Over- and underestimation in different product domains

Nava Tintarev and **Judith Masthoff** [1]

Abstract. This paper investigates the effects of over and underestimation on the perceived Effectiveness (helpfulness) of recommender systems. We consider four different product along two dimensions, degree of objectivity and investment. Overestimation was considered more severely than underestimation with regard to perceived Effectiveness. Overestimation was also considered more severely in high investment domains compared to low investment domains. In addition, we surveyed the effect of different gaps between initial (initial impression) and final ratings (true estimate). We found that for gaps which remained in the negative half of the scale were considered less Effective than gaps which crossed over from good to bad (or from bad to good), and gaps which remained in the positive half of the scale.

1 INTRODUCTION

Explanations of products play an important role in improving the user experience in recommender systems [10, 14, 15]. Among other things, good explanations could help users find what they want and/or persuade them to try or purchase a recommended product. Previous recommender systems with explanation facilities have been evaluated in a number of ways, reviewed and discussed in [19].

In this paper, we expand on the criterion of Effectiveness, or how helpful additional information is with regard to aiding users in making decisions about products. In this section, we define Effectiveness in more detail, describe different types of information skews that may occur in recommendations, and different product domains. In Section 2 we describe our experiment. We conclude with a summary of our results and discuss implications for related research in Section 4.

1.1 Effectiveness

In this paper, we consider the metric of Effectiveness, or decision support, with regard to recommendation information. Good decision support can in part be quantified by the metric suggested by Bilgic and Mooney [2]:

1. **(Rating1)** The user rates the product on the basis of the explanation
2. The user tries the product
3. **(Rating2)** The user re-rates the product

Effectiveness can then be measured by the discrepancy between Steps 1 and 3 (Rating1-Rating2). According to this metric, an Effective explanation is one which minimizes the gap between these two ratings. If an explanation helps users make good decisions, getting more (accurate and balanced) information or trying the product should not change their valuation of the product greatly.

[1] University of Aberdeen, Scotland, U.K., email: n.tintare@abdn.ac.uk

The difference between the two ratings may be positive (overestimation of the product) or negative (underestimation). Overestimation may result in false positives; users trying products they do not end up liking. Particularly in high investment recommendation domains such as holidays, a false positive is likely to result in a large blow to trust in the system. Underestimation may on the other hand lead to false negatives; user missing products they might have appreciated. If a user recognizes an underestimation due to previous knowledge or subsequent exposure, this may lead to a loss of trust as well. Likewise an underestimation may needlessly decrease an e-commerce site's revenue. For example, [16] argued that misperceptions which involve underestimating quality affect long term sales compared to perfect information, or even overestimation.

Our aim is to broaden the definition of Effectiveness suggested by [2]. The metric proposed by [2] does not give an indication of whether over or underestimation is preferable to users, or if this preference might be a domain dependent factor. As a consequence it also does not discuss whether skews of the same type, but with different starting points, are comparable. For example, skews can be represented in terms of the following three gap types: gaps which remain in the negative half of a Likert scale, gaps which cross over from good to bad (or from bad to good), and gaps which remain in the positive half of a scale.

1.2 Related work

1.2.1 Skews in valuations of recommendations

User valuations of recommended items can be skewed (either over- or underestimation) by a number of factors. For example if the quality of the information used to form a recommendation, or if the recommendation accuracy is otherwise compromised, this is likely to lead to poor Effectiveness. Likewise, the nature of the recommended object and presentation of the recommended items are likely to be contributing factors.

Firstly, the recommendation algorithm may be flawed. Other times, skewed recommendations are due to insufficient information, or a bias in data. [4] showed that manipulating a rating prediction can alter the user's valuation of a movie to cause either an over- or underestimation. For example, users rated movies lower than their initial rating when they saw a lower prediction for the movie, and vice versa. The study also suggests that users can be influenced to change their rating of a movie from negative to positive. [4] does not discuss whether over- or underestimation is considered more severely by users, but did find that users' valuations of movies changed more for lower predictions (underestimation) than for inflated predictions (overestimation). Also in the movie domain, [13] found that using the difference between the predicted rating (by similar users) for a given user and item, and the actual rating of the user for this item, could be used to increase recommendation accuracy. They considered the sign

of the error, and used this measure to define a prediction range which they used to improve recommendation accuracy. On average, errors based on underestimation were smaller than for overestimation, but were as such least effective for increasing accuracy.

Secondly, presentational choices for recommendations may skew a user's valuation of an item. For example, it has been argued that order of presentation [6], and the use of images [12] can have a persuasive effect on users. [6] found that users click more on highly ranked links, while [12] found that domain credible images could be used to increase credibility of websites.

Thirdly, assuming good algorithmic accuracy, additional information such as explanations can be used to either aid or hinder decision support. An explanation may contain both positive and negative information, and in that sense may have a polarity in a similar way to numerical ratings of a product. Modifying the polarity of an explanation is likely to lead to a similar skew to the one found by [4]. For example, in the study by Herlocker et al [5] participants were most likely to see a movie if they saw an explanation interface consisting of a bar chart of how similar users had rated the movie. This bar chart had one bar for "good", a second for "ok" and a third for "bad" ratings. Bilgic and Mooney [2] later showed that using this type of histogram causes users to overestimate their valuation for items when the dataset is skewed toward positive ratings.

Online reviews are another form of additional information and might sway user valuation of an item. Previous research considering the properties of helpful reviews has found a positive bias in the movie domain [18] as well as for cameras and mobile phones [7].

In our experiment, we study the effects of over- and underestimation due to additional information such as explanations. However, since the skew in the valuation of recommendations can be caused by any of these factors (e.g. limited algorithm, skewed or limited data, presentation, and additional information) the effects on evaluations of skews may be relevant to these causes as well.

1.2.2 Domains

In economics, there has been a great deal of debate about classification of products into different categories. [16] uses the distinction between experience goods, or goods that consumers learn about through experience, and "search goods" which they do not need to learn about through direct experience. Similarly, [3] distinguishes between sensory products and non-sensory products. We propose an interpretation of these categories which distinguishes between products which are easy to evaluate objectively and those which commonly require an experiential and subjective judgment.

Another common categorization in economics involves investment or cost. Often this is a complex construct. For example, [11] discusses *perceived* price in terms of the dimensions of risk and effort. This construct of risk includes financial risk but also psychological, physical, functional and social risk. The construct of effort considers purchase price, but also time that the purchase takes. [3] also discuss perceived price in terms of non-monetary effort and degree of involvement. [8] narrows down the definition of cost to the objective measure of the purchase price of an item. For simplicity, we will also use a definition of investment which only considers purchase price.

2 OVER- AND UNDERESTIMATION

In this experiment, we wanted to find out whether users are more accepting of underestimation or overestimation in general. We also investigated how the nature of a product domain can mitigate, or conversely, exacerbate faulty information.

2.1 Materials

The experiment was conducted using two questionnaires (one for overestimation and one for underestimation). The questionnaires considered four domains distributed over the dimensions of investment (low vs. high) and valuation type (objective vs. subjective) as shown in Table 1.

We defined investment in terms of price. By this definition cameras and holidays are high investment domains. Relatively to these domains, light bulbs and movies can be considered low investment domains.

We considered cameras and light bulbs as objective domains, and movies and holidays as subjective. Our definition of this dimension is based on the premise that while some domains are highly subjective, it is easier to give a quantitative judgment in others. For example, users might be able to reach a consensus as to what properties are important in a camera, and what generally constitutes good quality, while this might be harder for a movie. It might be easier to define good image resolution in a camera than define good acting in a movie. Note also that our choice of definition for this dimension does not preclude that different product features (such as resolution and shutter speed, or actors and director) may vary in terms of importance to different users in all four product domains.

Table 1. Choice of domains

	Low investment	**High investment**
Objective	Light bulb	Camera
Subjective	Movie	Holiday

2.2 Hypotheses

We expect that users will be more lenient toward underestimation, and consider it more helpful than overestimation in general. This hypothesis is based on the assumption that users would like to save money, and are wary of persuasion in commercial systems. Users may prefer being recommended only great items (and miss decent items) to buying more, and being recommended items that they will not like.

It also seems probable that users will have higher demands on accuracy in high investment domains such as movies and holidays. Likewise, users may respond more leniently to skews in subjective compared to objective domains as these are harder to gage.

We also consider that it is possible that the strength of an over- or underestimation may also depend on the starting point on a scale. Therefore, we also consider the effects of over- and estimations of the same magnitude, but with different starting points. For example, what is the effect of underestimation on perceived Effectiveness if a user's valuation of an item changes from negative to ok, and how does this compare to a change from ok to great? A user may consider an explanation least helpful when it causes them to perform an action they would not have performed if they had been given accurate information, e.g. when it changes their valuation of a product from good to bad, or from bad to good. Our hypotheses are thus:

- **H1:** Users will perceive overestimations as less Effective than underestimation.

- **H2:** Users will perceive skews as less Effective in high investment domains compared to low investment domains.
- **H3:** Users will perceive skews as less Effective in objective compared to subjective domains.
- **H4:** Users will perceive cross-over gaps which cross the line from good to bad and vice-versa as less Effective compared to other gap types.

2.3 Participants

Twenty participants (7 female, 12 male, one unknown) were recruited at the University of Aberdeen. They were all postgraduates or researchers in Computing Science. The average age was 31.95 (range 20-62).

2.4 Design

We used a mixed-design, with product domain as a within subject factor, and over- vs. underestimation as a between subject factor. Participants were assigned to one of two conditions. In the first, participants were given a questionnaire with overestimation scenarios, in the second underestimation scenarios.

In the underestimation condition participants saw *Paragraph A*:

Paragraph A: *"Assume you are on a website looking for a particular product to buy (such as a camera, holiday, light bulb, movie). Based on the information given, you form an opinion of the product, and decide* ***not*** *to buy it and to spend the money on something else. Later you talk to a friend who used the product, and your opinion changes."*

The user decides not to buy a product and spends the money on something else. This is to ensure that the choice (not to purchase) is perceived to be irreversible by the participants. Only later do they discover that the product was not as bad as they first thought.

For overestimation we considered situations in which the user initially rated the product highly, but then found the true value of the product lower after buying and trying it. *Paragraph A* is replaced with *Paragraph B* below:

Paragraph B: *"Assume you are on a website looking for a particular product to buy (such as a camera, holiday, light bulb, movie). Based on the information given, you form an opinion of the product, and decide to buy it. After using the product, your opinion changes."*

In both cases participants were asked to consider that they were viewing a new website for each scenario even for similar products. All participants considered products in all four product domains (cameras, light bulbs, movies and holidays) in randomized order. Each participant was given scenarios in which their valuation of the product changed by a magnitude of 2 on a scale from 1 (bad) to 5 (good). We varied the starting point for the initial valuation. The rating of the product can be either:

1. Positive, i.e. staying on the positive side ($3 \leftrightarrow 5$)
2. Negative, i.e. staying on the negative side ($1 \leftrightarrow 3$)
3. Cross-over, i.e. changing polarity ($2 \leftrightarrow 4$)

The order of the three starting points (positive, negative and cross-over) was randomized. The orders of the before and after values were reversed between over- and underestimation, e.g. $3 \rightarrow 5$ (underestimation) became $5 \rightarrow 3$ (overestimation). Given three different starting points and four product domains, each participant considered twelve scenarios.

For each of the twelve scenarios, participants rated how helpful they found the (presumed) information given on the website on a seven point Likert scale (1 = very bad, 7 = very good): *"How do you rate the information on this website given this experience?"*. While this *perceived* Effectiveness differs from true Effectiveness, it also differs from Persuasion. Persuasive information would give the user an initial impression (either positive or negative), but fails to consider the way the user finally rates the product once they try it. In this study the final rating is assumed to be known and true. Step 2 of the proposed metric (see Section 1.1), where the user would normally receive information about the product, is assumed to be a black box.

2.5 Results

2.5.1 Which is better?

Firstly we inquire if over- or underestimation is considered generally more helpful by users. Similarly we want to know just how harmful these skews are considered by users. As can be expected, in Table 2 we see that both over- and underestimation are considered unhelpful. Since it is arguable that the values on a Likert scale may not be equal in distance, we performed a Mann-Whitney non-parametric test which rendered a significant result ($p < 0.01$). Overestimation is considered to be less Effective than underestimation: H1 is confirmed.

Table 2. Perceived helpfulness (on a scale from 1 to 7) for over- and underestimation

	Mean	**StD**
Overestimation	2.59	1.065
Underestimation	3.08	1.212

2.5.2 Does the domain matter?

In Table 3 we offer an overview of perceived helpfulness, for all four domains.

Table 3. Mean (and Std) of perceived helpfulness (on a scale from 1 to 7) for the four domains

	Underestimation	**Overestimation**
Camera	2.87 (1.252)	2.37 (0.964)
Light bulb	3.15 (1.231)	2.63 (1.066)
Movie	3.30 (1.236)	3.00 (1.145)
Holiday	3.00 (1.145)	2.37 (0.999)

Low vs. High Investment Table 4 summarizes the perceived investment in low (light bulbs and movies) and high (cameras and holidays) investment domains. The perceived helpfulness was lower for high investment than for low investment domains (Mann-Whitney test, $p < 0.05$). A separate analysis for over- and underestimation shows a significant effect (Mann-Whitney test, $p < 0.05$ with Bonferroni correction) for overestimation, but not for underestimation. We also see that underestimation is considered as less Effective in high investment compared to low investment domains, but this trend is not statistically significant. It seems as if users are more sensitive to skews in high investment domains, but in particular with regard to overestimation. H2 is confirmed.

Table 4. Mean (and StD) of perceived helpfulness for low vs. high investment domains

	Underestimation	Overestimation
High	2.93 (1.191)	2.37 (0.974)
Low	3.23 (1.225)	2.82 (1.112)

Objective vs. Subjective In Table 5 we see that both over and underestimation are considered less Effective in objective compared to subjective domains, but the trend is not statistically significant. This hints that correct estimates may be more important in objective domains than subjective, regardless of direction of skew. User comments also confirm that some users are more forgiving of misleading information in subjective domains than objective: *"a wrong suggestion about 'subjective' evaluations of products (such as for movie or holidays) should not determine a severe bad judgment of the website."*, *"whether I like a movie (or holiday) is very subjective, and I would not blame my liking a movie less on the quality 1st description"*. The effect is however not sufficiently strong, and H3 is not confirmed.

Table 5. Mean (and StD) of perceived helpfulness for objective vs. subjective domains

	Underestimation	Overestimation
Objective	3.00 (1.239)	2.50 (1.017)
Subjective	3.15 (1.191)	2.68 (1.112)

2.5.3 Does the type of gap matter?

Table 6. Mean (and StD) of perceived helpfulness for different gap types

	Underestimation	Overestimation
Positive	3.90 (0.940)	3.02 (1.084)
Cross-over	3.03 (0.140)	2.68 (0.944)
Negative	2.31 (1.239)	2.05 (0.944)

We hypothesized that gaps which cross over between the positive and negative ends of the scale (cross-over gaps) are less helpful than the two other gap types. We found a significant effect of gap type on perceived Effectiveness in a Kruskal-Wallis test ($p < 0.05$). However, in a Mann-Whitney test we found no significant difference between cross-over gaps and the two other gap types combined. H4 is not confirmed.

Investigating the difference between gap types further, in Table 6 we see that participants found gaps on the negative end of the scale ($1 \leftrightarrow 3$) less helpful than gaps on the positive end ($3 \leftrightarrow 5$), and gaps which cross over between the positive and negative ends of the scale ($2 \leftrightarrow 4$), for data using both over and underestimation. Cross-gaps in turn were considered less helpful than positive gaps. Three Mann-Whitney tests comparing the three gap types pairwise were all found to be statistically significant ($p < 0.05$ with Bonferroni correction). Apparently, negative gaps damage perceived helpfulness the most out of the three gap types rather than cross-over gaps.

A similar series of Mann-Whitney tests were run for over and underestimation separately. All tests returned significant results (p < 0.05, with Bonferroni correction), except for the difference between positive and cross-over gaps for overestimation. That is, the difference in perceived Effectiveness between positive and cross-over gaps for overestimation is negligible.

2.6 Discussion

Our finding of user preference for underestimation compared to overestimation is in line with persuasive theory regarding expectancy violations and attitude change [17]. An audience's initial expectations will affect how persuasive they find a message. In a persuasive context, if expectations of what a source will say are disconfirmed, the message source can be judged to be less biased and more persuasive. For example, if a political candidate is expected to take a certain position with regard to an issue, but ends up advocating another position, their credibility rises.

Since it is a likely assumption that users expect a commercial recommender system to overestimate the value of an item, underestimation disconfirms this expectation and might cause users to find a recommender system less biased and more trustworthy. Two users stated expectations on an emphasis on high ratings in qualitative comments: *"I would expect the web to present items at their best and sometimes with some exaggeration."*, *"I expect there to be hype about a movie and to have to read between the lines to form a judgment for myself."*

The effect of gap type was surprising, we also were surprised to find that negative gaps were considered least helpful, and positive gaps most helpful, for *both* over and underestimation. This may reflect the way users distribute and assign ratings. The polar ratings of 1's and 5's are more uncommon and differently distributed from the other ratings, i.e. the 'distance' between 2 and 3 may be perceived as smaller than the distance between 2 and 1. So a user is much less likely to buy an item rated 1 rather than 2. Likewise, the probability of a user trying an item increases more between 4 and 5 than it does between 3 and 4. The lack of significant results for overestimation might be attributed to users' general expectation of overestimation in commercial recommender systems.

User comments also revealed some other interesting views on product categories. Two users left comments where they differentiate between holidays and the other products:
"Things like 'Holidays' matter more compared to goods, because holiday is a destination could be once in a life time thing.", *"A holiday is an experience of value that cannot be replaced or compensated for, knowledge should be accurate."*. One user found it difficult to imagine using a recommender system to buy light bulbs: *"I can't imagine going on to a web site to look for information on a light bulb!"*.

3 Reflections on the experimental setup

When considering the design of our experiment, two criticisms can be raised. In this section, we discuss what these criticisms are, and why we decided to perform the experiment in this particular way.

3.1 Why the wording for underestimation differs

In the scenario for overestimation the user changes their value judgment by experiencing the product directly. In contrast, in the underestimation scenario, the user changes their value judgment based on comments from a friend who experienced the product. So, why did we not let the user "experience" the product directly in the latter case, as this would have made the conditions more comparable? As the user did not buy the product, it was hard to devise a plausible story of how they ended up experiencing it after all. If somebody else bought it for them as a gift, the user is not likely to regret missing the item, and thus will not harbor feelings of resentment over poor information to the same degree. Experiencing the item by borrowing it from a friend is not possible for all domains (e.g. holidays).

3.2 Why the experiment is indirect

Instead of participants really experiencing the products, we only told them about their experience. What participants think they would do in such a situation may diverge from what they really would do [1]. We were however working on the basis of these assumptions:

- *Gap size matters.* Participants' perceived Effectiveness will depend on the size of the discrepancy between their first impression and their valuation after experiencing the item.
- *Gap position matters.* The influence of a skew will depend on the gap's position. For example, an under-estimation from 1 (first rating) to 3 (final valuation) may have a different effect than one from 3 to 5. Evidence for this was found in our experiment.

Given these assumptions, for a fair comparison between domains (H2, H3) we need to control for gap size and position. Practically, this would mean that participant's valuations (before and after) need to similarly distributed for all products. This would be very hard (if not impossible) to control rigorously. Even making the experiment a little more realistic, by giving participants particular information to form a first opinion, and then more information to form a final valuation, would be hard to control. Attempts in our earlier work to construct item descriptions with predictable ratings for all participants failed [9].

For a fair comparison between over- and underestimation (H1), we also need the gap size and position to be the same [2]. Suppose we knew that people *on average* like a particular item, and disliked another item. This may be hard to obtain in certain product domains, or limit us to a small subset of items where people converge on valuation. This is also likely to require a separate study to decide on suitable items. The estimated valuation allows us to know, on average, the real valuation (and in analysis, we would need to remove all subjects whose valuation differed from this average). We would still have to make the explanations such that they induce the right initial rating (namely the valuation for the liked item in the disliked item's case, and the other way around). Given that we also wanted to study gap types (H4), we would need multiple of these item pairs plus explanations per domain.

This does not mean that we will not do more direct experiments in the future. It is just that given the factors we wanted to investigate here, there were very clear benefits in doing an indirect experiment.

4 Conclusions

H1 is confirmed: overestimation is considered less helpful by users than underestimation. H2 is partially confirmed: overestimation is considered less helpful in high investment domains than in low investment domains. Underestimation in high investment domains is not considered significantly less helpful, even if there is a trend in this direction. The lack of significant result may be due to underestimation having a stronger effect on perceived Effectiveness. H3 is not confirmed, only a trend suggests that some users may be more critical in objective than subjective domains. H4 is disconfirmed: cross-gaps are not considered the least helpful by users, negative gaps are, for both over- and underestimation. For overestimation, positive gaps are not considered less helpful than cross-over gaps.

As mentioned in Section 1.2.1, recommendations can be skewed for a variety of reasons. The results of this study would be relevant for algorithmic correction as well as studies comparing different presentational interfaces. Understanding the role of factors such as gap type, domain type and over and underestimation will help better control for these factors when optimizing a recommender system for Effectiveness.

In light of our results we suggest an enhancement to the Effectiveness metric proposed by [2] and described in Section 1.1. We propose fine tuning this measure of Effectiveness by weighting it according to gap type, over/underestimation and degree of investment.

[2] We consider the gap '1 to 3' to be comparable to the gap '3 to 1' w.r.t. to position

REFERENCES

[1] Icek Ajzen and Martin Fishbein, 'Attitude-behavior relations: A theoretical analysis and review of empirical research.', *Psychological Bulletin*, **84(5)**, 888–918, (1977).

[2] Mustafa Bilgic and Raymond J. Mooney, 'Explaining recommendations: Satisfaction vs. promotion', in *Proceedings of the Beyond Personalization Workshop in association with IUI*, pp. 13–18, (2005).

[3] Yooncheong Cho, Il Im, and Jerry Fjermestad Starr Roxanne Hiltz, 'The impact of product category on customer dissatisfaction in cyberspace', *Business Process Managment Journal*, **9 (5)**, 635–651, (2003).

[4] Dan Cosley, Shyong K. Lam, Istvan Albert, Joseph A. Konstan., and John Riedl, 'Is seeing believing?: how recommender system interfaces affect users' opinions', in *Proceedings of the SIGCHI conference on Human factors in computing systems*, volume 1 of *Recommender systems and social computing*, pp. 585–592, (2003).

[5] Jonathan L. Herlocker, Joseph A. Konstan, and John Riedl, 'Explaining collaborative filtering recommendations', in *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, pp. 241–250, (2000).

[6] Thorsten Joachims, Laura Granka, and Bing Pan, 'Accurately interpreting clickthrough data as implicit feedback', in *ACM SIGIR conference on Research and development in information retrieval*, pp. 154–161, (2005).

[7] S.-M. Kim, P. Pantel, T. Chklovski, and M. Pennacchiotti, 'Automatically assessing review helpfulness', in *Empirical Methods in Natural Language Processing (EMNLP)*, pp. 423–430, (2006).

[8] David N. Laband, 'An objective measure of search versus experience goods', *Economic Inquiry*, **29 (3)**, 497–509, (1991).

[9] Judith Masthoff, 'Group modeling: Selecting a sequence of television items to suit a group of viewers', *User Modeling and User Adapted Interaction*, **14**, 37–85, (2004).

[10] David Mcsherry, 'Explanation in recommender systems', *Artificial Intelligence Review*, **24(2)**, 179 – 197, (2005).

[11] Patrick E. Murphy and Ben M. Enis, 'Classifying products strategically', *Journal of Marketing*, **50**, 24–42, (1986).

[12] Hien Nguyen and Judith Masthoff, 'Using digital images to enhance the credibility of information', in *Persuasive Technology symposium in association with the Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB)*, pp. 1–8, (2008).

[13] John O'Donovan and Barry Smyth, 'Eliciting trust values from recommendation errors', in *International Journal of Artificial Intelligence Tools (IJAIT)*, (2006).

[14] Pearl Pu and Li Chen, 'Trust building with explanation interfaces', in *International conference on Intelligent user interfaces*, Recommendations I, pp. 93–100, (2006).

[15] James Reilly, Kevin McCarthy, Lorraine McGinty, and Barry Smyth, 'Dynamic critiquing', in *European Conference on Case-Based Reasoning (ECCBR)*, volume 3155 of *Lecture Notes in Computer Science*, pp. 763–777, (2004).

[16] Carl Shapiro, 'Optimal pricing of experience goods', *The Bell Journal of Economics*, **14 (2)**, 497–507, (1983).

[17] James B. Stiff, *Persuasive Communication*, chapter 5, 94–98, Guilford Press, 1994.

[18] Nava Tintarev, 'Explanations of recommendations', in *ACM Recommender Systems*, pp. 203–206, (2007).

[19] Nava Tintarev and Judith Masthoff, 'A survey of explanations in recommender systems', in *WPRSIUI associated with ICDE'07*, pp. 801–810, (2007).

A Example questionnaire - underestimation

Experiment on product information

Age: _ _ _ _ Gender: M/F (please circle the one that applies)

All data gathered in this study will be treated confidentially, anonymized, and will only be used for the purpose of the research.

Assume you are on a website looking for a particular product to buy (such as a camera, holiday, light bulb, movie). Based on the information given, you form an opinion of the product, and decide not to buy it and to spend the money on something else. Later you talk to a friend who used the product, and your opinion changes.

Consider the following scenarios, and indicate how your experience in each case effects your perception of that particular website. Note: each scenario is about a different website, even for similar products.

Product	**Your opinion of the product based on info on the website**(1 to 5 scale with 1 being really poor and 5 really good)	**Your opinion of the product after talking to your friend** (1 to 5 scale with 1 being really poor and 5 really good)	**How do you rate the information on this website given this experience?**						
			Very unhelpful						Very helpful
Camera	3	5	1	2	3	4	5	6	7
Holiday	1	3	1	2	3	4	5	6	7
Light bulb	2	4	1	2	3	4	5	6	7
Movie	1	3	1	2	3	4	5	6	7
Camera	2	4	1	2	3	4	5	6	7
Holiday	3	5	1	2	3	4	5	6	7
Light bulb	1	3	1	2	3	4	5	6	7
Movie	3	5	1	2	3	4	5	6	7
Camera	1	3	1	2	3	4	5	6	7
Holiday	2	4	1	2	3	4	5	6	7
Light bulb	3	5	1	2	3	4	5	6	7
Movie	2	4	1	2	3	4	5	6	7

Would you like to explain your answers? Please do this here:

Thank you for your participation! If you would like to know more about this study, or receive a summary of the results please contact me at n.tintare@abdn.ac.uk