Recommendations, Credits and Discounts:
Essays in Behavioral Decision Making

by

Boris Maciejovsky

Mag. rer. nat. Psychology, University of Vienna, 1998
Mag. rer. soc. oec. International Business Administration, University of Vienna, 1998
Dr. rer. nat. Psychology, University of Vienna, 2000

Submitted to the Alfred P. Sloan School of Management
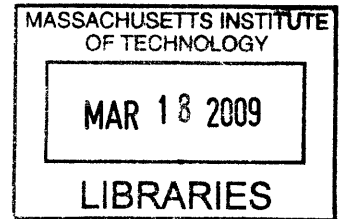in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Management

at the

Massachusetts Institute of Technology

February 2009

© 2009 Massachusetts Institute of Technology

Signature of Author
Alfred P. Sloan School of Management
November 7, 2008

Certified by .
Dan Ariely
James B. Duke Professor of Behavioral Economics
Thesis Supervisor

Accepted by .
Ezra W. Zuckerman
Nanyang Technology University Associate Professor of
Strategic Management and Economic Sociology
Chair, Doctoral Program

Recommendations, Credits and Discounts:
Essays in Behavioral Decision Making

by

Boris Maciejovsky

Submitted to the Alfred P. Sloan School of Management
on November 7, 2008 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy in
Management

ABSTRACT

Essay 1: Translation Errors in the Aggregation of Consumer Recommendations

There has been a substantial increase of websites providing consumers with recommendations about products and services. These recommendations are usually presented in the form of verbal reviews and numerical ratings. It is assumed implicitly that consumers can integrate adequately the information across the two presentation modes (verbal and numerical). However, research on the effects of compatibility between stimulus and response formats suggests that preference consistency is higher (lower) in cases of compatible (non-compatible) formats, implying that information aggregation across the two modes may be sub-optimal. The results of three experiments confirm this conjecture. Information aggregation and preference reversals were systematically affected by the compatibility of the stimulus and response format. Decision makers were not aware of this effect.

Essay 2: The Researcher as a Consumer of Scientific Publications: How Do Name Ordering Conventions Affect Inferences About Contribution Credits?

When researchers from different fields with different norms collaborate, the question arises how name ordering conventions are chosen, and how they affect contribution credits. In this paper we answer these questions by studying two disciplines that exemplify the two cornerstones of name ordering conventions: Lexicographical ordering (i.e., alphabetical ordering, endorsed in economics) and non-lexicographical ordering (i.e., ordering according to individual contributions, endorsed in psychology). Inferences about credits are unambiguous in the latter arrangement, but imperfect in the former, because alphabetical listing can reflect ordering according to individual contributions by chance.

We contrast the fields of economics and psychology with marketing, a discipline heavily influenced by both. Based on archival data, consisting of more than 38,000 journal articles, we show that the three fields have different ordering practices. In two empirical studies, with 351 faculty and graduate student participants from all three disciplines, as well as in a computer simulation, we show that ordering practices systematically affect and shape the allocation of perceived contributions and credit. While strong disciplinary norms in economics and psychology govern the allocation of contribution credits, a more heterogeneous picture emerges for marketing. This lack of strong norms has detrimental effects in terms of assigned contribution credits.

Essay 3: Performance-Contingent Discounts and Consumer Choice

Incentives affect individuals' attitudes and behaviors in a myriad of ways. In this paper we explore the effects of performance-contingent discounts on consumer choice. For that purpose we set up an online store for digital cameras. Half the subjects received a fixed rebate; the other half had to "earn" their rebate by learning about the products offered. The more information subjects remembered, as inferred from their answers to a short quiz, the higher their discounts. Our results indicate that subjects, who were offered performance-contingent discounts, found the online store more informative and reputable, were more likely to recommend the store to their friends, and were more likely to buy. The results cannot be attributed to a better performance in the quiz or a more thorough exploration of the products offered.

Thesis Committee:

Dan Ariely (Chair)
James B. Duke Professor of Behavioral Economics
Duke University, The Fuqua School of Business

David V. Budescu
Anne Anastasi Professor of Psychometrics and Quantitative Psychology
Fordham University, Department of Psychology

Birger Wernerfelt
J. C. Penney Professor of Management
Massachusetts Institute of Technology, Sloan School of Management

Acknowledgments

I would like to sincerely thank all the people who were instrumental in making this dissertation possible. To express my immense gratitude adequately, without forgetting anyone, is a difficult task, so I apologize in advance to everyone, who deserved to be included, but is not.

Most of all, I would like to thank my committee, who have encouraged and challenged me in my research. Dan has not only been a great advisor and mentor, with lots of great research ideas, but has also been a true friend, who supported me from the very first day I arrived at MIT. David and I have been collaborating already before my arrival to the US and it has always been a particular pleasure to learn how much more can be done with data and how to advance projects. I am particularly grateful to David's support and encouragement, both professionally and personally. Birger has been a pleasure, and a great privilege, to work with. I immensely benefited from his fantastic PhD seminars, which provided great insights into the research process.

I am also very grateful to the MIT Marketing family for their continued support throughout the years: Michael Braun, Sharmila Chatterjee, Shane Frederick, John Hauser, John Little, Drazen Prelec, Duncan Simester, Catherine Tucker, Glen Urban, and Juanjuan Zhang.

My friends, fellow students, and post-docs (many fall in more than one category) also supported me in many and important ways: Anat, Chun-Ming, Cristina, Daniel, Daria, Gui, Han, Kristina, Leonard, Matt, Mike, Nathan, Nina, Ray, Rebecca, Rene, Wan-Chen, and Zoe.

Finally, but most importantly, I would like to thank my family for their continued support and encouragement. I am particularly grateful to my wife Geeta Bhatia, who allowed me to make my professional development the priority in our lives for the past four years. I am also indebted to my mother Waltraut Elicker, my stepfather Peter Holyst, my mother-in-law Mohini Bhatia, my brother-in-law Leon Bhatia, his wife Sabine Bhatia-Riedl, my grandma Resi Kozarnik, and my grandpa Josef Kozarnik, who sadly passed away half-way through my program.

I would like to dedicate this dissertation to the memory of my father Eduard Maciejovsky.

Essay 1

Translation Errors in the Aggregation of Consumer Recommendations

ABSTRACT

There has been a substantial increase of websites providing consumers with recommendations about products and services. These recommendations are usually presented in the form of verbal reviews and numerical ratings. It is assumed implicitly that consumers can integrate adequately the information across the two presentation modes (verbal and numerical). However, research on the effects of compatibility between stimulus and response formats suggests that preference consistency is higher (lower) in cases of compatible (non-compatible) formats, implying that information aggregation across the two modes may be sub-optimal. The results of three experiments confirm this conjecture. Information aggregation and preference reversals were systematically affected by the compatibility of the stimulus and response format. Decision makers were not aware of this effect.

An increasing number of consumers purchase many items on the Internet rather than in brick and mortar stores. One important difference between traditional stores and virtual marketplaces is that the latter do not allow consumers to inspect directly the physical properties of the goods and commodities. This lack of immediate direct experience increases the uncertainty about the quality of the products.

One way to reduce this uncertainty is by allowing consumers to share their experiences in the form of reviews and recommendations (Roselius, 1971). Indeed, many of the leading Internet marketplaces (e.g., Amazon, EBay) offer consumers the opportunity to publicize their experiences in the form of verbal statements and numerical ratings. Verbal statements usually are free-format short essays, whereas numerical ratings provide a summary in the form of quantitative ratings. These two information formats - verbal and numerical - need to be integrated by consumers to adequately aggregate the available information.

The compatibility hypothesis (e.g., Slovic, Griffin & Tversky, 1990) suggests that correspondence between <u>stimulus</u> presentation (input) and <u>response</u> mode (output) facilitates the evaluation and the choice process. For example, compatibility decreases reaction time and increases accuracy of performance (Kornblum, Hasbroucq & Osman, 1990). The hypothesis implies that information is aggregated faster and more accurately for cases in which stimulus and response format match - or are compatible - as compared to cases of non-matching formats. For instance, consumers who encounter information in the form of verbal recommendations, and also process it in a matching format, are expected to integrate the information quicker and better than consumers who process the same information in a different format (e.g., numerical rating scales). The quality of

information integration is also expected to affect preference consistency. Preferences should be more consistent with compatible information formats (compared to non-compatible ones), because the information is more reliable (due to fewer "translation" errors from one mode to another), rendering the comparisons more direct and natural.

In this paper, we report empirical evidence for the compatibility hypothesis in information aggregation. Participants were presented with recommendations for fictitious products. The recommendations were presented in either verbal or numerical format. Participants first rated the products using a third modality (a graphical slider) individually, based on the advisors' recommendations, and were then asked to process the recommendations either in a matching or a non-matching format. Finally, they were shown pairs of (the same) products, and asked to pick their preferred ones. This procedure allowed us to investigate the internal consistency of preferences as a function of the stimulus and response information format and their compatibility.

Related literature

The importance of compatibility between stimulus and response has been first noted in the domain of industrial psychology. Responses to visual displays of information, such as an instrument panel, were faster and also more accurate if the structure of the response apparatus was compatible with the spatial arrangement of the stimuli (Fitts & Deininger, 1954; Fitts & Seeger, 1953). For instance, the response to a pair of lights was faster and more accurate if the left (right) light was associated with the left (right) key than if the left (right) light was associated with the right (left) key.

In the literature on judgment and decision-making, the concept of compatibility was invoked as a predictor in multi-attribute decision-making (Slovic & McPhillamy, 1974) and as an explanation for the preference reversal phenomenon (Tversky, Sattath & Slovic, 1988). This phenomenon refers to the inconsistencies in revealed preferences as a function of the elicitation method. For example, people tend to prefer a gamble with a higher probability of winning a relatively small amount of money to a gamble with a lower probability of winning a relatively large amount of money, yet they indicate a higher selling price for the latter as compared to the former gamble (Cubitt, Munro & Starmer, 2004; Lichtenstein & Slovic, 1971; Tversky, Slovic & Kahneman, 1990).

Tversky et al. (1988) propose the hypothesis of scale compatibility[1] to explain this finding. They suggest that the weight of a stimulus attribute is enhanced by its compatibility with a response mode.[2] For instance, payoffs of gambles are usually denoted in dollars, and are therefore more compatible with the elicitation of selling prices (which are also denoted in dollars) than with choices among gambles.[3] Nowlis and Simonson (1997) extend this work by showing preference reversals for different attribute-task combinations. Easy-to-compare attributes (e.g., price) are shown to receive more weight in direct comparisons, whereas "enriched" attributes (e.g., brand names) tend to receive greater weight in individual evaluations (e.g., ratings) of the same items. This result is also consistent with the distinction between joint and separate evaluation (e.g., Hsee, 2000).

González-Vallejo and Wallsten (1992) showed that the frequency of preference reversals is moderated by the format of the probability information. When probabilities were stated numerically, preference reversals were observed at a rate equal to previous

empirical studies. However, when probabilities were presented in verbal format, the frequency of preference reversals was significantly reduced. This result was driven by a reduction of risk aversion in the choice task.

The moderating role of the information format is somewhat surprising, since previous evidence suggests that despite the disagreements on the meanings of probability terms (Beyth-Marom, 1982; Wallsten, Budescu, Rapoport, Zwick & Forsyth, 1986), their use does not degrade performance in a variety of decision tasks (Budescu & Wallsten, 1990; Budescu, Weinberg & Wallsten, 1988; Erev & Cohen, 1990). For instance, Erev and Cohen (1990) asked sportswriters and basketball announcers to provide verbal and numerical estimates of the chances of specific events in upcoming basketball games with the promise of sharing a fraction of the profits earned by decision-makers, who relied on these estimates when betting. The results showed no difference in profits between the two information formats.

The present work

One domain, in which these two streams of literature - consistency of preferences and the effects of information format on decision-making - converge, is the aggregation of opinions. With the proliferation of Internet portals that allow (and in many cases, strongly encourage) consumers to publicize their experiences in the form of verbal summaries and/or numerical ratings, the question of how consumers integrate this information, and how it affects their preferences, becomes increasingly important.

Previous research has provided first clues by, for instance, examining how conflicting opinions are weighted (Budescu & Yu, 2006, 2007; Chevalier & Mayzlin,

2006; Gershoff, Mukherjee & Mukhopadhyay, 2003; West & Broniarczyk, 1998), how

decision makers (DMs) revise their judgments in light of the opinions of others (Soll &

Larrick, 2005; Yaniv & Kleinberger, 2000), and how DMs process and compare

numerical and verbal expressions of uncertainty (Jaffe-Katz, Budescu & Wallsten, 1989;

Wallsten, Budescu & Tsao, 1997).

In this paper we merge these separate streams of literature by studying how DMs

integrate opinions communicated numerically and verbally, and how this integration

affects preference consistency. More specifically, we explore whether preference

consistency is sensitive to (a) the format of recommendations provided by others (verbal

or numerical), (b) the modality the DMs used to process and summarize the information

(verbal or numerical) and, most importantly, (c) the compatibility between the two

modes. According to the compatibility hypothesis, we expect that if DMs are presented

with recommendations in the form of verbal statements, the rate of preference reversals

will be higher (lower) when they are asked to provide a numerical (verbal) summary of

these recommendations.

This prediction is based on the assumption that consumers actually switch

between verbal and numerical information when aggregating opinions and that they need

to "translate" some information across modalities. To illustrate this point we conducted a

pilot study. We asked 12 participants (approached at the library) to collect information on

Amazon.com about a product that they would like to purchase in the near future. Amazon

displays recommendations (numerical ratings and verbal summary evaluations)

sequentially for each source – a practice that is shared by other leading Internet stores and

portals (e.g., iTunes, epinions). Participants were free to collect as much or little

information as they wanted. During the entire process we tracked their eye-movements (with the software eyebox2 by Xuuk.com). The results show that all 12 consumers alternated between numerical and verbal information. For each product subjects first looked at the numerical rating, and then read (at least part of) the verbal evaluation, before moving on to the next review. This alternation between numerical and verbal information might make consumers susceptible to compatibility effects – the topic of our research.

We tested compatibility effects in three laboratory experiments, which will be described in the next section. In the first study we establish that non-compatibility between stimulus and response formats increases the frequency of preference reversals. In the second and third studies we show that DMs do not anticipate the detrimental effects of stimulus-response non-compatibility on preference consistency.

Study I

We obtained two individual preference orderings over products --- One is based on graphical ratings, the other is based on pair-wise comparisons among the same products using compatible and incompatible scales. Preference reversals are identified by comparing these two orderings. The compatibility hypothesis predicts higher rates of preference reversals when the stimulus format (verbal or numerical) does not match the response format (verbal or numerical). For example, if DMs are presented with recommendations in the form of verbal statements, the rate of preference reversals would be higher (lower) when they are asked to provide a numerical (verbal) evaluation based on these recommendations.

The study consisted of two stages. In stage 1, participants rated the desirability of 12 fictitious products on a sliding (graphical) scale (see Figure 1 for a screen shot of the task). Participants were instructed to assume that all products were from the same category - household appliances or entertainment products - and had equal prices. The products were shown in random order. For each product participants saw a list of 10 recommendations that were presented either as brief verbal statements or as numerical ratings (in the form of "stars") on a 5-point scale (see Table 1 for a complete listing). The verbal statements were pre-tested to match the 5-point numerical scale (see details below). We manipulated the mean (low, intermediate, or high), the standard deviation (low or high), and the presentation format (verbal or numerical) of the recommendations (see Table 2 for details). Thus, each of the 12 products represents a unique combination of the mean, the standard deviation, and the presentation format of the recommendations.

After a 10-minute filler task (involving the completion of consumer surveys), participants were shown pairs of the products used in stage 1 in random order, including the corresponding recommendations (see Figure 2, Panel a, for a screen shot). The subjects' task was to evaluate the two products. The evaluation was either verbal or numerical. In the verbal condition participants were shown five verbal statements, and asked to select a global evaluation of the product's recommendations by clicking the option that provided the best overall fit. The options ranged from "I am very unhappy with the product" to "I am very satisfied with the product" (see Figure 2, Panel b, for a screen shot of the task). In the numerical condition the participants were shown five numerical scale values (1 to 5 stars), and asked to evaluate the recommendations by clicking the best fitting option (higher number of stars indicated a stronger

recommendation). After evaluating the two products, participants were asked to pick their preferred product (see Figure 2, Panel c̲), and to indicate their confidence of having picked the superior one on a rating scale (see Figure 2, Panel d̲).

We provided incentives for consistent preferences across the two stages of the experiment. The probability of winning a prize, a $20 voucher for amazon.com, increased linearly in participants' preference consistency across both stages of the experiment.[4] The subjects were told that the closer their choices among product pairs in stage 2 replicated their initial preference orderings of stage 1, the higher was their probability of winning the prize.

Pretest

Seventy-one students, 29 women and 42 males, aged 19 to 29 years (M=21.14, SD=1.88), participated in the pretest. They were shown 50 short verbal statements, like "this product was alright" or "this product was amazing," and were instructed to assign each statement a numerical scale value of 1, 2, 3, 4, or 5, where a value of 1 indicated that the product was not at all recommended, and a value of 5 indicated that the product was highly recommended. We picked the 5 statements (one for each scale value) that yielded the highest rates of agreement across participants (see Table 1 for a listing of the selected statements and their corresponding agreement rates). We used these five verbal statements in the subsequent studies, reported below.

Experimental design and procedure

Sixty-eight students, 33 women and 35 males, aged 18 to 24 years (M=20.24, SD=1.45), participated in the study. We employed a 2 (stimulus format: verbal vs.

-14-

numerical) x 2 (response format: verbal vs. numerical) mixed-factorial design. The first factor was manipulated between-subjects, the second within-subjects.

In stage 1, participants rated the desirability of the 3 (means) * 2 (standard deviation) * 2 (format: verbal and numerical) = 12 products. After a filler task, the participants were shown pairs of products. All pairs were shown either in verbal or numerical format according to the between-subjects factor "stimulus format" (see Figure 1, Panel a). There were two blocks of (6*5/2=) 15 pairs of products (see Figure 2, Panel b). The response format in one block matched the format of the stimulus presentation (compatible condition), whereas in the other block it did not match the format of the stimulus presentation (non-compatible condition). Thus, each subject evaluated 30 pairs of products. The order of the two blocks was counter-balanced.

Figure 2 shows a (non-compatible) product pair where the stimulus format was numerical, and the response format was verbal. After evaluating the two products, participants were asked to pick their preferred product (see Figure 2, Panel c), and to rate the confidence of their choice (see Figure 2, Panel d).

Results

The desirability ratings in stage 1 allowed us to derive an individual preference ordering over the 12 products. Since this was done graphically, it does not favor either response mode (verbal or numerical) and serves as the "gold-standard." The pair-wise choices and the summary evaluations, among product pairs in stage 2 induced a second preference ordering. A comparison of these two orderings allowed us to identify preference reversals.

Does non-compatibility between the stimulus and response format increase the rate of preference reversals? To answer this question we computed for each participant the difference between the number of preference reversals under the non-compatible format condition and the corresponding number under the compatible formats condition, based on choices among products (see Panel c of Figure 2).[5] An ANOVA of this measure of compatibility with the factor stimulus format (verbal or numerical) indicates that the number of preference reversals is significantly higher with non-compatible cases, irrespective of the information format $(F(1, 64)=4.74, p<.05, \eta^2=.11)$. That is, there is no significant difference between the two stimulus modes.

What drives preference reversals, besides non-compatibility? To answer this question, we considered separately four different types of product combinations (see Table 3 for a complete enumeration of all cases): (1) pairs of products with different means but equal standard deviation (e.g., M=3.40, SD=1.90 and M=2.60, SD=1.90), which we label DMESD; (2) pairs of products with equal means but different standard deviation (e.g., M=3.40, SD=1.90 and M=3.40, SD=1.07), labeled EMDSD; (3) pairs of products were one has the higher mean and the higher standard deviation (e.g., M=3.40, SD=1.90 and M=3.00 and SD=1.05), labeled HMHSD; and (4) pairs of products where one has the higher mean and the lower standard deviation (e.g., M=3.40, SD=1.07 and M=3.00, SD=1.94), which we label HMLSD.

We ran a 3-way (2 x 2 x 4) mixed ANOVA with the between-subjects factor stimulus format (verbal or numerical) and two within-subjects factors – compatibility of response format (yes or no) and product combination (DMESD, EMDSD, HMHSD and HMLSD). The mean values are displayed in Table 4. We observed a significant main

effect for stimulus format ($F(1, 66)=5.39$, $p<.05$, $\eta^2=.02$), indicating that the average proportion of reversals was higher in the verbal than in the numerical format (M=0.30 vs. M=0.24). We also observed a significant main effect for product combination ($F(3, 64)=37.40$, $p<.05$, $\eta^2=.64$). Bonferroni post-hoc tests that control the experimenter-wise error rate indicate that the average proportion of reversals was higher for EMDSD than for the other combinations ($p<.05$).

To shed further light on the nature and the determinants of the reversals, we regressed the number of reversals for each pair of products on the following explanatory variables: The perceived distance between the members of the pair, measured by the absolute mean differences in the ratings of the two products in stage 1; same mean (1 = the two products have the same mean, 0 = otherwise); same standard variance (1 = the two products have the same variance, 0 = otherwise); compatible (1 = stimulus and response format match, 0 = otherwise); numerical stimulus (1 = stimulus format is numerical, 0 = otherwise); and numerical response (1 = response format is numerical, 0 = otherwise). Table 5 presents the correlation coefficients and the standardized regression coefficients of all 6 predictors.

The regression results indicate a systematic and highly predictable pattern (Adjusted R^2 = .74, $F(6, 53)$ = 29.61, $p < .05$). We observe that reversals are more likely when products are perceived as very similar, when they have identical means, when the stimulus format is non-compatible with the response format, and when products are represented numerically.

A dominance analysis (Azen and Budescu, 2003, Budescu, 1993) performed to identify the predictors' contributions to the overall fit (based on comparisons among them

in all subset regressions) indicates that most variance is explained by products with identical means followed by products that are perceived to be similar to each other. Compatibility explains more variance than the effects of stimulus and response mode combined (see last column of Table 5).

Is the self-reported confidence of having made the correct choice correlated with the number of preference reversals? The answer is no: We did not find significant correlations between these variables for compatible and for non-compatible formats $(r(68)=.22$ and $r(68)=-.11$, respectively).

Discussion

Our results provide strong evidence for the compatibility hypothesis in the aggregation of consumer recommendations. When DMs were asked to integrate recommendations and summarize them in a non-compatible format (e.g., verbal inputs and numerical evaluations), their preferences were significantly less consistent compared to compatible formats (e.g., verbal to verbal). Preferences were most inconsistent for product pairs with identical means but different standard deviations, suggesting that DMs choices are least stable and consistent when facing options with highly variable recommendations. To study the nature and source of the reversals, we classified each subject according to his/her modal preference for low or high variability in all tasks. We find that in stage 1, a slight majority of subjects (53%) favor high variance products when the recommendations are presented verbally, but most subjects (64%) favor low variance products when the recommendations are presented numerically. The corresponding rates are less extreme (closer to 50%) in the second stage. The reversals can be due to the change in the task (e.g., Tversky, et al, 1988) and/or format incompatibility. We found

that the rate of reversals across different tasks was 40% when the formats were compatible and 57% when they were incompatible. Thus, the net effect of format incompatibility was 17% (or about half the size of the preference-reversal baseline).

In the next study, we investigate whether DMs can anticipate compatibility effects.

Study II

Experimental design and procedure

Fifty-one students, 18 women and 33 males, aged 18 to 24 years (M=20.45, SD=1.55), participated in the study. We employed a two-group between-subjects design, varying the stimulus format (verbal vs. numerical). Subjects saw a product pair, described either by 10 verbal or 10 numerical recommendations each. They only had to indicate whether they believed consumers would summarize the information more accurately verbally or numerically.

Results

Most participants predicted that consumers would be able to summarize the information more accurately when using the numerical information *regardless* of the stimulus presentation ($\chi^2(1)=14.23$, p<.05). Twenty-one of the 25 (84%) subjects who saw the numerical stimulus presentation selected the compatible (numerical) response format, whereas only 8 of the 26 (31%) subjects who saw the verbal stimulus presentation selected the compatible (verbal) response format. These results suggest that DMs do not anticipate compatibility effects.

In the next study we investigate whether this observation holds also when DMs can choose their own response format. To this end we return to the experimental paradigm of study I.

Study III

We studied whether DMs anticipate the detrimental effects of non-compatibility between stimulus and response format on preference consistency by allowing participants to choose their response format (verbal, numerical, or random) for every pair of products. If participants anticipate the importance of compatibility between stimulus and response format, they would choose the response format that matches the stimulus format in a majority of cases.

We also test a major corollary of the compatibility effect: If DMs process the stimulus information in one format (e.g., verbal), and then provide a summary evaluation in a different format (e.g., numerical), it should take them longer to evaluate the products than if they evaluate the information in a matching format (e.g., verbal). The additional time is associated with the cross-modality translation (see Jaffe-Katz et al,, 1988). We test this prediction by measuring subjects' decision times (i.e., the time it took them to provide their summary evaluations).

Experimental design and procedure

Sixty-two students, 21 women and 41 males, aged 18 to 25 years (M=20.53, SD=1.47), participated in the study. The participants first rated the desirability of the 3 (means) * 2 (standard deviation) * 2 (format: verbal and numerical) = 12 products. After a filler task, they were shown 15 pairs of products, presented in verbal format, and 15

pairs of products, presented in numerical format. The 30 pairs were presented in random order. For each pair participants could select the response format for the product evaluation and their recommendations. They could choose to respond verbally, numerically, or randomly (in which case the program selected the verbal or numerical format at random). After providing their evaluations, participants were asked to pick their preferred product and to rate the confidence in their choice.

Results

Are subjects sensitive to format (in)compatibility? Most participants preferred the numerical format (67.6%), followed by the verbal format (17.6%), and the random determination of the response format (14.8%). This strong preference for the numerical response format suggests that our participants did not anticipate the detrimental effects of non-compatibility between stimulus and response format on preference consistency.

To investigate this claim more thoroughly, we calculated for each participant the proportion of cases, in which they selected the numerical response format for products that were represented numerically and products that were represented verbally. Table 6 displays the frequencies of the preferred response format as a function of the presentation mode. The average proportion of the numerical response format for products that were presented verbally was 0.56 as compared to 0.79 for products presented numerically $(t(61)=-4.90, p<.05)$. Excluding all instances where subjects were indifferent (and opted for a random choice of the response mode), the corresponding proportions are 0.67 and 0.92, respectively $(t(56)=-4.58, p<.05)$. These findings suggest a general preference for numerical responses for all products, but the intensity of the preference is stronger for compatible (numerical) cases.

Do subjects process compatible and incompatible cases alike? An analysis of subjects' decision times indicates that evaluations were considerably faster for matching stimulus-response formats (M=23.19 seconds, SD=6.00, for verbal-verbal and numerical-numerical) than for non-matching formats (M=29.57 seconds, SD=8.62, for verbal-numerical and numerical-verbal). This difference is significant (t(21)=3.21, p<.05.).[6]

Does incompatibility between stimulus and response formats increase the rate of preference reversals? We computed for each individual the proportion of reversals for compatible and non-compatible formats. Since we had many instances of few (or missing) observations, we used robust non-parametric tests. Wilcoxon ranked-sign tests indicate that (a) the number of preference reversals was significantly higher for non-compatible stimulus-response pairs for products presented verbally, compared to compatible stimulus-response pairs (z=2.66, p<.05), but (b) the number of preference reversals was not significantly different between compatible and non-compatible stimulus-response pairs for the products that were presented numerically (z=0.04, p>.05).[7]

What factors drive preference reversals for the various cases? We performed an ANOVA of the average proportion of reversals per subject with two repeated factors: stimulus format (verbal or numerical) and type of pair. Based on the results of the first study we only distinguished between pairs with equal means and different standard deviations (EMDSD) and all other pairs. The proportion of reversals is significantly higher for product pairs with identical means but different standard deviations as compared to all others (F1, 6)=18.58, p<.05; M=0.66 versus M=0.33). There was no effect for the format of presentation.

Is the self-reported confidence correlated with the rate of preference reversals?
The correlations between the confidence and the rate of preference reversals are not significantly different from 0 for the verbal stimulus format (r(61)=.11), the numerical stimulus format (r(61)=.09), the verbal response format (r(61)=-.08), and for the numerical response format (r(61)=.10).

Discussion

Our results demonstrate a general preference for the numerical response format, replicating the results of study II that asked for expectations about others, as well as the findings of Olson and Budescu (1997). One reason for this finding is that it is easier to compare numerical values, such as prices, than verbal information, such as brand names (Viswanathan & Narayanan, 1994). Preference consistency was impaired by the lack of compatibility between stimulus and response mode only when subjects opted to respond numerically to product pairs that were presented verbally. This combination of imprecise (verbal) input and precise (numerical) output is quite rare (see Wallsten et al., 1993).

In addition, we showed that – consistent with the compatibility effect – stimulus-response compatibility led to faster decision times than incompatability (see similar results for congruent and non-congruent judgments in Jaffe-Katz et al., 1989). This finding indicates that DMs require extra time to translate the information from one format to the other.

General discussion

The compatibility hypothesis states that congruence between stimulus presentation and response mode decreases reaction time and increases accuracy in

performance. This implies that information would be aggregated faster and more accurately for cases in which stimulus and response format match compared to cases of non-matching formats. Hence, we predicted that preferences would be more consistent for cases of compatible information formats.

We tested the compatibility hypothesis by presenting participants with recommendations for fictitious products. The recommendations were shown either in verbal or numerical format. Participants first rated the products individually using a (format-neutral) slider. Later, subjects were shown pairs of (the same) products, and asked to pick their preferred ones, while manipulating the match between the presentation and response modes. This allowed us to investigate the internal consistency of preferences as a function of the compatibility between the stimulus and response format.

In the first study, we showed that non-compatibility between stimulus and response formats increases the frequency of preference reversals. In the second and third study, we showed that DMs do not anticipate the detrimental effects of stimulus-response non-compatibility on preference consistency. We also showed that stimulus-response compatibility led to faster decision times than non-compatibility, suggesting that it is easier to process the information if the response involves the same format and there is no need to convert and switch formats. This finding is directly relevant to the ongoing debate on the mental representation of numerical information. Abstract-modular theories (e.g., McCloskey, Sokol & Goodman, 1986) postulate a central abstract and amodal representation of numbers. The compatibility effects observed in our studies, for both numerical and verbal representations, weaken this claim.

It is well established that DMs appear to reverse their prefaces across elicitation modes, and that the rate of reversals varies as a function of special features of the stimuli (e.g Tversky et al., 1990). We found that in the present context DMs showed the highest levels of reversals for products with identical means but different variances. The finding that DMs have more trouble exhibiting consistent preferences for products with different variances as compared to products with different means is in line with previous research on people's ability to extract statistical properties from various stimuli. For instance, Ariely (2001) and Chong and Treisman (2003, 2005) showed that DMs can accurately judge the mean sizes of groups of circles. Obrecht, Chapman and Gelman (2007) showed that in pair-wise comparisons DMs assigned the most weight in their decisions to mean differences, less weight to sample size, and little weight to standard deviation. Given the predominance of reversals associated with differences in variances we focused our attention only to these cases. We were able to determine that the net effect of format incompatibility is about half of the effect of the change in procedure (17% compared to 40%). This is a sizeable effect that deserves additional study. This finding, however, was moderated by the information format. When the information was presented both numerically and visually, DMs were more confident when the standard deviation was high.

Martin, Barron, and Norton (2006) showed that DMs sometimes prefer high variability options. For instance, they showed that movies with higher variability in ratings were more successful at the box office (controlling for mean differences in movie recommendations). Preferences for high variability options were found for other positive experiences, such as desserts, and jelly beans, but were reversed for negative experiences,

such as fear-factor foods, and dentist visits. West and Broniarczyk (1998) showed that DMs prefer variance in recommendations, i.e., disagreement, for alternatives that fall below an idiosyncratic reference or aspiration level. However, for alternatives that exceed such a level consensus among recommenders is preferred. Goldsmith and Amir (2008) investigate promotions that offer uncertain rewards. The authors find that a lottery between small and large rewards is perceived as more attractive than receiving the large reward for certain.

Our finding that compatibility between information presentation format and mode of processing matters, begs the question what is the most prevalent information format that consumers encounter, and how do they process it? The presentation format may vary systematically across sources. When consumers rely on the recommendations of friends, colleagues or peers, the information format is predominately verbal, because these recommendations are most likely voiced in direct verbal communication, and because this is most people's preferred mode (Wallsten et al., 1993). However, if consumers consult systematic surveys of consumer reports (e.g., www.consumerreports.org) the dominant information format is numerical, and most internet portals present mixtures of the two formats.

It is also reasonable to assume that the information format is a function of the product category. One could imagine that for products with high hedonic appeal consumers search for confirmatory cues in order to justify their purchase intent. These cues seem to be more likely to be verbal in nature, because words appear more ambiguous than numbers (Budescu et al., 1988; Windschitl & Weber, 1999). On the other hand, for products with high utilitarian appeal consumers might prefer numerical

information, which appears to be more precise, and more likely to identify the objectively best product of a category. Here consumers might be less emotionally involved with a specific, implicitly favored, product. And, of course, these two factors may interact in the sense that one is likely to rely more on verbal advice from friends for products with high hedonic appeal (e.g., entertainment), and gravitate towards formal, data driven numerical ratings for utilitarian products (e.g., appliances). Future research is needed to explore these important questions.

Our findings have practical implications for the design of recommendation systems and web portals. When consumers are asked to provide reviews about their experiences with certain goods and products, they should be given the option of either providing verbal summaries and/or numerical evaluations. This would ensure that consumers can adequately recall the originally processed information and provide an evaluation thereof in a matching format. Also, separating verbal summaries from numerical ratings might be beneficial to those consumers who wish to aggregate the available information on internet portals, like Amazon.com. This would ensure that consumers do not need to switch between different information formats (with sometimes conflicting evaluations) for single reviews, thereby reducing the likelihood of "translation" errors.[8]

Endnotes

[1] Fischer and Hawkins (1993) later differentiated between scale compatibility and strategic compatibility.

[2] In later explanations of the preference reversal phenomenon, Tversky et al. (1988, 1990) stress the relevance of both hypotheses; scale compatibility and prominence. Prominence refers to the attractiveness of the gamble with the higher probability of winning a small amount of money in choice.

[3] Schkade and Johnson (1989) use process measures to show that participants give differential attention to probabilities and payoffs in studies on preference reversals.

[4] We counted the number of violations of the initial preference orderings across the two information formats, resulting in a number between 0 (no violations) and 30 (all violations). We used this number to derive the probability of winning the $20 voucher: (30 - # of violations)/30. Thus, the probability of winning the prize is 1 in case of zero violations, and 0 in case of 30 violations.

[5] There was a 98% agreement between these choices and the evaluations (panel b of Figure 2), so we only report these results.

[6] This analysis is based only on those subjects who selected to answer, at least, once in each format for every presentation mode.

[7] There was a 97% agreement between these choices and the evaluations (panel b of Figure 2), so we only report these results.

[8] Diehl (2005) provides another example for the potentially negative effects of information display on consumer choice.

References

Ariely, D. (2001). Seeing sets: Representation by statistical properties. Psychological Science, 12, 157-162.

Azen, R., and Budescu, D. V. (2003). Dominance analysis: A method for comparing predictors in multiple regression. Psychological Methods, 8, 129-148.

Beyth-Marom, R. (1982). How probable is probable? A numerical translation of verbal probability expressions. Journal of Forecasting, 1, 257-269.

Brun, W., and Teigen, K. H. (1988). Verbal probabilities: Ambiguous, context dependent, or both? Organizational Behavior and Human Decision Processes, 41, 390-404.

Budescu, D. V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. Psychological Bulletin, 114, 542-551.

Budescu, D. V., and Wallsten, T. S. (1990). Dyadic decisions with numerical and verbal probabilities. Organizational Behavior and Human Decision Processes, 46, 240-263.

Budescu, D. V., Weinberg, S., and Wallsten, T. S. (1988). Decisions based on numerically and verbally expressed uncertainties. Journal of Experimental Psychology: Human Perception and Performance, 14, 281-294.

Budescu, D. V., and Yu, H.-T. (2006). To Bayes not to Bayes? A comparison of two classes of models of information aggregation. Decision Analysis, 3, 145-162.

Budescu, D. V., and Yu, H.-T. (2007). Aggregation of opinions based on correlated cues and advisors. Journal of Behavioral Decision Making, 20, 153-177.

Chevalier, J. A., and Mayzlin, D. (2006). The effect of word of mouth on sales: Online

    book reviews. Journal of Marketing Research, 38, 345-354.

Chong, S. C., and Treisman, A. (2003). Representation of statistical properties. Vision

    Research, 43, 393-404.

Chong, S. C., and Treisman, A. (2005). Attentional spread in the statistical processing of

    visual displays. Perception & Psychophysics, 67, 1-13.

Cubitt, R. P., Muro, A. and Starmer, C. (2004). Testing explanations of preference

    reversal. Economic Journal, 114, 709-726.

Diehl, K. (2005). When two rights make a wrong: Searching too much in ordered

    environments. Journal of Marketing Research, 42, 313-322.

Erev, I., and Cohen, B. C. (1990). Verbal versus numerical probabilities: Efficiency,

    biases, and the preference paradox. Organizational Behavior and Human Decision

    Processes, 45, 1-18.

Fischer, G. W., and Hawkins, S. A. (1993). Strategy compatibility, scale compatibility,

    and the prominence effect. Journal of Experimental Psychology: Human

    Perception and Performance, 19, 580-597.

Fitts, P. M., and Deininger, R. L. (1954). S-R compatibility: Correspondence among

    paired elements within stimulus and response codes. Journal of Experimental

    Psychology, 48, 483-492.

Fitts, P. M., and Seeger, C. M. (1953). S-R compatibility: Spatial characteristics of

    stimulus and response codes. Journal of Experimental Psychology, 46, 199-210.

Gershoff, A. D., Mukherjee, A., and Mukhopadhyay, A. (2003). Consumer acceptance of

    online agent advice: Extremity and positivity effects. Journal of Consumer

    Psychology, 13, 161-170.

Goldsmith, K., and Amir, O. (2008). Reflexive positivity: How uncertainty can improve

    promotions. Working Paper, University of California, San Diego.

González-Vallejo, C., and Wallsten, T. (1992). Effects of probability mode on preference

    reversals. Journal of Experimental Psychology: Learning, Memory, and

    Cognition, 18, 855-863.

Hsee, C.K. (2000). Attribute evaluability: Its implications for joint-separate evaluation

    reversals and beyond. In Kahneman, D., and Tversky, A. (Eds.) Choices, values

    and frames. Cambridge University Press (pp. 543-564).

Jaffe-Katz, A., Budescu, D. V., and Wallsten, T. S. (1989). Timed magnitude

    comparisons of numerical and nonnumerical expressions of uncertainty. Memory

    and Cognition, 17, 249-264.

Kornblum, S., Hasbroucq, T., and Osman, A. (1990). Dimensional overlap: Cognitive

    basis for stimulus-response compatibility – A model and taxonomy. Psychological

    Review, 97, 253-270.

Lichtenstein, S., and Slovic, P. (1971). Reversals of preferences between bids and

    choices in gambling decisions. Journal of Experimental Psychology, 89, 46-55.

Martin, J. M., Barron, G., and Norton, M. I. (2006). Preferences for high variance

    experiences: Risk-seeking in positive domains? Manuscript in preparation.

McCloskey, M., Sokol, S. M., and Goodman, R. A. (1986). Cognitive processes in verbal-number production: Inferences from the performance of brain-damaged subjects. Journal of Experimental Psychology: General, 115, 307-330.

Nowlis, S. M., and Simonson, I. (1997). Attribute-task compatibility as a determinant of consumer preference reversals. Journal of Marketing Research, 34, 205-218.

Obrecht, N., Chapman, G. B., and Gelman, R. (2007). Intuitive t-tests: Lay use of statistical information. Psychonomic Bulletin and Review, 14, 1147-1152.

Olson, M., and Budescu, D. V. (1997). Patterns of preferences for numerical and verbal probabilities. Journal of Behavioral Decision Making, 10, 117-132.

Roselius, T. (1971). Consumer rankings of risk reduction methods. Journal of Marketing, 35, 56-61.

Schkade, D. A., and Johnson, E. J. (1989). Cognitive processes in preference reversals. Organizational Behavior and Human Decision Processes, 44, 203-231.

Slovic, P., Griffin, D., and Tversky, A. (1990). Compatibility effects in judgment and choice. In: R. Hogarth (ed.). Insights in Decision Making. Chicago: University of Chicago Press (pp. 5-27).

Slovic, P., and McPhillamy, D. (1974). Dimensional commensurability and cue utilization in comparative judgment. Organizational Behavior and Human Performance, 11, 172-194.

Soll, J. B., and Larrick, R. P. (2005). Strategies for revising judgment: How (and how well) people use others' opinions. Manuscript submitted for publication.

Tversky, A., Sattath, S., and Slovic, P. (1988). Contingent weighting in judgment and choice. Psychological Review, 95, 371-384.

Tversky, A., Slovic, P., and Kahneman, D. (1990). The causes of preference reversals. American Economic Review, 80, 204-217.

Viswanathan, M., and Narayanan, S. (1994). Comparative judgments of numerical and verbal attribute labels. Journal of Consumer Psychology, 3, 79-101.

Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., and Forsyth, B. (1986). Measuring the vague meaning of probability terms. Journal of Experimental Psychology, 115, 348-365.

Wallsten, T. S., Budescu, D. V., Zwick, R., and Kemp, S. M. (1993). Preferences and reasons for communicating probabilistic information in verbal or numerical terms. Bulletin of the Psychonomic Society, 31, 135-138.

Wallsten, T. S., Budescu, D. V., and Tsao, C. J. (1997). Combining linguistic probabilities. Psychologische Beitraege, 10, 27-55.

West, P. M., and Broniarczyk, S. M. (1998). Integrating multiple opinions: The role of aspiration level on consumer response to critic consensus. Journal of Consumer Research, 25, 38-51.

Windschitl, P. D. and Weber, E. U. (1999). The interpretation of "likely" depends on the context, but "70%" is 70% - Right? The influence of associative processes on perceived certainty. Journal of Experimental Psychology: Learning, Memory, and Cognition, 25, 1514-1533.

Yaniv, I., and Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. Organizational Behavior and Human Decision Processes, 83, 260-281.

Table 1: Labels used in the verbal and numerical information format

Scale value	Verbal format	Numerical format	Agreement rate in pilot study
1	I am very unhappy with the product	★☆☆☆☆	88.73
2	The product was quite bad	★★☆☆☆	85.92
3	The product was okay	★★★☆☆	90.14
4	The product was quite good	★★★★☆	84.51
5	I was very satisfied with the product	★★★★★	90.14

Table 2: Mean (standard deviation) of the verbal and numerical recommendations

	High standard deviation	Low standard deviation
High mean	3.40 (1.90)	3.40 (1.07)
Intermediate mean	3.00 (1.94)	3.00 (1.05)
Low mean	2.60 (1.90)	2.60 (1.07)

Table 3: Classification of the 15 product pairs into 4 classes

Class of pairs			
DMESD	EMDSD	HMHSD	HMLSD
(6 pairs)	(3 pairs)	(3 pairs)	(3 pairs)
3.40 (1.07)	3.40 (1.07)	3.40 (1.90)	3.40 (1.07)
3.00 (1.05)	3.40 (1.90)	3.00 (1.05)	3.00 (1.94)
3.40 (1.07)	3.00 (1.05)	3.00 (1.94)	3.00 (1.05)
2.60 (1.07)	3.00 (1.94)	2.60 (1.07)	2.60 (1.90)
3.00 (1.05)	2.60 (1.07)	3.40 (1.90)	3.40 (1.07)
2.60 (1.07)	2.60 (1.90)	2.60 (1.07)	2.60 (1.90)
3.40 (1.90)			
3.00 (1.94)			
3.40 (1.90)			
2.60 (1.90)			
3.00 (1.94)			
2.60 (1.90)			

Note: The entries in the table denote possible product pairs, showing means plus standard deviations (in parenthesis) of each pair. DMESD denotes products with Different Mean, but Equal Standard Deviation. EMDSD denotes products with Equal Means, but Different Standard Deviation. HMHSD denotes products where one has the Higher Mean and the Higher Standard Deviation, and HMLSD denotes products where one has the Higher Mean and the Lower Standard Deviation.

Table 4: Percentage of preference reversals as a function of stimulus and response

compatibility (study I)

Format		Product pairs				
Stimulus	Response	DMESD	EMDSD	HMHSD	HMLSD	Overall
Numerical	Numerical	22.06	32.39	22.74	19.53	24.18
Numerical	Verbal	21.59	32.74	20.60	24.17	24.78
Verbal	Verbal	25.76	39.02	24.24	25.38	28.60
Verbal	Numerical	30.14	40.03	30.18	29.04	32.35
Overall		24.89	36.04	24.44	24.53	27.48

Table 5: Regression analysis on the frequency of reversals (study I)
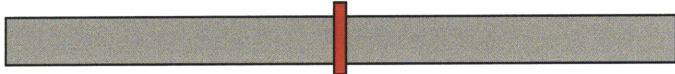
Variables	Zero-order correlations	Standardized coefficients	t (df=59)	General dominance
Perceived Distance	-.70	-.37	-4.39*	.267
Same Mean	.79	.51	5.69*	.364
Same Variance	-.42	-.12	-1.69	.076
Compatible	-.19	-.19	-2.84	.035
Numerical Stimulus	.12	.16	2.46*	.022
Numerical Response	-.07	-.07	-1.10	.005
Model's R^2 (Adjusted R^2)				.770 (.744)

$p < .05$

Table 6: Stimulus presentation format and preferred response format (study III)

Stimulus format	Preferred response format			
	Numerical	Random	Verbal	Total
Numerical	733	131	66	930
Verbal	525	144	261	930
Total	1,258	275	327	1,860

Figure 1: Schematic screen-shot of the desirability ratings task (stage 1)

PRODUCT #2

The product was okay.
The product was quite bad.
The product was quite good.
I was very satisfied with the product.
The product was quite bad.
The product was okay.
The product was quite good.
The product was okay.
I was very satisfied with the product.
The product was okay.

How desirable is the product to you?

Not at all desirable ▮ Highly desirable

Note: This product was shown in verbal format and was characterized by a high mean

(3.40) and a low standard deviation (1.07).

Figure 2: Schematic screen-shot of the pair-wise choice task (stage 2)

Note: Product A and B were represented in numerical format. Product A had an intermediate mean (3.00) and a low standard deviation (1.05), whereas product B had a high mean (3.40) and a high standard deviation (1.90). Panel a shows the stimulus presentation, b the response task, c the choice task, and d the confidence task. For this particular example, the stimulus presentation (numerical) was non-compatible with the response task (verbal).

Essay 2

The Researcher as a Consumer of Scientific Publications: How Do Name Ordering Conventions Affect Inferences About Contribution Credits?

ABSTRACT

When researchers from different fields with different norms collaborate, the question arises how name ordering conventions are chosen, and how they affect contribution credits. In this paper we answer these questions by studying two disciplines that exemplify the two cornerstones of name ordering conventions: Lexicographical ordering (i.e., alphabetical ordering, endorsed in economics) and non-lexicographical ordering (i.e., ordering according to individual contributions, endorsed in psychology). Inferences about credits are unambiguous in the latter arrangement, but imperfect in the former, because alphabetical listing can reflect ordering according to individual contributions by chance.

We contrast the fields of economics and psychology with marketing, a discipline heavily influenced by both. Based on archival data, consisting of more than 38,000 journal articles, we show that the three fields have different ordering practices. In two empirical studies, with 351 faculty and graduate student participants from all three disciplines, as well as in a computer simulation, we show that ordering practices systematically affect and shape the allocation of perceived contributions and credit. While strong disciplinary norms in economics and psychology govern the allocation of contribution credits, a more heterogeneous picture emerges for marketing. This lack of strong norms has detrimental effects in terms of assigned contribution credits.

Social norms and conventions establish implicit rules of conduct that facilitate and coordinate social interactions (Akerlof, 2007; Elster, 1989; Sherif, 1966). Applications range from food-sharing among small-scale societies (Kaplan and Hill, 1985) to demand for free consumer products (Shampanier, Mazar and Ariely, 2007) and personalized recommendations (Kramer, Spolter-Weisfeld and Thakkar, 2007), permeating the private and professional life. In academia, social norms provide guiding standards for academic integrity and methodological practices. Norms also govern more mundane areas, such as refereeing for academic journals, or deciding on the ordering of author names in academic publications.

Two recent trends highlight the importance of name ordering. First, many disciplines have noted a steady increase of collaborations, resulting in a sharp rise of multi-authored publications (e.g., Hudson, 1996; Mendenhall and Higbee, 1982) and higher citation rates (Wuchty, Jones and Uzzi, 2007). Second, the higher rate of interdisciplinary collaboration (e.g., Morillo, Bordons and Gómez, 2003) often leads to publications in journals different from the main outlets of an author's "home" discipline. When collaborators from different fields with different conventions determine name ordering, they can either accept a convention of one of the fields or create a new convention. In either case, inferences about contribution credits are imperfect due to a number of reasons:

1. Scientific disciplines cannot, and are not expected to, formally enforce their prevalent ordering conventions. Although the majority of researchers follow the customary norms, some researchers employ different conventions.

2. While inference about relative contributions for non-alphabetically ordered papers is unambiguous, alphabetical listing of authors can either reflect equal contributions or relative contributions, when the names reflect alphabetical order by chance.

3. Attribution ambiguity becomes progressively more important with the growth in interdisciplinary research, the rising number of publications in multi-disciplinary journals, and the increase of citations across disciplinary boundaries.

4. Attribution ambiguity may also affect how individual researchers are evaluated. Authors involved in collaborative work may employ multiple name ordering conventions in their scientific publications (i.e., using both alphabetical and non-alphabetical orderings), which might lead them to being penalized if they are evaluated according to the prevalent norms of their "home discipline."

In this paper we investigate the question of name ordering conventions -- and inferred contribution credits -- from the perspective of new and emerging fields. We consider two disciplines, which exemplify the two cornerstones of name ordering conventions in scientific publications: Alphabetical name ordering (endorsed in economics) and non-lexicographical ordering, according to individual contributions (endorsed in psychology). We study how name ordering conventions are shaped in a relatively young and emerging field -- marketing -- a discipline strongly influenced by the research culture of both economics and psychology (Simonson, Dhar, Drolet and Nowlis, 2001). Deciding about the name ordering on scientific publications is not only a question of which conventions or social norms to invoke. It also signals how researchers decide to "market" themselves. From this perspective, this paper can be seen as an

attempt to understand how the consumers of scientific research, i.e., fellow academics, view and evaluate authors' marketing efforts.

To illustrate the differential name ordering conventions, we asked 163 faculty members from all three disciplines (economics, marketing, and psychology) to participate in a survey ranking of the top journals in their respective disciplines. The selection of journals presented was based on previously published journal rankings (see Stigler, Stigler and Friedland, 1995, for economics; Hult, Neese and Bashaw, 1997, and Tellis, Chandy and Ackerman, 1999, for marketing; and Burgard, 2001, for psychology).

Those journals (see Table 1 for a complete listing) that received at least 80% agreement were included in our analysis, and all scientific articles published in these journals between January 1973 and December 2005 (except if noted otherwise) were downloaded from the Social Sciences Citation Index (SSCI). This generated a data set of 38,315 articles (16,765 in psychology, 15,424 in economics; 6,126 in marketing).

Table 1 shows the frequency of alphabetical name ordering as a function of the number of authors for economics, marketing, and psychology. The rate of alphabetical name ordering in multi-authored papers ranges from 33% in psychology to 87% in economics. As expected, marketing assumes an intermediate position with a rate of 50%. We compared the expected (E_i) and the observed (O_i) percentages of alphabetical name ordering for papers with i authors (i ranging from 2 to 6) for the various disciplines. The prevalence of alphabetical name ordering exceeds chance level by a factor of 2 in economics (see the row "Total"). A substantially different picture emerges in psychology where the prevalence of alphabetical name ordering closely matches the expectation. Marketing is in between these two extremes.

Analyzing the prevalence of alphabetical name ordering across years by discipline (see Figure 1) reveals a higher variance for marketing, which might reflect the existence of different research traditions (behavioral vs. quantitative) *within* the field. For instance, outlets which specialize in behavioral research might show ordering patterns that closely correspond to the existing norms of psychology, while others that specialize in quantitative research might correspond to the norms of economics. We find some directional support for this conjecture with respect to the behavioral tradition, but no support for the quantitative tradition.

The differential name ordering conventions in economics and psychology are reflected in the disciplines' research on this topic. In psychology, researchers emphasize the importance of visibility, inferred from author positions, as the driving motivation for non-lexicographical ordering (Over and Smallman, 1973). In economics most efforts have been devoted to identify conditions under which alphabetical name ordering is attainable (Engers, Gans, Grant and King, 1999; Joseph, Laband and Patil, 2005; Laband and Tollison, 2006). Recent research by Einav and Yariv (2006), however, has shown that researchers in economics with names earlier in the alphabet were more successful professionally. They suspect this finding to be linked to the prevalence of alphabetical name ordering in economics.

Previous research has identified differential norms of name orderings used in distinct academic fields, and the conditions under which they are sustainable, but not much is known about the "consumers" of academic output and the meaning they assign to name orderings. How do they perceive the contributions of individual authors, and what are the potential factors that shape and influence these inferences? Understanding these

factors is of crucial importance for the evaluation of individual scientific output (e.g., grants, promotions, awards) as well as for the motivation of the researchers in a given project. We address these questions in three studies.

STUDY 1: ALLOCATION OF CONTRIBUTION CREDIT

We study whether scholars in economics, marketing, and psychology departments endorse the prevalent conventions of their discipline when assigning contribution credit.

Method

Two hundred forty-seven faculty members and advanced graduate students from economics (n=45), marketing (n=150), and psychology (n=52) participated in an Internet study (80% of the participants held tenure-track or tenured positions and the median rank was assistant professor).[1] Participants were presented with two lists of authors, displayed side-by-side on the screen. In each list, one name (the target author) was highlighted, and the participants were asked to compare the contribution credit that the two authors deserve. First, they were asked to identify the target author that deserves more credit (or to indicate that both deserve equal credit). Next, they were asked for the relative contribution of each target author on a scale from 0% to 100% (for cases of equal contribution, this question was posed only once, applying to both authors).[2]

The lists of authors consisted of one to four names. In case of multiple authors, we presented the lists either in alphabetical or non-alphabetical order, and the target authors could assume positions 1 to 4. Overall, we designed 19 different name listings,

---

[1] Findings in this study, and in study 3, remain qualitatively unchanged when the group of graduate students is removed from the analyses. However, due to lower statistical power, some significant findings become insignificant.
[2] We used an abstract setting to isolate the role of norms in the contribution process, and acknowledge that the effect size of our findings might be different when using more realistic judgments.

resulting in 171 distinct pairs. A screen-shot of the task is shown in Figure 2 and a complete listing of names can be found in Table 2.

The author names were randomly sampled from a list of frequent British names (in order to eliminate prior opinions about particular researchers) and the co-authors from a list of real author names, obtained from the initial study. We instructed participants to assume that the papers being compared were published in the same high-quality journals, and that the lists of authors were hypothetical. Participants were asked to complete at least one block of 25 comparisons, but could complete more than one block. Incentives were provided by holding weekly lotteries raffling $100 coupons for Amazon.com, with the probability of winning proportional to the number of blocks completed. On average, respondents performed 29.77 comparisons.

Results and Discussion

We analyzed a total of 7,234 pairs of authors (4,544 by marketing scholars; 1,702 by psychologists, and 988 by economists). We included all respondents, who answered at least 15 pairs of lists. Applying the Bradley-Terry-Luce (BTL) choice model (Bradley & Terry, 1952; Luce, 1959) to the respondents' preferences, we estimated the scale values for the 19 target authors. The normalized solutions are displayed in Figure 3. We fixed the scale value for a single authored paper to unity in all cases. Thus, the values plotted represent credit assigned to a certain position in a multi-authored paper relative to a single-authored paper. Note that the values inferred from the psychologists' judgments span a wider range, and are considerably lower, than the other two disciplines, indicating that (a) psychologists differentiate more carefully between the various positions, and (b) they value single-authored papers more than the other two disciplines.

While there is high agreement in the ordering across disciplines, systematic differences are observed. Psychologists ordered the name listings lexicographically according to (a) the target author's position and (b) the total number of co-authors. This ordering identifies nine distinct clusters (starting with the first author in a pair and ending with the last author in a group of four). Each of these clusters includes two scenarios that vary in their (alphabetical and non-alphabetical) ordering (being first of two authors, being first of three authors,..., being fourth of four authors). No systematic ordering within these pairs was found, confirming the hypothesis that psychologists do not pay attention to this distinction. On the other hand, economists assign credit differentially for alphabetical and non-alphabetical orderings of authors. They cluster all positions under alphabetical ordering together, but reproduce the psychologists' differentiation for non-alphabetical ordering. The marketing solution is quite similar to psychology, with a few deviations from the lexicographical ordering that, in most cases, are consistent with the economists' judgments.

We regressed these estimates on the following explanatory variables: number of "authors" (1, 2, 3, 4), "position" of the target author (1, 2, 3, 4), and the dummy-variables "alphabetical" ordering (1=yes, 0=no), "economist" (1=yes, 0=no), and "psychologist" (1=yes, 0=no), and interactions between these variables. The model selected includes only those interactions that contributed significantly to the fit. Table 3 presents the correlation coefficients and the standardized regression coefficients of all 9 predictors.

The regression results indicate a systematic and highly predictable pattern (Adjusted R^2=.97, $F(9; 44)$=161.58, $p<.05$), and confirm our impressions from the BTL-analysis. We observe significant differences between the disciplines (lower contributions

for psychologists) and, not surprisingly, significant effects for the number of authors and their position in the author list.[3] A significant interaction between the position of an author and the name ordering of the paper suggests that when the names are ordered alphabetically the assigned contribution credit is higher for later positions. More important for our purposes, we found an interaction between the position and the academic discipline of psychology, indicating that psychologists strongly discount the contribution of authors whose names appear in later positions.

A dominance analysis (Azen and Budescu, 2003; Budescu, 1993) performed to identify the predictors' contributions to the overall fit (based on comparisons among them in all subset regressions) indicate that the author's position, the distinction between psychologists and the other disciplines, and the interaction of these two factors account for the largest share of the variance in the model (see last column of Table 3).

We also analyzed the average contributions assigned by the three academic disciplines, as a function of the name ordering (alphabetical in the top panel and non-alphabetical in the bottom panel), and the number and position of authors (see Figure 4). Although economists tend to assign equal contributions to all authors in the case of alphabetical name ordering, they display a slight decrease in credit for authors in later positions. In case of non-alphabetical name ordering, however, the three disciplines show very similar patterns, suggesting a strong first-author advantage and relatively steep discounting of contributions in later positions.

Next, we analyze the first-author advantage across disciplines. To achieve comparability across differential numbers of authors, we divided the assigned

---

[3] In a similar vein, Stremersch, Verniers and Verhoef (2007) found that author visibility, as captured - among other things - by the number of co-authors, predicts citation counts for marketing articles.

contributions by the expected contributions, assuming that all authors contributed equally (i.e., 50% for two-authored papers, 33.3% for three-authored papers, and 25% for four-authored papers). This value is greater than 1 in all cases (grand mean =1.33, SD=0.22), suggesting that it always pays to be the lead author. We performed a 3-way ANOVA with the factors "academic discipline" (economics, marketing, psychology), "number of authors" (2, 3, 4), and "name ordering" (alphabetical, non-alphabetical) on this measure and found significant main effects for academic discipline ($F(2; 162)=52.74$, $p<.05$, $\eta^2=.39$), number of authors ($F(2; 162)=102.67$, $p<.05$, $\eta^2=.56$), and name ordering ($F(1; 162)=40.41$, $p<.05$, $\eta^2=.20$). We observed a significant 2-way interaction between discipline and number of authors ($F(4; 162)=9.13$, $p<.05$, $\eta^2=.18$), suggesting that as the number of authors increases, the perceived difference in the "first author advantage" between economists and psychologists increases (see Figure 5a). Another 2-way significant interaction between discipline and name ordering ($F(2; 162)=13.58$, $p<.05$, $\eta^2=.14$) indicates that the perceived advantage of being first author is more pronounced in cases with non-alphabetical ordering (psychologists are insensitive to this factor, as shown in Figure 5b). Marketing is between economics and psychology, showing considerably less differentiation of assigned contribution credits.

In summary, study 1 showed that researchers apply their respective disciplinary conventions when deciding about contribution credits. Next we report results of a computer simulation designed to explore potential implications of these findings.

STUDY 2: COMPUTER SIMULATION

We seek to determine how economists, marketing researchers, and psychologists would assess the aggregate contribution credits of colleagues from their own, and from neighboring disciplines. To compare the contribution credits of authors within their own discipline, and across fields, we performed a computer simulation in which we applied the inferred contribution weights from study 1 to the empirical distribution of author names and papers identified in our initial survey.

Method

The simulation was programmed in Matlab and consisted of the following steps:

> *Step 1a*: Select the discipline of the target author (economics, marketing, or psychology).

> *Step 1b*: Select the discipline of the evaluator (economics, marketing, or psychology).

> *Step 2*: Draw randomly a surname initial from the distribution of author names of the particular discipline chosen in step 1a. The distributions of surname initials were based on the journals selected in study 1 for the time period 1973 to 2005. It consisted of 34,561 author names in psychology, 23,820 names in economics, and 10,874 names in marketing.

> *Step 3*: Given the empirical distribution of single-authored papers, two-authored papers, three-authored papers, and four-authored papers (see Table 1), as well as the distribution of alphabetical and non-alphabetical orderings, k papers are drawn at random for the target author identified in step 2. If the target author's discipline (step 1a) is economics, k=20; if it is marketing, k=23, and if it is psychology,

k=27. The differential number of papers across fields reflects our assumption that every author (in every discipline) puts the same amount of time and/or effort into the work, but that time/effort per paper is inversely proportional to the number of authors (meaning that in an m-authored paper each author invests $1/m$ of the amount of time/effort that she would have invested in a single-authored paper. Since the average number of authors differs across fields (1.54 in economics, 1.78 in marketing, and 2.08 in psychology), we allowed the number of papers (k) to differ between disciplines too, reflecting the differential average numbers of authors.

Step 4: For each target author in each of the k papers, the standardized contribution weights from study 2 (see Figure 3) are recorded.[4]

The procedure was repeated 5,000 times for each of the nine cases (three disciplines of the target author x three disciplines of the evaluator).

Results and Discussion

Table 3 displays the average simulated contribution credits (across the 5,000 replications) as a function of the author's discipline as well as the evaluator's discipline. Higher numbers imply higher credits.

Two findings are suggested: First, marketing scholars assign colleagues from their own field lower contributions (30.40) than economists (31.21) and psychologists (32.86)! Second, the contribution credits that marketing scholars assign to colleagues from their field (30.40) are lower than what economists (31.84) and psychologists (32.87) would

---

[4] We standardized the contribution weights with mean equal to 0 and standard deviation equal to 1 within each discipline to render them comparable across fields.

assign to the same target authors! Marketing is the only field for which these two observations hold.

These suggestive findings extend the results of our survey and of study 1 by showing that the lack of a well-differentiated name ordering norm (in interdisciplinary fields such as marketing) can have detrimental effects in terms of overall contribution credits. One shortcoming of this simulation is its implicit assumption that researchers evaluate one's record as a simple summation of all the *individual* papers, without considering potential spillovers from one paper to another. We tackle this possibility in the next study, where we consider publication records with *multiple* publications and conflicting norms; a setting most representative of real-world publication records.

STUDY 3: INFERENCE BASED ON CONTRIBUTION CREDIT

In this study we investigate inferences based on mixed signals about name ordering conventions. Mixed signals occur when researchers with multiple publications list their names according to at least two different conventions. We predict that in such cases, inferences are moderated by the relative position of an author's surname initial. To see this consider, for instance, two scenarios involving a researcher (B) with two publications: one two-authored paper and one three-authored paper (note that B is the second author in all the papers):

Scenario I: 1. A, *B*   2. A, *B*, C

Scenario II: 1. A, *B*   2. C, *B*, A

In scenario I, the authors are listed alphabetically in both papers, whereas in scenario II, they are listed non-alphabetically in the second paper. For simplicity, assume

that B is assigned 1/2 and 1/3 of the contribution credit in scenario I. Given the mixed

signals about author B's name ordering practices in scenario II, inference about the first

paper is ambiguous: it could be that A contributed more than B or that both contributed

equally, justifying alphabetical ordering. We expect that, on average, author B's assigned

contribution to the first paper gets discounted ($< 1/2$) because of this ambiguity.

The reverse pattern is also conceivable: Consider author A in scenarios III and IV.

Author A, the target author in scenarios III and IV, is always the lead author.

Scenario III: 1. *A,* B   2. *A,* B, C

Scenario IV: 1. *A,* B   2. *A,* C, B

Let's assume that A is assigned 1/2 and 1/3 of the contribution credit in scenario

III. In scenario IV author A sends a mixed signals about the name ordering practice, and

inference about the first paper is ambiguous: it could be that A contributed more than B

or that both contributed equally. This ambiguity leads us to predict that, on average,

author A's assigned contribution to the first paper would increase ($> 1/2$).

Our four scenarios highlight the potential interplay of name ordering conventions

and the relative position of an author's surname initial when making contribution

inferences.

Method

One hundred four faculty members and advanced graduate students from economics

(n=21), marketing (n=46), and psychology (n=37) participated in an Internet study (71%

of the participants held tenure-track or tenured positions and the median rank was

assistant professor). Participants were shown pairs of candidates for a prestigious

scholarship. The six major publications of the two applicants were presented side-by-

side. For each publication we presented the list of author names (with the candidate's name highlighted), a classification of the journal (application, methods, or theory), and the length of the paper (8, 9, 10, 11, 12, or 13 pages). Participants were asked to select the more deserving candidate and to rate the strength of their preference on a scale from 1 (barely preferred) to 7 (strongly preferred).

The publications of each candidate consisted of one single-authored paper, two two-authored papers, one three-authored paper, and two four-authored papers. The positions of each candidate in the list of authors were identical: 1st and 2nd position in the two-authored papers, 3rd position in the three-authored paper, and 3rd and 4th position in the four-authored papers. The only differences between candidates were (a) the number of alphabetically listed papers, and (b) whether the initial of the candidate's last name was in the first or the second half of the alphabet (initials D to G vs. R to U). We asked participants to complete all 28 distinct pairs. A screen-shot of the task is shown in Figure 6 and a complete listing of candidate pairs can be found in Table 5.

The names of the candidates were randomly sampled from a list of frequent British names and the co-authors' names were taken from the list of authors identified in study 1. The order of the papers, the journal types, and the length of the papers were sampled randomly (without replacement). We instructed participants to assume that the papers of the two candidates being compared were published in the last three years in high-quality journals. We also informed them that the publications were hypothetical. Incentives were provided by holding weekly lotteries raffling $100 coupons for Amazon.com.

Results and Discussion

We analyzed a total of 2,078 paired comparisons among candidates (1,016 by marketing scholars, 664 by psychologists, and 398 by economists), including all respondents who made at least 10 comparisons. We computed the average "strength of preference for candidate L (left author)" for each participant, and regressed this measure on the following explanatory variables: "economist" (1=yes, 0=no), "psychologist" (1=yes, 0=no), "first half" (1=if candidate L and candidate R's names are in the first half of the alphabet, 0 otherwise), "mixed" (1=if candidate L's name is in the first half of the alphabet and candidate R in the second half, 0 otherwise), and "alphabetical" (number of alphabetically ordered papers for candidate L minus the corresponding number for candidate R). We also considered interactions between these variables. The model selected includes only those interactions that contributed significantly to the fit (see Table 6).

The results indicate a reasonably high fit (Adjusted R^2=.55, $F(9; 74)$=12.33, $p<.05$). The most interesting finding is the interaction of academic discipline and the quality of the ordering signal (as evinced by the number of alphabetically ordered papers). Economists favored candidates with a higher number of papers with alphabetical listing of co-authors, whereas psychologists picked more often the candidate with a higher number of non-alphabetically ordered papers. These findings were not qualified by the position of the candidates' surname initials in the alphabet. The results of a dominance analysis (see last column of Table 6) confirm that the interaction between economists and alphabetical order, and the interaction between psychologists and alphabetical order, account for the largest share of the variance in the model.

Contrary to expectation, inferences were invariant across the surname initials of the authors. This finding demonstrates the power of norms. Researchers in economics and psychology relied on their professions' norms as a heuristic to resolve the conflict induced by the mixed signals: they preferred those candidates with "familiar" publication records, which matched the prevalent ordering patterns of their disciplines. Because of marketing's weaker disciplinary norms these researchers could not rely on simple rules to resolve the mixed signals.

An obvious question then is whether marketing scholars *either* adhere to the norms of economics or psychology, or whether they employ their own norms. To answer this question we calculated the mean ratings of the eight candidates (D to G and R to U, see Table 7), as inferred from the expressed preferences. On average, economists preferred candidates with a higher number of alphabetical papers and psychologists those with a higher number of non-alphabetical papers. The mean marketing rating shows much less variance, and lacks a clear ordering, suggesting that the marketing respondents do not fall in two distinct sub-groups that either follow the norms of economics or psychology.[5]

GENERAL DISCUSSION

Name ordering norms allow for inferences about contributions to joint research by unifying the standards of evaluation, and help coordinate effort among authors. For example, if the authors agree to list names alphabetically, the equal-contribution norm is invoked, leading authors to exert the same level of effort. Conversely, a decision to list authors non-alphabetically might create incentives to exert effort as a function of how

---

[5] A similar result is suggested by multi-dimensional scaling.

important it is for an author to be listed in various positions. Researchers who place a high value on being listed in a prominent position are likely to exert more effort than those who place less value on these positions.

Various conventions have different strategic implications. For example, ordering authors by relative contributions provides incentives to include more authors as a research project progresses, and allows rewarding these authors accordingly at a smaller cost to oneself than in the case of alphabetical ordering. This might be particularly conducive to interdisciplinary research or to projects that require different types of skills (analytical tools, statistical analysis, clinical skills, etc.). An obvious problem of this norm stems from the fact that researchers often have different perceptions of the importance of their own contribution to joint work (Fine and Kurdek, 1993; Floyd et al., 1994),[6] possibly resulting in fierce arguments about the final ordering of author names on research papers. This, of course, is precisely the problem that alphabetical ordering intends to solve. However, alphabetical ordering is not immune from manipulation: A shrewd and strategically sophisticated researcher who understands the universal first author advantage would seek collaborators whose last names are later in the alphabet to place her in a position of "first among equals."[7]

Although it is perfectly feasible to use a set of different norms -- as a coordination device -- for each particular project, this actually might have detrimental effects on the evaluation of a collection of works, such as in promotion and tenure decisions. These credit inferences are the focus of our work. Our results show that disciplines with strong

---

[6] See Bohlmann, Rosa, Bolton and Qualls (2006) for an example of how group interaction may affect satisfaction levels.
[7] See Mishra, Mishra and Nayakankuppam (2007) for a recent account of the interplay, and interference, of strategic, deliberate, and affective components in information aggregation.

norms, economics and psychology, rely on conventions to resolve conflicting contribution signals. Marketing scholars lacked this opportunity due to weak ordering norms. What are the implications of this?

First, strong norms render inferences about contribution credits more accurate (i.e., more likely to reflect the authors' intentions). Second, when the norms are violated, researchers are penalized in terms of their assigned contribution credits. Third, the penalty for mixed contribution signals is more complex when the norms are alphabetical, because a non-alphabetical ordering on one paper questions the ordering norm of other (alphabetically ordered) papers. Fourth, the signal about individual contributions is least clear for two-authored papers where alphabetical ordering either reflects equal contribution or relative contribution (when the ordering happens to coincide with the surnames' alphabetical order). In marketing, 64% of all multi-authored papers are two-authored, and 60% of these papers list authors alphabetically.

Our results suggest that marketing scholars could benefit from clearer norms. Moreover, the increasing rate of interdisciplinary research, joint-projects, and mixing of alphabetical and non-alphabetical papers suggests that it would be best for the discipline to adopt the individual-contribution norm. If this is not possible, researchers could increase the accuracy of inferences about their contributions by specifying, for example, in an author's note (a) what conventions the authors employed to order names (e.g., alphabetical) and (b) how much each of the authors contributed to the research.[8] These explicit clarifications render later evaluations of joint research less ambiguous and

---

[8] A similar policy is currently implemented for publications in the Proceedings of the National Academy of Sciences (http://www.pnas.org/misc/iforc.shtml). A paper that follows this recommendation is, for instance, Syam and Kumar (2006).

therefore increase the likelihood of accurate and fair evaluations (e.g., for tenure review) of a researcher's output.

Finally, it is worth pointing out that the importance of norms is not limited to academia. Credit-sharing is also important for business and public policy, such as for team-work, press-releases, patents, and the contribution to public goods. In these domains, norms can highlight individual contributions, thereby serving as an implicit motivator and reward, facilitating coordination. Along these lines, Weber and Camerer (2003) show that declines of productivity after company mergers can be partly attributed to conflicting cultures and norms, and Sell and Wilson (1991) show that individual visibility improves contributions to public goods.

References

Akerlof, George A. (2007), "The Missing Motivation in Macroeconomics," *American Economic Review*, 97 (March), 5-36.

Azen, Razia and David V. Budescu (2003), "Dominance Analysis: A Method for Comparing Predictors in Multiple Regression," Psychological Methods, 8 (2), 129-148.

Bradley, Ralph A. and Milton E. Terry (1952), "Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons," *Biometrika*, 39 (December), 324-345.

Bohlmann, Jonathan D., José A. Rosa, Ruth N. Bolton and William J. Qualls (2006), "The Effect of Group Interactions on Satisfaction Judgments: Satisfaction Escalation," *Marketing Science*, 25 (4), 301-321.

Budescu, David V. (1993), "Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression," *Psychological Bulletin*, 114, 542-551.

Burgard, Daniel E. (2001), "Journals of the Century in Psychology," *The Serials Librarian*, 39 (3), 41-56.

Einav, Liran and Leeat Yariv (2006), "What's in a Surname? The Effects of Surname Initials on Academic Success," *Journal of Economic Perspectives*, 20 (Winter), 175-188.

Elster, Jon (1989), "Social Norms and Economic Theory," *Journal of Economic Perspectives*, 3 (Autumn), 991-117.

Engers, Maxim, Joshua S. Gans, Simon Grant and Stephen P. King (1999), "First-Author Conditions," *Journal of Political Economy*, 107 (August), 859-883.

Hudson, John (1996), "Trends in Multi-Authored Papers in Economics," *Journal of Economic Perspectives*, 10 (Summer), 153-158.

Fine, Mark A. and Lawrence A. Kurdek (1993), "Reflections on determining authorship credit and authorship order on faculty-student collaborations," *American Psychologist*, 48 (November), 1141-1147.

Floyd, Steven W., Dean M. Schroeder and Dale M. Finn (1994), "Only If I'm First Author": Conflict over Credit in Management Scholarship," *Academy of Management Journal*, 37 (3), 734-747.

Hult, G. Tomas, William T. Neese and R. Edward Bashaw (1997), "Faculty Perceptions of Marketing Journals," *Journal of Marketing Education*, 19 (1), 37-52.

Joseph, Kissan, David N. Laband and Vivek Patil (2005), "Author Order and Research Quality," *Southern Economic Journal*, 71 (3), 545-555.

Kaplan, Hillard and Kim Hill (1985), "Food Sharing Among Ache Foragers: Tests of Explanatory Hypotheses," *Current Anthropology*, 26 (2), 223-246.

Kramer, Thomas, Suri Spolter-Weisfeld and Maneesh Thakkar (2007), "The Effect of Cultural Orientation on Consumer Responses to Personalization," *Marketing Science*, 26 (2), 246-258.

Laband, David N. and Robert D. Tollison (2006), "Alphabetized Coauthorship," *Applied Economics*, 38 (14), 1649-1653.

Luce, R. Duncan (1959), *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley.

Mendenhall, Mark and Kenneth L. Higbee (1982), "Recent Trends in Multiple

Authorship in Psychology," *Psychological Reports*, 51 (3), 1019-1022.

Mishra, Himanshu, Arul Mishra and Dhanajay Nayakankuppam (2007), "Seeing

Through the Heart's Eye: The Interference of System 1 in System 2," *Marketing

Science*, 26 (5), 666-678.

Morillo, Fernanda, Maria Bordons and Isabel Gómez (2003), "Interdisciplinarity in

Science: A Tentative Typology of Disciplines and Research Areas," *Journal of

the American Society for Information Science and Technology*, 54 (13), 1237-

1249.

Over, Ray and Susan Smallman (1973), "Maintenance of Individual Visibility in

Publication of Collaborative Research by Psychologists," *American Psychologist*,

(February), 161-166.

Persson, Olle (2001), "All Author Citations versus First Author Citations,"

Scientometrics, 50 (2), 339-344.

Sell, Jane and Rick K. Wilson (1991), "Levels of Information and Contributions to

Public Goods," *Social Forces*, 70 (1), 107-124.

Simonson, Itamar, Ziv Carmon, Ravi Dhar, Aimee Drolet and Stephen M. Nowlis

(2001), "Consumer Research: In Search of Identity," *Annual Review of

Psychology*, 52, 249-275.

Shampanier, Kristina, Nina Mazar and Dan Ariely (2007), "Zero as a Special Price: The

True Value of Free Products," *Marketing Science*, 26 (6), 742-757.

Sherif, Muzafer (1966), *The Psychology of Social Norms*. New York: Harper

Touchbooks.

Stigler, George J., Stephen M. Stigler and Claire Friedland (1995), "The Journals of

    Economics," *Journal of Political Economy*, 103 (April), 31-359.

Stremersch, Stefan, Isabel Verniers and Peter C. Verhoef (2007), "The Quest for

    Citations: Drivers of Article Impact," *Journal of Marketing*, 71 (July), 171-193.

Syam, Niladri B. and Nanda Kumar (2006), "On Customized Goods, Standard Goods,

    and Competition," *Marketing Science*, 25 (5), 525-537.

Tellis, Gerard J., Rajesh K. Chandy and David S. Ackerman (1999), "In Search of

    Diversity: The Record Major Marketing Journals," *Journal of Marketing*

    Research, 36 (1), 120-131.

Weber, Roberto and Colin F. Camerer (2003), "Cultural Conflict and Merger Failure: An

    Experimental Approach," *Management Science*, 29 (4), 400-415.

Wuchty, Stefan, Benjamin F. Jones and Brian Uzzi (2007), "The Increasing Dominance

    of Teams in Production of Knowledge," *Science*, 316 (5827), 1036-1039.

Table 1: Frequency of alphabetical name ordering as a function of the academic discipline and the number of authors

No. of authors	Economics Frequency	Frequency of alphabetic order	Percentage of alphabetic order O_i	E_i	Marketing Frequency	Frequency of alphabetic order	Percentage of alphabetic order O_i	E_i	Psychology Frequency	Frequency of alphabetic order	Percentage of alphabetic order O_i	E_i
1	8,676 (56.25%)	-	-	-	2,812 (45.90%)	-	-	-	6,756 (40.30%)	-	-	-
2	5,328 (34.54%)	4,773	89.58	50.00	2,120 (34.61%)	1,276	60.19	50.00	5,445 (32.48%)	2,704	49.66	50.00
3	1,229 (7.97%)	996	81.04	16.67	1,010 (16.49%)	333	32.97	16.67	2,674 (15.95%)	500	18.70	16.67
4	166 (1.08%)	122	73.49	4.17	148 (2.42%)	36	24.32	4.17	1,125 (6.71%)	77	6.84	4.17
5	19 (0.12%)	13	68.42	0.83	22 (0.36%)	6	27.27	0.83	446 (2.66%)	12	2.69	0.83
6	2 (0.02%)	0	0	0.14	8 (0.13%)	0	0	0.14	149 (0.89%)	9	6.04	0.14
>6	4 (0.03%)	0			6 (0.10%)	1			170 (1.01%)	5		
Total	15,424	5,904	87.49	42.65	6,126	1,652	49.85	37.33	16,765	3,307	33.04	32.72

Note: O_i denotes the observed percentage of alphabetically ordered author names, whereas E_i denotes the probability in percent that alphabetic order occurs purely by chance. The value in column "E_i", row "Total," denotes the conditional probability Prob(alphabetical | No. of authors)*Prob(No of authors).

For economics, we included the American Economic Review, Econometrica, Quarterly Journal of Economics, Journal of Political Economy, and Journal of Economic Theory. For psychology, we included Psychological Review, Journal of Experimental Psychology: General (1975-), Journal of Personality and Social Psychology, Psychological Science (1990-), American Psychologist, and Cognitive Psychology. For marketing, we included the Journal of Marketing Research, Journal of Marketing, Journal of Consumer Research (1974-), and Marketing Science (1991-).

Table 2: Experimental design of the author lists of study 1

Identification number	Number of authors	Order	List	Position of target author
1	1	-	A	1
2	2	A	A, B	1
3	2	A	A, B	2
4	3	A	A, B , C	1
5	3	A	A, B , C	2
6	3	A	A, B , C	3
7	4	A	A, B , C, D	1
8	4	A	A, B , C, D	2
9	4	A	A, B , C, D	3
10	4	A	A, B , C, D	4
11	2	NOTA	B, A	1
12	2	NOTA	B, A	2
13	3	NOTA	C, A, B	1
14	3	NOTA	C, A, B	2
15	3	NOTA	C, A, B	3
16	4	NOTA	B, D, A, C	1
17	4	NOTA	B, D, A, C	2
18	4	NOTA	B, D, A, C	3
19	4	NOTA	B, D, A, C	4

Note: A denotes alphabetical and NOTA denotes non-alphabetical ordering of authors, and A, B, C, and D represent the initials of authors' last names. Actual names were sampled from a distribution of real author names.

Table 3: Regression analysis of the 19 estimated scale values (study 1)

Variables	Zero-order correlations	Standardized coefficients	t (df=44)	General Dominance
Economist	0.25	-0.16	-2.30*	0.026
Psychologist	-0.65	-0.38	-5.43*	0.211
Authors	-0.49	-0.22	-5.54*	0.111
Position	-0.66	-0.58	-10.69*	0.224
Alphabetical	-0.01	-0.04	-0.36	0.015
Authors x Alphabetical	-0.12	-0.23	-1.91	0.017
Position x				0.029
Alphabetical	-0.22	0.31	4.27*	
Economist x Position	0.06	0.06	0.83	0.022
Psychologist x				0.317
Position	-0.82	-0.38	-5.16*	
Model's R^2				0.971

Note: * $p < 0.05$.

Table 4: Mean simulated contribution credits as a function of the author's discipline and the evaluator's discipline (study 2)

	Evaluator from			
Candidate from	Economics	Marketing	Psychology	Mean
Economics	31.21^b	29.34	31.99	30.85
Marketing	31.84^c	30.40^a	32.87^c	31.70
Psychology	32.36	30.35	32.86^b	31.86
Mean	31.80	30.03	32.57	31.47

Note: $^{a<b:}$ Marketing scholars assign colleagues from their field lower contributions than what economists and psychologists assign their respective colleagues

$^{a<c:}$ Marketing scholars assign colleagues from their field lower contributions than what economists and psychologists would assign them

Table 5: Pairs of candidates used in study 3

No. of authors	Order	Candidate's position	1st half of the alphabet (D)	(E)	(F)	(G)	2nd half of the alphabet (R)	(S)	(T)	(U)
1	-	-	X	X	X	X	X	X	X	X
2	A	1	X	X			X	X		
2	A	2	X	X			X	X		
2	NOTA	1			X	X			X	X
2	NOTA	2			X	X			X	X
3	A	3	X	X	X		X	X	X	
3	NOTA	3				X				X
4	A	2	X				X			
4	A	3	X				X			
4	NOTA	2		X	X	X		X	X	X
4	NOTA	3		X	X	X		X	X	X
No. of alphabetical papers:			5	3	1	0	5	3	1	0

Note: A denotes alphabetical and NOTA denotes non-alphabetical ordering of author names. D, E, F, G, R, S, T, and U represent the initials of candidates' last names. X and X denote possible candidates and their papers, with X representing alphabetical ordering and X representing non-alphabetical ordering.

Table 6: Regression analysis of the standardized strength of preference for candidate L (study 3)

Variables	Zero-order correlations	Standardized coefficients	t (df=44)	General Dominance
Economist	0.47	0.24	2.53*	0.108
Psychologist	-0.28	0.06	0.67	0.024
First half	-0.01	-0.01	-0.06	0.002
Mixed	-0.07	-0.10	-0.51	0.008
Alphabetical	0.03	-0.07	-0.20	0.020
Economist x Alphabetical	0.61	0.60	5.10*	0.283
Psychologist x Alphabetical	-0.45	-0.34	-3.03*	0.138
First half x Alphabetical	-0.03	-0.09	-0.40	0.004
Mixed x Alphabetical	-0.01	-0.05	-0.21	0.012
Model's R^2				0.776

Note: * p < 0.05.

Table 7: Mean ratings (standard deviation) of the authors as a function of the number of alphabetical papers and discipline (study 3).

		Ranking by		
Author's initial	Number of alphabetical papers	Economics	Marketing	Psychology
D	5	0.97 (1.98)	-0.28 (1.59)	-1.05 (2.35)
R	5	1.30 (1.34)	-0.12 (1.79)	-0.43 (2.82)
E	3	0.36 (0.80)	-0.45 (1.97)	-0.55 (1.94)
S	3	0.21 (1.46)	0.64 (1.63)	-0.67 (1.86)
F	1	-0.12 (1.54)	0.07 (1.73)	1.21 (2.12)
T	1	-0.48 (1.19)	-0.08 (1.67)	0.57 (2.15)
G	0	-0.98 (0.95)	-0.02 (1.70)	0.50 (2.77)
U	0	-1.65 (1.13)	0.13 (1.52)	0.91 (2.70)

Figure 1: Mean proportion (and 95% confidence interval) of alphabetical name ordering as a function of publication year and academic discipline

Figure 2: Schematic screen-shot of study 1

Figure 3: Normalized parameter estimates of the 19 author lists by academic discipline (study 1)
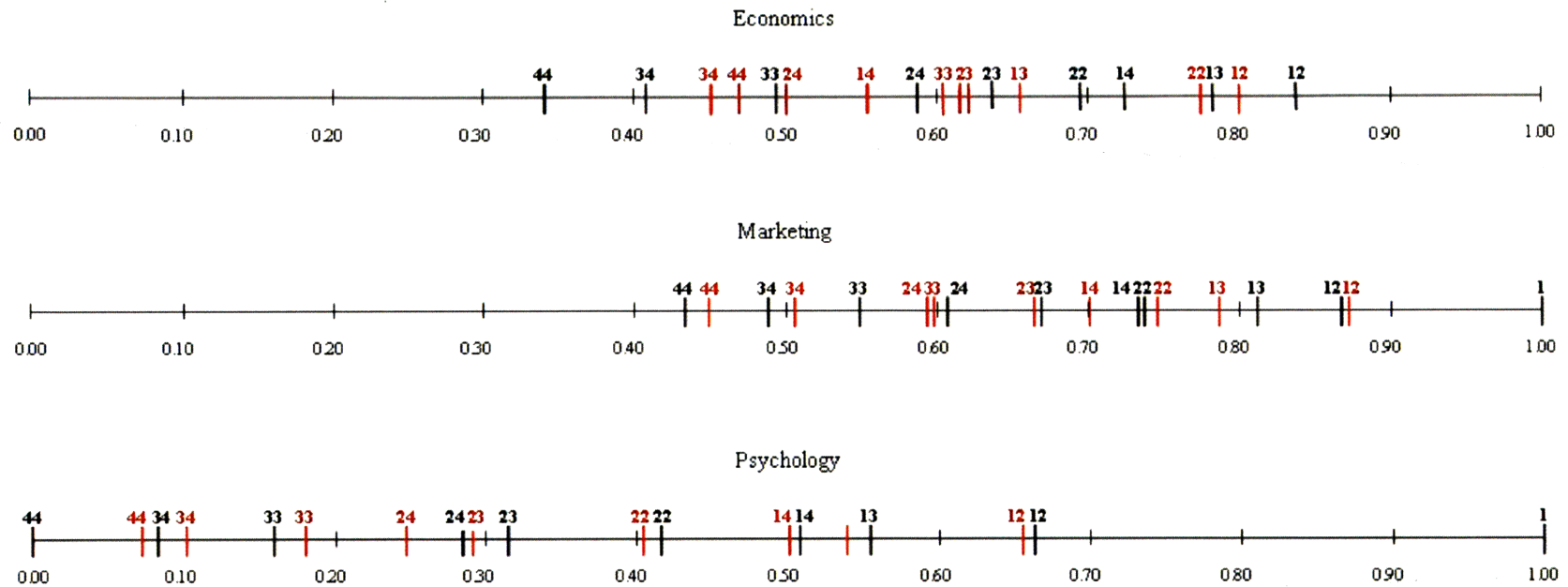
Note: 1 denotes single-authored papers. Alphabetically ordered papers are depicted in red and non-alphabetically ordered papers in black. Papers are identified by an ordered pair of numbers. The first digit denotes the author position, and the second digit identifies the total number of authors. For example, an alphabetically (non-alphabetically) ordered paper with two authors, in which the target author assumes position 1, is denoted by 12 (12).

Figure 4: Average assigned contribution credit as a function of academic discipline, name ordering, and the number and position of authors (study 1)
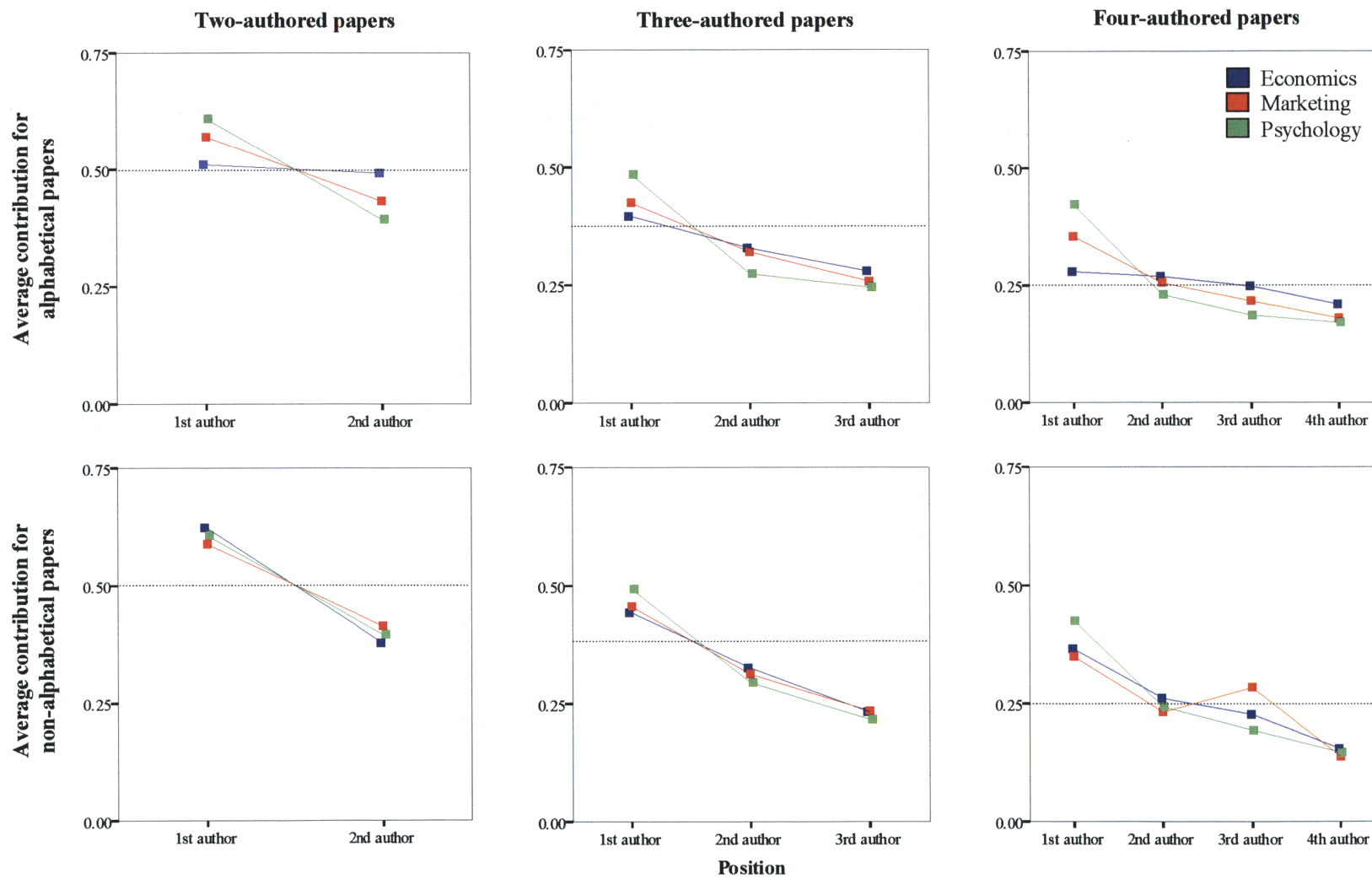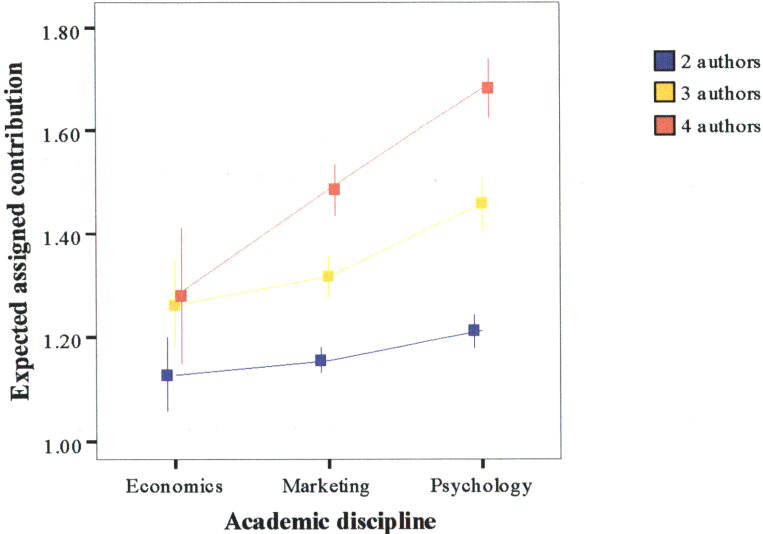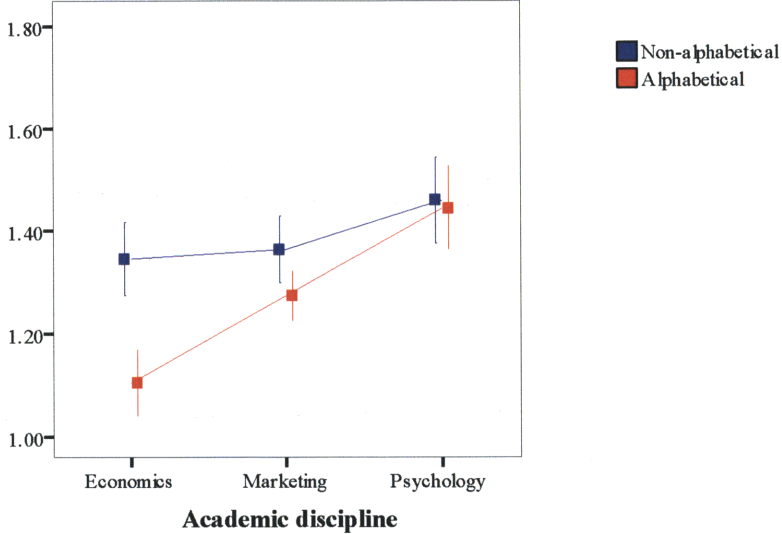
Figure 5: Mean assigned contribution (and 95% confidence interval) to the first author by academic discipline (study 1)

5a: Contribution by discipline
        and number of authors

5b: Contribution by discipline
        and ordering practice

Figure 6: Schematic screen-shot of study 3

In your opinion, which of the two candidates is more deserving of the scholarship?

Sands

Paper	Authors	Journal Type	Length
1	Sands, Schwartz	Methods	13
2	Galvin, Sands	Theory	12
3	Henrekson, Holmlund, Sands	Theory	9
4	Sands	Application	11
5	Vandervoort, Sands, Hofacker, Jordan	Methods	10
6	Thomas, Vanijzendoorn, Sands, Bloomfield	Application	8

Tester

Paper	Authors	Journal Type	Length
1	Wittenbrink, Tester	Methods	8
2	Tester	Theory	11
3	Wheatley, White, Tester, Gaerner	Application	9
4	Tester, Abihashem	Theory	12
5	Venables, Tester, Reznick, Farina	Methods	13
6	Mayes, Sternin, Tester	Application	10

Please rate your preference on a scale from 1 to 7

○1　○2　○3　○4　○5　○6　○7

Barely preferred　　　　　　　　　　　Strongly preferred

Next

Essay 3

Performance-Contingent Discounts and Consumer Choice

ABSTRACT

Incentives affect individuals' attitudes and behaviors in a myriad of ways. In this paper we explore the effects of performance-contingent discounts on consumer choice. For that purpose we set up an online store for digital cameras. Half the subjects received a fixed rebate; the other half had to "earn" their rebate by learning about the products offered. The more information subjects remembered, as inferred from their answers to a short quiz, the higher their discounts. Our results indicate that subjects, who were offered performance-contingent discounts, evaluated the online store as more informative and reputable, were more likely to recommend the store to their friends, and were more likely to buy. The results cannot be attributed to a better performance in the quiz or a more thorough exploration of the products offered.

Businesses face stiff competition for consumers' attention. With too many companies, using similar marketing and advertising strategies, the need for innovative marketing has become crucial. As a consequence, advertising has expanded from traditional communication channels to bathroom stalls, doctors' offices, Broadway plays, Hollywood movies, novels, and even shelf-eggs in supermarkets (Dang, 2006).

But advertising clutter is not the only challenge businesses face today. Ever increasing options on the supply side require consumers to sift through tens or hundreds of products before having explored the market's offerings. With this increase of supply, the probability of picking a suboptimal product has increased sharply. Indeed, Iyengar and Kamenica (2008) showed that an increase in the number of 401(k) funds shifts choices towards money market and bond funds relative to equity funds and Iyengar and Lepper (2000) showed that extensive arrays of jam discourage purchases altogether.

However, with consumer satisfaction being a major determinant of retention and loyalty, companies have an inherent interest in enabling consumers to pick the product that's right for them. Advertising clutter coupled with consumers' risk of choosing suboptimal products pose a challenge to companies and marketers. In this paper, we aim to provide a solution to that challenge. We design an incentive scheme that increases attention rates by encouraging consumers to learn about the market's offerings, allowing consumers to better understand their preferences. The basic principle for the effectiveness of the incentive scheme is the relation between consumers' effort and their motivation to learn about the market's offerings.

Traditional sales promotions, like mail-in rebates, price discounts or coupons, differ in the degree to which they depend on consumers' effort to apply them (Jolson,

Weiner & Rosecky, 1987; Silk & Janiszewski, 2008; Tat, William & Babakus, 1988). For instance, price discounts granted for preferred methods of payment, like cash discounts, require hardly any effort on part of the consumer at the point of purchase. Mail-in rebates, on the contrary, require a sequence of actions for the consumer: the completion of a rebate redemption form, the attachment of the original receipt, proof of purchase from the product pack, and mailing all of the above to the company that offered the rebate (Soman & Gourville, 2005).

The effort, associated with redeeming the rebate, creates uncertainty of the reward, namely the reduced purchase price. Consequently, the consumer has to evaluate whether the attractiveness of the uncertain reward justifies a purchase or not. One way to perform this evaluation is to discount the rebate by the perceived probability of redeeming it (Soman & Gourville, 2005). Prior research indicates that consumers are too optimistic (Soman, 1998) about the redemption probability of rebates, leading them to buy products without ever applying their rebates (Jolson et al., 1987). One factor that is responsible for why consumers do not redeem rebates is that future effort is not very salient (Akerlof, 1991; Loewenstein, 1996).

Past research on the redemption of rebates mainly focused on face-value immediate incentives (Blattberg & Neslin, 1990) and on future delayed incentives (Soman, 1995). However, due to the perceived reduced salience of *future* effort its role on the attractiveness of rebates remains opaque. We contribute to the literature by studying the importance of effort on the *immediate* redemption of rebates by varying the effort level required to "earn" rebates. Subjects in our experiments explore an online store that offers various digital cameras. Afterwards they participate in a quiz that tests their

knowledge of the cameras. Subjects either receive a fixed discount of $50 towards any of the offered cameras or a performance-contingent discount of up to $50. Performance is measured by the number of correctly solved quiz questions. By tying the rebate to the number of correctly solved questions, we increase the effort level that participants put into the exploration of the store and its cameras.

The theoretical predictions on the consequences of increased effort on the attractiveness of the store and its products are ambiguous. Research on the perceived ease, with which information comes to mind (Kruger, Wirtz, Van Boven & Altermatt, 2004; Schwarz, Bless, Strack, Klumpp, Rittenauer-Schattka & Simons, 1991), and the self-perception theory (Bem, 1967), suggest that consumers might interpret and justify the expanded effort in terms of increased liking: "If I worked so hard to explore all the products, and to do well in the quiz, then it must be that I really liked the products." The corresponding prediction would be that subjects in the performance-contingent discount condition evaluate the store, and its products, in a *more positive* way than subjects in the fixed-discount condition, because in the latter, the rebate is a mere windfall gain that does not involve effort.

The opposite prediction is suggested by research on the effects of monetary incentives on internal motivation (Frey & Jegen, 2001). According to this account, explicit rewards, like payments, change the experience of the task, thereby moving the focus of consumers from playful exploration of the online store to being in a business or consulting setting (Heyman & Ariely, 2004). This shift of focus might lead to the "crowding out" of intrinsic motivation and interest in the store, leading subjects in the performance-contingent discount condition to evaluate the store, and its products, in a

less positive way than subjects in the fixed-discount condition, because in the latter, the link between effort/performance and rebate is stronger (and more explicit).

We test these two competing hypotheses in two studies.

Study 1

Experimental design and procedure

Sixty students participated in the study. We employed a one-factorial between subjects design. Subjects either received a fixed discount of $50 towards any of the offered cameras or a performance-contingent discount of up to $50. Performance was measured by the number of correctly solved test questions. This test quizzed subjects about the various features and attributes of the offered cameras. Overall, subjects (in both conditions) had to answer 20 test questions. In the fixed-discount condition, performance had no impact on the offered discount – it was always $50. In the performance-contingent discount condition, however, each correctly solved question translated into an additional discount of $2.50 – with a maximum discount of $50 (if all 20 questions were solved correctly).

In the following, we explain the sequence of events in the experiment in more detail:

Stage 1 – Product & attribute exploration: Subjects could explore the features of 10 different digital cameras in a computer interface. If subjects were interested in a specific camera, then they had to click on it (see Figure 1 for a screen-shot). If they did, then subjects could learn the specific details of the following 6 attributes of the camera: price, pixels, optical zoom, digital zoom, weight, and screen size. In order to obtain this

information, subjects needed to click on a box (see Figure 2a for a screen-shot). This procedure is an application of the so-called mouselab technique (see Johnson, Payne, Bettman & Schkade, 1998).[9] Figure 2b shows the case, where a subject selected the Canon PowerShot A550, and explored its pixels number. Once subjects learned about the specific details of a camera, they could return to the introductory screen, which displayed all 10 cameras, and decide either to explore another camera or to end the exploration stage and to move on to stage 2. Subjects were free to explore as many – or few – of the cameras, and their attributes, as they wanted.

Stage 2 – Preference & quiz: Subjects were asked which of the cameras they preferred. Then they had to answer 20 multiple-choice questions (with four answer categories each) regarding the 10 cameras, encountered in stage 1. For each question, subjects also had to indicate their subjective degree of confidence of having provided the correct answer. As indicated earlier, each correctly solved question translated into an additional discount of $2.50.- in the performance-contingent discount condition, whereas subjects in the fixed-discount condition received a discount of $50, irrespective of the number of quiz questions answered correctly.

Stage 3 – Store evaluation: Subjects were asked how informative and reputable the online store was. Also, subjects were asked how much they trusted the site and the information on the site. Finally, subjects were asked whether they would recommend the online store to their friends and how likely they were to buy their preferred camera – even without the discount. All questions were asked on 11-point scales, where higher values indicated higher rates of affirmation.

---

[9] This technique provides process data of the information subjects attended to.

Stage 4 – Discount information & purchase decision: Subjects were informed about their discount. Then they were asked to indicate whether or not they wished to purchase their preferred camera for the list price minus the discount.

Results and discussion

Stage 1 – Product & attribute exploration: Overall, subject could explore 10 (cameras) * 6 (attributes) = 60 different pieces of information.[10] The number of information pieces viewed did not differ significantly between the two conditions (Mann-Whitney test, z=1.40, p=.16). The mean number of information pieces viewed in the performance-contingent discount condition was 51.44 (SD=14.71) and 45.55 (SD=19.53) in the fixed-discount condition. Sixty-seven percent of the subjects viewed all 60 information pieces in the performance-contingent discount condition and 45% in the fixed-discount condition. There was no significant difference in the average time in seconds that subjects viewed the information (Mann-Whitney test, z=0.73, p=.86).

Stage 2 – Preference & quiz: There were no significant differences in the average number of questions solved, and the average confidence of having given the correct answer, between the conditions. The average number of questions solved was 8.96 (SD=2.65) in the performance-contingent discount condition and 8.24 (SD=2.43) in the fixed-discount condition (t(55)=1.08, p=.29). The average confidence of having given the correct answer was 52.58% (SD=21.14) in the performance-contingent discount condition and 52.34% (SD=18.55) in the fixed-discount condition (t(55)=0.05, p=.96).

Stage 3 – Store evaluation: Subjects in the performance-contingent discount condition found the online store significantly more informative (t(58)=2.98, p<.01), marginally more reputable (t(58)=1.93, p=.06), were more likely to recommend the site

---

[10] However, subjects were free to revisit previously consulted information pieces as often as they liked.

to their friends (t(58)=2.38, p=.02), and were more likely to buy the camera even if they had not been offered a discount (t(58)=2.66, p=.01). Table 1 displays the associated means.

Generally, the six items were highly correlated (Cronbach's alpha = .84), we therefore repeated the analysis with a composite measure of the six items. On this measure, subjects in the performance-contingent discount condition achieved significantly higher values (i.e., higher affirmation) than those in the fixed-discount condition (t(58)=2.91, p=.01).

Stage 4 – Discount information & purchase decision: Three participants in the performance-contingent discount condition decided to purchase their preferred cameras for prices minus discounts of $10, $10, and $15, respectively. In the fixed-discount condition, despite discounts of $50, no-one decided to purchase their preferred cameras. However, none of the three participants followed up on their decision to buy the cameras, so no transactions actually took place.

Despite the increased evaluations of the online store (results from stage 3) in the performance-contingent discount condition, effort in terms of exploration of cameras as well as performance in the quiz did not differ across the two conditions. So the actual mechanism responsible for the increased evaluation does not seem to depend on effort or performance. Before endorsing this result, we replicated study 1 with a few modifications, which will be explained next.

Study 2

Experimental design and procedure

Ninety-eight students participated in study 2, which was a replication of study 1 with the exception that subjects could now explore 20 cameras with 6 attributes each in stage 1. Everything else was identical to study 1.

Results and discussion

Stage 1 – Product & attribute exploration: Overall, subject could explore 20 (cameras) * 6 (attributes) = 120 different pieces of information. This time, in contrast to study 1, the number of information pieces viewed differed significantly between the two conditions (Mann-Whitney test, z=1.97, p<.05). The mean number of information pieces viewed in the performance-contingent discount condition was 88.31 (SD=30.23) and 65.32 (SD=37.35) in the fixed-discount condition. Thirteen percent of the subjects viewed all 120 information pieces in the performance-contingent discount condition and 7% in the fixed-discount condition. Subjects in the performance-contingent discount condition spent marginally more time (in second) viewing the information pieces than subjects in the fixed-discount condition (Mann-Whitney test, z=1.74, p=.08).

Stage 2 – Preference & quiz: This time, subjects in the performance-contingent discount condition solved significantly more quiz questions correctly than subjects in the fixed-discount condition (t(95)=2.761.08, p<.01). The average number of questions solved was 6.50 (SD=2.41) in the performance-contingent discount condition and 7.80 (SD=2.22) in the fixed-discount condition. The average confidence of having given the correct answer was 53.22% (SD=14.71) in the performance-contingent discount condition and 48.47% (SD=20.15) in the fixed-discount condition (t(95)=1.33, p=.19).

Stage 3 – Store evaluation: Subjects in the performance-contingent discount condition found the online store significantly more informative ($t(95)=2.98$, $p=.03$), marginally more reputable ($t(95)=1.90$, $p=.06$), were more likely to recommend the site to their friends ($t(95)=2.52$, $p=.02$), and were more likely to buy the camera even if they had not been offered a discount ($t(95)=2.18$, $p=.03$). Table 2 displays the associated means.

Generally, the six items were highly correlated (Cronbach's alpha = .87), we therefore repeated the analysis with a composite measure of the six items. On this measure, subjects in the performance-contingent discount condition achieved significantly higher values (i.e., higher affirmation) than those in the fixed-discount condition ($t(95)=2.66$, $p=.01$).

Stage 4 – Discount information & purchase decision: No-one wished to purchase their preferred cameras.

We predicted that store evaluations in the performance-contingent discount condition would be moderated by how much effort subjects put into the exploration of the camera features and how well they subsequently did in the quiz. To test this prediction we ran an ANOVA with the composite measure of the store evaluations as dependent variable, the experimental condition (performance-contingent discount vs. discount rebate) as a between-subjects factor, and the total information pieces explored as well as the total number of questions solved in the quiz as covariates. The results suggest that neither exploration ($F(1, 92)=0.12$, $p=.73$) nor performance in the quiz ($F(1, 92)=0.01$, $p=.99$) moderates store evaluation.

General discussion

Sales promotions differ in the degree to which they depend on the effort of consumers to make use of them. Some sales promotions, like for instance mail-in rebates, require considerably more action, planning, and consistency on the part of consumers, as compared to, for instance, cash discounts. From past research little is known on the effects of effort – associated with rebates - on the perceived attractiveness of products. In fact, two prominent streams of literature on the consequences of increased effort on the attractiveness of choice options lead to opposite predictions. The literature on the perceived ease, with which information comes to mind (Kruger et al., 2004; Schwarz et al., 1991), and the self-perception theory (Bem, 1967), suggest that consumers might interpret and justify the expanded effort in terms of increased liking, whereas the literature on the crowding out of internal motivation (see Frey & Jegen, 2001, for an overview) suggests that increased effort is associated with decreased liking.

We test these two predictions experimentally in an online store, varying the effort that is associated with a cash discount. Half of the consumers, visiting the store, received a fixed discount, whereas the discounts of the other half depended on how well the consumers did in a quiz, testing their knowledge of the products offered. Our results indicate that performance-contingent discounts led consumers to evaluate the online store as more informative and marginally more reputable. Also, consumers were more likely to recommend the store to their friends and indicated that they were more willing to purchase their preferred product even without the discount offered. The results cannot be attributed to a better performance in the quiz or to a more thorough exploration of the products offered.

Future research is needed to uncover the mechanism that is responsible for the increased evaluations that we observed. Possible explanations are that the learning and testing phase increased the purposefulness of the exploration and evaluation experience. Explicit incentives might have shifted the mindset of consumers, leading to more positive rather than neutral evaluations. Also, the testing phase might have increased the general arousal level of our online consumers, which in turn might have affected their evaluations.

A possible limitation of our study is the fact that the vast majority of our subjects did not express a desire to buy a digital camera within the next three months of the time the experiment was conducted. This raises serious concerns on subjects' motivation to learn about the products offered. So, future tests of our incentive scheme should vary the product domain to assess the impact of consumers' motivation on product exploration, and subsequent evaluation, more directly.

References

Akerlof, G. A. (1991), "Procrastination and obedience," *American Economic Review*, 81, 1-19.

Bem, D. J. (1967), "Self-perception: An alternative interpretation of cognitive dissonance phenomena," *Psychological Review*, 74, 183–200.

Blattberg, R. C. and S. Neslin (1990), *Sales Promotion: Concepts, Methods and Strategies*. Englewood Cliffs, NJ: Prentice Hall.

Dang, D. T. (2006), "Little is sacred when advertisers vie for consumers' attention: Competition is fierce to just get recognized," *San Francisco Chronicle*, August 6, F4.

Frey, B. S. and R. Jegen (2001), "Motivation crowding theory: A survey of empirical evidence," *Journal of Economic Surveys*, 15, 589-611.

Heyman, J. and D. Ariely (2004), "Effort for payment: A tale of two markets," *Psychological Science*, 15, 787-793.

Iyengar, S. S. and E. Kamenica (2008), "Choice proliferation, simplicity seeking, and asset allocation," *Working Paper*. University of Chicago.

Iyengar, S. S. and M. Lepper (2000), "When choice is demotivating: Can one desire too much of a good thing?" *Journal of Personality and Social Psychology*, 79, 995-1006.

Johnson, E. J., J. W. Payne, J. R. Bettman and D. A. Schkade (1998), *Monitoring Information Processing and Decisions: The Mouselab System*. Unpublished Manuscript. Defense Technical Information Center OAI-PMH Repository.

Jolson, M., J. Weiner and R. Rosecky (1987), "Correlates of rebate proneness," *Journal of Advertising Research*, 27, 33-43.

Kruger, J., D. Wirtz, L. Van Boven and T. W. Altermatt (2004), "The effort heuristic," *Journal of Experimental Social Psychology*, 40, 91-98.

Loewenstein, G. (1996), "Out of control: Visceral influences on behavior," Organizational Behavior and Human Decision Processes, 65, 272-92.

Schwarz, N., H. Bless, F. Strack, G. Klumpp, H. Rittenauer-Schatka and A. Simons (1991), "Ease of retrieval as information: Another look at the availability heuristic," *Journal of Personality and Social Psychology*, 61, 195–202.

Silk, T. and C. Janiszewski (2008), "Managing mail-in rebate promotions," *Working Paper*. University of Florida.

Soman, D. (1998), "The Illusion of Delayed Incentives: Evaluating Future Effort-Money Transactions," *Journal of Marketing Research*, 35, 427-437.

Soman, D. and J. T. Gourville (2005), "The consumer psychology of mail-in rebates: A model of anchoring and adjustment," *Working Paper*. Harvard Business School.

Tat, P., C. William and E. Babakus (1988), "Consumer perceptions of rebates," *Journal of Advertising Research*, 28, 45-50.

Figure 1: Screen-shot of the camera display

Please click on a camera to get more information about it

Done with Site -> Continue to Next Level

Figure 2a: Screen-shot of the attributes display (covered)

Canon PowerShot A550

Price	Pixels

Optical Zoom	Digital Zoom

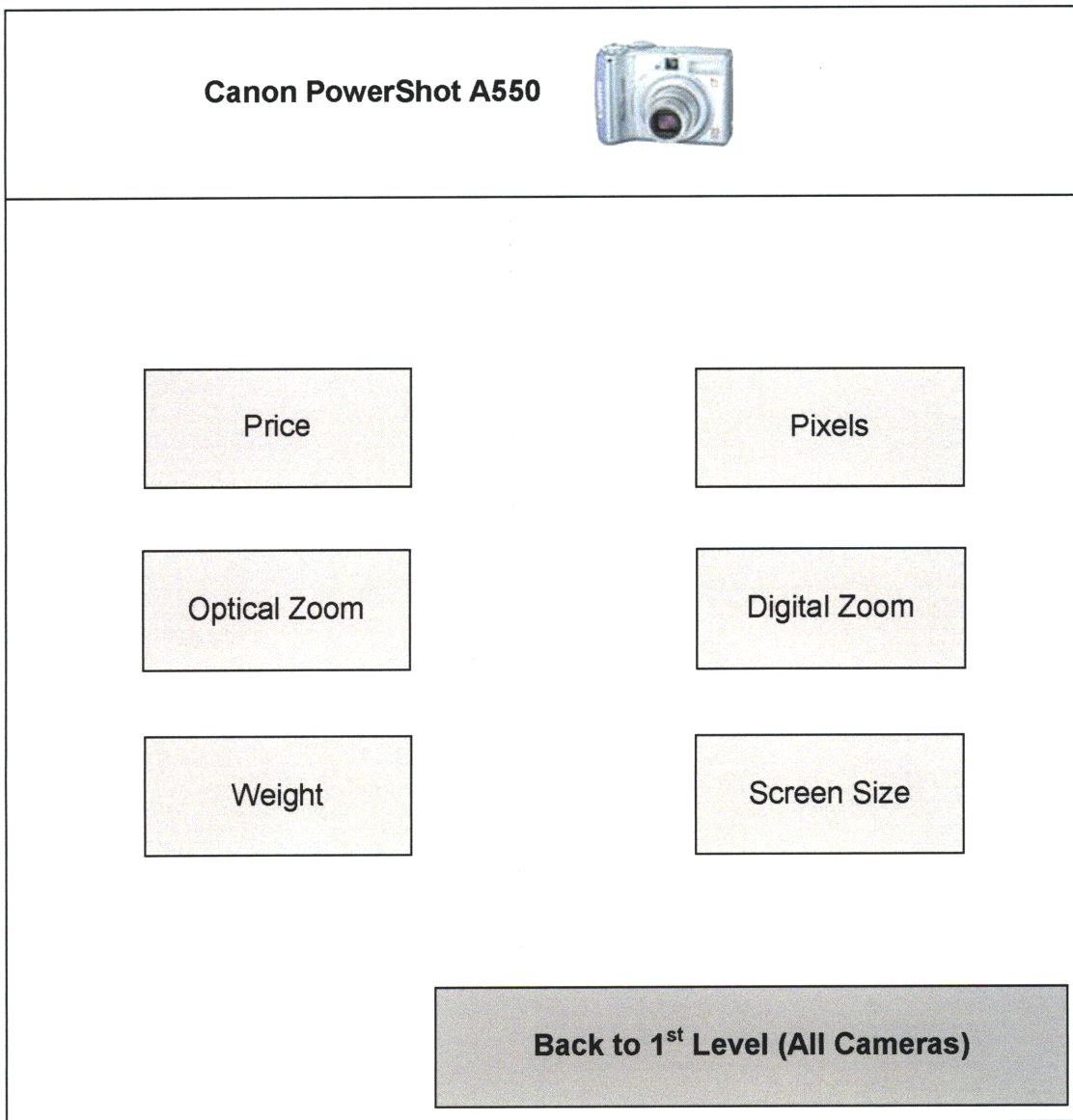| Weight	Screen Size

Back to 1st Level (All Cameras)

Figure 2b: Screen-shot of the attributes display (partially uncovered)

Canon PowerShot A550

Price

7.1 Megapixels

Optical Zoom

Digital Zoom

Weight

Screen Size

Back to 1st Level (All Cameras)

Table 1: Means and test statistics for the online store evaluation (study 1)

Item	Discount	Mean	SD	t	p
Informativeness	Performance-contingent	7.87	1.76	2.98	.004
	Fixed	6.40	2.04		
Reputation	Performance-contingent	7.07	1.98	1.93	.059
	Fixed	6.07	2.03		
Trust in Information	Performance-contingent	7.40	2.36	1.45	.154
	Fixed	6.60	1.90		
Trust in Site	Performance-contingent	7.10	2.38	1.30	.199
	Fixed	6.33	2.19		
Recommendation	Performance-contingent	5.83	2.48	2.38	.021
	Fixed	4.43	2.06		
Likely to Buy	Performance-contingent	4.17	2.69	2.66	.010
	Fixed	2.63	1.65		

Table 2: Means and test statistics for the online store evaluation (study 2)

Item	Discount	Mean	SD	t	p
Informativeness	Performance-contingent	7.32	2.45	2.24	.027
	Fixed	6.23	2.38		
Reputation	Performance-contingent	6.62	2.28	1.90	.056
	Fixed	5.75	2.19		
Trust in Information	Performance-contingent	7.59	2.12	1.58	.119
	Fixed	6.75	2.60		
Trust in Site	Performance-contingent	7.10	1.89	1.66	.098
	Fixed	6.40	2.30		
Recommendation	Performance-contingent	6.20	2.53	2.52	.020
	Fixed	4.88	2.97		
Likely to Buy	Performance-contingent	4.38	1.96	2.18	.031
	Fixed	3.48	2.14		