Tradeoffs in the Dark: The Effect of Experience on Extrapolated Consumer Preferences

Yanliu Huang
The Wharton School
University of Pennsylvania

Robert J. Meyer
The Wharton School
University of Pennsylvania

Working Paper # 2009-05-07

Risk Management and Decision Processes Center
The Wharton School, University of Pennsylvania
3730 Walnut Street, Jon Huntsman Hall, Suite 500
Philadelphia, PA, 19104
USA
Phone: 215-898-5688
Fax: 215-573-2130
http://opim.wharton.upenn.edu/risk/

THE WHARTON RISK MANAGEMENT AND DECISION PROCESSES CENTER

Established in 1984, the Wharton Risk Management and Decision Processes Center develops and promotes effective corporate and public policies for low-probability events with potentially catastrophic consequences through the integration of risk assessment, and risk perception with risk management strategies. Natural disasters, technological hazards, and national and international security issues (e.g., terrorism risk insurance markets, protection of critical infrastructure, global security) are among the extreme events that are the focus of the Center's research.

The Risk Center's neutrality allows it to undertake large-scale projects in conjunction with other researchers and organizations in the public and private sectors. Building on the disciplines of economics, decision sciences, finance, insurance, marketing and psychology, the Center supports and undertakes field and experimental studies of risk and uncertainty to better understand how individuals and organizations make choices under conditions of risk and uncertainty. Risk Center research also investigates the effectiveness of strategies such as risk communication, information sharing, incentive systems, insurance, regulation and public-private collaborations at a national and international scale. From these findings, the Wharton Risk Center's research team – over 50 faculty, fellows and doctoral students – is able to design new approaches to enable individuals and organizations to make better decisions regarding risk under various regulatory and market conditions.

The Center is also concerned with training leading decision makers. It actively engages multiple viewpoints, including top-level representatives from industry, government, international organizations, interest groups and academics through its research and policy publications, and through sponsored seminars, roundtables and forums.

More information is available at http://opim.wharton.upenn.edu/risk.

Tradeoffs in the Dark: The Effect of Experience on Extrapolated Consumer Preferences

YANLIU HUANG
ROBERT J. MEYER*

*Yanliu Huang (email: yanliuh@wharton.upenn.edu) is a doctoral candidate in marketing and Robert Meyer (email: meyerr@wharton.upenn.edu) is the Gayfryd Steinberg Professor of Marketing in the Wharton School, University of Pennsylvania, Suite 700 Jon M. Huntsman Hall, Philadelphia, PA 19104-6371. The authors thank the Wharton Behavioral Laboratory for support in data collection and the research grant from the Russell Ackoff Doctoral Student Fellowship program of the Wharton Risk Management and Decision Processes Center.

ABSTRACT

The dynamics of how consumers use preference knowledge gained over one range of product-attribute levels to predict their preferences for new, un-experienced, attribute combinations and levels is explored. Data from six studies supports a hypothesis that as judgmental experience in a core domain grows the process of predictive inference evolves from a reliance on meta-cognitive "guessing" rules to exemplar-based rules and finally to functional prediction rules. One of the consequences is that consumers will often be prone to under-estimating the utility they will draw from novel product combinations that are objectively superior to familiar options and over-estimating the utility they will draw from inferior options. The boundary conditions for this effect are explored in terms of the depth of prior judgment experience in a category and the form of the "true" affective response policy.

Keywords – Preference extrapolation, judgmental experience, exemplar-based rule, meta-cognitive rule.

Many choices require consumers to extrapolate preferences formed in one domain of experience to an unfamiliar new one. Examples include decisions to adopt new goods or services (Hoeffler 2003; Zhao, Hoeffler, and Dahl 2009), judgments of the potential attractiveness of novel product-attribute combinations (Gregan-Paxton and John 1997), or evaluations of extension product from the parent brand (Boush and Loken 1991; Broniarczyk and Alba 1994). While there has been considerable work investigating such related problems as inference formation and behavior in prediction tasks (e.g., DeLosh, Busemeyer, and McDaniel 1997; Juslin, Olsson, and Olsson 2003; Meyer 1987; Shanks and Darby 1998; Shirai and Meyer 1997), less is known about either the process individuals use to make trade-offs among unfamiliar ranges of attributes or the algebraic structure of such trade-offs.

Understanding errors or biases people are generally prone to making when predicting their preferences for new attribute level combinations, the processes underlying these biases, and how previous relevant judgmental experience will influence evolvement of these processes has important implications for marketers. In particular, it will help firms identify, based on consumers' product experience, the appropriate forms or amount of information they want to reveal about their new products in order to elicit desirable transfer processes and responses. Furthermore, this research can benefit preference measurement researchers by providing insights regarding how different types of new product preference measurements (e.g., measuring attractiveness vs. unattractiveness of the products) will lead to different prediction biases; and in what circumstances, these biases can be either corrected or exacerbated given consumers' previous relevant experience and the form of the underlying product utility generating rule.

The goal of this work is to investigate dynamics of how consumers use preference knowledge gained over one range of product-attribute levels to predict their preferences for new and un-experienced attribute combinations and levels. Prior work has suggested that given the limited direct experience in consuming products, predictive inferences may involve a blend of two processes: exemplar-based and rule-based. Exemplar-based policies are those that forecast utilities for new stimuli based on the experienced utility of similar options (e.g., Brooks 1978; Juslin, Olsson, and Olsson 2003; Kruschke 1992; Meyer 1987), and rule-based policies are those that exploit generalized beliefs about the functional relationship between attribute values and utility such as monotonicity and additivity (Anderson 1990; Ashby and Gott 1988; Delosh, Busemyer, and McDaniel 1997). In this project, we focus on two central research questions: 1) how does the use of these two processes of hedonic prediction (exemplar-based or rule-based) evolve as consumers develop judgmental experience in a product category; and (2) what does the answer to (1) imply about the both the likely accuracy of consumers' hedonic forecasts for unfamiliar product-attribute combinations?

INFERENCE STRATEGIES FOR ESTIMATING PREFERENCES FOR NEW PRODUCT ATTRIBUTE COMBINATIONS

Studies on unfamiliar product attribute combinations can be dated back to several decades ago when conjoint analyses in consumer research emerged (Green and Srinivasan 1978). However, this research stream focuses on how to construct a set of unfamiliar stimuli in the conjoint experiment design without looking at how consumers react to these novel combinations. (Green and Srinivasan 1978; Steckel, DeSarbo, and Mahajan 1991).

In terms of the strategies adopted by consumers to infer their preferences for new product attribute combinations, consider a consumer who is moving from New York City to a small town in the Midwest and is faced with the problem of choosing an apartment to rent. What makes the task difficult is that almost all apartments take on attribute values that are unlike those with which she had experience in New York: rents are cheaper, sizes are larger, views (perhaps) are better, and safety seems higher. How will the knowledge she gained evaluating apartments in New York be used to make trade-offs in the Midwest? While we are not aware of prior work that speaks to this question directly, previous research on knowledge effects on judgment (e.g., Meyer 1987; Shanks and Darby 1998; Sujan 1985) suggests that the answer will depend on two considerations:

1. The nature of existing knowledge about apartments in New York (whether it is limited to a set of exemplars or is comprised of generalized beliefs about the functional relationship between attribute variation and utility); and
2. The perceived *sufficiency* this knowledge for evaluating apartments in the Midwest.

Specifically, if the consumer believes that apartments in the Midwest are, at the end of the day, quite similar to those in New York (only larger and cheaper), judgments will presumably be made by directly applying existing knowledge to the new setting. For example, if elevator quality is critical in New York, this same attribute will be looked to when initially judging apartments in the Midwest. In contrast, if the consumer believes that existing knowledge is *insufficient* — either because experience with apartments in New York is overly limited or the two locations are perceived as radically different — the consumer would be assumed to evaluate new apartments using more abstract and higher-level rules that are not specific to apartments — such as beliefs that lower prices are better than higher, and more is better than less (e.g., Sujan 1985).

These two strategies of making inferences about new stimuli's utility are well documented in literature. Specifically, the former approach of applying experience with similar options to predict new stimuli's utility is called exemplar-based heuristics, and the latter approach of forming functional relationship between attribute values and applying these rules to estimate new alternatives' utility is called rule-based policies or functional learning (Delosh, Busemyer, and McDaniel 1997; Gawronski and Bodenhausen 2006; Juslin, Olsson, and Olsson 2003; Kruschke 1992; Thompson, Gentner, and Loewenstein 2000). Rule-based approach is comparable to schema-based transfer process where an abstract knowledge structure is created as a means of transporting knowledge from the base to the target (Gregan-Paxton and John 1997). As a relatively high-order process, the rule-based strategy often leads to better performance in judgment and decision making when compared with the exemplar-based approach (Brooks, Norman, and Allen 1991; Rips 1990; Shanks & Darby 1998; Sloman 1996).

THE EVOLUTION OF INFERENCE STRATEGIES AND ITS IMPLICATIONS

Researchers have tried to identify factors determining which of these two processes is more likely to arise in a given context. For instance, rule-based strategy is more likely to happen when individuals are given plenty of time and pay full attention during the learning (Smith and Kemler Nelson, 1984; Smith and Shapiro 1989). However, previous literature has spoken little regarding the dynamics of how individuals use these two strategies to predict product preferences as they develop relevant experience in one domain. Furthermore, among the existing findings on this topic, no consistent conclusions have been achieved (Meyer 1987; Shanks & Darby 1998; Sujan 1985). For example, Sujan (1985) found that when compared with novices, experts are faster at deciding whether the incoming information belongs to a category they are familiar with and which information processing strategies, category-based or piece-meal, should be used. However, this research did not really contrast exemplar-based approach with rule-based strategies and also did not examine directly the dynamic effect of experiences on skill extrapolation. As another example, in their consumer knowledge transfer model, Gregan-Paxton and John (1997) described that novices rely on similarity-to-exemplar transfer process for all domain comparisons because rule-based transfer is beyond their ability when they have only lower levels of base domain knowledge. In contrast, experts have well-developed structures and are capable of going through more sophisticated rule-based process. However, their learning model did not speak directly to the dynamic evolvement of these two processes as a function of experience development; and they also did not provide empirical evidence to support their claims.

Now back to our example, how *good* will extrapolation judgments be given underlying processes consumers will rely on based on their previous experience? The answer, we suggest, depends on two interrelated factors:

1. The skill with which consumers make judgments about sufficiency; and
2. The sophistication of relevant rule-based knowledge of the drivers of utility.

The role of the first factor is transparent: if consumers are too quick to assume that experience in one domain is relevant to a new one when, it is not, then extrapolation judgments will be poor predictors of the actual experienced utility of an option (the new resident who focuses on elevator quality when they should be focusing on access to convenient parking). The role of the second is more subtle: even if the old environment is relevant for predicting valuations in the new, the quality of extrapolations still depends on the completeness of the consumer's understanding of the functional drivers of utility in the original environment. As an example, if our resident's knowledge of apartments in New York consists of the attractiveness of a few residences, she may feel that it is sufficient to make predictions by extrapolating from these several examples (e.g., using a pattern-matching heuristic); whereas an individual whose knowledge is about only 2 or 3 apartments may perceive that she does not have sufficient examples to rely on pattern-matching heuristic; as a result, she might try to guess the possible rules the target is using to form the preference. Therefore, given that rule-based processes generally lead to better performance than the exemplar-based ones because the latter relies entirely on attribute level overlap (Brooks, Norman, and Allen 1991; Gregan-Paxton and John 1997; Novick 1988; Shanks & Darby 1998; Sloman 1996), when the relationship between product attributes and preference is easy to guess, the decision maker with virtually no or little experience in a judgment category may be able to provide more accurate utility predictions in novel domains than those with moderate amounts of experience.

THE CURRENT RESEARCH

In the sections below we describe the results of six studies designed to explore how consumers make extrapolation judgments given varying levels of experience in an original context, and the quality of these extrapolations. All six studies focus on contexts where experience in the original context *is* relevant for prediction in the new; that is, a common ecological rule determines quality for all alternatives. Our focus is on how knowledge gained in an original context is utilized by consumers to make inferences about the new, and biases that characterize such predictions.

Studies 1A to 1C explore the nature of extrapolation judgments when decision makers have reasonably high levels of experience in making judgments in a given setting and are asked to predict likely utility of options that have systematically superior or inferior attribute values. Here we find that extrapolation skills display a scale-goal congruence bias. Specifically, when participants were asked to predict the likely attractiveness of a novel apartment, they over-estimated the likely value of inferior options, but were accurate in their hedonic forecasts for superior ones. In contrast, when participants were asked to predict the unattractiveness of an apartment, participants under-predict the attractiveness of superior options.

Studies 2 and 3 explored the boundary conditions of these prediction biases based on the amount of prior judgmental experience and the underlying “true” utility-generating rule. As predicted, the forecasting bias was most pronounced among participants with moderate amounts of judgmental experience (Study 2) but this disadvantage of moderate experience vanishes when the underlying composition rule is a complex one unlikely to be naively guessed (Study 3). Finally in Study 4 we analyze concurrent written protocols to provide more direct evidence for how process of evaluation changes as experience in a judgment category grows.

STUDY 1A: EXTRAPOLATION WHEN RELEVANT EXPERIENCE IS HIGH

The purpose of the first study was to obtain an initial look at biases that arise when consumers who are trained in making judgments in one domain and are asked to extrapolate that knowledge to unfamiliar new domains. Prior work on inference (e.g., DeLosh, Busemeyer, and McDaniel 1997; Juslin, Olsson, and Olsson 2003) suggests that the nature of such biases will depend on the process that decision makers use to make inferences. If predictions are primarily made using case-based heuristics – judging the likely value of new options based on their similarity to familiar ones – extrapolations should display an anchoring bias in which predictions are skewed in the direction of the mean valuation of the set of familiar cases (see, e.g., Juslin, Olsson, and Olsson 2003). In contrast, if decision makers make predictions using generalized functional knowledge about the relationship that exists between attributes and the criterion, there would be no such bias as long as a common functional relationship determines value over all attribute values. To explore this issue we adopted an agent-learning approach in which participants are asked to act as agents to learn through feedback a hypothetical target consumer's apartment preference. They are then tested on a new set of stimuli without feedback.

Method

Materials. The study involved an agent-prediction task set in the context of apartments. Participants were asked to play the role of a real-estate agent whose goal was to predict the degree to which a hypothetical client would like a series of one-bedroom apartments described by a level on each of four attributes: appearance, safety, travel time to work or school, and rent. Appearance and safety were described by ratings on a 1-100 scale provided by the agent (11, 24, 37, 50, 63, 76, or 89), travel time was expressed in minutes (5, 10, 15, 20, 25, 30, or 35 minutes), and rent was described by a monthly amount in dollars (500, 700, 900, 1100, 1300, 1500, or 1700 dollars). To avoid heterogeneity in respondents' beliefs about the affordability of apartments, participants were told that while the client would prefer a lower-priced apartment to a more expensive one, he or she could afford to spend up to $1800 a month for rent. In addition, to make the ratings scales for appearance and safety be more meaningful to participants, vivid descriptions and photos were provided to illustrate the different rating values. For example, a rating of 10 for location security was described as "Borders on a high-crime area; safe to walk in during the day, but security systems are recommended and you are advised not to be unaccompanied later at night," and a rating of 90 was described as "No known crime problems; walking is safe at all hours, and few residents have found the need for security systems because of a strong local police presence." In addition, participants were instructed that since these apartments come from different city locations, higher rents did not always mean better apartments or safer neighborhoods.

The client customer's preferences were given by a linear-additive deterministic multi-attribute preference function unknown to the participant. Specifically, the weights were 0.3 for travel time, 0.4 for apartment appearance, 0.1 for security of the location, and 0.2 for rent. Through a linear transformation, preferences for all the combinations were represented by a value between 1 and 90.

Design. As the objective was to explore participants' extrapolations of preferences formed in one domain of experience to an unfamiliar new one given moderately-large bases of experience, we sought a design that would allow us to compare preferences formed in both familiar and unfamiliar domains. In particular, during the training session, participants received apartment combinations of the three middle attribute levels for all the four attributes (e.g., 15, 20, and 25 for the time attribute). But in the testing session, participants were randomly assigned to one of three conditions: a "better" condition where profiles had greater values on all attributes (e.g., 5, 10, and 15 for the time attribute), a "poorer" condition where profiles had poorer attribute values (e.g., 25, 30, and 35 for the time attribute), and a "similar" condition where profiles had similar values (e.g., 15, 20, and 25 for the time attribute again).

Since during both training and testing each attribute had three levels, a fractional factorial design were used in which participants always received 12 combinations to allow orthogonal estimation of all main effects of the four attributes. The order of the presentation of 12 combinations was randomized for both training and testing and was same for all the participants during training.

Participants and Procedure. One hundred and twenty-six students from an East Coast university received $10 monetary compensation for participating in the experiment, which included this study among others.

All materials and instructions were presented on a computer located in individual cubicles. Participants were asked to imagine that they just found a job as an agent in a local famous apartment agency, and they now had their first client for whom they would need to find an appropriate apartment. Learning was achieved in a training session where participants were presented with 12 apartment profiles that were described by the 4 attributes described earlier. After viewing each of these 12 profiles, participants entered a preference rating between 1 and 90, and were then shown the true preference score of the target client. Following the training session, participants made judgments on another 12 new apartments according to their experiment condition without feedback.

In order to encourage participants to take the task as seriously as possible, they were told that the two participants who made the most accurate predictions over 12 new apartments would be given a $50 cash award at the end of the experiment.

Results

Our primary focus was on participants' judgment patterns as revealed in their evaluation of a new set of 12 profiles that were either similar with, better, or poorer than the training set. The key findings are displayed in Figure 1, which plots the true versus estimated marginal means for all the four attributes. The data provide mixed evidence on the process participants seemed to be using to make extrapolations. On one hand, when participants were asked to extrapolate to unfamiliar negative attribute ranges, judgments displayed a positive anchoring bias – a systematic error that would be consistent with the use of a case-based or pattern-matching heuristic to make predictions. On the other hand, when they were asked to extrapolate to more extreme positive values we see no similar bias, suggestive of the use of accurate functional knowledge of the relationship between attributes and values.

Insert Figure 1 about here

To analyze this difference more rigorously, we subjected the absolute deviations between predicted and true apartment preferences to a mixed repeated measures ANOVA, and it was found that when compared with the condition with worse attribute values, this deviation was lower for both the conditions where the attribute values were better (M_{better} = 6.06 vs. M_{poorer} = 9.42; F (1, 123) =7.63, p = .007) and similar ($M_{similar}$ = 6.17 vs. M_{poorer} = 9.42; F (1, 123), p = .009). The estimations in both “better” and “similar” conditions were also more homogeneous than that where attribute values were poorer (SD_{better} = 9.09, $SD_{similar}$ = 10.64, and SD_{poorer} = 12.29). Finally, there was additional evidence that participants’ ability to extrapolate was decidedly asymmetric: the correlation between estimated and true preferences were 0.58 and 0.66 respectively when the new apartments took on attribute values that were similar to or better than those in the training set, respectively, but only 0.33 when the attribute values were worse.

Discussion

The first study showed that extrapolation judgments were marked by an asymmetric anchoring effect, where participants were able to make reasonably accurate predictions about the utility of apartments that had more positive attribute values than a training set, but made predictions that were too optimistic when the attribute values were more negative. What drove this asymmetry? A possible answer might be found in earlier work by Meyer (1987) who argued that such effects could be explained by consumers using pattern-matching rules to make judgments, where consumers have better-formed internal representations a prototypical "good" option would look like as opposed to a "bad" option. What complicates the application of this idea here is that we also observed that judgments of poorer options were also marked by higher variance — an effect that will statistically drive average responses away from the true values to the scale mean. Hence, an alternative explanation for the asymmetry bias is that judgments accrued to a heterogeneous mix of ill-formed beliefs about the shape of the utility function near the low end of the scale and that drove the mean judgmental prediction close to the scale mid-point.

In Study 1B we attempt to tease apart these explanations by priming exemplar-based versus rule-base processing of the training set. If the primary source of the asymmetric learning bias was indeed the tendency for participants to use pattern-matching rules to make judgments, we should see an amplification of the bias under exemplar priming and a deflation under rule-based priming.

STUDY 1B: PRIME EXEMPLAR OR RULE-BASED PROCESSING

Method

One hundred and sixty-four students participated in the experiment in return for $10 monetary compensation. The materials and procedure in Study 1B were the same as in Study 1A except for the following changes. First, before participants started the training session, they were primed to focus on either exemplar-based or rule-based strategy. Specifically, in the "good exemplar priming" condition, people were asked to pay attention to the specific good apartments in the training set because their client really preferred those best apartments. They were also told that after training they would be asked to recall the *best* apartment they have seen. People in the "bad exemplar priming" condition received the similar instructions except that they were asked to focus on those specific bad apartments. Finally, people in the "rule priming" condition were asked to attend to the rule that seemed to be driving the target client's apartment rating and be prepared to state how important each attribute is as a predictor of apartment attractiveness after the training.

After the training session and before the test phase, participants performed the corresponding reinforcing task where people in the "exemplar priming" conditions (both good and bad) were asked to recall and select the exact attribute values of the apartment that was given the highest rating by the target client in the training set, whereas people in the "rule priming" condition were asked to rate on a scale (from 0 to 1) the relevant importance of each attribute. Finally, participants were tested on either a "better-attribute" condition where profiles had greater values on all attributes when compared with the training combinations or a "poorer" condition where profiles had poorer attribute values than the training set.

Results

In order to test our hypothesis that once consumers successfully identify the correct rule to predict new stimuli's utility, the positivity bias as demonstrated in Study 1A would disappear, we first focus on the predictions made by those participants in the "rule-priming" condition who displayed evidence of learning the rule driving preferences in their training set. Basically, participants were declared "rule correct" if they displayed correct weak rank-order knowledge of attribute importance in the judgment task that followed the training rounds. Specifically, if they could identify the two most important attributes from the two least important, then they would be classified as learning the rule correctly. As predicted, for the 23 of 55 participants who satisfied this knowledge criterion, there was no positivity bias, as the deviation between estimated and true values not statistically different between the “bad attribute values” and the “good attribute values” conditions (8.04 and 8.19 respectively, F (1, 21) = 0.04, p= .96). Furthermore, the correlations between estimated and true preferences were 0.51 and 0.46 respectively for the “bad” and “good” attribute conditions. For those 32 people who did not get the rule correctly, positivity bias still existed because the deviation between true and estimated values was much smaller for the “good attribute values” than the “bad attribute values” (7.15 and 11.89 respectively, F (1, 30) = 8.52, p = .007; correlations between true and estimated values were 0.59 and 0.28 respectively).

In contrast, the positivity bias *was* clearly evident among participants in the exemplar-priming conditions. A potentially surprising result, however, was that positivity bias was observed regardless of whether a “best apartment” or “worst apartment” exemplar was primed. Specifically, the mean deviation between true and estimated preferences for the “good exemplar priming” condition was M_{better} = 7.72 vs. M_{poorer} = 11.30; F (1, 105) = 51.66, p < .0001; while for the “bad exemplar priming” condition the respective means were M_{better} = 8.95 vs. M_{poorer} = 10.58; F (1, 105) =11.03, p = .0009).

Discussion

Study 1B offers mixed evidence for the process that was driving extrapolation judgments in Study 1A. On the one hand, supporting the idea that the judgments were driven by exemplar-based pattern-matching rules, the positivity bias observed in Study 1A vanished when participants were explicitly given the goal to learn the functional rule that was used to drive preferences; whereas this bias returned when participants were given the goal to learn exemplars. But an exemplar-based process explanation for Study 1A would also predict that we should see a *reversal* of the positivity bias when the locus of knowledge was shifted to learning what was most prototypical of a “bad” apartment as opposed to a “good”. In fact, what we observed was a positivity bias in both cases.

What might explain this result? One possibility is that participants *were* using exemplar based rules, but the attempt to manipulate the exemplar was mitigated by the fact that the response scale continued to be one where higher numbers meant better apartments. Hence, it is possible that what we were seeing in the first study was not a positivity bias *per se*, but rather a *scale-goal congruence* bias, where participants asymmetrically develop knowledge of what is most prototypical of options that earn the highest marks on whatever the scale is at hand. This proposition is consistent with the “task compatibility” literature where weighting of alternatives increases by their match with task (Nagpal and Krishnamurthy 2008; Shafir 1993). For example, attractive and unattractive aspects of alternatives will be focused on when one is choosing and when one is rejecting, respectively. If such is the case, an instruction that simply asked participants to imagine the worst possible apartment would not reverse the positivity bias, but *reversing the scale* — such that the worst apartments now get the highest scores — would. In study 1C we tested this hypothesis.

STUDY 1C: PREDICT UNATTRACTIVENESS OF OPTIONS

Method

Ninety-five students participated in the experiment in return for a monetary compensation. The design, materials and procedure in Study 1C were the same as in Study 1B except that this time participants were asked to enter a rating between 1 and 90 for each apartment in terms of how much the target client will *dislike* it rather than how much he will like it. Target client's true rating for each apartment was still generated by same linear additive rule but reflected the "mirror" version of the original preference ratings (e.g., an original preference rating of 10 became 80 this time in terms of how much the apartment would be disliked).

Results

We first plotted the true versus estimated marginal means for all the four attributes as displayed in Figure 2. Consistent with our hypotheses, the previous positivity bias was reversed. Specifically, when participants were asked to extrapolate to unfamiliar positive attribute ranges, judgments displayed a big negative anchoring bias; and when they were asked to extrapolate to more extreme negative values, much smaller such bias was demonstrated, especially when participants were asked to pay attention to those bad exemplars.

Insert Figure 2 about here

To more rigorously analyze these data we subjected the absolute deviations between predicted and true apartment preferences to a mixed repeated measures ANOVA, and for both the "bad exemplar priming" and "rule priming" conditions, this deviation was lower for worse attribute values than for better attribute values, although only the difference in the former condition was significant ("bad exemplar priming", M_{better} = 16.81 vs. M_{poorer} = 10.08; F (1, 89) = 39.54, p <.0001; "rule priming", M_{better} = 16.12 vs. M_{poorer} = 15.03; F (1, 89) =1.09, p =.29). For the "good exemplar priming" condition, this deviation was comparable between good and bad attribute values (M_{better} (1, 89) = 16.23 vs. M_{poorer} = 17.39; F (1, 89) =1.06, p =.30). Finally, the correlation between estimated and true dislike estimations were 0.29 when the new apartments took on worse attribute values than those in the training set, but only 0.14 when the attribute values were better.

Discussion

Taken together, Studies 1B and 1C offer suggestive evidence that when developing moderate amounts of knowledge in one domain individuals tend to predict the attractiveness of options in a new domain by referring to exemplar rather than rule-based knowledge acquired in the original domain. The primary source of this evidence was the finding that predictions displayed a "scale-goal congruence bias" where participants were much better able to predict the value of options that were more prototypical of an option that would earn a superior score on a scale than an inferior one — but this effect vanished when participants were explicitly instructed to focus on learning a rule, and were successful at doing so.

One interesting side finding that arises when comparing Studies 1B and 1C is that they suggest that participants seemed to have a better innate ability to develop knowledge of what is prototypical of "likes" as opposed to "dislikes". Specifically, asking participants to imagine a "worst" apartment did little to mitigate the positivity bias observed in Study 1B when the scale was such that better apartments earned higher scores. But asking participants to imagine a "best" apartment *did* mitigate the negativity bias in Study 1C when the scale was such that *worse* apartments earned the higher scores.

A caveat that comes in attempting to generalize these results is that Studies 1A-1C all focused on cases where participants had a moderately large amount of experience (on training instances) on which to base judgments. In natural settings, of course, there will be times when experiential knowledge will be more limited, to the point where a decision maker would feel that it would form an insufficient basis for making predictions about unfamiliar stimuli. A wine novice whose only exposure to wines was three bottles of Australian Cabernets would presumably feel that the knowledge gained from tasting these cannot be extrapolated to predicting, say, the quality of a South African Chardonnay. But if he or she had to make the prediction, how would it be done? As noted at the outset, the general presumption based on prior work is that he or she would turn to generalized meta-rules rather than exemplar-based policies to make the prediction (e.g., “white wines tend to be tasty”). Most critically, if these naïve guessing rules have ecological validity (i.e., white wines do tend to be robustly tasty), then the naïve judge may survive quite well despite the lack of experiential knowledge.

But here is the rub: how would the naïve judge know whether he or she is better off extrapolating from personal experience or using a naïve guessing rule? Most often will be unable to, and therein lies a potential paradox in predictive ability: if an individual is overconfident in the generalizability of his or her experiences or *under*confident in the ecological validity of guessing rules, it would be easy to construct scenarios where less knowledge is better than more. That is, naïve judges who have almost no experience — hence are *forced* to use guessing rules — may be better skilled at making predictions to new environments than those with moderately small amounts, and turn to exemplar-based extrapolation policies prematurely. We investigate this possibility in Study 2.

STUDY 2: EXTRAPOLATION WHEN RELEVANT EXPERIENCE IS LOW

Method

One hundred and fifty-six students received $10 monetary compensation for participating in the experiment. The materials and procedure in Study 2 were the same as in Studies 1A with the following modifications. First, the number of training cases was manipulated and participants learned either 3 or 6 apartment combinations prior to getting tested. In order to capture the whole attribute surface, these apartment combinations had attributes that were from either good, bad, or middle attribute levels. Using travel time as one example, people who received 3 "good" apartment combinations had travel time attribute levels of 5, 10, or 15 for all the 3 options; people in the "middle" condition encountered 3 combinations with the travel time values of either 15, 20, or 25; and the "bad" condition people saw the time attribute values of either 25, 30, or 35. The selection for both 3 and 6 training cases was random and at the same time satisfied the criterion of being non-orthogonal. The second change made in Study 2 was that participants were then tested on the whole attribute surface after training.

Results

In order to test the effect of the number of training cases on the accuracy of the judgment, we subjected the absolute value of the difference between predicted and true preferences to a mixed repeated ANOVA. Consistent our hypothesis we observed a somewhat paradoxical relationship between experience and forecasting skill: participants performed better in the 3-training case condition than in the 6-training case condition as measured by the deviation between true and predicted preferences (M = 10.22 vs. 11.90; F (1, 149) =2.80, p = .08).

To provide insights into the process that might have generated this result, in the left panel of Figure 3 we plot prediction accuracy as a function of the number of training cases and the prediction task — whether it was to predict better or worse attribute combinations. In the 6-

training case condition we replicated the key result of Study 1A that participants were much better at predicting the likely value of better attribute values than worse. However, we see no such bias when there were only 3 training instances. A plausible explanation is that when participants recognized that their experience in the 3-training-case condition was insufficient to make informed predictions, they turned to the use of the meta-rule of “more is always better” that, in this case, provided quite accurate predictions. This interpretation of the data is further supported by the right side of the Figure 3, which shows that in addition to facilitating more accurate judgments, fewer training cases was also associated with more homogeneous predictions ($SD_{3\text{-}case}$ = 14.95 vs. $SD_{6\text{-}ase}$ = 17.44).

Insert Figure 3 about here

Combined Data from both Studies 1 and 2. In order to obtain a more comprehensive view of the effect of the amount of relevant experience on preference extrapolation, we took the data from Study 2 where participants received either 3 or 6 training combinations of middle attribute levels and combined them with the data from Study 1A where people learned 12 apartments with the similar attribute values. We examined the prediction accuracy (i.e., the absolute deviation between predicted and true values) as a function of the number of training exemplars. As shown on Figure 4, here we see a U shape relationship between prediction accuracy and the number of training exemplars (F (2, 279) = 4.65, p = .01). The accuracy difference was significant between the 6-case and 12-case conditions (F (1, 279) = 5.49, p = .02) but not significant between the 3- and 12-case conditions (F (1, 279) = 1.16, p = .28).

Insert Figure 4 about here

However, although we have demonstrated that in some situations less learning experience may be better as they encourage the rule-based learning, there should be circumstances under which the rule-based strategy can harm the performance when people fail to identity the correct rule with so few examples. For instance, what will happen when the true multi-attribute rule is a much more complex one? We explore this question in Study 3.

STUDY 3: EXTRAPOLATION WHEN UTILITY RULE IS COMPLEX

The explanation that we offered for the U-shaped relationship between training experience and prediction skill that was observed in Study 2 was a simple one: there will be times when simple guessing rules will outperform more informed — but imperfect — thoughtful rules. Hence, one would *not* expect to see such an effect in settings where the true rule that drives valuation is more complex, and not likely to be guessed by a naïve judge. Under such circumstances we would expect to observe the more intuitive result that more experience is strictly better than less. To test this we conducted a variation of Study 2 in which true apartment value was driven by a more complex (quadratic) response function.

Method

The participants were 176 students from an East Coast university. The materials and procedure were similar to the previous experiments with the following changes. First, instead of the simple linear additive rule as used before, a curvilinear or quadratic rule was adopted in the current study where the apartment associated with the middle attribute values (i.e., 20 minutes for the travel time attribute, a rating of 50 for both appearance and location attributes, and $1100 for the rent attribute) was set to be the best option for the target client, and the more extreme attribute values were associated with the lower level of preferences. The second change involved the number of the training and testing combinations. To be particular, participants learned either 3 or 6 or 12 examples before testing, and we expanded the number of test apartments to 28 attribute combinations.

Results

We first subjected the absolute value of the difference between predicted and true preferences to a mixed repeated ANOVA to explore the effect of the number of training exemplars on judgment accuracy. As can be seen from the left part of Figure 5, the estimation error was significantly smaller for the 12-training instance condition than for the 3-instance and the 6-instance conditions (21.60 vs. 27.22, F (1, 174) = 73.12, p<.0001 and 21.60 vs. 27.34, F (1, 174) = 77, p<.0001 respectively), suggesting that, given enough experience, participants could learn to extrapolate even for a complex data-generate rule. In addition, consistent with our expectations, we no longer see a superiority in the predictions of participants with negligible (3) versus moderate (6) levels of judgmental experience. In this case, neither naïve guess rules nor imperfect informed rules provide a useful means of predicting the true value of novel apartments

----

Insert Figure 5 about here

STUDY 4: PROCESS EVIDENCE FOR STRATEGY EVOLUTION

Implicit to our discussion of the above findings was a suggestion that as experience in judgment grows the process of evaluation evolves from the use of naïve meta-rules given little or no experience in a category, to imperfect case-based or pattern-matching rules given moderate experience, to informed functional rules given more extensive experience. The evidence for this, however, was highly indirect. To provide more direct evidence of the processes that were driving evaluations we conducted a forth study that analyzed the content of concurrent written protocols provided by decision makers engaged in the prediction task.

Method

The participants were 85 students who received $10 monetary compensation for participating in the experiment. This experiment differed from Study 3 on the following aspects. First, in order to maximize the number of participants in each condition and at the same time assure the completeness of the data, there were a total of 4 experiment conditions including the 3 conditions in Study 3 (i.e., complex quadratic rule, either 3 or 6 or 12 training instances) and 1 condition from Study 2 (the simple linear additive rule, 3 training instances). These four conditions were referred to as “3-quadratic”, “6-quadratic,”, “12-quardatic”, and “3-simple” respectively in the result session. Second, following the training session, right after making the first new judgment without feedback, all participants were asked to recall how they made this previous prediction. Particularly, they were instructed to use the computer keyboard to type everything that went through their mind when they made the prediction and ignore spelling, grammar, and punctuation. They were asked to be as specific as possible and try to avoid listing simple statements like "just guessing". Among these 85 participants, 39 of them performed the same task again after they were tested on the sixth combination, and the remaining 46 participants did this after the twelfth test prediction. These three positions where the protocols were collected were referred to as “Testing 1”, “Testing 6”, and “Testing 12” respectively in the result session.

Coding

In order to assess the extent to which the protocols were indicative of rule-based versus exemplar-based approaches to judgment, two lists of words were first identified as characterizing either process. In particular, the 9 words those are typically present in the rule-oriented statements include “seems”, “seemed”, “concerned”, “care”, “cares”, “appears”, “affect”, “values”, and “matter”; and the 9 words those usually appear in the exemplar-based statements are “comparable”, “comparison”, “same”, “better”, “higher”, “lower”, “equal”, “similar”, and “remember”. One typical example of rule-oriented verbal report that included several of these words was “I figured that from the training, Mr. Smith *values* travel time less than he does appearance and safety.” In contrast, exemplar-based processes would be suggested by statements such as, “based off of previous *comparable*s...the rent was *lower* than what he wants and the other factors are in his range,” and “*Similar* to one apartment in trials.”

Results

In Figure 6 we plot the relative frequency of words suggestive of rule-based versus exemplar-based strategies by the number of training instances and the number of prediction instances. Because relative frequencies were similar for the 1- and 6-predictioon cases, they are pooled in the graphical displays. The data provide consistent support for the hypothesized evolution of judgment strategies. Specifically, given extremely limited experience participants appeared to recognize the insufficiency of their training as a basis for judgments, and made predictions using rule-based strategies that presumably reflected their naïve guesses about how the attributes would combine to yield apartment quality. Given moderate (6 cases) level of experience, however, we now see an increased use of exemplar-based strategies, as participants attempt to make use of what they have learned from the early judgments to make predictions. Finally, given more extensive levels of experience, we again see a rise in rule-based strategies, though here these rules presumably reflect the associations that have actually been learned through repeated judgments.

. ---

Insert Figure 6 about here

GENERAL DISCUSSION

The goal of this research was to investigate dynamics of how consumers use preference knowledge gained over one range of product-attribute levels to predict their preferences for new attribute combinations. Across six studies we find evidence that these judgments are made using an evolving mix of meta-cognitive and exemplar-based rules, with the former being invoked when the set of referents is too sparse to allow reliable inferences, and, at the other extreme, when knowledge in a domain is more extensive. Specifically, given limited experience, predictions will rely on meta-cognitive "guessing" rules that make no attempt to utilize what has been observed about product values in a core domain. As experience grows, however, knowledge gained in the core domain will be increasingly utilized, first by using exemplar-based prediction rules that draw analogies between novel products and previously-viewed ones, then by using functional prediction rules that exploited learned continuous associations between attribute levels and valuations. One of the consequences is that consumers are prone to under-estimating the utility they will draw from novel product combinations that are objectively superior to familiar options and over-estimating the utility they will draw from inferior options. The magnitude of these errors, however, is in some cases U-shaped, where decision makers with virtually no experience at all in a judgment category may be able to provide more accurate utility predictions in novel domains than those with moderate amounts of experience.

The data also suggest that there are no simple “switch points” between different types of extrapolation policies, and, in fact, as knowledge grows judgments involve a blend of both. To illustrate, in Study 1A we argued that the scale-goal congruence bias observed in participants’ predictions was consistent with their using exemplar-based policies to make extrapolative judgments. The “smoking gun” for this explanation was the findings in Studies 1B and 1C that this bias occurred when participants were asked to develop knowledge of the best or worst apartments, but vanished when they were primed with the goal of learning a generalized rule. But this explanation would seem to conflict with the protocol data reported in Study 4 where participants revealed a dominant use of rule-based polices when given 12 training instances, the same as those used in Study 1. A natural explanation is that extrapolative judgments in Study 1 involved a *blend* of both processes: exemplar-based rules were used to gain a first approximation to the prediction through categorization (which would explain the asymmetry bias), and then rule knowledge was invoked to fine tune the prediction.

A natural objective of future research would be to further explore the boundaries of the results reported here. For example, there is some evidence from prior research that as expertise in a product category becomes extensive (as opposed to the moderate levels studied here) cognitive efficiency increasingly again favors the use of exemplar-based judgments strategies over rule-based strategies (e.g., Meyer 1987). If this is indeed the case, we might observe that extrapolation abilities are poorer among those with the most extensive expertise in a category. Hence, the "best" level of knowledge to have if one's goal is extrapolation could lie in intermediate ranges — large enough to allow the detection of functional associations, but small enough to dissuade judgment by complex taxonomies.

REFERENCES

Anderson, John R. (1990), *The Adaptive Character of Thought*. Hillsdale, NJ: Erlbaum.

Ashby, F. Gregory and Ralph E. Gott (1988), “Decision Rules in the Perception and Categorization of Multidimensional Stimuli,” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14 (1), 33-53.

Boush, David and Barbara Loken (1991), “A Process Tracing Study of Brand Extension Evaluation,” *Journal of Marketing Research*, 28 (February) , 16–28.

Broniarczyk, Susan and Joseph Alba (1994a) , “The Impor tance of the Brand in Brand Extension,” *Journal of Marketing Research*, 31 (May) , 214–228.

Brooks, Lee R. (1978), “Non-analytic concept formation and memory for instances,” In: E. Rosch and B. Lloyd, Ed. *Cognition and categorization*, Erlbaum, Hillsdale (1978), 169–215.

Brooks, Lee R., Geoffrey R. Norman, and Scott W. Allen (1991), “Role of Specific Similarity in a Medical Diagnostic Task,” *Journal of Experimental Psychology: General*, 120 (3), 278-87.

DeLosh, Edward L., Jerome R. Busemeyer, and Mark A. McDaniel (1997), “Extrapolation: The Sine Qua Non for Abstraction in Function Learning,” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23 (4), 968-86.

Gawronski, Bertram and Galen V. Bodenhausen (2006), “Associative and Propositional Processes in Evaluation: An Integrative Review of Implicit and Explicit Attitude Change,” *Psychological Bulletin*, 132, 692-731.

Green, Paul E. and V. Srinivasan (1978), “Conjoint Analysis in Consumer Research: Issues and Outlook,” *Journal of Consumer Research*, 5 (September), 103-23.

Gregan-Paxton, Jennifer and Deborah Roedder John (1997), “Consumer Learning by Analogy: A Model of Internal Knowledge Transfer,” *Journal of Consumer Research*, 24 (December), 266–84.

Hoeffler, Steve (2003), “Measuring Preferences for Really New Products,” *Journal of Marketing Research*, 40 (November), 406–420.

Juslin, Peter, Henrik Olsson, and Anna-Carin Olsson (2003), “Exemplar Effects in Categorization and Multiple-Cue Judgment,” *Journal of Experimental Psychology: General*, 132 (1), 133-56.

Kruschke, John K. (1992), “ALCOVE: An Exemplar-Based Conectionist Model of Category Learning,” *Psychological Review*, 99 (1), 22-44.

Meyer, Robert J. (1987), “The Learning of Multiattribute Judgment Policies,” *Journal of Consumer Research*, 14 (September), 155-73.

Nagpal, Anish and Parthasarathy Krishnamurthy (2008), “Attribute Conflict in Consumer Decision Making: The Role of Task Compatibility,” *Journal of Consumer Research*, 34 (February), 696-705.

Novick, Laura R. (1988), “Analogical transfer, problem similarity, and expertise,” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(July), 510-20.

Rips, Lance J. (1990), “Reasoning,” *Annual Review of Psychology*, 41, 321-53.

Shafir, Eldar (1993), “Choosing Versus Rejecting: Why Some Options are Both Better and Worse than Others,” *Memory and Cognition*, 21 (4), 546-556.

Shanks, David R. and Richard J. Darby (1998), “Feature- and Rule-Based Generalization in Human Associative Learning,” *Journal of Experimental Psychology: Animal Behavior Processes*, 24 (4), 405-15.

Shirai, Miyuri and Robert J. Meyer (1997), “Learning and the Cognitive Algebra of Price Expectations,” *Journal of Consumer Psychology*, 6 (4), 365-88.

Sloman, Steven A. (1996), “The Empirical Case for Two Systems of Reasoning,” *Psychological Bulletin*, 119 (1), 3-22.

Smith, J. David, and Kemler Nelson, D. G. (1984), “Overall similarity in adults' classification: The child in all of us,” *Journal of Experimental Psychology: General*, 113, 137-159.

Smith, J. David and Shapiro, June H. (1989), “The occurrence of holistic categorization,” *Journal of Memory and Language*, 28, 386-399.

Steckel , Joel H., Wayne S. DeSarbo and Vijay Mahajan (1991), “On the creation of acceptable conjoint analysis experimental designs, ” *Decision sciences*, 22(2), 435-442.

Sujan, Mita (1985), “Consumer Knowledge: Effects on Evaluation Strategies Mediating Consumer Judgments,” *Journal of Consumer Research*, 12(June), 31-46.

Thompson, Leigh, Dedre Gentner, and Jeffrey Loewenstein (2000), “Avoiding missed opportunities in managerial life: Analogical training more powerful than individual case training,” *Organizational Behavior and Human Decision Processes*, 82 (May), 60-75.

Zhao, Min, Steve Hoeffler, and Darren Dahl (2009), “The Role of Imagination-Focused Visualization on New Product Evaluation,” *Journal of Marketing Research*, 46 (February), 46–55.

FIGURE 1

ESTIMATED VS. TRUE MARGINAL MEANS FOR TRAVEL TIME, APPEARANCE, LOCATION, AND RENT AT TESTING (STUDY 1A)

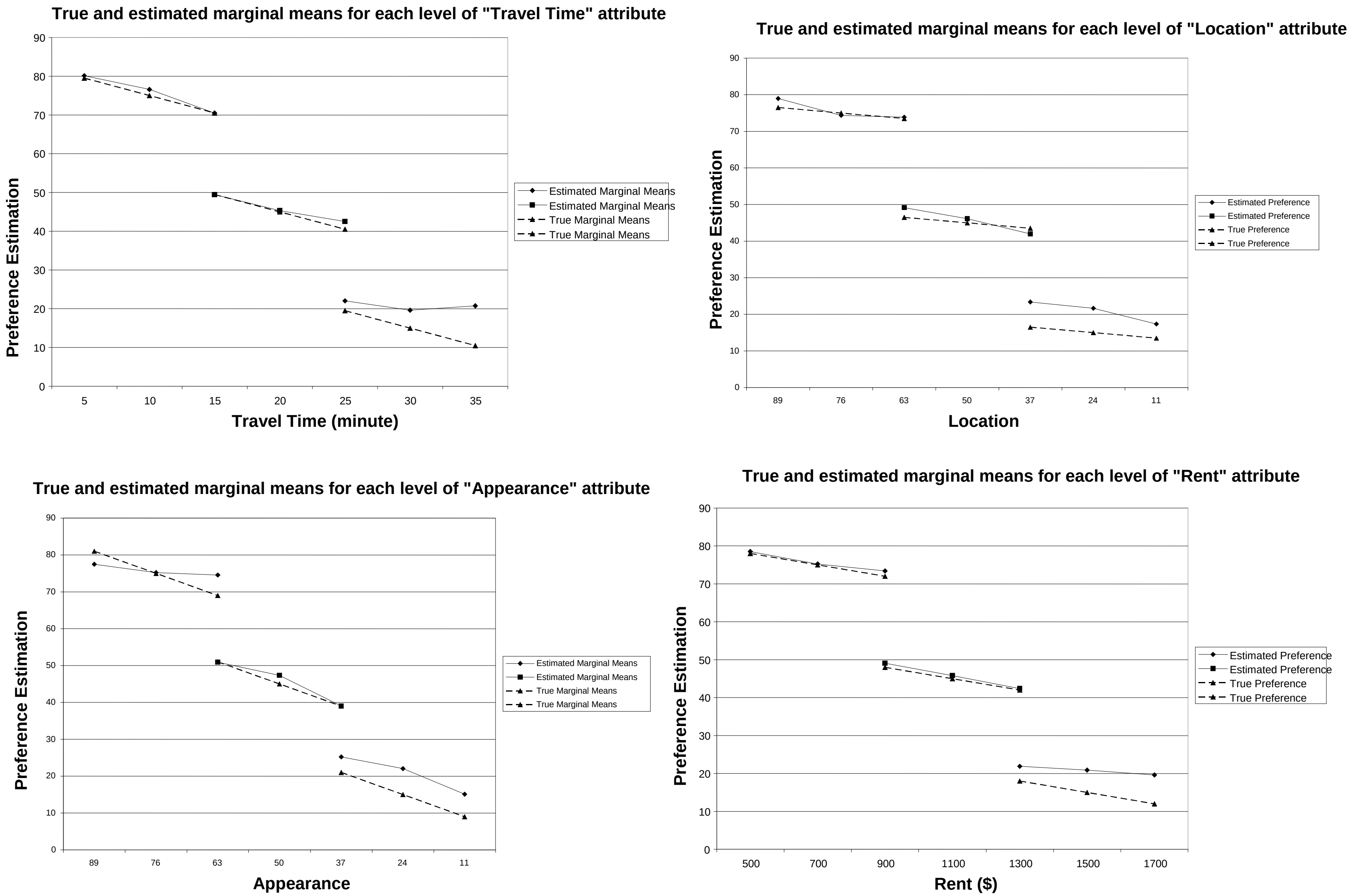


FIGURE 2

ESTIMATED VS. TRUE MARGINAL MEANS FOR TRAVEL TIME, APPEARANCE, LOCATION, AND RENT AT TESTING WHEN PEOPLE JUDGEED HOW MUCH EACH APARTMENT WOULD BE DISLIKED (STUDY 1C)

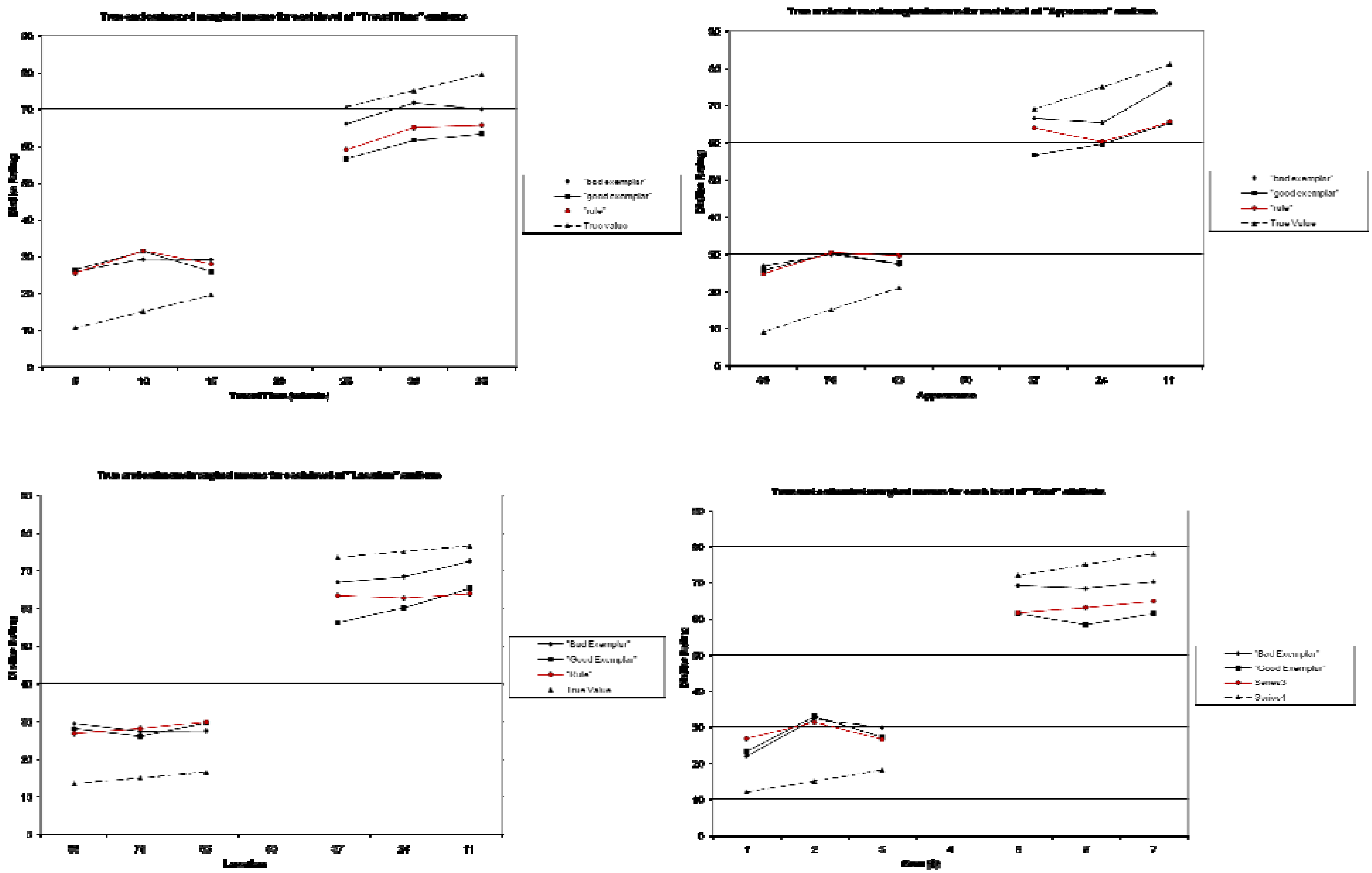
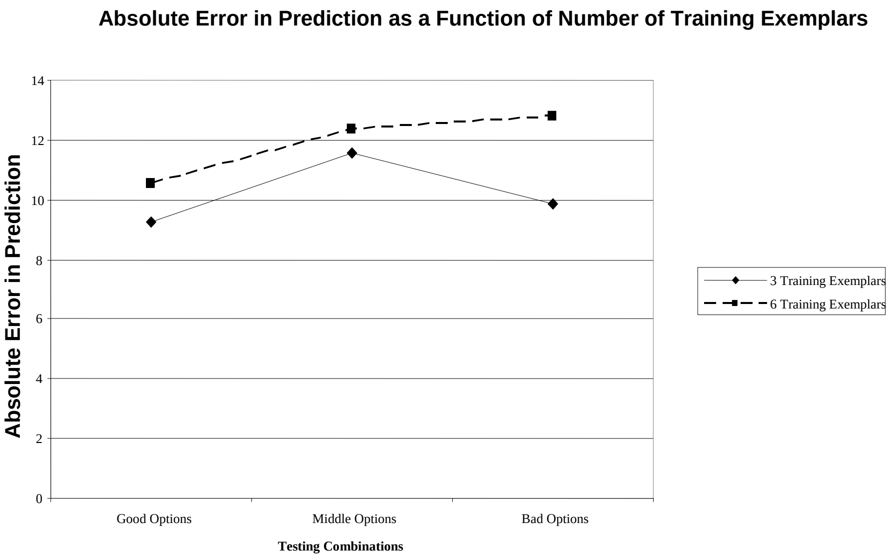


FIGURE 3
JUDGMENT ACCURACY AND HOMOGENIETY AS A FUNCTION OF THE NUMBER OF TRAINING CASES (STUDY 2)



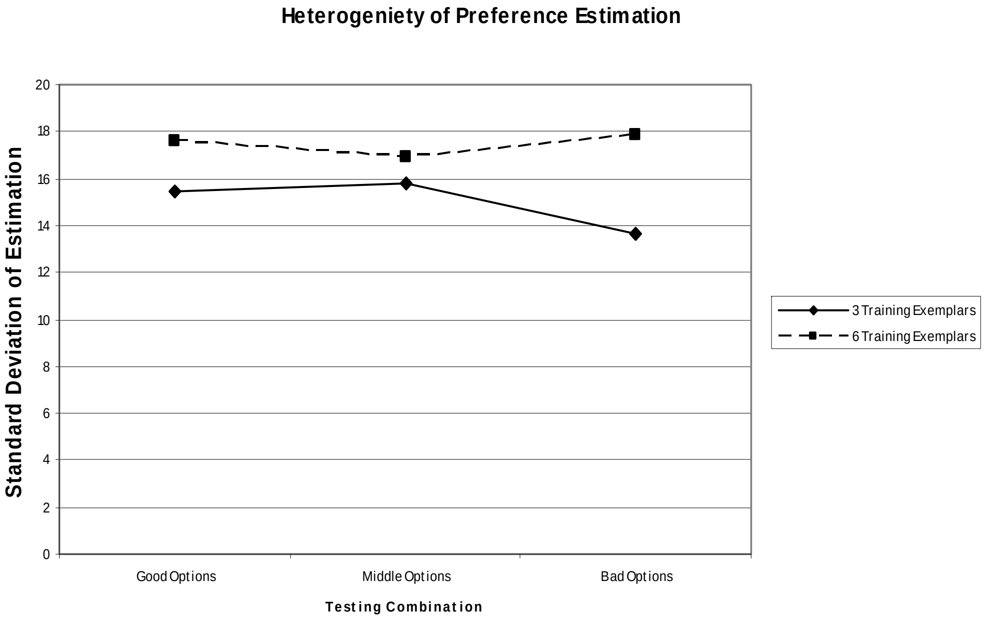


FIGURE 4

THE DEVIATION BETWEEN TRUE AND ESTIMATED PREFERENCES AS A FUNCTION OF THE NUMBER OF TRAINING CASES (STUDIES 1A AND 2)

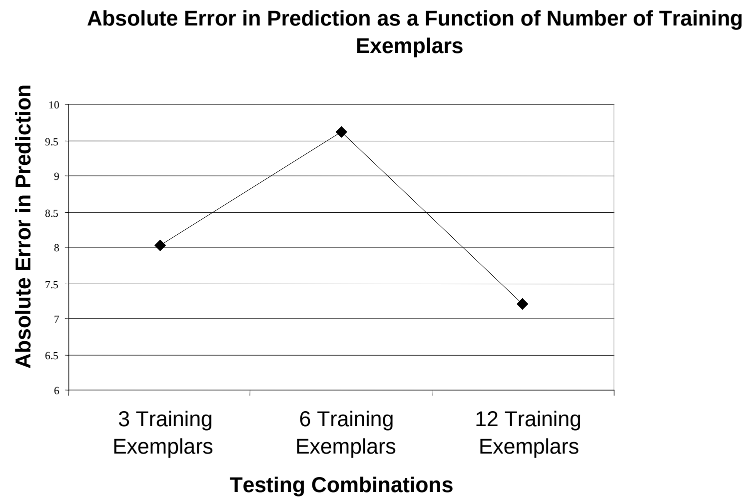


FIGURE 5
JUDGMENT ACCURACY AS A FUNCTION OF THE NUMBER OF TRAINING CASES (STUDY 3)

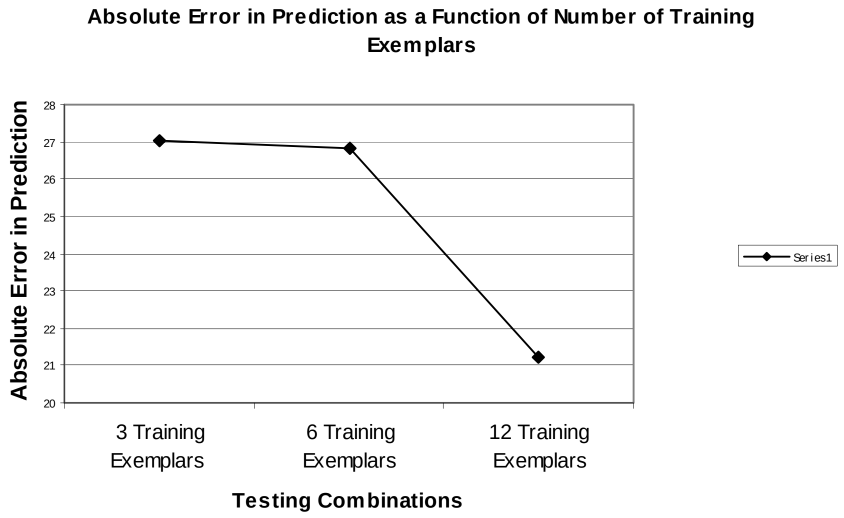


FIGURE 6

NUMBER OF WORDS PER PERSON CHARACTERIZING EACH EXTROPOLATION STRATEGY AS A FUNCTION OF THE NUMBER OF TRAINING CASES (STUDY 4)