more extreme after one year rather than one week, and more extreme in the bad-good condition than the good-bad condition. No evidence is provided for a main effect of purchase context \(F(1,284)=1.8; p<.18\) or for any interaction (highest \(F=1.9; p<.17\)).

Figure 4-3. The observed pattern of results as a function of time delay and order in which the outcomes are presented.

Because this is the only experiment with inconsistent outcomes, the coding scheme for the cognitive responses is different in this experiment. Judges classified the responses either as evidence for statistical reasoning or for a lay theory (other than statistical reasoning). “Statistical reasoning” responses refer to a 50/50 chance of the next experience being positive/negative. A lay theory response was defined as a reference to one experience dominating the other, such as ‘the company is improving’, or ‘the negative experience was a fluke’. The judges agreed in 94% of the cases. References to a lay theory tend to dominate as generalization became more extreme. For instance, in the bad-good condition with a one-year time interval, 85% of the responses were classified as referring to a trend or an outlier, while 14% were classified as statistical reasoning (1%
was not classified in either category). References to statistical reasoning were relatively more dominant when generalization was less extreme. For instance, in the good-bad condition with a one-week interval, 57% of the responses were classified as a lay theory, and 36% as statistical reasoning (7% were not classified in either category).

**Discussion**

Experiment 1 failed to support the typicality account, but could not rule out a base rate account as an explanation for its findings. Indeed, consumers may have had knowledge about the overall percentage of positive and negative purchase outcomes and used that knowledge exclusively to make their generalization estimate. In a similar way, participants in Experiment 2 could have plugged base rate knowledge into their estimate. For instance, if 60% of purchases can be objectively classified as having a positive outcome, at least a positive number should have appeared in each of the four conditions without significant differences across the conditions. However, the observed pattern of results is very different.

In addition, the pattern of results deviates significantly from what could be expected if consumers reason statistically. Aggregate generalization levels were strongly different from zero in all four conditions, despite the strong overt heterogeneity. Clearly, the combination of one positive and one negative purchase experience is not perceived in the same way as a head-and-tail result from two coin flips. In other words, the results in Experiment 2 cannot be fully explained by the objective heterogeneity of the sample or by the objective heterogeneity of the larger population of positive and negative purchase experiences; that is, the base rates.

Still, considerable systematic variance is left: A combined recency and positivity effect is even more pronounced as the time interval increases. What is causing this
variance? One possible interpretation is that consumers try to give meaning to--come up with a lay theory about--the sequence of events encountered and that this meaning is reflected in their generalization estimate. For instance, and in line with the results of the cognitive responses, participants in the bad-good condition may believe that this sequence of outcomes occurred because the company is improving. Understandably, there is more room for improvement in a time span of one year as opposed to one week. The result is therefore more extreme in the one-year condition. The opposite seems true for the good-bad condition, but to a lesser degree.

These results, then, may provide the first piece of evidence, albeit indirect, in favor of a lay theory perspective, while at the same time making a simple base rate interpretation less likely. The Law of Small Numbers is supported in that people do not tend to reason statistically, even in the face of strong heterogeneity cues, but it is contradicted in that consumers do not find a small sample representative.

The specific content of the lay theories is not a first concern, but it is certainly not inappropriate to ask why these theories and not others? One reason positive rather than negative information may dominate theories is that consumers may assume the company they interacted with is in business. Given this assumption, it is more logical to expect that a company is improving or performing well (bad-good condition), rather than deteriorating or performing badly (good-bad condition).

The other question pertains to why lay theories are dominated by recent rather than primacy information. Research on cognitive--rather than commercial--skills has shown a primacy effect (Jones et al. 1968). The same pattern dominates in the social realm (Nisbett and Ross 1980) and was anticipated in an impressive review on recency and
primacy effects (Einhorn and Hogarth 1992). Einhorn and Hogarth expect a recency effect when more complex and cognitively demanding tasks are involved or when two estimates are made (one after each outcome) instead of one. The task in Experiment 2 is classified as short and easy, and participants make only one estimate, as soon as they received information about the valence of both outcomes. The only way to align our findings with Einhorn and Hogarth’s framework is to assume that participants made two estimates--one implicitly after the first experience and the other explicitly after the second experience. Future research may want to explore this possibility, for instance, by explicitly having participants make two estimates--one after each outcome. If the result does not change, Einhorn and Hogarth’s framework is supported. However, it is not inconceivable that the procedural change might induce a primacy--instead of a recency--effect, given that consumers may make up their mind quickly after a first experience.

Taken together, the results of Experiment 2 suggest (1) that consumers do not reason statistically in the face of overt heterogeneity, (2) that simple base rates do not (always) drive generalization, and (3) that, instead, lay theories may impact generalization. A recurring observation is that generalization is more extreme on the positive than on the negative side. Experiment 3 tests the boundaries of this positivity effect.

**Experiment 3**

Experiment 3 explores the boundaries of the positivity effect. It also tests the impact of lay theories about purchase experiences in a product versus a service context.

Experiments 1 and 2 showed that generalization is more extreme on the positive than on the negative side. An emerging question is how persistent this positivity effect is.
Are consumers incurably positive? In other words, is even a set of multiple negative experiences perceived as an exception and is the updating process consequently slow, as suggested by Boulding et al. (1999)? Or does a second negative experience lead rather quickly to the same level of extreme generalization as does one positive experience? Such a finding would not be inconsistent with a lay theory perspective. It would suggest that consumers perceive one negative experience as an exception, but two negative experiences as an indication of stable negative performance.

The second goal of Experiment 3 is to explore whether consumers hold different theories about various sets of purchase experiences and, more importantly, whether those lay theories drive generalization. For instance, many scholars may agree with Zeithaml et al. (1985) that heterogeneity is higher in a service context than in a product context. If so, consumers should generalize to a larger degree in a product than in a service context. The specific theories underlying generalization may recognize that products are the output of an automated production process whereas services are generated by a much more variable mechanism. A final question is whether the positivity effect observed in a restaurant scenario in Experiment 1 can be replicated in a different context.

**Method**

In this experiment, 93 students at the University of Florida were asked to imagine they had had one or two experiences with “a product” or “a service,” with the description left vague. If two experiences occurred, the outcomes were consistent. In both the product and the service conditions, the scenario showed four examples of a product/service in parentheses (Appendix C for the full stimuli).
This resulted in a 2x2x2 design with purchase context (product vs. service), number of experiences (one vs. two) and outcome valence (positive vs. negative) as three between-subject factors. As in Experiment 2, the task was to estimate the likelihood that the next experience would be positive or negative, and cognitive responses were recorded after the generalization estimate had been submitted.

**Predictions**

A persistent positivity effect and thus slow updating, as suggested by Boulding et al. (1999), would be evidenced by a pattern similar to the one depicted in Figure 4-4A.

![Figure 4-4A](image1)

![Figure 4-4B](image2)

![Figure 4-4C](image3)

![Figure 4-4D](image4)

Figure 4-4. Pattern of results as anticipated by several accounts (A, B and C) and as observed in Experiment 3. A) Expected pattern if the positivity effect is to persist when n=2, B) Pattern as anticipated by work of Zeithaml et al. (1990), C) Expected pattern if a positivity effect occurs in a service context and a negativity effect in a product context. D) The actually observed means as a function of number of experiences and outcome valence.
Figure 4-4B shows the possible impact of a product-service effect: The direction of generalization is identical in both conditions, but the magnitude is higher in the product condition (Zeithaml et al. 1990). If, however, a positivity effect occurs in the service condition, and a negative effect in the product condition (Folkes and Patrick 2003), the pattern should look like the one depicted in Figure 4-4C.

**Results**

Descriptively, Figure 4-4D shows a replication of the positivity effect when one experience is involved, but no evidence for a positivity effect when two consistent episodes are experienced. A 2x2x2 [valence (positive-negative) x number of exposures (one-two) x purchase context (product-service)] ANOVA confirmed this pattern. The number of exposures by valence of outcome interaction is significant \[F(1,85)=7.8; p<.01\], thereby confirming the annihilation of the positivity effect after two exposures \[t(44)=1.2; p<.24\] for the difference between positive and negative conditions after two episodes on absolute values. Although the main effects of number of exposures \[F(1,85)=4.7; p<.05\] and valence \[F(1,85)=540.7; p<.001\] are significant, the main effect of purchase context is not \(F<1\). Overall, generalization is more extreme after two experiences than after one, and more extreme on the positive than negative side. Whether the experience involves a product or service, makes no difference. In analyzing the cognitive responses, the two judges agreed in 96% of the cases. The overall correlation between generalization and stability is \(r=.39\) \((p<.053)\).
Discussion

Experiment 3 was designed to test boundaries of the positivity effect and thus the speed with which consumers update their beliefs. At the same time, a first test probed the impact of product- and service-related lay theories on generalization.

Despite the changed context and the use of a slightly different dependent measure, the positivity effect observed in Experiment 1 is compellingly replicated in Experiment 3, thus supporting the robustness of the phenomenon.

In addition, the results show that consumers confronted with two experiences generalize equally extreme on the negative as on the positive side. That is, the positivism of consumers has clear limits. One implication is that consumers may switch easily from an unstable to a stable mechanism when generalizing. In other words, the increased generalization in the negative outcome condition may be suggestive of a fast updating process, which is inconsistent with Boulding et al.’s (1999) predictions but not with a lay theory perspective.

The other interesting observation is the lack of differentiation between the product and the service condition. Experiment 3 suggests that consumers do not hold the same theories as academic scholars. Another possibility is that consumers do hold the same lay theories, but that these do not impact generalization. One could argue that this is evidence against the lay theory perspective. Still another interpretation, however, is that these lay theories are held by consumers, but that the manipulation in Experiment 3 was too weak to elicit them.

The full pattern of results in Experiment 3, then, seems at least partially inconsistent with each of the three frameworks that reasonably could make predictions.
The Law of Small Numbers has difficulties with the low level of generalization after one negative experience, while both the heterogeneity and the lay theory framework would have anticipated differences as a function of purchase context.

When taken together, the full pattern of results observed in Experiments 1-3 cannot easily be accounted for by any of the reviewed frameworks. An unmodified Law of Small Numbers and the typicality account cannot be retained, while the heterogeneity and lay theory perspectives receive mixed support. Because the heterogeneity account does not add much to the lay theory perspective, the next set of experiments targets the validity of the lay theory perspective and further explores the degree to which consumers hold theories about positive and negative experiences. They also further examine purchase experiences in a product versus a service context.
CHAPTER 5
IN SEARCH OF LAY THEORIES AND THEIR IMPACT ON GENERALIZATION

The major goals of experiments reported in this chapter are to obtain direct evidence for the existence of lay theories in a consumer context and for their impact on generalization.

Experiment 4

Given that the evidence in favor of lay theories has been indirect and even then inconclusive at best, the main goal of Experiment 4 is to provide direct evidence. Do consumers believe that positive outcomes are caused by stable mechanisms, such as the expertise of the chef, and negative outcomes by unstable mechanisms, such as the chef having a bad day? Such beliefs are necessary to interpret the positivity effect in Experiments 1 and 3 in terms of a lay theory perspective. A variant of the restaurant scenario employed in Experiment 1 is used in Experiment 4.

Method

This experiment included 55 students, 27 of whom were randomly assigned to the positive outcome condition. All participants were presented with a scenario in which a restaurant critic goes to a particular restaurant to write a review. The critic’s meal is said to be “very good” for half of the participants and “not very good” for the other half. As part of the review, the critic wants to explain why the restaurant produced a good/bad meal. The task of each participant is to provide reasons that could be used in the critic’s review (Appendix D for full scenario). In a second phase, two independent judges who
were unfamiliar with the hypothesis of the experiment coded each reason as either a “stable” or “unstable” mechanism.

**Results**

There were 120 separately codable units (causes) in the negative condition and 109 in the positive condition, all coded as “stable,” “unstable” or “non-codable.” The two judges agreed in 74% of the cases. Inconsistencies were resolved through deliberation by the two judges.

![Figure 5-1. Classification of causes as either stable or unstable.](image)

Of the codable responses, 94% was classified as “stable” (6% as “unstable”) in the positive outcome, compared to 51% in the negative outcome condition, where 49% were coded “unstable” (Figure 5-1). The difference in the number of stable/unstable causes between the positive and negative outcome condition is statistically significant \[\chi^2=23.7 ; p<.001\]. Non-codable responses comprised 52% of all responses in the positive outcome condition and 61% in the negative outcome condition.

**Discussion**

Experiment 4 provides the first direct evidence for lay theories. Consumers are more likely to believe that good outcomes (e.g., a good meal) are caused by a stable
mechanism and bad outcomes (e.g., a bad meal) by an unstable mechanism. This result is consistent with the positivity effect observed in Experiments 1 and 3. However, as interesting and important as this observation is, it is not helpful unless consumers use such theories when generalizing. Experiments 5 and 6 further investigate the impact of lay theories on generalization by probing the product-service distinction employed in Experiments 2 and 3.

**Experiment 5**

Experiment 5 takes both Experiments 3 and 4 one step further. Having provided evidence for the existence of lay theories (Experiment 4), the next step is to show the impact of those theories on generalization. The context of product versus service (Experiment 3) is selected again, with the manipulation slightly strengthened. The generic product and service description in Experiment 3 is made more specific in Experiment 5. The rationale is that the more specific the description, the more likely that specific theories will be triggered, and therefore the more likely that differences in generalization will occur between the product and the service condition.

**Method**

The goal in the stimuli selection was to include products that typically show little variance in the production process and to choose service stimuli in which more variability can be perceived. Toothpaste, a printer cartridge, and a battery were used as product stimuli; financial advice, a hotel stay, and delivery service were used as service stimuli (Appendix E for full stimuli).

In this experiment, all the purchase experiences were negative. The 127 participants imagined they had had either one or two negative experiences. The outcome
information was still rather general, just as in Experiment 3. A 2x2 [number of exposures (one vs. two) x purchase context (product vs. service)] was employed.

Results

![Figure 5-2. Average generalization as a function of number of experiences (one/two) and purchase context (product/service).](image)

The results are shown in Figure 5-2. Once again, despite the stronger manipulation, no significant difference is observed between the product and the service condition (F<1). Instead, the main effect of “number of exposures” is highly significant [F(1,123)=50.5; p<.001]. The interaction is not significant (F<1) and neither is the single main effect of replicate in the product [F(2,62)=1.6; p<.21] or service condition (F<1). Agreement in coding the cognitive responses was reached in 91% of the cases and the correlation was r=.39 (p<.05).

Discussion

Given the intuitively plausible difference in heterogeneity between a product and a service context, the lack of any significant difference is striking. Again, one possible explanation is that consumers do not hold the same theories about products and services that most academics do. Or, consumers may hold those theories but not apply them when generalizing. Experiment 6 further probes a third possibility: Consumers hold lay theories
about products and services similar to those that academics hold, but a stronger manipulation is needed to elicit them.

**Experiment 6**

The rationale behind Experiment 5 was taken one step further in Experiment 6. The purchase scenario was made even more specific. This time participants were not only provided with specific product or service information, they were also informed about the specific dimension on which the product or service did not perform appropriately.

**Method**

A pool of 95 students evaluated either four service or four product replicates. Replicates in each condition were selected as intuitive representatives of exemplars in their respective categories and to fall short in a specific characteristic. Replicates in the product condition are “a razor blade that provides a somewhat rough shave,” “a bad tasting chocolate bar,” “a wristwatch that is running behind,” and “a pen that does not distribute the ink evenly.” “An undercooked meal,” “rude service in a coffee shop,” “your lawn is not carefully mowed,” and “being placed on hold for a long time at a call center” are the replicates in the service condition (Appendix F for the full scenario). The question to the participant was, “If you purchased 100 ‘replicate,’ what percentage would perform equally poorly?”

**Results**

Figure 5-3 shows the means in the product and service condition for each of the replicates. Descriptively, average generalization seems higher in the product condition than in the service condition. Statistical analyses confirm the main effect of purchase
context when the replicate is treated as a repeated measure, \([F(1,93)=11.5; p<.001]\) and almost when treated as a between-subject factor \([F(1,87)=3.1; p<.09]\).

![Figure 5-3. Mean generalization per replicate in the product and service condition.](image)

The main effect of replicate in the product \([F(3,141)=20.3; p<.001]\), but also in the service condition \([F(3,138)=19.5; p<.001]\) was highly significant when the replicate was treated as a repeated measure in two one-way ANOVAs, which indicates high variability even within the product and service contexts. Agreement among judges in the coding task was 88%. The correlation was \(r=.57 (p<.001)\).

**Discussion**

The results of Experiment 6 extend those observed in Experiments 3, 4, and 5 in various ways. Relative to Experiments 3 and 5, overall generalization in the negative condition of Experiment 6 increased in the product, but not in the service, condition. That is, the specific description of the purchase situation in Experiment 6 led to the pattern inferred from Zeithaml et al. (1990): higher generalization in a product context than in a service context. A product-service distinction in Experiment 6 that is absent in Experiments 3 and 5 suggests that, even if consumers hold generic lay theories about purchase experiences in a product versus service context (e.g., the “assembly line”
theory), more particular lay theories about specific purchase situations, rather than broad product-service differences, may be elicited more naturally.

By showing evidence for the product-service distinction, Experiment 6 extends the results of Experiment 4--consumers hold lay theories--in multiple ways. First, it provides evidence for the existence of another set of lay theories, a set reflecting the work of Zeithaml et al. (1985). More importantly, it reinforces the lay theory perspective by showing that those lay theories do affect generalization.

However much Experiment 6 provides important support for the lay theory perspective and answers crucial questions, it certainly sparks new ones. It is informative about the specific level of product- or service-related lay theories that consumers may apply when generalizing. Although the observed main effect of product-service is interesting and in line with the proposition inferred from Zeithaml et al. (1985), one should not disregard the considerable systematic variance within both the product and service conditions. This variance suggests that the specific dimension and/or the specific product or service category may be as crucial as the broad product-service distinction. Indeed, if just one of the replicates had been selected for each of the two conditions, it would have been possible to observe a main effect of purchase context that is opposite to the observed pattern: higher generalization in a service than in a product context. Such an outcome is not anticipated by the classical heterogeneity distinction between products and services (Zeithaml et al. 1990) and may add an important insight to the existing theoretical literature, as well as to the knowledge base of the manager. Whether the specific dimension or the specific category or a combination of both determines the
specific lay theory is impossible to infer from the design employed in Experiment 6. Experiment 7 aims at disentangling the impact of both factors.

**Experiment 7**

Although Experiment 6 shows differences in generalization between a product and a service context, in line with Zeithaml et al. (1990), the considerable variance within the product and service conditions suggests that this distinction does not suffice to predict an appropriate level of generalization. Instead, the specific product/service category may be crucial just like the dimension on which the product/service performs inadequately.

Experiment 7 disentangles the impact of the two factors by keeping the category constant and varying the dimension. Because more overall variance is expected in a service than in a product context, only service replicates are included in Experiment 7. Dimensions that had been considered relevant by previous researchers (Zeithaml et al. 1990; Coulter and Coulter 2003) were selected and crossed with three service categories.

**Method**

Four dimensions of service quality were selected from Zeithaml et al. (1990): competence, reliability, courtesy, and credibility. Each dimension was crossed with each of three replicates: a restaurant scenario, a car repair scenario, and a painter scenario. Operationalization for the four dimensions in the restaurant scenario was as follows (Appendix H for the full stimuli): “The waiter mixed up the orders and therefore no one gets exactly what he ordered” (competence); “The waiter is slow to take your order” (reliability); “The waiter seems to be abrupt and unfriendly” (courtesy); “The waiter has overcharged you” (credibility). The dependent measure was worded as follows: “For each
set of 100 customers, what percentage do you feel would have the same experience with
this waiter/mechanic/painter?” The study participants were 259 students.

Results

Descriptively, the results in Figure 5-4 show that generalization is a function of a
combination of the replicate (category) and the dimension. This observation is confirmed
by the statistical analyses in a 3x4 (replicate x dimension) ANOVA, with both factors
treated as between-subject factors. The interaction between dimension and replicate is
significant [F(6,247)=7.6; p<.001] and so are the main effects of dimension
[F(3,247)=16.7; p<.001] and replicate [F(2,247)=14.6; p<.001].

Figure 5-4. Average generalization as a function of replicate and dimension.

Agreement among judges in coding the cognitive responses was 89%. The overall
correlation between the numeric level of generalization and the dichotomous level of
stability was r=.43 (p<.001). To illustrate, the extremely low generalization in the
credibility-restaurant scenario is backed up by an “unstable” rating for 14 out of 15
codable responses.
Discussion

No specific directional predictions were advanced at the outset of this experiment. Instead, the main goal was to explore whether generalization varies as a function of the specific dimension and/or category. The interaction between dimension and category confirms that both the category and the dimension matter. Together with the results of Experiment 6, the data in Experiment 7 indicate that generalization can vary as a function of the broader category (product vs. service: Experiment 6) but also as a function of the specific category and dimension within a service context. The suggestion is that consumers come up with a different theory, depending on the very specific purchase situation in which they find themselves. For instance, overcharging seems to be perceived more of an exception when performed by a waiter than when performed by a mechanic or painter. Overcharging by a waiter also seems to be perceived as more of an exception than unfriendliness of a waiter. That is, not only is the product-service distinction too broad to determine an accurate generalization level, so is the specific product or service category. The performance dimension is also needed to predict generalization.

An even more extreme example of how specific lay theories can be is suggested by a comparison across experiments. Experiment 7 includes a restaurant/unfriendly (courtesy) scenario while Experiment 6 describes a coffee shop/rude scenario. Although both scenarios seem highly similar and even interchangeable--especially against a background of broad product-service differences--generalization is considerably higher in the restaurant/unfriendly scenario (mean = 59) than in the coffee shop/rude scenario (mean = 38). It is entirely possible that this difference is not systematic and is attributable
to error. However, it would be interesting to see whether such subtle differences spark differential theories. Experiment 8 pursues this possibility.

**Experiment 8**

Context (restaurant/coffee shop) and behavior (unfriendly/rude) are orthogonally manipulated. The suggestion is that consumers may perceive negative outcomes to be caused by more stable mechanisms in a coffee shop than in a restaurant and that unfriendly behavior is seen as more stable than rude behavior. It would be interesting if consumers are sensitive to these subtle differences, but not to such seemingly obvious marketing variables as “product” and “service” at the broad generic level (Experiments 3 and 5).

**Method and results**

Sixty-five undergraduate students participated in this 2x2 experiment that crosses context (restaurant/coffee shop) with behavior (rude/unfriendly) as between-subject factors. The students were told to imagine an encounter with a rude/unfriendly waiter in a restaurant/coffee shop and were asked what percentage of customers they feel would have the same experience in this place (Appendix I for the full stimuli). Confirming the across-experiment differences twice (Figure 5-6), the two main effects are significant. Generalization is higher in a coffee shop than in a restaurant \[ F(1,61)=4.7; \ p<.04 \], and higher with the unfriendly than the rude behavior \[ F(1,61)=4.1; \ p<.05 \]. There is no evidence for an interaction \( F<1 \). Coding of the cognitive responses resulted in 89% agreement and a correlation of \( r=.66 \) (\( p<.001 \)).
Discussion

The differences observed across experiments between rudeness and unfriendliness on the one hand and a restaurant and coffee shop context on the other hand are replicated in a well-controlled environment. This suggests that consumers perceive true differences between seemingly interchangeable contexts or behaviors. This finding, together with the positive correlation between generalization and stability, is important from a marketer's perspective and is interesting because it suggests that rudeness/unfriendliness are perceived as being caused by a more unstable mechanism in a restaurant than in a coffee shop. At the same time, rudeness is perceived as being caused by a more unstable mechanism than unfriendliness. However, one cannot rule out a simple base rate explanation: Rude behavior may be less likely in a restaurant than in a coffee shop and rude behavior may occur less frequently than unfriendly behavior. If so, future research may want to investigate the interplay between causal lay theories and base rates in determining generalization.

Before turning to Chapter 6, I summarize the evidence collected in Experiments 4-8. The second set of experiments supports the lay theory account where the first set did not. Experiment 4 shows that consumers hold lay theories consistent with the positivity
effect observed in the first set. Experiment 6 indicates that consumers not only hold lay theories but also apply them when generalizing. Experiments 7 and 8 suggest that those lay theories tend to be more specific than could reasonably be anticipated. The suggestion--when \( n=1 \)--is that generalization is dependent on the very specific lay theory that is applied, and not easily predictable. To what degree the same is true as the sample size increases is not clear. The question is especially pertinent in those situations when \( n=1 \) and where an unstable mechanism is assumed. The third set of experiments explores this issue more systematically.
CHAPTER 6
MULTIPLE “UNSTABLE” OBSERVATIONS

Experiment 9

Given that consumers seem to be unwilling to generalize in some situations, it becomes important to understand the boundaries of this reluctance as the number of observations increases. It is possible that Boulding et al.'s (1999) prediction holds and that an unstable mechanism dominates even after multiple negative outcomes have been experienced. Alternatively, consumers may quickly assume a stable mechanism when \( n=2 \) after initially having perceived an unstable mechanism. In other words, the assumption of an unstable mechanism may be short-lived. If true, the implication may well be that the Law of Small Numbers holds when \( n=2 \), if not when \( n=1 \).

Method

Ninety-five students participated in the study. The two scenarios producing the lowest mean generalization in Experiment 7 were included in Experiment 9: the wristwatch running behind and the undercooked meal. In addition, two scenarios from Nisbett et al. (1983) were included for comparison: the element floridium burning with a green flame on an imaginary island and obese members of a tribe (Barratos) on this island. Extreme generalization has been observed with floridium, even after one trial, while moderate to low generalization has been observed for obesity (Appendix I for full stimuli).
Results

Descriptively, the low levels of generalization observed after one episode are replicated for the marketing stimuli (Figure 6-1). Similarly, the high level of generalization for the floridium element in Nisbett et al. (1983) is replicated, as is the moderate level of generalization for the obese Barratos. Generalization after two episodes increases dramatically for the two marketing stimuli, but much less or not at all for the other replicates.

![Figure 6-1. Mean generalization as a function of replicate and number of exposes.](image)

These observations are confirmed by a replicate by number of exposures interaction [F(3,87)=6.1; p<.01]. The single main effect for number of exposures is significant for the marketing stimuli [F(1,46)=43; p<.001], but not for the two other replicates [F<1]. Overall, the main effect of number of observations is significant [(F(1,87)=28.7; p<.001)], as is the main effect of replicate [F(3,87)=9.2; p<.001]. Coding of the cognitive responses led to agreement in 93% of the responses and an overall r=.63 correlation (p<.01).

Discussion

All four cells with a sample of n=1 replicate generalization previously observed either in Experiment 6 or in Nisbett et al. (1983). Generalization is low in the wristwatch,
restaurant, and obese scenario, while extremely high in the floridium condition. In addition, generalization based on two observations in the obesity and floridium cases seems in line with the results that Nisbett et al. reported. The observed pattern in the obese condition seems to correspond to the slow updating process anticipated by Boulding et al (1999). However, the contrast with the marketing stimuli is stark. Whereas generalization was extremely low after only one observation, the addition of a second observation induced a huge increase. In this case, the updating process does not seem to be slow but extremely fast. The suggestion is that even though consumers may initially surmise an unstable mechanism, a stable mechanism is assumed as soon as the same episode is experienced twice. The lesson for the marketer may not be that consumers learn and update slowly, as argued by Boulding et al., but that they infer consistent low quality easily, even when high quality is initially expected. However, it is also possible that the high generalization when n=2 is the result of participants’ compliance with what they think is the hypothesis pursued by the experimenter. Experiment 10 pursues this possibility.

**Experiment 10**

Thus far all experiments that included marketing stimuli showed high generalization in the two-experience conditions. Experiment 10 explores whether the high generalization when n=2 should be attributed to “demand” or “blind generalization.” To test this contention, one condition is included in which the target service behavior for the second negative experience is explicitly said to be performed by a different rather than the same waiter. The actual outcome is identical to that when the second experience is caused by the same waiter. If the previous results are attributable to “blind
generalization” when n=2, high generalization may reasonably be expected in Experiment 10 when n=2, even when the waiter in the second episode is different. Alternatively, if lay theories--instead of “demand”--drive generalization, one might expect low generalization when the waiter is different in the second episode.

**Method and Results**

Sixty participants were asked to imagine they went to a coffee shop and had one or two negative service experiences with a rude waiter. The second experience either involved the same waiter or a different waiter. When the waiter was the same, participants were asked, “For each set of 100 customers, what percentage do you feel would have the same experience with *this waiter*?” When the waiter of the second experience was a different one, the question was, “For each set of 100 customers, what percentage do you feel would have the same experience in *this coffee shop*?” (Appendix J for the full stimuli) This resulted in a 2x2 design that crosses the number of experiences with the target of generalization (this waiter/this coffee shop). Because the dependent measure is different in the two “target of generalization” conditions, the data are analyzed separately. For purposes of presentation, the means are presented in one figure.

![Figure 6-2. Mean generalization in each of the four conditions in Experiment 10.](image)
The mean levels of generalization per condition indicate that generalization is considerably higher after two experiences than after one in the “individual” condition but much less in the “coffee shop” condition (Figure 6-2). T-tests confirm that the former is highly significant \[t(28)=-4.7; p<.001\] while the second does not even approach significance \[t(28)=-1.1=p<.3\]. The correlation between level of generalization and stability is \(r=.47\) (\(p<.01\)). The judges agreed in 97% of the responses.

**Discussion**

When rude behavior was displayed in the second experience by the same waiter, generalization increased dramatically, just as in previous experiments. However, when the target behavior was associated with a different waiter, generalization did not increase relative to the one-experience condition. This result suggests that participants did not generalize “blindly” in previous experiments when \(n=2\) and that the Law of Small Numbers does not always hold, even when \(n=2\). Instead, the suggestion is that generalization can be low when \(n=2\) if the appropriate theory is cued, e.g., a different waiter is responsible for the negative outcome.
"When do consumers (not) generalize?" and "How should one understand generalization in the marketplace theoretically?" are the central questions in this dissertation. The experiments show that generalization is high in a product context rather than a service context and when the outcome is positive rather than negative, but not when the sample is typical rather than atypical. Even within a product or service context, considerable variance indicates that generalization can be high in a service context and low in a product context. Although two consistent outcomes tend to be perceived as representative of the larger population, two inconsistent outcomes tend not to be.

The results are interpreted in line with a causal lay theory perspective. Generalization is high when the assumed mechanism is stable, low when the mechanism is unstable. Direct evidence for the existence of such lay theories is provided by Experiment 4, which shows that consumers tend to believe that stable mechanisms cause a positive outcome while unstable mechanisms cause a negative outcome. Evidence for the impact of lay theories on generalization is provided by Experiment 6, where generalization is higher in a product context than in a service context. Across experiments, correlational evidence shows that generalization tends to be higher when a stable rather than unstable mechanism is assumed.

The level of correspondence between these results and previous research differs depending on the domain selected for comparison. Correspondence is lowest with the
majority of research in the generalization and induction domain, but surprisingly high
with research in the social psychology literature and with category formation research.
Relevant literatures are discussed and suggestions for future research are incorporated in
the following discussion that focuses first on the claim that lay theories drive
generalization and then on the actual lay theories.

The Lay Theory Account

A major inconsistency occurs when the focus of comparison is the “calculation”
account suggested by the induction literature (Osherson et al. 1990; Sloman 1993; Heit
2000). In many induction studies, typicality exerts a pervasive and reliable impact, yet it
failed to influence generalization in this research. Does this discrepancy suggest that two
qualitatively different mechanisms drive generalization in a biological (induction)
context compared to a marketing or social context? Do lay theories drive generalization
in marketing and social contexts, while calculations predominate in a biological context?
Or is the lay theory account a special case of the calculation account or perhaps the other
way around?

Another look across the borders of strict generalization, induction, and even social
psychology research might be instructive. A seminal paper on category formation by
Murphy and Medin (1985) suggests a way to resolve the apparent discrepancy. Murphy
and Medin rejected the long-held belief that similarity calculations determine which
objects are grouped together to form a category (Medin and Schaffer 1978; Posner and
Keele 1968; Rosch and Mervis 1975). They argued that the concept of similarity is too
unconstrained (Goodman 1972). Instead, they proposed that “concepts are coherent to the
extent they fit people's background knowledge and naive theories about the world”
(Murphy and Medin 1985, p.289; also Rehder and Hastie 2001; Rips and Collins 1993). Simply put, even if similarity calculations drive category formation, lay theories drive and constrain similarity judgments and thus category formation. Transposed to the domain of generalization, Murphy and Medin's claim suggests that even typicality findings in the induction literature are the result of a specific set of theories about biological categories rather than the output of simple and “objective” calculations. The suggestion is that generalization in the induction literature is therefore not qualitatively different from generalization in a consumer context or in a social context. The paradox is resolved by considering the typicality effect as induced by subjective lay theories rather than by objective calculations. Future research may want to investigate this empirically.

Still, even if valid, Murphy and Medin’s (1985) suggestion does not explain why generalization seems hypersensitive to typicality information in the induction literature, but not at all in our context. Are there fundamental differences between lay theories about the biological world and lay theories in the consumer and social world? This is certainly an interesting hypothesis to be investigated in future research. Alternatively, there may be differences between the paradigms that are responsible for the discrepancy. Perhaps the induction paradigm favors typicality more than does the paradigm used in this research project. Induction scenarios tend to include both a typical and atypical sample. As such, one can argue that typicality information is made salient and a typicality effect is more likely. Should one go as far as implying that the induction paradigm induces artificial effects that do not exist in the real world? Probably not. Instead, the induction paradigm may simply highlight certain aspects of a situation, and historically this has often been typicality information. As such, the paradigm can be considered a valuable tool for
detecting specific sets of theories. In fact, an interesting challenge for future research would be to employ the induction paradigm as a tool to test the impact of theories not based on typicality. For instance, managers may want to know if consumers hold different lay theories according to whether a company owns or franchises its stores (Agrawal and Lal 1995). Vertically integrated systems (owning) may induce more generalization across stores than does the franchising format because more independence, and potentially instability or heterogeneity, is allowed in the latter. Similar questions can be addressed regarding “make” or “buy” decisions (Anderson and Weitz 1986). Do consumers infer more stability when something is made by the company rather than bought from another company? Also, consumers may hold different theories about “direct” salespeople versus “representatives” (Anderson 1985). Perhaps theories about representatives imply lower levels of generalization than do theories about “direct” salespeople. Systematically including both options in a stimuli set, may give a sense of the degree to which consumers hold different theories in each of these situations.

My results are not necessarily inconsistent with the hypothesis generation literature and with Hoch and Deighton’s (1989) suggestion that consumers generate very few hypotheses--often only one--about why a certain outcome occurred. However, my research extends theirs by positing that the stability of the underlying mechanism is important in determining the level of generalization, even if only one hypothesis is generated.

The results are at least partially inconsistent with the Law of Small Numbers. Although generalization tends to be high when the outcome is positive, and in a product context even when the outcome is negative, it is often low in a service context when the
outcome is negative. Even an overtly heterogeneous sample is not perceived as representative.

Interestingly, many of the findings that are inconsistent with the Law of Small Numbers are in line with Nisbett et al.’s (1983) view that people, in at least some contexts, perceive heterogeneity and thus refrain from generalizing. Our results extend Nisbett et al.’s research by identifying factors that do and do not seem sufficient to explain the variance in generalization. Neither the objective heterogeneity of the population (base rates) nor the objective heterogeneity of the sample proved sufficient to understand variance in generalization. That even typicality calculations failed to impact perceived heterogeneity is perhaps even more perplexing. Instead, causal lay theories seem to determine the level of perceived heterogeneity and thus generalization.

My results seem most consistent with findings in social psychology that people hold and apply causal lay theories in a variety of judgments. Research on stereotypes has long agreed that people hold beliefs--mostly stable--that impact judgment dramatically (Kunda and Spencer 2003). Research on the correspondance bias (Gilbert and Malone 1995), on person versus group perception (Hamilton and Sherman 1996) and on valenced behaviors (Ybarra 2002) implies that people perceive more stability in some situations than in others.

The results are also consistent with research on causal attribution, which states that people seek out causal mechanisms in developing an explanation for a specific event and that they do not necessarily need covariation information (Ahn et al. 1995).

Thus far, the discussion has concentrated on consistency between relevant literatures and my claim that lay theories drive generalization. At a more specific level, it
may be instructive to compare the actual lay theories that have been observed with what could reasonably be expected on the basis of previous research.

**Specific Lay Theories**

The number of lay theories that can be applied is virtually countless. We have started to document a few of them--some about individual purchase experiences, others about a sequence of two experiences.

First, consumers tend to believe that a positive outcome is generated by a stable mechanism, while a negative outcome is believed to be caused by an unstable mechanism. Consequently, positive outcomes lead to higher generalization than do negative outcomes. This finding is consistent with what Folkes and Patrick (2003) observed in a service context, but is opposite to what an entire body of research in the psychology literature anticipated. For instance, Ybarra (2002) argued that negative human behavior tends to be seen as the consequence of a stable mechanism--a personality trait--whereas positive behavior is perceived as the consequence of an unstable mechanism, e.g., situational factors. Together, these results suggest a perplexing dissociation between the social and the commercial world regarding the valence-stability relationship: Whereas people seem to believe that their own peer is inherently bad, a commercial entity seems to be perceived as inherently positive. Although this dissociation sounds like a scary thought in a world of ever-present and sometimes aggressive marketers, it is comforting to know that such positive beliefs about a commercial entity are not long-lasting, as suggested by Boulding et al. (1999), but change rapidly as multiple negative outcomes are experienced.
Second, the results suggest that consumers tend to believe that a product is generated by more stable mechanisms than is a service experience. Consequently, generalization tends to be higher in a product than in a service context, as suggested by the work of Zeithaml et al (1990). However, this seems true only when the outcome is negative. When the outcome is positive, stable mechanisms are assumed across the board—in line with the positivity effect—and generalization is high.

Another deviation pertains to the specificity of product- and service-related lay theories. Contrary to marketing scholars, consumers do not seem to hold differential theories regarding heterogeneity in a product versus service context at the generic level. Indeed, the product-service difference did not occur until very specific scenario descriptions were introduced. The implication is that, even though the product-service distinction may be an interesting guideline, generalization can vary greatly even within a product or service context, possibly to such a degree that generalization can be higher in a service context than in a product context. Equally important in determining the specific level of generalization is the specific product/service category and even the specific dimension on which the product/service is performing inadequately.

Still, the product-service distinction is an interesting parallel to findings in the social psychology literature about person versus group perception (Hamilton and Sherman 1996). Group behavior and a service experience seem to induce perceived heterogeneity (instability) much more than individual behavior and a product experience. My research suggests that this is true only when the outcome is negative. It would be interesting for future research to investigate whether valence moderates the effect in the social realm as well. If it does, one may wonder about the direction of the moderation...
since a negativity rather than a positivity effect has been reported in a social context (Ybarra 2002).

While theories about individual experiences seem to vary as a function of the valence and the nature of the purchase (product-service), theories about multiple inconsistent experiences tend to vary as a function of the delay between the outcomes and the order in which they appear. In addition, in line with theories about individual experiences, generalization tends to be more stable as function of a positive rather than negative experience. The order effect is interesting because it indicates that the most recent--rather than the first--outcome is more likely to be considered as a stable indicator of true quality. This is inconsistent with the majority of findings in the social psychology literature where a primacy-- instead of a recency--effect has more often been observed (Nisbett and Ross 1980). Whether the discrepancy between my findings in a marketing context and the bulk of evidence in the social psychology literature is attributable to procedural or more fundamental differences may be the target of future research. One possibility is that my paradigm induced recency by probing only one response right after information about the recent outcome had been processed, rather than two responses--each of them administered after each outcome was experienced. More fundamental differences between the two domains cannot be ruled out, and neither can a third possibility.

In the social realm, a primacy effect has been observed in person perception but a recency effect in group perception (Hamilton and Sherman 1996). Given the earlier analogy between person perception and generalization in a product context on the one hand and group perception and generalization in a service context on the other hand, it is
possible that a primacy effect would be observed in a product context and a recency effect in a service context. One reason this pattern did not occur in Experiment 2 may be the lack of specificity in the scenario.

**Conclusion**

Analogous to a hypothesis on concept coherence (Murphy and Medin 1985) and to work in the social realm (Hamilton and Sherman 1996; Ybarra 2002), but not anticipated by induction (Heit 2000) or generalization research (Tversky and Kahneman 1971), this dissertation claims that generalization is primarily driven by people’s lay theories about how the world functions---rather than by typicality calculations. To understand how and when consumers generalize, this account suggests, one should know which theory the consumer invokes in the specific purchase situation rather than (1) probe the typicality of the purchase situation or (2) assume that each first experience with a vendor is representative of all future experiences with the vendor. More specifically, one wants to know the stability of the underlying mechanism. A virtually endless set of theories can be applied, but it is the one-dimensional stability of the underlying mechanism that determines the level of generalization.
APPENDIX A
STIMULI FOR EXPERIMENT 1

Imagine it's Friday evening and you're going out for dinner with a group of friends. You decide to go the new restaurant downtown. You’ve arrived in the restaurant and you're about to explore the menu.

Look at it carefully.

--- New screen ---

- New York steak of Black Angus Beef with Crushed Black Peppercorn
- Tenderloin Filet of Beef
- Baby Back Ribs slow cooked with soy, beer and garlic, glazed with the house made barbecue sauce
- Beef Kebab In Teriyake Marinade
- Flat Iron Steak of Black Angus Beef Topped with Fresh Herb Butter
- Pasta with your choice of Tomato, Pesto or Clam Sauce
- Prime Rib of Beef
- Medaillons of Buffalo Tenderloin wrapped in Apple-Wood Smoked Bacon, served on a Grilled Portobello Mushroom
- All entrees come with a salad, bread and one side dish

--- New screen ---

You decided to have the Pasta / Flat Iron Steak of Black Angus Beef

It turns out you liked the meal a lot. / It turns out you didn’t think it was a very good meal compared to other restaurants where you’ve eaten at.
--- New screen ---

What percentage of all meals in this restaurant do you expect to be of the same quality as the meal you had?

Please move the sliding scale below to give your answer.

--- New screen ---

Why did you guess this percentage of other meals that would have the same quality, rather than a lower or higher number?

Write down your answer in the box below. Don't press Continue before you have finished writing.
APPENDIX B
STIMULI FOR EXPERIMENT 2

Imagine that you purchased a product/service (e.g. coffee, toothpaste, battery, cartridge for printer, disposable lenses, financial advice, hairdresser, hotel stay, delivery service, babysitting) from a particular firm for the first time. Your assessment of the product/service was positive/negative. A week/year later you decide to try the brand again. This time, however, your experience is negative/positive.

--- Next screen ---

Now, imagine that you have the opportunity to purchase a third time from the same firm. Using the scale below, please express your belief regarding the probability that the third experience will be good versus bad.

--- Next screen ---

Why did you give the answer you gave? That is, please describe the rationale behind your response.

Write down your answer in the box below. Don't press Continue before you have finished writing.
APPENDIX C
STIMULI FOR EXPERIMENT 3

Imagine that you purchased a product/service (e.g. coffee, toothpaste, battery, cartridge for printer, disposable lenses, financial advice, hairdresser, hotel stay, delivery service, babysitting) from a particular firm for the first time. Your assessment of the product/service was positive/negative. [Sometime later you decide to try the brand again. Your experience is positive/negative again.]

--- Next screen ---

Now, imagine that you have the opportunity to purchase once more from the same firm. Using the scale below, please express your belief regarding the probability that your next experience will be good versus bad.

--- Next screen ---

Why did you give exactly the rating you gave, instead of a lower or higher number? That is, please describe the rationale behind your response.

Write down your answer in the box below. Don't press Continue before you have finished writing.
APPENDIX D
STIMULI FOR EXPERIMENT 4

Restaurants try to provide good meals to their customers. Imagine that you just went to a particular restaurant and you had a meal that was (not) very good. Please list as many reasons as possible as to why this could have happened.

Write down your answer in the box below. Don't press Continue before you have finished writing.

Please separate different reasons with a star (*).
APPENDIX E
STIMULI FOR EXPERIMENT 5

Imagine that you purchased a tube of toothpaste / battery / cartridge for a printer // financial advice / hotel stay / delivery service from a particular firm for the first time. Your assessment of toothpaste / battery / cartridge for a printer // financial advice / hotel stay / delivery service was negative. [Sometime later you decide to try the brand again. Your experience is negative again].

--- Next screen ---

Now, imagine that you have the opportunity to purchase once more from the same firm. Using the scale below, please express your belief regarding the probability that your next experience will be good versus bad.

--- Next screen ---

Why did you give exactly the rating you gave, instead of a lower or higher number? That is, please describe the rationale behind your response.

Write down your answer in the box below. Don't press Continue before you have finished writing.
Imagine that you purchased a brand of razor blade for the first time. You try it and find that it provides a rough shave. If you purchased this brand 100 times, what percentage of the blades do you think would perform at the same level as the first one you purchased?

Imagine you bought a brand of chocolate candy bar for the first time. You try it and find that it doesn't taste very good. If you purchased 100 of these candy bars, what percentage do you think would taste the same as the first one you tried?

Imagine that you purchased a new wristwatch. After a week you find out that the clock is running behind. If you tried 100 watches of this model, what percentage do you think would give an inaccurate time?

Imagine that you purchased a brand of ballpoint pen that you have never purchased previously. You discover that the pen does not distribute the ink evenly, leaving the page with blobs and smears. If you purchased 100 pens of this model, what percentage do you think would perform equally poorly?

Imagine that you try a restaurant for the first time. You order your meal and find that it is undercooked. If you visited this restaurant 100 times, what percentage of all the meals you have do you think would be improperly cooked?

Imagine you go to a coffee shop and find the service to be rude. If you visited this coffee shop 100 times, what percentage of the service encounters do you think would be rude?
Imagine you just bought a new home and hired a company to take care of the garden. The company will send someone to your house once a week to mow the lawn. After the first week you find that the lawn is not mowed carefully. If the company mowed your lawn 100 times, what percentage of times do you think the lawn would be mowed poorly?

Imagine that you just bought a cell phone with a new service provider. You have some questions about your plan and decide to call the help desk. It turns out that you are placed on hold for 15 minutes before someone gets to your call. If you were to call the help desk 100 times, for what percentage of your calls do you think the waiting time will be (at least) as long as during your first call?
APPENDIX G
STIMULI FOR EXPERIMENT 7

Imagine that you go to a restaurant with a few friends. Each person orders a meal. When the meals are delivered, you find that the waiter has mixed up the orders and that no one gets exactly what they ordered in the way they wanted it prepared. For each 100 set of customers, what percentage do you feel would have the same experience with this waiter?

Imagine that you go to a restaurant with a few friends. The waiter is slow to take your order and to deliver the check at the end. For each 100 set of customers, what percentage do you feel would have the same experience with this waiter?

Imagine that you go to a restaurant with a few friends. The waiter takes your order but seems abrupt and not very friendly. For each 100 set of customers, what percentage do you feel would have the same experience with this waiter?

Imagine that you go to a restaurant with a few friends. At the end of the meal you receive your check and find that the waiter has overcharged you. For each 100 set of customers, what percentage do you feel would have the same experience with this waiter?

Imagine that you go to a mechanic for an oil change and some other standard maintenance. Afterward, you pick up your car and when you get home you find that your car is leaking oil. For each 100 set of customers, what percentage do you feel would have the same experience with this mechanic?

Imagine that you go to a mechanic for an oil change and some other standard maintenance. The mechanic promises that the car will be ready in 2 hours but it actually
takes much longer. For each 100 set of customers, what percentage do you feel would have the same experience with this mechanic?

Imagine that you go to a mechanic for an oil change and some other standard maintenance. The mechanic agrees to fix it but you find him to be somewhat unfriendly. For each 100 set of customers, what percentage do you feel would have the same experience with this mechanic?

Imagine that you own a home and would like to have some rooms painted. You hire a painter who does the job. Afterward, you notice that the painter had been sloppy in places and that the end result is not as nice as you expected. For each 100 set of customers, what percentage do you feel would have the same experience with this painter?

Imagine that you own a home and would like to have some rooms painted. You talk to a painter who is willing to do the job. You find him to be somewhat unfriendly.
For each 100 set of customers, what percentage do you feel would have the same experience with this painter?

Imagine that you own a home and would like to have some rooms painted. You hire a painter who does the job. When the job is finished, you get a bill that is higher than the price that the painter originally quoted. For each 100 set of customers, what percentage do you feel would have the same experience with this painter?
APPENDIX H
STIMULI FOR EXPERIMENT 8

Imagine that you go to a restaurant / coffee shop. The restaurant / coffee shop employs a lot of waiters and you are being served by one of them. You find him/her to be somewhat rude / unfriendly. For each set of 100 customers, what percentage do you feel would have the same experience in this place?
APPENDIX I
STIMULI FOR EXPERIMENT 9

Imagine that you purchased a new wristwatch. After a week you find out that the clock is running behind. [You decide to test a second watch of the same model. The performance of the second watch is identical to the first.] If you tried 100 watches of this model, what percentage do you think would give an inaccurate time?

Imagine that you try a restaurant for the first time. You order your meal and find that it is undercooked. [You decide to try the restaurant again. You order a meal and get the same result.] If you visited this restaurant 100 times, what percentage of all your meals do you think would be improperly cooked?

Imagine that you are an explorer who has landed on a previously unknown island in the Southeastern Pacific. You encounter several new animals, people and objects. Suppose you encounter a native who is a member of a tribe called the Barratos. He is obese. [You encounter a second member of the tribe and he is also obese.] If you encountered 100 male Barratos, what percentage do you think would be obese?

Imagine that you are an explorer who has landed on a previously unknown island in the Southeastern Pacific. You encounter several new animals, people and objects. Suppose you encounter a sample of a new element you call floridium. Upon being heated to a very high temperature, it burns with a green flame. [You encounter a second sample and it also burns with a green flame.] If you encountered 100 samples of floridium, what percentage do you think would burn with a green flame?
APPENDIX J
STIMULI FOR EXPERIMENT 10

Imagine that you go to a restaurant with a few friends. The restaurant employs a lot of waiters and you are being served by one of them. S/he takes your order but seems abrupt and not very friendly. [Sometime later you visit the restaurant again. It turns out that you are served by the same/a different waiter]. For each set of 100 customers, what percentage do you feel would have the same experience with this waiter/in this restaurant?
LIST OF REFERENCES


BIOGRAPHICAL SKETCH

Wouter Vanhouche graduated with a Masters of Science degree in psychology from the University of Leuven in Belgium in 1995. He held several positions as a research assistant in his hometown university before he decided to take his academic career one step further and pursue a doctorate in the United States in 2001. Four more years of rigorous training at the Marketing Department of the University of Florida culminated in the defense of his dissertation in the summer of 2005.