

Investigating Explanations to Justify Choice

Ingrid Nunes^{1,2}, Simon Miles², Michael Luck², and Carlos J.P. de Lucena¹

¹ Pontifical Catholic University of Rio de Janeiro - Rio de Janeiro, Brazil
{ionunes, lucena}@inf.puc-rio.br

² King's College London, Strand, London, WC2R 2LS, United Kingdom
{simon.miles, michael.luck}@kcl.ac.uk

Abstract. Many different forms of explanation have been proposed for justifying decisions made by automated systems. However, there is no consensus on what constitutes a *good* explanation, or what information these explanations should include. In this paper, we present the results of a study into how people justify their decisions. Analysis of our results allowed us to extract the forms of explanation adopted by users to justify choices, and the situations in which these forms are used. The analysis led to the development of guidelines and patterns for explanations to be generated by automated decision systems. This paper presents the study, its results, and the guidelines and patterns we derived.

Keywords: User Explanation, Guidelines, Patterns, Recommender Systems.

1 Introduction

The popularity of recommender systems has increased significantly in the last decade, with many commercial applications already adopting them. For many years, the main goal of research into such systems has been to improve their *accuracy*, associating this measure with the quality of the recommendation. However, as argued by McNee et. al. [6], the most accurate systems (based on standard metrics) may not be those that provide the most useful recommendations to users. Other aspects, such as trust and transparency, have also been considered, and many of these can be improved by providing users with *explanations* [12]. Such explanations justify the choice of a particular recommendation to users, and their applicability extends to decision support systems [4] and over-constrained problem solvers [3].

There are different existing approaches to generating explanations, from exposing the rationale of the underlying recommendation technique to selecting the essential attributes on which the decision is based. However, there is no consensus on what constitutes a *good* explanation, and what kinds of information must be presented to users in such explanations. Even though existing work [12] provides qualitative arguments that characterise good explanations, there is no extensive research into the kinds of explanation that users expect and need to understand and accept recommendations or decisions made on their behalf and, where work does exist, it is particular to a specific system.

In response, this paper presents a study whose main objective is to give guidance for explanation generation. The study performed consisted of a survey, from whose results we extract types of explanation that people use to justify a choice from a set of available

(a) Research Questions.

RQ1. Do users adopt a pattern to justify an option chosen from the set of those available?
RQ2. Is there a relationship between the type of explanation given to support the decision and the chosen option?
RQ3. Do users use a pattern to justify the rejected (not chosen) options?
RQ4. Is there a relationship between the type of explanation given to reject options and the rejected or chosen option?

(b) Evaluation Approaches.

EA1. Analysis of the arguments given to justify the chosen option and identification of commonalities among arguments given by different users.
EA2. Comparison among the arguments given to justify each different chosen option.
EA3. Analysis of the arguments given to reject options and identification of commonalities among arguments given by different users.
EA4. Comparison among arguments given to reject options according to each different chosen and rejected option.

Table 1. Research questions and their evaluation approach.

options. As, based on the design of the study, we can assume that the explanations provided by study participants are those that the users would expect to receive, we derive a set of *guidelines* and *patterns*, which are a basis for generating explanations for users as to why particular options are chosen by a recommender system or decision support systems. Therefore, this paper presents three contributions: (i) the design and results of a study into what explanations users expect when justifying choices made; (ii) guidelines for the qualities and forms of explanation needed to best meet user expectations; and (iii) patterns for explanations to be given under different circumstances. The aim of such explanations is to expose to users why a system chose a particular option, thus improving user *satisfaction* and *trust* in the decision.

2 Description

As outlined above, our goal is to identify explanations to be provided to users by recommender systems or automated decision making systems. This section describes our study, including the research questions, procedure and participants. We addressed four research questions, presented in Table 1(a). By answering these questions, we are able to extract patterns for user explanations to be generated by the relevant systems (*RQ1* and *RQ3*), and also the context in which each pattern is adopted (*RQ2* and *RQ4*). These explanations are associated with both chosen and rejected options — the first two questions focus on patterns and their context for explaining the chosen option; and the last two focus on determining why other options were rejected (or not chosen).

2.1 Procedure

Our study consists of collecting information provided by participants through a web-based questionnaire, and its analysis. Our aim is to obtain a high number of participants, so anyone with Internet access could access the questionnaire. The study concerns decision-making and explanation of the decisions made. The choices to be made were between hotels in New York city, a domain chosen because most people are aware of the attributes that characterise hotels; they will have preferences over individual attributes of hotels; New York is a widely known tourist destination, so participants are more likely to have a *known* set of preferences [5]; we have relevant knowledge of the

city, so were able to select appropriate options; and there are many New York hotels with available *real data*. The questionnaire was divided into three parts, as follows.

User data. Our study does not investigate if explanations depend on people’s characteristics, such as age or gender. However, we collected some information about the participants in case it was informative, and also to be able to provide demographic information of the participants as a group. The collected participant data is: (i) age; (ii) gender; (iii) location (city and country); and (iv) field of work or study.

Product choice. Participants are requested to imagine the scenario in which they go to New York on vacation, and must choose a hotel for staying there from a set of options. Hotels are described in terms of attributes associated with hotels and their rooms available at the `booking.com` website, presented in a table that allows a side-by-side comparison. We took five existing hotels — *Hotel 91*, *Econo Lodge*, *The Hotel at Times Square*, *Comfort Inn*, *Renaissance*, viewing these options as forming three groups (not known to participants), below.

G-1 Dominated option. Although a dominated option (one that has no advantage and at least one disadvantage with respect to another) is generally not chosen, we add such an option (or at least something close to it) to capture arguments used to reject them. If we ignore small differences in room size, and discount parking price (which typically does not appear in catalogues of features), we can identify one hotel (*Comfort Inn*) dominated by another (*The Hotel at Times Square*) even though *Comfort Inn* actually has better parking price and a slightly better room size than *The Hotel at Times Square*. The assumption (subsequently confirmed by our study) is that most participants focus on the main attributes and ignore small differences, so that *Comfort Inn* is dominated.

G-2 Extreme options. Extreme options compromise one attribute, e.g. quality, too much in order to improve another, e.g. price. In general, people avoid such options, known as the extremeness aversion principle [11]. We select two extreme options: (i) much lower quality and much lower price (*Hotel 91*); and (ii) much higher quality and much higher price (*Renaissance*).

G-3 Options with Trade-off. Two options that have relative pros and cons require a trade-off to be made. As this may require a different form of explanation from either category above, we include options that clearly illustrate such a need for trade-off, *Econo Lodge* and *The Hotel at Times Square*.

Reasons for choice. The participant is asked to state why they choose a particular option, and why they reject the remaining options — we assume that if participants do not choose an option, they automatically reject it. In order to obtain useful responses, we highlight for the participant that *complete* answers should be provided and that arguments should be sufficiently strong to convince another person about the choice made.

In all this, the most important information collected is the provided justifications, expressed in natural language. The analysis part of the study consists of carefully investigating these justifications to identify patterns and define explanation types so that, based on this initial analysis, we can extract quantitative data. Table 1(b) shows our approach to answering our research questions, which is mainly based on a classification of explanation types. In summary, the collected *qualitative* data is: justifications for acceptance, justifications for rejection, explanation types, and additional characteristics of justifications. The *quantitative* data consists of: chosen hotel, chosen hotel vs. ex-

Gender	Male	Country	Brazil	United Kingdom	Age (years)	16-25	26-35	Field of Work or Study	Informatics	Education
	58 (58%)		78 (78%)	8 (8%)		4 (4%)	61 (61%)		36-45	>45
Female		Canada	Other		11 (11%)	24 (24%)		Management	Other	
42 (42%)		5 (5%)	9 (9%)					7 (7%)	28 (28%)	

Table 2. Demographic Characteristics of Participants.

planation types for acceptance, chosen hotel vs. explanation types for rejection of other hotels, rejected hotels vs. explanation types for their rejection.

2.2 Participants

The participants in our survey are selected using convenience sampling, obtained through the social network of the researchers involved in this study, by means of two forms of publishing the survey: (i) by e-mail, using the contact list of the researcher; and (ii) by Facebook (<http://www.facebook.com>), the widely known social network. The distributed message consists of an invitation to participate in the survey and a request to forward the invitation for others. The survey was available for participation from 12th to 24th October, 2011 and was completed by 100 people. The demographic characteristics of the participants that completed the survey are described in Table 2. Because we adopted the social network of the lead researcher to perform the study, most participants are aged between 26 and 35 years (61%) and are Brazilians (78%).

3 Results and Analysis

Our collected data consists mainly of justifications expressed in natural language and, as these are qualitative data, we analyse them in a systematic way to extract quantitative information. This section explains how we perform this analysis and provides the results obtained, according to the research questions we aim to answer. Note that, at various points, we label some findings with “*Evidence X*,” in order that we can later refer to them to support our proposed guidelines.

Before proceeding, however, we enumerate the hotels chosen by our participants: *Hotel 91* (18%), *Econo Lodge* (52%), *The Hotel at Times Square* (19%), *Comfort Inn* (7%) and *Renaissance* (4%). As expected, the majority of participants choose a hotel from group G-3. This information is relevant to understanding the relationship between the chosen option and justifications, as indicated by research questions *RQ2* and *RQ4*.

RQ1: Do users adopt a pattern to justify an option chosen from the set of those available? Each participant has to provide five justifications for their choice, from which one explains why they choose a particular hotel. With the analysis of all provided justifications and the principles of grounded theory [2], we derive a classification, which we refer to as *explanation types*, consisting of *six* different types that are described below. We illustrate each of these explanation types for the acceptance scenario in Table 3.

Critical attribute. For some participants, a single attribute plays a crucial role in the decision-making process, *price* in most cases. In these situations, the justification focuses only on this crucial attribute, and the remaining ones are omitted. The same attribute is used to justify the chosen and all rejected options.

Explanation Type	Example of Justification for Acceptance	Example of Justification for Rejection
Critical attribute	H_i is the cheapest option.	There are other options cheaper than H_i .
Dominance	H_i is better in all aspects.	There is no reason for choosing H_i , as it is worse in all aspects than H_j .
Main reason	I chose H_i because it offers the benefit a_i .	I did not choose H_i because it does not offer the benefit a_i .
Minimum requirements	From the hotels that satisfy my requirements, H_i is the cheapest.	H_i is too expensive.
One-sided Reasons	I chose H_i because it provides the benefits a_i and a_j .	I did not choose H_i because it has the disadvantages a_i and a_j .
Pros and Cons	Even though H_i is not the cheapest, it provides the benefits a_i and a_j .	Even though H_i provides the benefits a_i and a_j , its price does not compensate it.

Table 3. Example of Justification for Acceptance and Rejection.

Dominance. The domination relationship can be used as an argument to justify a decision, but the acceptance of an option is justified using dominance only when it dominates all other options. This is an uncommon situation when choosing among products because, due to seller competition, there is typically a trade-off to be resolved, with options presenting both pros and cons. However, if domination *does* arise, the decision is extremely easy: one option may dominate another from a particular participant’s perspective, as they might not care about a set of attributes, and the remaining ones create this ideal scenario to make the decision.

Main reason. Some participants take into account many attributes to make a decision, but a particular option may be chosen (or rejected) when there is one attribute value that, together with its importance, is decisive for the choice. This most important attribute is specific to each option.

Minimum requirements. People usually have hard constraints, used to filter available options by discarding those that do not satisfy all of them — this can be seen as the establishment of cut-off values. If only one option satisfies all requirements, the decision becomes easy as the justification for option acceptance is that it satisfies all requirements. Furthermore, some participants provide a justification based on minimum requirements but, since more than one option satisfies these requirements, the participants also provide some criterion to distinguish between them, e.g. minimum price.

One-sided reasons. Instead of only providing the main reason for acceptance, many participants focus on exposing only positive aspects (or negative, in case of rejection) of the option, even though the chosen option has disadvantages (or advantages) with respect to other options in relation to their preferences. This indicates the existence of a minimal set of attributes that caused the option to be chosen (or rejected).

Pros and cons. The most complex type of explanation consists of making the option pros and cons explicit, and showing the reasoning process behind the choice. Based on an evaluation of these pros and cons, the participant states that the pros compensate for the cons (or do not, in case of rejection). In some cases, participants do not enumerate pros and cons, but only state “*this is (not) the best cost-benefit relationship.*”

These explanation types indicate that justifications for choosing an option *do* follow patterns, and these can be used in systems for explanation generation. The right hand side of Figure 1 (which shows the explanation types used to justify each hotel) represents the total number of the different explanation types adopted by the participants, who mostly adopt *one-sided reasons* and *pros and cons* to explain their choices.

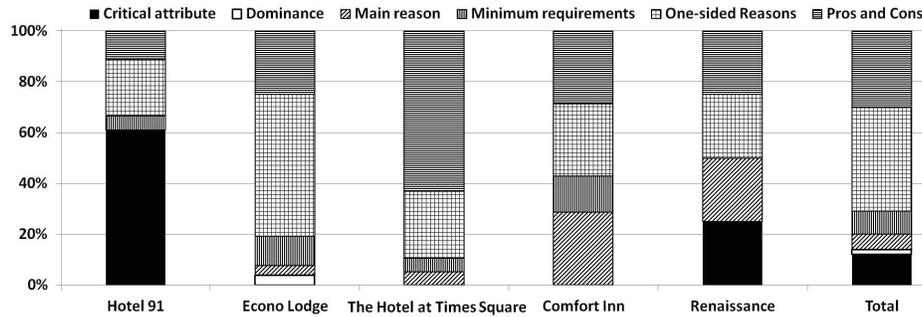


Fig. 1. Explanation types used to justify each chosen hotel.

RQ2: Is there a relationship between the type of explanation given to support the decision and the chosen option? Given that we have identified patterns used to justify why a particular hotel is chosen, we now investigate if there is any relationship between the type of explanation given and the chosen option. Figure 1 shows how much each explanation type is adopted for each individual hotel.

The distribution of explanation types indicates three norms. First, most of the participants that choose *Hotel 91* (61.11%) justify their decision by referring to a critical attribute, price, indicating that what matters for them is that this hotel is the cheapest. Some participants provide further positive information about the hotel (*one-sided reasons*, 22.22%), in addition to stating that it is cheapest, i.e. they indicate that even though the hotel is the cheapest, the quality that they require is not compromised.

As can be seen in Figure 1, the main adopted explanation types for choosing hotels of the G-3 group are *one-sided reasons* and *pros and cons*, which together has a total of 80.77% for *Econo Lodge* and 89.47% for *The Hotel at Times Square*. The first explanation type is used to show that a whole set of hotel characteristics is responsible for the choice made. In general, participants that choose *Econo Lodge* exclude the cheapest hotel from the set of hotels being considered in the decision, and explain the benefits of this hotel to show that it was suitable for them; i.e. there is no reason to pay more for another option if this hotel already provides what the participant wants. Conversely, participants that choose *The Hotel at Times Square* make a detailed analysis of this hotel against *Econo Lodge*; i.e. they discuss the *pros and cons*, and show that the higher price of the former justifies the benefits it provides, when compared against the latter. With respect to these two options, we make one last observation: two participants (3.85%) use dominance to justify why they choose *Econo Lodge*, and ignore attributes that are not relevant for them, creating a scenario in which this hotel dominates all others.

Finally, we discuss the results obtained for the dominated option and the most expensive option. It can be seen that there is no explanation type that is most adopted, with participants adopting different explanation types for justifying them. Few participants choose these two options and, since it is not obvious why these options should be chosen, the participants give their particular explanations to justify this decision. In the first case, *Comfort Inn*, some participants are vague and say that they choose this hotel because it has the best cost-benefit relationship, but do not give details. The remaining

participants use as arguments the two attributes that this hotel is better than *The Hotel at Times Square*, i.e. parking price and room size. The room size argument is also used as an expression of *intuition*: as the room is bigger, and the price is higher, the hotel “apparently” provides more comfort. For this same reason, some participants choose the 4-star *Renaissance*, as comfort is the most important issue for them, and they are not concerned with price.

RQ3: Do users use a pattern to justify the rejected (not chosen) options? By analysing justifications for rejecting options, we have observed the same explanation types used for justifying the chosen option. The description given for our set of explanation types shows that they can also be applied to reject options. In Table 3, we show examples of how each of these explanation types is used in the context of option rejection.

RQ4: Is there a relationship between the type of explanation given to reject options and the rejected or chosen option? In order to understand how participants choose a particular explanation type, we analyse the relationship between the types adopted to justify rejected options from two perspectives. The first consists of analysing justifications for rejection given for each hotel (Figure 2). The second perspective groups justifications according to the chosen hotel; i.e. we observe which explanation types are adopted to reject other options according to a particular chosen hotel (Figure 3).

Many interesting aspects can be observed. *Critical attribute* is the type of explanation used when the decision is guided by it. For instance, if the participant wants to minimise price, the justification for the rejected hotels is that they are more expensive (than the chosen hotel). Similarly, this situation occurs with the more expensive hotel, when the participant wants to maximise the price (as a proxy for comfort maximisation). *Dominance*, on the other hand, is adopted when the chosen option dominates the rejected option; i.e. the comparison made in the explanation is always for the chosen option against the others. In many situations, preferences (hidden in justifications) of participants who choose *Econo Lodge* indicate that *The Hotel at Times Square* dominates *Comfort Inn*; however, this is not given as an argument to discard the latter, but the participants seek an explanation why *Econo Lodge* is better than *Comfort Inn* (*Evidence A*). Some participants have hard constraints that they require to be satisfied by the chosen hotel, such as a maximum price that they are willing to pay, or a minimum distance from the city centre. In these situations, an option is rejected regardless of the remaining options, and the justification given is that the option does not satisfy the participant’s *minimum requirements*.

Main reason and *one-sided reasons* indicate that there is an attribute (or a set of them) that is especially important for the participant and, even though it is not part of a hard constraint, it plays a *decisive role* in the decision; i.e. because of this (these) attribute(s), the option is rejected. This set of attributes is *kept as simple as possible* (*Evidence B*); e.g. some participants that choose *Econo Lodge*, reject *The Hotel at Times Square* and *Comfort Inn* because they do not have a refrigerator and are more expensive (than the chosen hotel). But, to justify the *Renaissance* (which also does not have a refrigerator), they argue only that it is more expensive. It is important to note that the explanations given for *The Hotel at Times Square* and *Comfort Inn* are exactly the same, and there are many other cases in which the same explanation is given for different options rejected for the same reason (*Evidence C*). Finally, *pros and cons* are given as

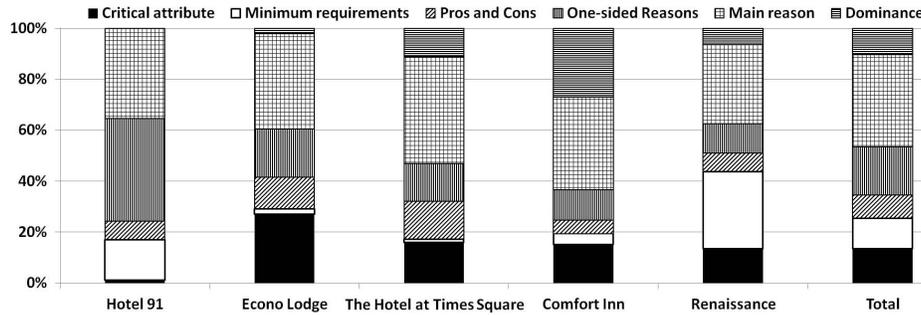


Fig. 2. Explanation types used to justify the rejection of each hotel.

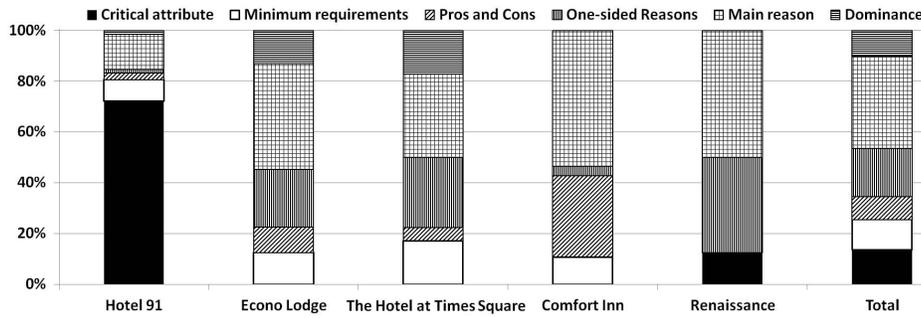


Fig. 3. Explanation types used to justify the rejection of other hotels given a chosen hotel.

rejection arguments by participants when the decision between two (or three) options is difficult, so they expose these options' pros and cons to show that the chosen option has the best cost-benefit relationship. Thus, *pros and cons* are used only in the *absence of a decisive subset of attributes (Evidence D)*.

In this way, the justification given for rejecting an option depends on both the chosen and rejected options, as the explanation given typically justifies why the rejected option is worse than the chosen one. Only in those cases in which the option is rejected due to a hard constraint (*minimum requirements*), the rejection explanation depends only on the option being rejected.

Further Observations. While analysing the collected data, we also identify other relevant characteristics present in the provided justifications. We describe each of these characteristics below, most of which can be used to suggest informal arguments for systematic approaches to decision-making.

Explicit trade-off. Some participants (34%) state that the chosen hotel has the best cost-benefit relationship (or not the best, for rejecting a hotel), and sometimes just provide the argument without any details; e.g. "For a trip like this, it seems the best cost-benefit among the 3-star hotels."

Preferences mentioned. Only a few participants (14%), when requested to justify their decision, provide arguments based on their preferences (*Evidence E*); for example,

a participant argued the “*absence of a fitness centre*” to justify a rejection, but this is due to the participant’s preference for a hotel *with* a fitness centre — and in some cases, participants make their preference explicit.

Price as a first class attribute. The majority of participants (92%) mention the attribute *price* in their justifications, and evaluate options by comparing this attribute with all others. This indicates that *cost* (which can also be time, effort, etc.) is not seen as a disadvantage of an option when compared to another, but as a fixed attribute that should be treated differently in the provided explanations (*Evidence F*).

Irrelevant attributes. When participants choose a hotel that does not offer as many benefits as the others, they state that those benefits are not important for them and, as a result, there is no reason to pay more for something that will not be used. Irrelevant attributes are mentioned in both acceptance and rejection justifications (34%).

These observations show that cost should be treated as a first class attribute in explanations, as it is a crucial factor considered in the decision. In the cases when a higher price is chosen, and this difference is very small, many participants acknowledge this fact. When the chosen option has a lower price, the benefits provided by other options may be relevant to be mentioned, even though the decision maker does not care about it. In cases where the pros and cons of a set of options make the decision hard, an explicit statement that a particular option has the best cost-benefit relationship might be helpful. Finally, participants typically do not support their arguments with their preferences.

4 Guidelines and Patterns

This study provides us with a means of understanding how users construct arguments to justify a choice, by explaining why an option is chosen and why the remaining ones are rejected. Moreover, based on the results from this study, we are able to contribute to our ultimate goal of providing guidance that serves as a basis for the development of explanation approaches. To this end, we introduce guidelines and patterns derived from our study in this section. For each guideline, we indicate the evidence that supports it.

4.1 Guidelines

1. Provide chosen-option-oriented explanations. (*Evidence A*) The explanation generation process must be guided by a previously chosen option. The goal of the explanation is not to expose all the reasoning process used to make the decision, but to provide the main arguments that justify a chosen option and reject the remaining ones. After the choice is made, the explanations given should answer two main questions: (i) what makes the chosen option better than the others; and (ii) what makes other options worse than the chosen option.

2. Keep it simple. (*Evidence B*) The explanation given to a user should be as simple as possible, even justifying the decision with a single sentence; e.g. *A is the cheapest option*. Therefore, the less complex the explanation, the better. The next three guidelines are associated with this, and provide concrete ways of keeping the explanation simple.

3. Focus on the most relevant criteria. (*Evidence D*) In the given explanation, only the *decisive criteria* should be mentioned; i.e. the minimum set of attributes that causes

Decisive Criteria Pattern
<p>Classification: <i>Explanation goal:</i> both; <i>Target:</i> option; <i>Position:</i> absolute.</p> <p>Context: even though there are other attributes that contribute for the option acceptance (or rejection), there is a subset of them that would confirm this decision regardless of the values of the other attributes.</p> <p>Template: <i>Option option was [chosen rejected] because of its set of decisive attributes .</i></p> <p>Description: options, when compared, might have different pros and cons. However, some attributes are the most decisive in the decision (according to their value and importance), while others — which can make a difference in particular cases — do not impact on the decision between two options. Therefore, the only attributes that must be part of the explanation are those that impact on the decision, leaving aside remainder.</p> <p>Example: three hotel options are given for a user: (i) hotel <i>A</i> is a 3-star hotel, cheaper than the other two options and has a refrigerator in the room; (ii) hotel <i>B</i> is also a 3-star hotel, more expensive than the former, with a better location; and (iii) hotel <i>C</i> is a luxury 4-star hotel, much more expensive than the others and, like hotel <i>B</i>, does not have a refrigerator. While the rejection of <i>B</i> is justified by the absence of the refrigerator <i>and</i> its price; the rejection of <i>C</i> is justified only because of its price, as this is the decisive criterion for not choosing it.</p> <p>Back up preference: preferences over the set of decisive attributes.</p>

Table 4. Decisive Criteria Pattern.

an option to be selected or rejected. These decisive criteria should be derived from comparison of the chosen option against the others.

4. Group similar options. (*Evidence C*) An explanation to reject an option can also be given to reject other options. So, rejected options should be grouped when they are rejected for the same reason, and presented as a group and not individually.

5. Back up explanations with preferences, but provide them only if asked. (*Evidence E*) Characteristics mentioned in explanations are relevant, because of the preferences being considered in the decision-making process; e.g., “I chose this option as it is the cheapest,” (explanation) and “I want to minimise costs” (preference). People usually do not explicitly state their preferences to justify their decisions but, if a decision is made on someone’s behalf, it is fundamental to back up an explanation with their preferences. As this information is not always needed, and as simpler explanations are better, preferences must be provided as part of explanations only upon request.

6. Use cost as a first class attribute. (*Evidence F*) An option is chosen by an individual when they believe that the cost being paid for that option compensates for the benefits it provides. The trade-off between benefits and costs is the key issue in the process of decision-making, so the option attributes that define the option costs should be made explicit and used as first class attributes in the explanation provided to justify a decision.

4.2 Patterns

Based on our study, we derived patterns of explanations, which can be used for supporting a decision made by a software system. Moreover, we identified the components these patterns must have, which comprise a template for an explanation pattern catalog. These components are: (i) a *classification*; (ii) a *context* in which the pattern should be applied; (iii) a *template* for the explanation; (iv) the pattern *description*; (v) an *example*; (vi) *preferences* that back up the explanation; and (vii) optionally, *extensions* to the pattern. Patterns are classified (item (i)) according to three attributes, explained below.

- *Goal: accept/reject/both.* An explanation can justify a chosen option (*accept*), a rejected or not chosen option (*reject*), or both (*both*).

- *Target: decision/option.* A pattern can provide guidance to justify the decision as a whole (*decision*), or the acceptance or rejection of a single option (*option*).

Pattern	Goal	Target	Position	Template
Critical Attribute	both	decision		Option <u>chosen option</u> was chosen because it has the best value for <u>critical attribute</u> .
Cut-off	reject	option	absolute	Option <u>rejected option</u> was rejected because it does not satisfy constraints associated with <u>attribute</u> .
Domination	reject	option	relative	There is no reason to choose option <u>rejected option</u> , as option <u>chosen option</u> is better than it in all aspects, including <u>cost</u> .
Minimum Requirements ⁻	reject	option	relative	Even though option <u>rejected option</u> satisfies all your requirements, it has a worse value for <u>attribute</u> than option <u>chosen option</u> .
Minimum Requirements ⁺	accept	option	absolute	Besides satisfying all your requirements, option <u>chosen option</u> has the best value for <u>attribute</u> .
Decisive Criteria	both	option	absolute	Option <u>option</u> was [<u>chosen</u> <u>rejected</u>] because of its set of decisive attributes .
Trade-off Resolution	both	option	absolute	Even though option <u>rejected option</u> provides better <u>pros</u> than the chosen option, it has worse <u>cons</u> . Even though option <u>chosen option</u> does not have the best value for <u>cons</u> , its values for <u>pros</u> compensate its cons.

Table 5. Explanation Pattern Classification.

- *Position: absolute/relative.* When a pattern target is *option*, the explanation given can be based solely on the target option (*absolute*), or make a statement that explicitly compares the option to another (*relative*).

Due to space restrictions, we do not describe each pattern, but present just one as illustration in Table 4. We also summarise all the patterns in Table 5 (complete description available elsewhere [8]), in which patterns are ordered according to their complexity; i.e. the simpler the explanation associated with a pattern, the earlier it is presented. According to our second guideline, the explanation should be as simple as possible so, if two patterns can be used in a particular situation, the simplest must be applied.

5 Related Work

Recommender systems have different aims that should be achieved beyond accuracy, such as trust, effectiveness and satisfaction, and these have been addressed through explanations that expose the rationale behind the adopted recommendation approach (content-based, collaborative, or hybrid). For example, if a collaborative approach is adopted, the user may receive as an explanation a histogram of ratings of the product given by similar users. McSherry [7] focused on case-based reasoning approaches, in which products are seen as cases from which one should be selected when it is similar to the case provided by the user, and the explanation is based on selected similar cases. Another direction is explanation interfaces [10], which organise recommended products in a way that causes trade-off situations to be resolved explicitly for users, thus facilitating the decision making process.

Even though explanations improve recommender systems, they currently focus on explaining the *means* used to obtain recommendations, but stating that “someone like you chose this product” or “you like similar products” is not sufficient for justifying a recommendation and for users to accept it. This can be seen in a taxonomy recently proposed for *classifying explanation generation approaches* [1], whose dimensions focus on the information used to generate the explanation and the underlying recommendation model. Characteristics of *good* explanations have been defined based on the analysis

of existing approaches [12], and these can be used as metrics to *evaluate existing approaches*. Our work, on the other hand, identifies *good* explanations that should be given to users, which can be used as guidance for elaborating new explanation approaches, and proposes a template *for classifying explanations*. The challenge of obtaining these explanations from existing recommendation approaches, however, still remains. As our patterns indicate explanations based on option attributes, it may be more straightforward to generate such explanations from approaches based on the relevance of attribute values for users.

6 Conclusion

In this paper, we have presented a study performed to understand how people justify their decisions, by giving explanations why they choose a particular option from the set of those available, and why remaining options are rejected. The study consisted of providing participants (100 people) with a set of carefully chosen hotel options, and requesting them to give reasons for the choice. Based on collected data, we have identified explanation types that are patterns of justifications given by people, and how they are selected to be given as explanation — for both chosen and rejected options. Assuming that explanations given by people are the explanations that users expect to receive as reasons for a choice, our study allowed us to propose a set of guidelines and patterns for the development of explanation approaches. Future work involves producing explanations for choices made by our decision-making technique [9], which takes into consideration this guidance derived from our study.

References

1. Friedrich, G., Zanker, M.: A taxonomy for generating explanations in recommender systems. *AI Magazine* 32(3), 90–98 (2011)
2. Glaser, B.G.: *Emergence vs. Forcing: Basics of Grounded Theory Analysis*. Sociology Press, Mill Valley, USA (1992)
3. Junker, U.: Quickxplain: Preferred explanations and relaxations for over-constrained problems. In: *AAAI'04*. pp. 167–172. AAAI Press, USA (2004)
4. Labreuche, C.: A general framework for explaining the results of a multi-attribute preference model. *Artif. Intell.* 175, 1410–1448 (May 2011)
5. Lichtenstein, S., Slovic, P.: *The construction of preference*. Cambridge Univer. Press (2006)
6. McNee, S.M., Riedl, J., Konstan, J.A.: Being accurate is not enough: how accuracy metrics have hurt recommender systems. In: *CHI '06*. pp. 1097–1101. ACM (2006)
7. McSherry, D.: Explanation in recommender systems. *Artif. Intell. Rev.* 24, 179–197 (2005)
8. Nunes, I., Miles, S., Luck, M., Lucena, C.: A study on justifications for choices: Explanation patterns and guidelines. Tech. Report CS-2012-03, University of Waterloo, Canada (2012)
9. Nunes, I., Miles, S., Luck, M., Lucena, C.: User-centric preference-based decision making. In: *AAMAS'12* (2012), (to appear)
10. Pu, P., Chen, L.: Trust-inspiring explanation interfaces for recommender systems. *Knowledge-Based Systems* 20, 542–556 (2007)
11. Simonson, I., Tversky, A.: Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research* 29(3), 281–295 (1992)
12. Tintarev, N., Masthoff, J.: A survey of explanations in recommender systems. In: *ICDE Workshop*. pp. 801–810. IEEE Computer Society, USA (2007)