

8

Guidelines and Patterns for Explanations

As presented in the previous chapter, there are many challenges that need to be overcome in order to produce effective explanations that support decision making performed by software systems. One of them is the identification of the kinds of explanation that users expect and need to understand the rationale behind decisions made in order to accept the them. In this chapter, we address this issue by describing an exploratory study performed to capture patterns that describe how people justify their choice among a set of available options. As we assume that the explanations provided by people are those that users expect to receive, we derive from our study results a set of guidelines and patterns, which serves as a basis for approaches whose aim is to explain to users why a particular option is chosen based on multi-attribute decision-making. We first describe this study in Section 8.1, then detail and discuss its results in Section 8.2 and our interpretation in Section 8.3. Finally, we present the guidelines and patterns derived from this study in Section 8.4, concluding in Section 8.5.

8.1

Study Description

The exploratory study presented in this chapter follows the same overall procedure adopted in our previous study, which was presented in Chapter 2. The framework that provided guidance for the study elaboration was proposed by Basili et al. (Basili et al. 1986), which includes the GQM template (Basili and Rombach 1988), used to define the goal of the study — and later to define research questions and select metrics for answering those questions. The goal of the present study, following the GQM template, is presented in Table 8.1. As highlighted previously, the work of Basili et al. focused on experimentation in the context of SE, but it is sufficiently generic to be applied to other areas, and the reason for using their guidelines and template is due to the experience of the researchers with SE studies.

In order to achieve this goal, the study we designed is based on a questionnaire made available online, in which participants (individuals that are part of the social network of the researchers) had to make a choice from a set of available options

Definition element	Our experiment goal
Motivation	To identify the kinds of explanations users expect to receive from software systems,
Purpose	characterise and evaluate
Object	explanations to justify a choice
Perspective	from a perspective of the researcher
Domain:people	as they are provided by people
Scope	of the context of the social network of the researcher.

Table 8.1: Goal Definition (GQM template).

and later justify their decision. The arguments given by participants were carefully analysed to understand their common characteristics and also the dependency between options and the arguments given. The options given in our study consist of hotels located in New York City, USA, and this decision was made due to three reasons.

- (i) *New York is a widely known touristic city*, therefore participants are more likely to have a broad idea of close to where they would like to stay, prices they are accepting to pay, and so on.
- (ii) *Researchers knowledge about the city*, thus we are able to select appropriate options for being part of the study.
- (iii) *Massive amount of available hotels*, which is important as our study is based on real hotel data so participants take it more seriously.

The next sections provide further details about our study. We start by presenting the research questions of the study in Section 8.1.1, and then detail the study procedure in Section 8.1.2. We describe the participants of our study in Section 8.1.3, to later proceed to the analysis and interpretation of the results.

8.1.1 Research Questions

As shown in the previous section, our main objective while performing this study is to give guidance for explanation generation. It is accepted by the research community that explanations improve software systems that automate decision-making or provide recommendations by making its reasoning process or reasons for the choice more explicit, thus enhancing user acceptance and trust. Nevertheless, there is no consensus on what constitutes a *good* explanation, and what kind of information must be provided to users. Therefore, the present study aims to solve this issue in a user-centric way by identifying the kinds explanation

RQ1. Do users use a pattern to justify an option chosen from the set of those available?	EA1. Analysis of the arguments given to justify the chosen option and identification of commonalities among arguments given by different users.
RQ2. Is there a relationship between the type of explanation given to support the decision and the chosen option?	EA2. Comparison among the arguments given to justify each different chosen option.
RQ3. Do users use a pattern to justify the rejected (not chosen) options?	EA3. Analysis of the arguments given to reject options and identification of commonalities among arguments given by different users.
RQ4. Is there a relationship between the type of explanation given to reject options and the rejected or chosen option?	EA4. Comparison among the arguments given to reject options according to each different chosen and rejected option.
(a) Research Questions.	(b) Evaluation Approaches.

Table 8.2: Research questions and their evaluation approach.

that people give — and we assume that they are those that users expect to receive — to justify a decision, and then providing guidelines and patterns that allow defining “good” explanations from a user perspective. The aim of such explanations is to expose to users why a system chose a particular option, thus improving *effectiveness* and user *trust* in the decision. In our study we addressed four different research questions, presented in Table 8.2(a).

By answering these research questions, we are able to extract patterns for user explanations to be generated by decision-making systems (based on *RQ1* and *RQ3*), and also the context in which each pattern is adopted (based on *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 address explaining why other options were rejected (or not chosen).

8.1.2 Procedure

In a nutshell, our study consists of collecting information provided by participants through a web-based questionnaire, and later analysing the collected data. Our aim was to obtain a high number of participants, and therefore anyone with Internet access could access the questionnaire (more details about the set of participants are given in next section). Our study involves decision-making and explanation about this process, and we chose hotels as the domain associated with the decision. The main reason for this design choice is that most of people are aware of the attributes that characterise hotels, and have preferences for individual attributes. Moreover, we chose to provide hotels in New York city for the reasons already presented. The applied questionnaire, which can be seen in Appendix C, consists of three parts, and each of which is explained next.

User Information Data. Our study does not assume that explanations depend on people characteristics, such as age or gender; however we collect some information about the participants to obtain demographic information (as we made the questionnaire available online, any individual can access it). The collected participant data is: (i) age; (ii) gender; (iii) location (city and country); and (iv) working/studying field.

Choosing Product. The study participant is then requested to imagine the scenario in which she is going to spend holidays in New York, and must choose a hotel for staying there from a set of options that we made available for her. As hotel rates for double rooms are very similar to those for single rooms, and people usually spend holidays with at least one friend, we include in this hypothetical scenario that the participant would travel with a friend and does not mind to share a bed with him or her. In order to make our scenario more realistic, we have selected real existing hotels to offer to participants. Hotels are described in terms of attributes associated with hotels and their rooms available at the [booking.com](https://www.booking.com) website, presented in a table that allows a side-by-side comparison. We have selected *five* different hotels (*Hotel 91*, *Econo Lodge Times Square*, *The Hotel at Times Square*, *Comfort Inn Times Square*, *Renaissance New York Hotel 57*), viewing these options as forming three groups (not known to participants), as described below. Complete details about each hotel can be seen in Appendix C.

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 Times Square) dominated by another (The Hotel at Times Square) even though “Comfort Inn Times Square” 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 Times Square” is dominated.

G-2 Extreme options. Extreme options compromise too much one attribute (e.g. quality) to improve another (e.g. price). People in general avoid such options, as stated by the extremeness aversion principle (Simonson and Tversky 1992), so we also selected extreme options to understand how participants explain their rejection or, if they choose

Subjective	Objective
• Justifications for acceptance • Justifications for rejection • Explanation types • Additional characteristics of justifications	• Chosen hotel • Chosen hotel vs. Explanation types for acceptance • Chosen hotel vs. Explanation types for rejection of other hotels • Rejected hotels vs. Explanation types for their rejection

Table 8.3: Data collected in our study.

them, why they do so. There are two extreme options: (i) much lower quality and much lower price (Hotel 91); and (ii) much higher quality and much higher price (Renaissance New York Hotel 57).

G-3 Options that Require Trade-off Resolution. 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 Times Square” 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 of the study consists of carefully investigating these justifications to identify patterns and define explanation types from which, based on this initial analysis, we can extract quantitative data. Table 8.2(b) shows our approach to answering our research questions, which is mainly based on a classification of explanation types. In summary, the collected subjective and objective data are presented in Table 8.3.

8.1.3 Participants

The participants of our survey were selected using convenience sampling, which reached a total number of 100 participants. The sample was obtained based on 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;

Gender	Male 58 (58%)		Female 42 (42%)	
Country	Brazil 78 (78%)	United Kingdom 8 (8%)	Canada 5 (5%)	Other 9 (9%)
Age	16-25 years 4 (4%)	26-35 years 61 (61%)	36-45 years 11 (11%)	>45 years 24 (24%)
Field of Work of Study	Informatics 54 (54%)	Education 11 (11%)	Management 7 (7%)	Other 28 (28%)

Table 8.4: Demographic Characteristics of Participants.

and (ii) by Facebook,¹ which is a widely known social network. The distributed message consists of an invitation to participate of the survey and a request to forward the invitation for other people.

The survey was available for participation on October 12–24, 2011 and was initiated by 191 people, who answered at least one of the steps of the survey, from which 100 (52.36%) finished all the survey steps — the remaining surveys were discarded. The demographic characteristics of the participants that completed the survey are described in Table 8.4. Because we adopted the social network of this researcher to perform the study, most of the participants are aged between 26 and 35 years (61%) and are Brazilians (78%). Non-Brazilian participants are from 8 other countries: United Kingdom, Canada, Germany, United States of America, Switzerland, China, France and Netherlands, and the latter six were grouped into the “Other” category in Table 8.4, as only there are only a few participants from these countries.

8.2

Results and Analysis

Our collected data consists mainly of justifications expressed in natural language and, as these are qualitative data, we analysed them in a systematic way to extract quantitative information. In the section, we explain how we performed this analysis and show results obtained from our study, separating our findings according to the research questions we set out to answer. We focus on describing the obtained data and our qualitative analysis, and we leave for the next section further discussions and our interpretation. Note that, at various points, we label some findings with “*Evidence X*,” so that we can later refer to them to support our proposed guidelines.

Before proceeding to this detailing, we present the hotels chosen by our participants. This information is relevant for understanding the relationship between the chosen option and the corresponding justifications, as indicated by our research questions *RQ2* and *RQ4*. Figure 8.1 shows how many participants selected each

¹<http://www.facebook.com>

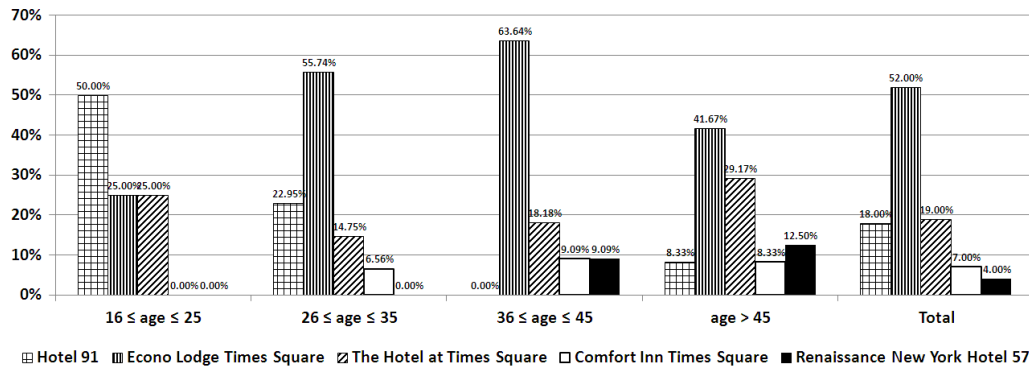


Figure 8.1: Hotel Choice.

hotel and, as expected, the majority of participants chose a hotel from group G-3. We also show the choice distribution according to age, and it can be observed that younger participants prefer cheaper options.

RQ1: Do users use a pattern to justify an option chosen from the set of those available? Each participant had to provide five justifications for their choice, being one of them a justification for why they chose a particular hotel. We analysed all provided justifications and derived from them a classification, which we refer to as *explanation types*, consisting of *six* different types that are described as follows. This classification emerged from the qualitative analysis of collected data, supported by the principles of content analysis from the social sciences (Braun and Clarke 2006). We exemplify each of these explanation types for the choice scenario in Table 8.5.

Critical attribute. For a group of participants, there is an attribute that plays a crucial role in the decision-making process, being in most of the cases the attribute *price*. 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.

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.

Explanation Type	Example of Justification for Acceptance
Critical attribute	<i>H_i is the cheapest option.</i>
Dominance	<i>H_i is better in all aspects.</i>
Main reason	<i>I chose H_i because it offers the benefit a_i.</i>
Minimum requirements	<i>From the hotels that satisfy my requirements, H_i is the cheapest.</i>
One-sided Reasons	<i>I chose H_i because it provides the benefits a_i and a_j.</i>
Pros and Cons	<i>Even though H_i is not the cheapest, it provides the benefits a_i and a_j.</i>

Table 8.5: Example of Justification for Acceptance.

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 used in this kind of justification is specific to each option, differently from the critical attribute explanation type. We can observe that other attributes contribute for the decision by analysing the other justifications.

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.*”



Figure 8.2: Explanation types used to justify each chosen hotel.

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 8.2 (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.

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 8.2 shows how much each explanation type is adopted for each individual hotel.

The distribution of explanation types indicates three trends. First, it can be seen that most of the participants that chose “Hotel 91” (61.11%) justified their decision by giving information about a critical attribute. As price is an attribute extremely relevant for these participants, and what matters for them is basically that this hotel is the cheapest one. Some participants provided further positive information about the hotel (*one-sided reasons*, 22.22%), besides stating that it is the cheapest one — they provided other positive aspects that complement the fact that it is cheapest, i.e. they showed that even though the hotel is the cheapest, the quality that they require is not compromised.

The second observation is related to the hotels of G-3 group. As expected, the main adopted explanation types for choosing them are *one-sided reasons* and *pros and cons*, as it can be seen in Figure 8.2 and as it is shown in more detail in Table 8.6. 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 chose the “Econo Lodge Times Square” had excluded the cheapest hotel from the set of hotels being

Hotel	One-sided Reasons	Pros and Cons	Total
Econo Lodge Times Square	55.77%	25.00%	80.77%
The Hotel at Times Square	26.32%	63.16%	89.47%

Table 8.6: Main explanation types used for justifying hotels of the G-3 group.

considered in the decision, and they explained the benefits of this hotel to show that this hotel is suitable for them, i.e. there is no reason to pay more for another option if this hotel already provides what the participant wants. On the other hand, participants that chose “The Hotel at Times Square” made a detailed analysis of this hotel against the “Econo Lodge Times Square,” i.e. they discussed their *pros and cons*, and showed that the higher price of the former justifies the benefits it provides, when compared against the latter. With respect to these two options, we point out one last comment: there are two participants (3.85%) that used dominance to justify why they chose “Econo Lodge Times Square.” The participants ignored attributes that are not relevant for them, creating a scenario in which this hotel dominates all the others.

Finally, we discuss the results obtained for the dominated option and the most expensive option. It can be seen that there is no most adopted explanation type, and participants adopted different explanation types for justifying them. Only few participants chose these two options and, as it is not obvious why these options should be chosen, the participants gave their particular explanations to justify this decision. In the first case, “Comfort Inn Times Square,” some participants were vague and said that they chose this hotel because it has the best cost-benefit relationship without giving further details. The remaining ones used 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 was also used with the 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 chose the 4-star “Renaissance New York Hotel 57,” as comfort is the most important issue for them, and they are not concerned with price, and in their justification they explained this situation, i.e. for them the price of the hotel justifies the possible comfort it offers, and this is assumed because of the hotel stars. In one case, a participant said that she prefers the most expensive (critical attribute), as she wants to maximise comfort.

RQ3: Do users use a pattern to justify the rejected (not chosen) options?

Now, that we have already addressed the research questions related to choosing an option, we focus on the rejected options. By analysing justifications for rejecting options, we have observed the same explanation types used for justifying the chosen

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

Table 8.7: Example of Justification for Rejection.

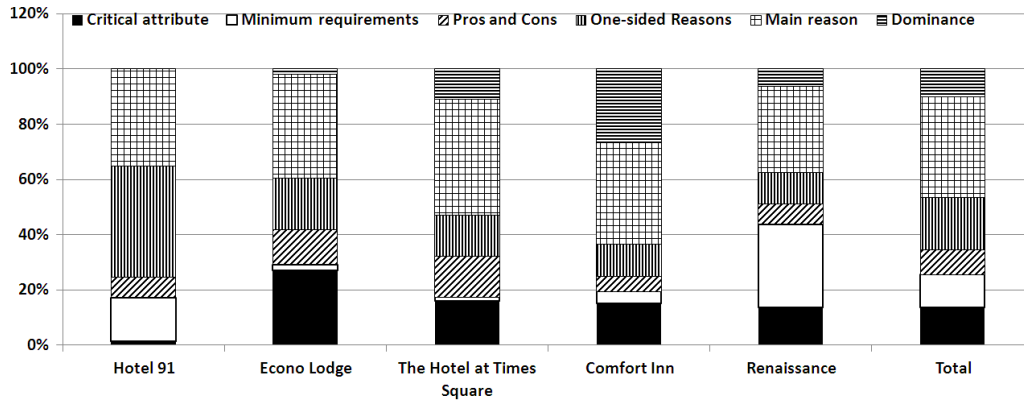
option. The description given for our set of explanation types show that they can also be applied for rejecting options, e.g. if an option does not satisfy the minimum requirements, then it is rejected due to this reason. In Table 8.7, we show examples of how each of these explanation types is used in the context of option rejection.

As it is the case with justifications for accepting an option, we also concluded that participants do use patterns for constructing arguments to reject options, and we next analyse the relationship between the adopted explanation types and the options involved in the decision-making process.

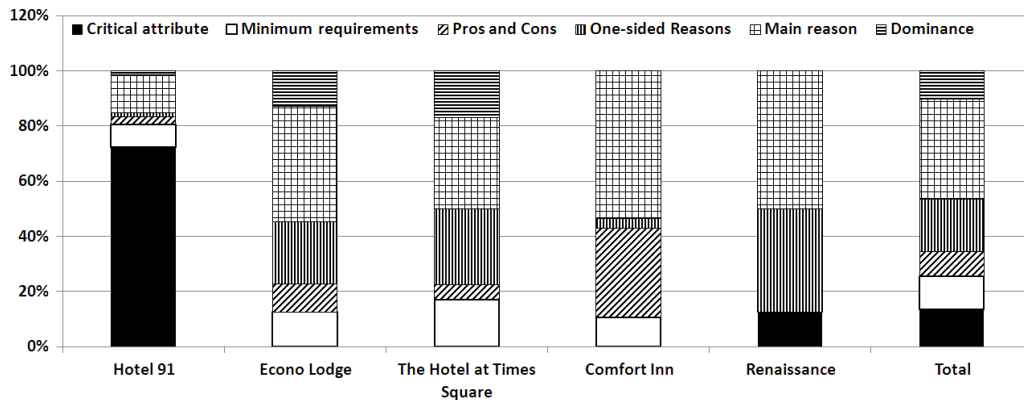
RQ4: Is there a relationship between the type of explanation given to reject options and the rejected or chosen option? In order to understand how participants choose a particular explanation type, we analyse the relationship between the types adopted to justify rejected options from two perspectives. The first consists of analysing justifications for rejection by relating them to the hotel that was rejected, i.e. we observe which explanation types were adopted to reject a particular hotel. The second perspective groups justifications according to the chosen hotel, i.e. we observe which explanation types were adopted to reject other options according to a particular chosen hotel. These two discussed views of justifications for rejection are presented in Figures 8.3(a) and 8.3(b).

There are many interesting aspects that can be observed in the collected data. *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 acceptance is that the chosen hotel is the cheapest, and the justification for the remaining rejected hotels is that they are more expensive (than the chosen hotel). Similarly, this situation happens with the more expensive hotel, in which the participant wanted to maximise the price (as a proxy to the 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 comparing



8.3(a): Explanation types used to justify the rejection of each hotel.



8.3(b): Explanation types used to justify the rejection of other hotels given a chosen hotel.

Figure 8.3: Rejection explanation types.

the chosen option with the others. In many situations, preferences (hidden in justifications) of participants, who chose “Econo Lodge Times Square,” indicate that “The Hotel at Times Square” dominates “Comfort Inn Times Square;” however this is not given as an argument to discard the latter, but the participants seek for an explanation why “Econo Lodge Times Square” is better than “Comfort Inn Times Square” (*Evidence A*). Dominance was used as argument by participants that chose “Econo Lodge Times Square” when the set of attributes that matter for the participants indicated that this option dominates both “The Hotel at Times Square” and “Comfort Inn Times Square.”

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, a maximum distance from the city centre or a minimum number of stars. In these situations, an option is rejected regardless the remaining options, and the justification given is that the option does not satisfy the participant *minimum requirements*.

Main reason and *one-sided reasons* indicate that there is an attribute (or a set thereof, in case of one-sided reasons) that is really important for the participant that, even though it is not part of a hard constraint, plays a *decisive role* in the

decision, i.e. because of this (these) attribute(s), the option is being rejected. This set of attributes is *kept as simple as possible* (*Evidence B*); e.g., some participants that chose “Econo Lodge Times Square” rejected “The Hotel at Times Square” and “Comfort Inn Times Square” because they do not have a refrigerator and are more expensive (than the chosen hotel). But, for justifying the “Renaissance New York Hotel 57” (which also does not have a refrigerator), they argued 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 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 have also identified other relevant characteristics present in the provided justifications. We describe each of these characteristics below, and most of them can be used to suggest informal arguments to be used in systematic approaches for decision-making.

Explicit trade-off (TO). As already mentioned in the description of the *pros and cons* explanation type, some participants stated that the chosen hotel has the best cost-benefit relationship (or not the best, for rejecting a hotel), and sometimes just provided this argument without any details, e.g. “*For a trip like this, it seems the best cost-benefit among the 3-star hotels.*”

Preferences mentioned (PREF). Participants, when requested to justify their decision, provided arguments that are constructed based on their preferences (*Evidence E*); for example, a participant argue “*Absence of a fitness centre*” to justify a rejection, but this is due to the participant preference for a hotel *with* a fitness centre — and in some cases, participants made their preference explicit.

Insignificant difference (ID). The “Econo Lodge Times Square” has a US\$5.00 difference from “The Hotel at Times Square.” While some participants argue

that the benefits provided by the second does not compensate the price difference, others, who have chosen the second, stated that the price difference is insignificant, as it is very small, and both hotel prices can be considered the same. The same applies for room size or location, from the perspective of some participants.

Intuition (INT). One interesting characteristic of some provided justifications consists of inferring information of the hotel without any basis, i.e. some participants used their intuition to choose a hotel. For instance, one participant that chose “Econo Lodge Times Square” justified the rejection of “The Hotel at Times Square” by saying “*The name The Hotel seems to provide quality and, consequently, high price.*”

Price as a first class attribute (PRICE). The majority of participants (92%) mentioned the attribute “price” in their justifications, and evaluated options by comparing this attribute with all the other ones. This indicates that *cost* is not seen as any disadvantage that an option has when compared to another, but a fixed attribute that should be treated differently in the provided explanations (*Evidence F*).

Irrelevant attributes (IRR). When participants chose a hotel that does not offer as many benefits as the others, mainly when they chose the cheapest hotel, they used as an argument that those benefits are not important for them and, as they do not care about them, there is no reason for paying more for something that will not be used. Irrelevant attributes were mentioned in both acceptance and rejection justifications. For supporting a choice, participants state: “*Even though hotel H_i does not offer attribute a_i , this is not important to me,*” and for rejecting an option, they say “*Even though hotel H_i offers attribute a_i , this is not useful to me.*”

In Table 8.8, we show the percentage of participants whose justifications presented these identified characteristics. The table is split into each chosen hotel, and rows of each separate sub-table is related to the justification provided for each separate hotel. We highlight in gray the hotel that was chosen, therefore the row of a highlighted first cell is associated with justifications for acceptance.

As mentioned before, it can be seen that price should be treated as a first class attribute in explanations, as it is a crucial factor considered in the decision. In cases that a higher price is chosen, but this difference is very small, many participants acknowledge this fact. When the chosen option has a lower price, benefits provided by other options may be relevant to be mentioned, even though the decision maker does not care about it. In cases in which pros and cons of a set of options make the

Hotel 91						
Reason for	TO	PREF	ID	INT	PRICE	IRR
Hotel 91	0.00%	0.00%	0.00%	0.00%	100.00%	16.67%
Econo Lodge	5.56%	0.00%	0.00%	0.00%	77.78%	11.11%
The Hotel	0.00%	0.00%	0.00%	0.00%	83.33%	16.67%
Comfort Inn	0.00%	0.00%	0.00%	0.00%	83.33%	16.67%
Renaissance	0.00%	0.00%	0.00%	0.00%	94.44%	11.11%
Econo Lodge Times Square						
Reason for	TO	PREF	ID	INT	PRICE	IRR
Hotel 91	1.92%	3.85%	0.00%	3.85%	15.38%	0.00%
Econo Lodge	19.23%	1.92%	5.77%	1.92%	76.92%	1.92%
The Hotel	3.85%	3.85%	7.69%	1.92%	82.69%	13.46%
Comfort Inn	3.85%	1.92%	1.92%	0.00%	88.46%	7.69%
Renaissance	0.00%	0.00%	0.00%	0.00%	96.15%	13.46%
The Hotel at Times Square						
Reason for	TO	PREF	ID	INT	PRICE	IRR
Hotel 91	5.26%	0.00%	0.00%	0.00%	5.26%	0.00%
Econo Lodge	0.00%	5.26%	47.37%	5.26%	42.11%	0.00%
The Hotel	36.84%	10.53%	5.26%	5.26%	63.16%	0.00%
Comfort Inn	0.00%	5.26%	5.26%	0.00%	68.42%	5.26%
Renaissance	5.26%	0.00%	0.00%	0.00%	73.68%	0.00%
Comfort Inn Times Square						
Reason for	TO	PREF	ID	INT	PRICE	IRR
Hotel 91	28.57%	14.29%	0.00%	14.29%	28.57%	0.00%
Econo Lodge	28.57%	14.29%	0.00%	0.00%	14.29%	0.00%
The Hotel	28.57%	0.00%	0.00%	0.00%	28.57%	0.00%
Comfort Inn	28.57%	0.00%	0.00%	0.00%	42.86%	14.29%
Renaissance	14.29%	0.00%	0.00%	0.00%	85.71%	0.00%
Renaissance New York Hotel 57						
Reason for	TO	PREF	ID	INT	PRICE	IRR
Hotel 91	0.00%	25.00%	0.00%	25.00%	25.00%	0.00%
Econo Lodge	0.00%	25.00%	0.00%	0.00%	25.00%	0.00%
The Hotel	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Comfort Inn	0.00%	0.00%	0.00%	0.00%	25.00%	0.00%
Renaissance	0.00%	0.00%	0.00%	25.00%	0.00%	0.00%

Table 8.8: Results for additional characteristics observed in justifications.

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.

8.3

Interpretation

Our study investigates explanations given by people to justify their choices, providing reasons to accept or reject options. In this section, we provide an interpretation for our analysis, which explains how participants, and more generally people, choose a particular type of explanation for their decisions.

8.3.1

Explanation for Choice

We provided five different options for participants, chosen in order to characterise options with certain particularities. Options that have quality lower than most of the other options available, but also lower price, are justified by the “critical attribute,” which in this case is price. Therefore, other option details are not relevant, and do not need to be part of the explanations. In some cases, as the cheapest option being offered in our study already provides some comfort (as the hotel has 2 stars, shower, etc.), some participants mentioned that they chose the cheapest hotel, as it satisfies all their minimum requirements. Moreover, irrelevant attributes can be mentioned as part of explanations for this kind of option, in order to make the argumentation stronger. Explanations for the most expensive hotel (and with highest quality), although also characterised as an extreme option, does not follow this same reasoning. This kind of option is justified with all positive aspects it provides, or a main one — that is generally only offered by expensive options. In this scenario, price is typically not mentioned, as it is not a concern.

The majority of the participants have reduced their choice to options that require trade-off, by first discarding some options due to a minimum requirement, such as maximum price. There are mainly two kinds of justifications given for supporting a choice from this set of options. When the chosen option is the cheapest in this set, people use only the main reason or the benefits provided by the chosen option. On the other hand, when the option is not the cheapest, a more detailed explanation is needed, therefore the pros and cons related to the chosen option should be exposed, with the aim of showing that pros justify cons. For making this argument stronger, explicitly mentioning that the option provides the best cost-benefit relationship is helpful.

The last kind of option, namely the dominated option, is never chosen. In our study, the option that represents this group of options (G-1) has few advantages

with respect to the others, which are the room size and parking price. As these attributes are extremely relevant for some of the participants, they supported their choice evaluating pros and cons or by stating these essential aspects, in summary, this option was considered as an option of the G-3 group by those participants.

8.3.2

Explanation for Rejection

Having described how explanations are constructed to justify a chosen option, we now discuss how explanations for rejecting an option are built. A very important aspect of the explanation for rejecting an option, is that the *explanation is driven by the chosen option* in many cases.

As in explanations for acceptance, when an extreme option is chosen, explanations for rejecting options have a different behaviour for the two different extreme directions. When the chosen option is the cheapest and with lowest quality, the explanation given for all the remaining options are the same, which says that the other options are more expensive than the chosen one. So the typical explanations for this scenario are “*because of the price*” or “*because this option is more expensive [than the chosen option]*.” When the chosen option is the most expensive option and with the highest quality, the arguments used to reject the remaining options are the most decisive aspects that are not offered by these options. For example, if one individual considers “fitness centre” and “bar” as important, but the former is more important than the latter, this individual would justify the rejection of “Econo Lodge Times Square” by stating that it does not have “fitness centre,” even though it also does not have a “bar.” Both aspects should be mentioned in the explanation if they, individually, would not change the choice made. For this kind of extreme option, dominance is never used because, as the chosen option is the most expensive, it does not dominate any other option.

Next, we discuss explanation for rejection when the chosen option is part of the G-3 group, which has a different explanation for each kind of rejected option. When there is a dominated option, this is the argument to be given, but *only if the chosen option dominates this option*. In addition, people that choose an option from the G-3 group, typically discard some options due to a cut-off value, i.e. options that do not satisfy minimum requirements, which are often part of the G-2 group. In this situation, the reason for cutting this option off should be given as the explanation, such as “*too expensive*” or “*too far away*.” Finally, to reject other acceptable options, the reasoning is similar to that described for the most expensive option, i.e. the option decisive aspects must be exposed. If pros and cons of the rejected option have a similar balance to the chosen option, then this should be discussed in detail in the explanation, in order to show that pros do not compensate cons.

8.4

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.

8.4.1

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 the other options worse than the chosen option.

An example of the application of this guideline is the case of domination, which is used as a justification only if the chosen option dominates the rejected 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 an option to be selected or rejected. These decisive criteria should be derived from the comparison of the chosen option against the others.

For example, the case described above in which the rejection of the “Econo Lodge Times Square” is justified only by the fact that it does not have a fitness centre, although it also does not have a bar.

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 user 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. Benefits is a subset of all possible positive characteristics that an option can have, for example, a hotel that provides breakfast, “big” room, “good” location, etc.; nevertheless benefits always come with a cost, which in the hotel case is its price, but in other scenarios it can be time or effort. 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.

8.4.2 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.

- *Explanation 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.* An explanation pattern can provide guidance to generate an argument that justifies the decision as a whole (*decision*), or the

generation of an argument that supports the acceptance or rejection of a single option (*option*).

- *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*).

Next we present each of our proposed patterns. Patterns are presented ordered according to their complexity, i.e. the simpler the explanation associated with a pattern is, 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.

Pattern 1: Critical Attribute

Classification:

- *Explanation goal:* both
- *Target:* decision

Context: this pattern is applied in two situations: (i) there is an attribute that is extremely important for the user and this is the only one to be taken into consideration; and (ii) all available options satisfy all constraints and there is one criterion to choose the best.

Template:

(Option) chosen option was chosen because it has the best value for critical attribute.

Description: some users have a single criterion to choose an option, and may additionally have a set of constraints that is satisfied by all options. In these situations, the decision becomes trivial, as well as its associated explanation, which consists of stating that the chosen option was selected according to this single criterion.

Example: the user wants to buy the cheapest flight from Rio de Janeiro to London, regardless number of stops, flying time, airline company and so on.

Back up preference: preference that establishes criterion used to make the choice, and possibly other preferences satisfied by all options.

Extensions: rejected options might have characteristics not present in the chosen option, which are typically considered as benefits; however, for the decision maker, these characteristics are irrelevant. The explanation in this case can be extended by stating that the chosen option does not have such characteristics, but these are not relevant given the provided preferences.

Pattern 2: Cut-off**Classification:**

- *Explanation goal*: reject
- *Target*: option
- *Position*: absolute

Context: an option does not satisfy a user requirement (hard constraint), or does not satisfy a constraint that is very relevant for the user, but there are other options that satisfy it.

Template:

(Option) rejected option was rejected because it does not satisfy constraints associated with attribute.

Description: in many situations, users have a set of requirements that **must** be satisfied (or hard constraints), and therefore an option that does not satisfy at least one of these requirements cannot be chosen. As not satisfying at least one of these constraints is enough for rejecting an option, only the constraint associated with the most important attribute for the user is part of the explanation. In addition, there are cases in which some requirements are not hard constraints, because the user might accept options that do not satisfy it when there is no other available options. When there are options that satisfy these “almost-hard” constraints, options that do not satisfy it can be rejected with an explanation of a not satisfied hard constraint.

Example: the maximum amount of money that a user will spend in a hotel room is US\$300.00 for two nights. A hotel that costs US\$350.00 for two nights is rejected because it is too expensive, i.e. due to a cut-off value.

Back up preference: not satisfied hard-constraints (or “almost-hard”).

Pattern 3: Domination**Classification:**

- *Explanation goal*: reject
- *Target*: option
- *Position*: relative

Context: the chosen option dominates a rejected option.

Template:

There is no reason to choose (option) rejected option, as (option) chosen option is better than it in all aspects, including cost.

Description: when an option dominates another, there is no argument that justifies considering the latter better than the former. Therefore, exposing this fact is enough to explain the rejection of the dominated option. However, the domination argument must be used only if the option that dominates the rejected option is the chosen one.

Example: there are two hotel rooms available for the user: standard queen room and superior queen room. The difference between them is that the superior queen room is bigger, it has a sitting area, it has a bath besides the shower, and, because of these extra features, it is more expensive. All user constraints are satisfied by both options, and she does not care about these three extra features, but she cares about the price. Therefore, according to the preferences of this user, the superior queen room is dominated by the standard queen room.

Back up preference: preferences that establish that individual attributes of the chosen option are considered better than the rejected one.

Extensions: in this pattern, attributes that users do not care about might be mentioned to support the domination relationship. See more details in the extensions of Pattern 1: Critical Attribute.

Pattern 4: Minimum Requirements⁻

Classification:

- *Explanation goal:* reject
- *Target:* option
- *Position:* relative

Context: user established a set of minimum requirements for options, and a way of choosing from those that satisfy it. According to the requirements, some options were discarded. Other option attributes might have been used for making the decision, but only one of them makes the difference.

Template:

Even though (option) rejected option satisfies all your requirements, it has a worse value for attribute than (option) chosen option .

Description: this pattern addresses justifying rejected options that differ only by a single attribute (that matters for the user) from the chosen option, and these rejected options, and also the chosen one, satisfy a set of user requirements. This scenario might happen when provided preferences consist only of these minimum requirements plus a preference to choose among them or when, even though when the decision involves a careful evaluation of pros and cons of each individual option, after choosing a particular option, one or more options are distinguished from it by a single attribute, relevant for the decision. Options that are in this category

can be rejected using a simple explanation — rather than a more complex one, possibly used in the decision-making process — consisting only of the requirements satisfaction and a single attribute that makes the difference.

Example: a user wants to stay in a 2-star hotel, whose price is up to US\$150.00 per night, and is within the city centre, with breakfast. Given these requirements, the cheapest one. The chosen hotel costs US\$120.00 per night, there is one that costs US\$130.00; and all the remaining attributes have the same values for both options. The rejection of the second hotel is explained using this pattern, as there are other hotels that do not satisfy the price constraint.

Back up preference: user requirements and preference used to choose from the options.

Pattern 5: Minimum Requirements⁺

Classification:

- *Explanation goal:* accept
- *Target:* option
- *Position:* absolute

Context: user established a set of minimum requirements for options, and a way of choosing from the ones that satisfy it. According to those requirements, some options were discarded. Other option attributes might have been used for making the decision, but only one of them made the difference.

Template:

Besides satisfying all your requirements, (option) chosen option has the best value for attribute.

Description: users, in certain decisions, establish a set of minimum requirements that reduces the set of available options to a subset, in which options differ only by a single criterion from the chosen option. This is a situation in which the decision becomes easy, and also the explanation, which consists of acknowledging users that, from those options that satisfy their requirements, the chosen option is the preferred one according to a particular criterion.

Example: a user wants to stay in a 2-star hotel, whose price is up to US\$150.00 per night, and is within the city centre, with breakfast. Given these requirements, the cheapest one. The chosen hotel costs US\$100.00 per night, and other available hotels provide the same features and are more expensive. Therefore, the explanation for the chosen hotel is that it is the cheapest from those that satisfy requirements.

Back up preference: user requirements and preference used to choose from the hotels.

Extensions: in this pattern, attributes that users do not care about might be mentioned to support the decision. See more details in the extensions of Pattern 1: Critical Attribute.

Pattern 6: Decisive Criteria

Classification:

- *Explanation goal:* both
- *Target:* option
- *Position:* absolute

Context: even though there are other attributes that contribute to the option acceptance (or rejection), there is a subset of them that would confirm this decision regardless the values of the other attributes.

Template:

(Option) option was [*chosen* | *rejected*] because of its set of decisive attributes.

Description: options, when compared, might have different pros and cons. However, there are attributes that are the most important ones in the decision, and other attributes — 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 the remaining attributes out of it.

Example: three hotel options were given to a user. One is a 3-star hotel, cheaper than the other two options and has a refrigerator in the room. The second is also a 3-star hotel, more expensive than the former, with a better location. The last is a luxury 4-star hotel, much more expensive than the others and, as the second option, also does not have a refrigerator in the room. While the rejection of the second is justified by the absence of the refrigerator *and* its price; the rejection of the third is justified only because of its price, because this was the decisive criteria for not choosing it.

Back up preference: preferences over the set of decisive attributes.

Pattern 7: Trade-off Resolution**Classification:**

- *Explanation goal*: both
- *Target*: option
- *Position*: absolute

Context: there is no set of attributes that are decisive.

Template:

Template for rejected options:

Even though (option) rejected option provides better pros than (option) chosen option, it has worse cons.

Template for the chosen option:

Even though (option) chosen option does not have the best value for cons, its values for pros compensate its cons.

Description: a set of decisive attributes does not exist in all situations. Options might provide different pros and cons in a way that all attributes are important for making the decision, therefore, all option attributes that differ for these options have to be evaluated, and their evaluation has to be informed to the user.

Example: a user is provided with two hotel options. Both of them are 3-star hotels, the first is cheaper (US\$115.00 per night) and its room has a refrigerator, and the second is more expensive (US\$130.00 per night), better located (two blocks closer to the city centre), and has breakfast included. As, according to the user preference in our example, location has a higher priority than price, and their difference is strong enough when considering the included breakfast and the refrigerator, the second hotel is chosen. The explanation thus states that even though the second hotel has a lower price and a refrigerator, it has a worse location and does not include breakfast.

Back up preference: all user preferences used to evaluate pros and cons of options that require trade-off analysis.

Extensions: in situations in which pros and cons of the chosen option create a balance that is very similar to the one of another option, it might be not obvious for the user why pros compensate cons. So, additionally, it can be explicitly told to the user, i.e. inform the user that the chosen option provides the best cost-benefit relationship.

8.5

Final Considerations

In this chapter, we presented a study performed to understand how people justify their decisions, by giving explanations why they chose a particular option from the set of those available, and why remaining options are rejected. The study consisted of providing participants (a hundred people) with a set of carefully chosen hotel options, and requesting them to give reasons for the choice. Based on collected data, we 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. We now will show how we produce explanations for choices made by our decision-making technique, which take into consideration this guidance that we derived from our study.