European Journal of Operational Research 221 (2012) 593–602

Contents lists available at SciVerse ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Decision Support

# When conflict induces the expression of incomplete preferences

Stéphane Deparis [a], Vincent Mousseau [a,*], Meltem Öztürk [b], Christophe Pallier [c], Caroline Huron [c]

[a] Laboratoire Génie Industriel, Ecole Centrale Paris, Châtenay-Malabry 92290, France
[b] LAMSADE, Université Paris Dauphine, Paris 75016, France
[c] INSERM Cognitive Neuroimaging Unit, CEA Neurospin, Gif-sur-Yvette F91191, France

ARTICLE INFO

*Article history:*
Received 19 April 2011
Accepted 23 March 2012
Available online 2 April 2012

*Keywords:*
Multicriteria decision
Preference elicitation
Experimental investigation
Conflict
Incomplete preferences

ABSTRACT

Multicriteria conflict arises in pairwise comparisons, where each alternative outperforms the other one on some criterion, which imposes a trade-off. Comparing two alternatives can be difficult if their respective advantages are of high magnitude (the *attribute spread* is large). In this paper, we investigate to which extent conflict in a comparison situation can lead decision makers to express incomplete preferences, that is, to refuse to compare the two alternatives, or to be unable to compare them with confidence. We report on an experiment in which subjects expressed preferences on pairs of alternatives involving varying conflicts. Results show that depending on whether the participants are allowed to express incomplete preferences or not, attribute spread has a different effect: a large attribute spread increases the frequency of incomparability statements, when available, while it increases the use of indifference statements when only indifference and preference answers are permitted. These results lead us to derive some implications for preference elicitation methods involving comparison tasks.



## 1. Introduction

The notion of preference is central in decision-making and accounts for the way decision makers compare the alternatives when facing a multicriteria decision problem. The field of preference modeling has steadily grown over the last fifty years and numerous models have been proposed, such as, utility-based models (Keeney and Raiffa, 1976) or outranking-based models (Roy, 1996). Preference models define a rationality from a normative point of view, that is, norms that the decision maker (DM) should conform to. For several decades, in addition to this normative perspective, many descriptive works have studied the behaviour of DMs in real situations or laboratory experiments. The existing literature addresses the cognitive limitations of DMs (Simon, 1955), the lability of preferences (Lichtenstein and Slovic, 2006), or the biases due to loss aversion and perception of probabilities (Kahneman and Tversky, 1979), etc.

Reciprocation between the normative and the descriptive views has enriched the models, in the sense that descriptive limitations of normative models have been pointed out and have been continually improved upon. The mutual improvement of descriptive and normative approaches has been particularly fruitful in the fields of decision under risk, and multicriteria decision analysis, particularly with respect to utility-based models (von Winterfeldt and Edwards, 1986).

In this paper, we study the incompleteness of preferences, a topic on which few descriptive work exist. Incomplete preferences occur in comparisons, where the DM is not able to select one of the two alternatives as the best one, despite not considering them as equal. In incomplete preferences models, the pairs of alternatives that are not in the relation of *preference* or *indifference* correspond to *incomparability*.

One potential source of incompleteness is multicriteria conflict. Multicriteria conflict can arise in the evaluation of a single item, when this item is good on certain attributes and bad on others. Fischer et al. (2000) experimentally showed that, in a single alternative evaluation task, *within-alternative conflict* results in what they called *preference uncertainty*, measured through judgement times and consistency of evaluations over time. Our work involves *between-alternative conflict*, that is, the conflict when comparing two alternatives, caused by the discrepancy in their evaluations on attributes. Such a conflict is present as soon as no alternative dominates the other one. Conflict is what makes trade-offs necessary and imposes to compensate between options' strengths and weaknesses.

Our work aims at empirically studying hypotheses linking the expression of incomplete preferences to conflict in multicriteria comparisons. More precisely, we test the hypothesis that increasing the intensity of conflict in a multicriteria comparison increases the likelihood that DMs consider two alternatives as incomparable.

This paper is organized as follows: in the next section, we examine how incompleteness is considered in preference modeling, and we distinguish conflict from other semantics for preference incompleteness. Then, in Section 3, we formulate hypotheses about conflict-based incomparability. The experimental design is de-

* Corresponding author. Fax: +33 1 41 13 1272.
*E-mail addresses:* stephane.deparis@ecp.fr (S. Deparis), vincent.mousseau@ecp.fr (V. Mousseau), ozturk@lamsade.dauphine.fr (M. Öztürk), christophe.pallier@cea.fr (C. Pallier), caroline.huron@cea.fr (C. Huron).

0377-2217/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ejor.2012.03.041

scribed in Section 4 and the results are presented in Section 5 and discussed in Section 6. In the last section, we draw some conclusions and propose further research avenues.

## 2. Incomplete preferences and multicriteria comparison conflict

### 2.1. Modeling incomplete preferences

Many existing models of decision theory are grounded on the construction of a value function (see Keeney and Raiffa, 1976). Given such a value function $v$, preferences can be defined on $A$, such that: $\forall a, b \in A$, $aPb$ iff $v(a) > v(b)$ and $aIb$ iff $v(a) = v(b)$, where $P$ denotes a preference relation (asymmetric and irreflexive) and $I$ an indifference relation (symmetric and reflexive). Such a representation assumes the existence of a *complete* preference structure, i.e. the DM is supposed to be able to compare any pair of alternatives. However, the descriptive validity of the completeness axiom in the utility approach has been questioned as early as in Von Neumann and Morgenstern (1944).

Various alternative models have been proposed to generalize the complete weak order one, e.g. partial preference structures (see Fishburn, 1985). More recently, utility-based models without the completeness axiom have been proposed (e.g. Dubra et al., 2004). In these models, the incompleteness of preferences yields an incomparability relation $R$ (irreflexive and symmetric), which holds for pairs of alternatives $(a,b)$ for which neither $aPb$, nor $bPa$, nor $aIb$ holds.

### 2.2. Reasons for incomplete preferences

Various motives to build models that allow for incomplete preferences have been put forward. Some researchers point out that the data obtained in a revealed preferences paradigm is usually incomplete. Danan, 2003 proposed a model to represent incomplete preferences and Danan and Ziegelmeyer (2004) provided a way to measure incompleteness compatible with revealed preferences and empirically showed the existence of incompleteness in the context of risky decisions.

From a descriptive perspective, DMs can express incomplete preferences in order to avoid trade-offs with implications for moral rules of behavior, like pain, friendship or safety. Luce (1998) showed that such difficult trade-offs lead to negative emotion, which results in emotion-coping behaviors, like choosing the status quo, choosing an alternative that dominates another in the set, or prolonging the search.

A DM can also be inclined to express incomplete preferences when she is afraid to endorse the responsibility for a high-stake decision. Tetlock and Boettger (1994) showed that blame avoidance and loss aversion reinforce each other in accountable decisions implying to give up the status quo.

In this paper, we specifically address incomplete preferences caused by between-alternative conflict. Such a between-alternative conflict occurs when comparing two alternatives which both have an advantage over the other. When a decision situation comes with a conflict, the DM is faced with "mutually exclusive courses of action such that each satisfies some goal relatively better than another" (Luce et al., 2000). Intensity of the between-alternative conflict can be defined by the amplitude of differences between the alternatives through the set of criteria.

### 2.3. Between-alternative conflict in multicriteria comparisons

In a situation of dominance, i.e., when one alternative is better than the other with respect to all criteria, the comparison is conflict-free: the dominant alternative should be preferred.

Consider now a bicriteria comparison, where no dominance occurs. The decision implies a trade-off between two performances: "*does the difference on the first attribute compensate for the difference on the second attribute?*". The intensity of the multicriteria conflict refers to how much performances differ between the two alternatives. The greater these differences, the harder it is for the DM to settle a trade-off with confidence, and to express her resulting preferences.

For instance, consider a DM who has to compare two job opportunities $a$ and $d$ (see Fig. 1). *Job interest* describes the diversity of tasks and the degree of autonomy associated with the job. Job $a$ comes with a low salary but is very interesting. At the opposite, job $d$ is well paid but has a very low interest. Since the two jobs have very different profiles, choosing one job requires to sacrifice one goal in order to achieve the other. Now consider the choice between jobs $b$ and $c$. The DM is sacrificing much less, be it money or job interest, because the compensation she has to estimate involves smaller differences: there is little conflict when comparing $b$ and $c$. Hence, the regret or the threat associated to this decision is much more bearable.

When comparing $a$ and $d$ (high conflict), a DM can have difficulties to express a strict preference in favor of one of them, or indifference between them. Such difficulty can be due to the fact that both jobs appear as unattractive (such a phenomenon has been pointed out by Chatterjee and Heath (1996)). However, even when both jobs $a$ and $d$ are viewed as attractive, a strong conflict can induce a difficulty in comparing the two jobs.

According to the standard utility model, it is assumed that all alternatives having the same utility are indifferent. All options belonging to the same isopreference curve are considered indifferent, regardless of the differences of performance between these options. In other words, a transitivity of the indifference relation is assumed along any utility curve. The incomparability caused by between-alternative conflict that we are pointing at contradicts this implicit assumption of transitive indifference. A DM may indeed accept a sequence of trade-offs concerning small differences of performance, while having difficulties accepting indifference when comparing the first and the last option in the sequence. The between-alternative-conflict-based incomparability described here can be modeled using veto thresholds, as proposed by Roy (1991), in outranking methods.

Consider the situation described in Fig. 1 and let us suppose that the following chain of indifference statements is obtained from the DM: $aIb$, $bIc$, and $cId$. These statements imply that the alternatives are equal in overall value. If we assume a utility model, they lie on the same indifference curve. The difficulty caused by the high conflict when comparing $a$ and $d$ is then reinforced by the fact that they are judged as equal in value. On the contrary, consider $a'$ which is dominated by $a$. $a'$ is therefore on a lower indifference

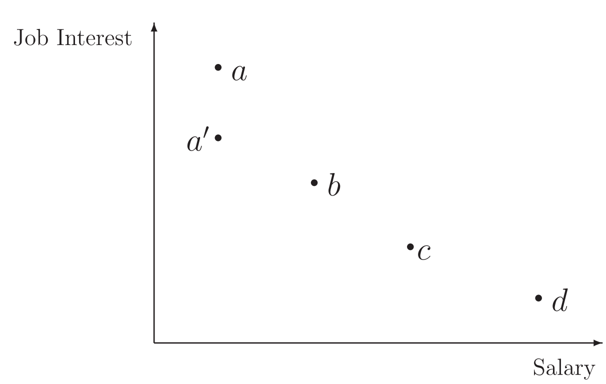


**Fig. 1.** Compared alternatives in the bicriteria space.

chain than $a$ and $d$. In this case, the difficulty caused by the between-alternative conflict in comparing $a'$ and $d$ is mitigated by the fact that one alternative is higher in overall value. Hence, even if $a'$ and $d$ differ largely on both criteria, the DM is likely to prefer $d$ over $a'$ (and therefore will not consider them incomparable).

The above example leads us to distinguish two variables which will play an important role in our experiment:

**attribute spread**, which refers to the magnitude of differences, on each attribute, in the comparison of two alternatives. The attribute spread is defined *independently* from the DM.

**overall value difference**, which refers to the degree by which two alternatives differ in overall value. This difference is low when alternatives belong to close isopreference curves. The overall value difference is grounded on a value representation of the DM's preferences and thus is defined *relative* to some preference information.

## 3. Hypotheses

The specific cause for incomplete preferences that we investigate is related to between-alternative conflict. Our first and second hypotheses respectively deal with the two conflict-related variables: *attribute spread* and *overall value difference*.

Our first hypothesis predicts that a large *attribute spread* or trade-off size makes comparisons difficult, which leads to incomparabilities.

**Attribute spread effect:** the proportion of decision-makers who consider two alternatives as incomparable increases with the attribute spread between the two alternatives under consideration.

*Overall value difference* refers to how much the alternatives differ considering a value representation. Our second hypothesis is based on the intuitive idea that the judgement of the DM should be facilitated when one of the alternatives has a notably higher overall value.

**Overall value difference effect:** the proportion of decision-makers who consider two alternatives as incomparable decreases with the overall value difference between the two alternatives under consideration.

Our third hypothesis describes the behaviour of DMs when modeling of their preferences requires that they express their judgement using only *indifference* and *strict preference* statements.

**Indiscriminate indifference:** if incomparability is not proposed as an answer (indifference is the only symmetric answer available), decision makers will use indifference to express both incomparability and indifference.

This hypothesis is based on the fact that indifference (I) and incomparability (R) are both symmetric relations, whereas strict preference (P) is asymmetric. Both I and R do not favour any alternative over the other. Therefore it is more natural for a DM to use *indifference* than *strict preference* when facing a comparison task, where he would explicitly express *incomparability* between the alternatives if permitted. Our last hypothesis can be understood as follows: when no incomparability answer is available, some stated indifference answers correspond to semantics of incomparability (refusal to compare or lack of confidence in one's judgement), instead of semantics of indifference (equality in value).

## 4. Method

In this section, we describe a behavioral experiment designed to test the hypotheses described above. Students individually compared apartments which differed on two attributes: rent and distance to city center. The experiment was divided into two blocks. The first block, called *isopreference chains construction*, aimed at generating three sets of pairwise equalised options (see Fig. 3), for each participant. In the second block, called the *main task*, all the alternatives to be compared belonged to one of the isopreference chains obtained in the first block. In this block, attribute spread and overall value of the alternatives were randomly varied for each comparison. This allowed us to test for the influence of attribute spread, as well as overall value difference, on the expression of incomparability.

During the construction of the isopreference chains, participants had three available answers to express their preferences: *strict preference* (P) for one of the alternatives and *indifference* (I). The indifference was defined as follows: "*I feel indifferent about the choice of one of these two apartments because they provide me with the same level of satisfaction*". Participants were also informed that indifference implied that they would agree to have an apartment randomly picked out of the pair. At the beginning of the main task, participants were randomly split into two groups: IP group and IRP group, and *incomparability* (R) was added to the available answers for all participants in the IRP group only. The description given for incomparability was: "*I cannot compare these two apartments. Choosing between them feels very hard to me. I'd rather not give an opinion.*" Participants in the IRP group were told that this answer implied that they considered the three other answers as inadequate to describe their opinion about the comparison.

### 4.1. Participants

Forty-two graduate students in the final year of engineering at the Ecole Centrale Paris participated in this experiment. They were paid 20€. All were native French speakers. The experimental instructions (available as Supplementary data) were read to them by the experimenter, and the stimuli and additional instructions were delivered on a computer controlled by a program written using the E-Prime software.

Although the experiment does not involve incentives, it constituted a realistic decision situation that preoccupied the participants at the time of the experiment: students in the final year were chosen because they have to achieve a training course abroad and look for an apartement. There has long been a debate about hypothetical and real tasks (see Kühberger et al., 2002). Although a certain number of authors believe that only revealed preferences should be considered in experimental studies, our position is that hypothetical setups allow to obtain good insights on behavioral phenomena, provided that the task is meaningful to subjects, and that there are proofs of consistency in the results. This position is common in decision-aiding. For instance, such an influential article as Kahneman and Tversky (1979) reports only hypothetical tasks (choosing between prospects, deciding to pay an insurance,...). Finally, we chose to avoid incentives because it would require to define a "revealed incomparability", which raises serious difficulties.

### 4.2. Stimuli

In both the *isopreference chains construction* block and in the *main task*, the participants received a series of trials, where they had to compare two apartments described in terms of their respective *rent* and *distance to city center*, and otherwise equal in every respect. They used the arrows on the keyboard to indicate their answer (left and right for P, down for I and up for R). The distance to city center was displayed using a schematic map of the city, divided in five nested areas, numbered from the center (area 1) to the most remote part (area 5). The area where the apartment was located was highlighted (See Fig. 2). The instructions informed the participants that the three features they had chosen during a preliminary survey (see Section 4.3) were located in the center, as well as their workplace. Travelling across any area using public transports took approximately fifteen to twenty minutes. Thus the location area of an apartment indicated roughly the time it would take for the student to reach area 1, where items important to him were located. The rent was displayed as an amount in Euros per month.

### 4.3. Procedure

#### 4.3.1. Context

At the beginning, participants were asked to imagine that their training course would take part in a big city, remote from their family. The amount of their training wage was 1600€/month, and they had to cover all common costs, including rent, using this money. Hence, the more they would spend on rent, the less they would be able to spend on anything else.

After receiving instructions, participants were asked to fill in an on-screen survey about their habits and tastes regarding the choice of accommodation. The aim of this questionnaire was to acclimate each participant to the context of evaluating an apartment on its rent and distance to places important to him/her. Then we elicited rent levels: five monetary amounts typical of their feeling about rent, from very unattractive to very attractive. These levels were used to anchor the *rent* attribute scale. The average duration time for delivering instructions and completing the survey was 24 min.

#### 4.3.2. Isopreference chains construction

Next, they practised on three comparisons intended to check that they were not choosing a dominated alternative nor using a lexicographic order on attributes. Then they took on a first series of pairwise comparisons, aimed at eliciting twelve values of rent, distributed over three isopreference chains $I_1$, $I_2$ and $I_3$. Each chain $I_j$ was constructed by matching an apartment with a rent value being attractive ($j = 1$), neutral (2) or unattractive (3) and of medium distance to city center, with apartments of one-area-lower and one-area-higher distance to city center. These very apartments were then matched with apartments of even higher and lower distance to city center, until each chain was comprised of five apartments distributed over the *distance to city center* scale (see Fig. 3).

The difference on the *distance to city center* attribute between alternatives was equal to one for all comparisons of the first series. This constraint was aimed at minimizing the difficulty during the elicitation of isopreference chains. The alternatives that make up an isopreference chain are considered as very close in overall value to each of their neighbors in the set. For any pair of alternatives on the same isopreference chain, there exists an indifference path joining them together.

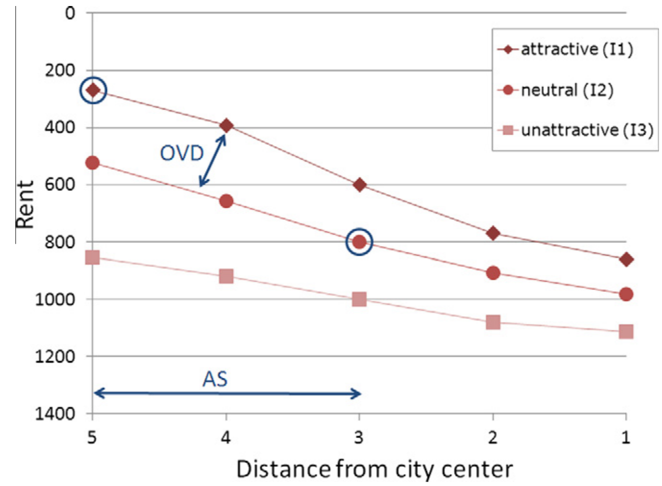


**Fig. 3.** A participant's isopreference chains obtained at the end of the first block (*isopreference chains construction*). OVD: overall value difference and AS: attribute spread.

The elicitation of the chains was based on *I* and *P* statements by the DM in response to successive comparisons, using a hidden matching, similarly to the procedure described in Fischer et al. (1999). Each of the twelve rent values were determined by a comparison-based matching: the matching algorithm is grounded on a dichotomy principle, calculating bounds on the matching rent value updated with each response from the DM to a comparison. Moreover, the matching sequences were intertwined (by group of six) in order to hide the matching to participants, as recommended in Fischer et al. (1999).

The indifference situation elicited for a pair of alternatives through a given comparison sequence could result from two events: either the DM directly expressed indifference at a given comparison, or the inferior and superior bounds on the matching value were equal by 15€. In each case, the event ended the matching sequence. In addition, some properties of monotonicity of preferences were used when updating bounds in order to minimize the number of comparisons needed for each matching. The algorithm and the monotonicity properties are formally described in Supplementary data.

The nature of the procedure implies that the number of comparisons varied across participants. The first series was completed in 17 min on average. Participants were then encouraged to take a break.

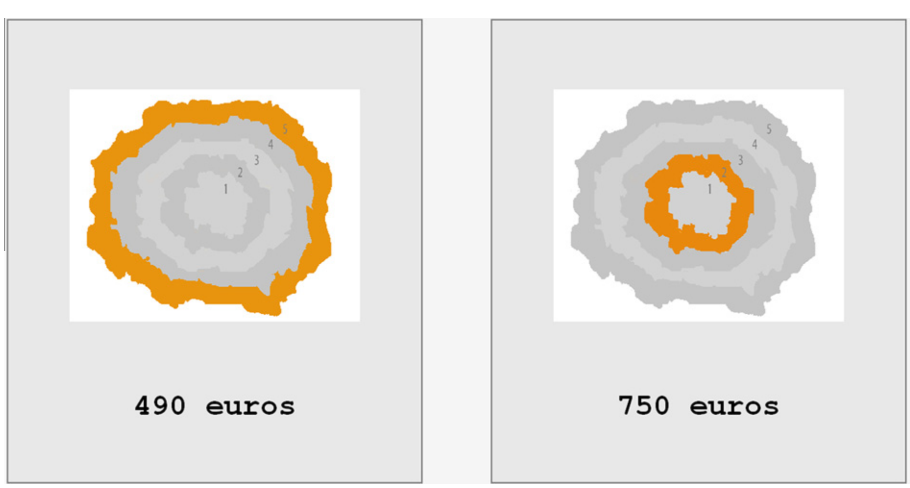


**Fig. 2.** Example of a stimulus during the main task.

**Table 1**
Results for IRP group, $n = 18$, 1006 comparisons.

| OVD | 0 | | | | 1 | | | | 2 | | | | Mean | SD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AS | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | | |
| %P | 74 | 70 | 73 | 57 | 88 | 73 | 66 | 48 | 99 | 91 | 80 | 53 | 73 | 15 |
| %I | 24 | 20 | 4 | 12 | 9 | 16 | 9 | 7 | 0 | 1 | 1 | 0 | 9 | 8 |
| %R | 3 | 10 | 22 | 31 | 4 | 11 | 25 | 45 | 1 | 7 | 19 | 47 | 19 | 15 |

**Table 2**
Results for IP group, $n = 14$, 759 comparisons.

| OVD | 0 | | | | 1 | | | | 2 | | | | Mean | SD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AS | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | | |
| %P | 80 | 76 | 75 | 74 | 91 | 75 | 77 | 66 | 99 | 93 | 80 | 66 | 79 | 10 |
| %I | 20 | 24 | 25 | 26 | 9 | 25 | 23 | 34 | 1 | 7 | 20 | 34 | 21 | |

**Table 3**
Incomparability depending on attribute spread (IRP group).

| AS | 1 | 2 | 3 | 4 | Mean | SD |
|---|---|---|---|---|---|---|
| %R | 3 | 10 | 22 | 41 | 19 | 15 |

### 4.3.3. Main task

Finally, after three filler items to reacclimate them to the experiment, participants completed the second series, made of sixty pairwise comparisons, between alternatives that all belonged to one of the isopreference chains elicited during the first series. Conflict was controlled using the two variables described below. The second series was completed in 8 min on average. The average total duration time of the experiment was 53 min.

## 4.4. Main task design

Our statistical tests involve the difference in overall value between the two alternatives, called *overall value difference* or OVD, and the difference of value on the attribute *distance to city center* between the two alternatives in comparison, called *attribute spread* or AS (see Fig. 3).

- The three elicited chains are characterized by their associated overall value: attractive ($I_1$), neutral ($I_2$) or unattractive ($I_3$). Thus, the observed OVD come in three possible levels: the apartments share the same isopreference chain (OVD = 0), belong to two close isopreference chains (OVD = 1) or belong to the unattractive chain and the attractive chain respectively (OVD = 2). For instance, a comparison between an apartment of attractive overall value (chain $I_1$) and an apartment of neutral overall value (chain $I_2$) has an associated OVD equal to one.
- The attribute spread as defined in the previous sections involves *both* attributes. However, when moving an alternative along one of the isopreference chains, any increase in the difference on *distance to city center* implies an increase in the difference on *rent*. Therefore we used the difference on *distance to city center* as a proxy for attribute spread, and call it *attribute spread* in the following[1]. Since *distance to city center* was a five level attribute, AS had four possible levels.

[1] We chose *distance* rather than *rent* because the distance levels were common to all participants and because discrete AS classes were required to do statistical treatment.

Five stimuli were chosen for each of the twelve possible (AS, OVD) couple, which amounted to sixty comparisons. For some values of (AS, OVD), for instance (4, 0), the number of possible comparisons was less than five. In this case, some randomly chosen comparisons were repeated in order to display exactly five comparisons by possible (AS, OVD) case. The sixty comparisons were displayed in random order during the second series.

## 4.5. Data filtering

Raw data consisted in answers to sixty consecutive pairwise comparisons for each participant. Although each group originally included 21 participants, the data from all participants was not used (see below).

We analysed the effect of AS and OVD on the proportion of each type of answers (P, I, and for the IRP group, R). These proportions were calculated for each participant, broken down by (AS, OVD) values. Then we pooled this data across participants of the same group, and performed a transformation in order to cancel the inter-individual differences.

These data were analysed with Analyses of Variance, including two within-subject factors, AS with four levels and OVD with three levels, and one between subject factor, Group (IP vs. IRP).

A filtering on the data obtained during the main task was made prior to the analysis. Firstly, three participants (2 IP, 1 IRP) were rejected because they used lexicographic preferences under a given rent value. Secondly, we assured that all comparisons taken into account in the analysis fulfilled two conditions:

1. *Distinct levels condition*: For a given *distance to city center* value, two amounts corresponding to adjacent overall values must differ by at least 30€. This condition strengthens the monotonicity imposed in the hidden matching procedure used to elicit alternatives in the first series. We considered any violation of this condition as a preference reversal.
2. *Minimal rent condition*: an amount of rent cannot be worth less than 50€. If a participant's preference indicated that she matched a given amount with a rent that was less than 50€, the matching procedure marked this amount as "over bound" and considered an amount of 50€ in the further steps. This situation occurred typically for some comparisons involving an apartment with bad distance to city center (fifth area) but high overall value.

All comparisons which did not fulfil these two requirements were excluded from further analyses. After this filtering, if there was any empty cell in the 12 (AS, OVD) cases of a participant, the participant was rejected. This excluded two participants from the IRP group and five from the IP group. Finally, the data points comprised 1006 comparisons for the IRP group and 759 comparisons for the IP group.

## 5. Results

Tables 1 and 2 indicate the proportion of each type of answer (P, I, R) as a function of attribute spread (AS) and overall value difference (OVD).

### 5.1. Incomparability

The answer R (incomparability) was used 19% of the time. AS exerted a significant positive effect on the proportion of incomparability ($F(3,51) = 27.1$, $p < 0.0001$, see Table 3).

Attribute spread is the only variable to have a significant impact on R. No significant effect of OVD or significant interaction of OVD with AS could be found ($p = 0.53$ and $p = 0.38$ respectively). This invalidates our hypothesis that the frequency of incomparabilities would decrease with overall value difference. It is sufficient for the apartments to differ strongly on their attribute values for them to be considered as incomparable, even if one of them has been evaluated overall as well better than the other when building the isopreference chains.

### 5.2. Indifference

The distributions of indifference responses depending on AS and OVD radically differ between the two groups (IP and IRP). First, the average proportion of indifference is significantly larger in the IP group than in the IRP group (21% vs 9%, $F(1,30) = 6040.2$, $p = 0.005$).

Second, there is a very different influence of AS depending on the group (interaction: $F(3,90) = 8.0$, $p < 0.0001$, see Fig. 4). In the IP group, AS increases the proportion of indifference while it lightly decreases the proportion of indifference in the IRP group. An ANOVA restricted to each group confirms that each of these effects are significant (increase in the IP group: $F(3,39) = 4.8$, $p = 0.006$; decrease in the IRP group: $F(3,51) = 4.0$, $p = 0.01$).

Third, the near absence of indifference in the IRP group for OVD2 comparisons is consistent with the definition of isopreference chains. On the contrary, the relatively high proportion of indifference in the IP group for such comparisons (16%) is in contradiction with the notion of overall value difference (see Fig. 4).

### 5.3. Strict preferences vs. symmetric responses

In this section we examine how AS, OVD and Group affect the split between asymmetric responses (strict preferences) and symmetric responses (merging indifference and incomparability when available). As any response falls into one of those two categories, we chose to speak about strict preferences here, but the results could be expressed in terms of the proportion of symmetric responses.

First, the higher the difference on the attribute *distance to city center*, the less often participants choose an apartment over the other ($F(3,90) = 20.1$, $p < 0.0001$). When AS is minimal, 12% of statements do not favor an apartment over the other, while this proportion raises to 39% when AS is maximal (see Fig. 5).

Second, consistent with the definition of isopreference chains, the participants choose one apartment more often when its overall value was considered well better than the other during the first series, i.e., when the overall value difference is maximal. ($F(2,60) = 6.4$, $p = 0.003$, see Fig. 5).

Finally, there is no significant effect of Group ($p = 0.21$) nor interaction of the predictor Group with AS ($p = 0.23$), OVD ($p = 0.70$) or AS $\times$ OVD ($p = 0.98$)). Considering the symmetric responses (I in IP and $I \cup R$ in IRP), this result means that the distributions for symmetric answers along AS and OVD are not significantly distinguishable. In other words, the hypothesis stating that the use of I in the IP group and of $I \cup R$ in the IRP group follow the same pattern when OVD and AS varies, cannot be rejected.

### 5.4. Learning and response times

In order to investigate learning effects, we divided the sixty comparisons made by each participant into four quartiles of fifteen consecutive comparisons. Variation of the mean response time across the quartiles was then tested for each type of response (I, R, P).

It turns out that a statistically significant effect of quartile occurs for all types of responses, in both groups, except strict preferences in the IP group. Mean response times tend to decrease globally for each type of response (merging types, 9.4 s for the first quartile vs 7.6 s for the last). This shows that participants got accustomed to the task during the experiment. However, the main results reported in the previous section (in particular, the effect of AS on R and on I) remain significant when neglecting the first 30 comparisons (out of 60) in the series. Therefore, the diminishing response times should not be interpreted as some weariness on the part of participants, that would have lead to automatic or thoughtless responses.

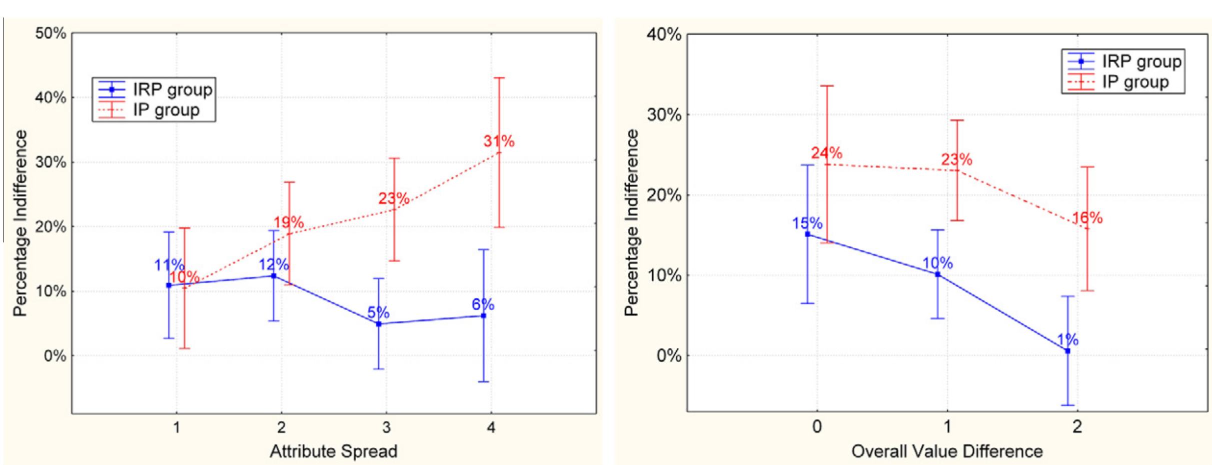


**Fig. 4.** Proportion of indifference I depending on group and (a) AS and (b) OVD.

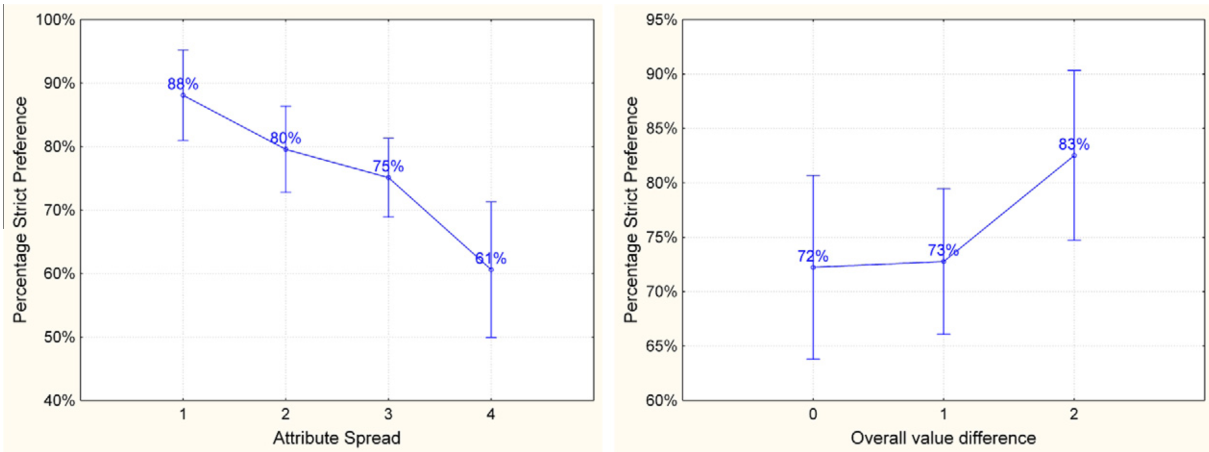


**Fig. 5.** Proportion of strict preferences P depending on (a) AS and (b) OVD, collapsing groups.

## 6. Discussion

The main findings of our experiment are the following: the proportion of incomparability increases significantly with attribute spread, while no effect of overall value difference is observed. There is a significant interaction between the group (IP or IRP) and attribute spread on the proportion of indifference: the proportion of indifference significantly increases in the IP group, whereas it significantly decreases in the IRP group. The proportion of indifference in the IP group behaves, with respect to attribute spread and overall value difference, like the summed proportion of indifference and incomparability in the IRP group.

### 6.1. Conflict-related expression of incomplete preferences

The results support the hypothesis that people can exhibit incomplete preferences in the multicriteria comparison of alternatives. This incompleteness is positively stated by DMs when an explicit incomparability statement is provided in the procedure: according to them, the *incomparability* response reflects their judgement about the comparison of the two alternatives better than the *preference* or *indifference* responses. Participants in the IRP group used the response incomparability in a proportion that is far from negligible (187 responses out of 1006, 19% total) and that did not decrease along the experiment (53 out of 245, 22% when considering only the 15 last comparisons for each participant). This shows that incomparability statements should not be interpreted as a lack of experience about the comparisons, but correspond to structural aspects of decision behavior.

Our attribute spread conflict hypothesis is strongly supported by the data analysis. The proportion of incomparability answers increases with the magnitude of the difference on the attribute *distance to city center* ($F(3,51) = 27.1$, $p < 0.0001$). Affirmative statements of incomplete preferences are more frequent when pairwise comparisons involve more conflictual attribute values. When the options have very contrasted evaluations, DMs have difficulty comparing them. This result extends the work of Tversky and Shafir (1992) to pairwise comparisons. These authors showed that experimental subjects, facing the choice between two products, seek additional options more often when facing conflictual pairs than dominant ones.

Results support the *indiscriminate indifference* hypothesis. Subjects who had no incomparability statement available seem to have used the indifference response both to express indifference (i.e. equality in value) and incomparability (i.e. refusal to make a trade-off). Indeed, we found a proportion of indifference significantly higher in the IP group than in the IRP group (21% vs. 9% on average, $F(1,30) = 9.1$, $p = 0.005$). This difference is largely due to the significant interaction of Group (i.e. the availability of an incomparability statement) on the effect of attribute spread ($F(3,90) = 8.0$, $p < 0.0001$). When participants had indifference as the only symmetric relation, the proportion of indifference increased from 10% to 31% with the level of attribute spread, while it decreased from 11% to 6% in the other group. The increase of indifference in the IP group should be related to the increase of incomparability with attribute spread in the IRP group.

In addition, attribute spread impacts the indifference response time in opposite ways depending on the group (significant interaction: $F(1,14) = 5.8$, $p = 0.03$, see Table 4). When AS is maximal, I statements take more time in the IRP group, and less time in the IP group. Again, a parallel can be drawn, in terms of response time, between the behavior of participants in the IP group, when stating an indifference in response to high conflict, and that of participants in the IRP group when stating an incomparability in the same situation: high AS values significantly decreased the average time to state R in the IRP group ($F(1,13) = 8.6$, $p = 0.01$, see Table 4).

Finally, no test could reject the assertion that symmetric responses were used in the same way in both groups, in spite of the particular meaning of incomparability. The distributions for I and for I ∪ R, in the IP group and IRP group respectively, are not significantly different. No significant effect of Group ($p = 0.208$) nor any interaction of Group with AS ($p = 0.231$), OVD ($p = 0.700$) or AS × OVD ($p = 0.977$) is observed.

Restricting to comparisons between options on the same isopreference chain, the effect of attribute spread on the proportion of stated incomparability remains significant ($p < 0.001$). We interpret this rise of incomparability along isopreference chains as demonstrative of an intransitivity of indifference. Intransitivity of indifference was initially raised by Luce (1956): it concerns the intransitivity of indifference on a single criterion, leading to a strict preference. A sequence of negligible differences can be judged as significant. Here we demonstrate an effect concerning intransitivity of indifference in a multicriteria decision. A sequence of indifference statements leads to a situation of incomparability. That is, cumulating small acceptable trade-offs may result in an unacceptable trade-off.

**Table 4**
Average response time (*s*).

| AS | 1–3 | 4 |
|---|---|---|
| R | 13.3 | 10.9 |
| I (IP group) | 11.7 | 8.5 |
| I (IRP group) | 9.5 | 11.5 |

Such empirical finding deeply questions the concept of indifference curves itself. Indeed, the very meaning of such isopreference curves is grounded on the principle that any pair of alternatives belonging to the same isopreference curve are equal in value and thus are judged indifferent. Our empirical finding shows that such a principle is only valid locally.

The possible intransitivity of indifference along isopreference curves is considered in outranking methods. The indifference relation is not allowed beyond a certain difference in evaluations. This is done by using veto, as has been proposed by Roy (1991). In outranking methods, a threshold defines the maximum difference in value admissible on any criterion. When an alternative $a$ beats an alternative $b$ on a criterion, the difference in evaluation being beyond that threshold, the assertion $a \succcurlyeq b$ is vetoed, which prevents $a$ and $b$ from being considered as indifferent.

Our hypothesis that the proportion of incomparability would decrease with OVD is not supported by the data. The frequency by which subjects express incomparability is not related to the fact that the two alternatives in comparison belong to the same isopreference or differ in overall value. To explain such phenomenon, we suggest that attribute conflict "blurs" isopreference curves, which means that such curves may not be relevant to represent preferences between alternatives that differ radically on at least one attribute. It is difficult for DMs to judge the relative overall values of two alternatives when there is a high attribute conflict.

Chatterjee and Heath (1996) showed an effect of the global attractivity of the alternatives on the decision difficulty. Decisions involving avoidance-avoidance conflict (i.e. alternatives both unattractive) are more difficult than decisions involving approach-approach conflict (i.e alternatives both attractive). Nagpal and Krishnamurthy (2008) showed that this effect is related to the task of choosing and that, symmetrically, unattractive pairs make a rejecting task easier. Using a comparison task, we also observe an effect of the attractivity of the alternatives on the decision difficulty: restricting to pairs of same overall value, incomparability is significantly more frequent when overall value of the options decreased ($F(2,30) = 3.6$, $p = 0.04$). Our experiment replicates these previously observed effects, independently of the exhibited effect of attribute conflict. All the statistical tests of the effect of AS remain significant when restricted to each isopreference set (attractive, neutral and unattractive).

### 6.2. Consistency of the expressed preferences

Considering OVD0-AS1 comparisons, corresponding to pairs of alternatives determined as indifferent during the first block (IPC contruction), a quite high proportion of strict preference can be observed during the second block: 74% IRP group, 80% IP group. This shows a high lability of the preferences expressed by subjects.

Nevertheless, the chains of indifference do bear a definite meaning, as shown by the significant decrease of indifference with OVD in the IRP group (see Fig. 4, $F(2,34) = 12.3$, $p = 0.0001$). Indifference also decreases significantly with OVD in the IP group, when restricting to AS $\leqslant 2$, in order to cancel the $I$ statements with ambiguous interpretation due to high conflict (for OVD = 0,1,2 respectively: 22%, 18%, 5%, $F(2,26) = 4.4$, $p = 0.02$). Despite the observed lability of the expressed preferences, the main results of the experiment remain statistically valid.

Recall that in the first phase of our experiment, series of preference statements are used to bound the rent value of an equivalent until the bounds differ by 15€ or until the subject states an indifference (see subsection 3). Considering indifference pairs elicited by a *stated* indifference only (and not derived from P statements), reliability is better: P proportions are 59% IRP group, 70% IP group.

In order to test the consistency of the isopreference chains, it is also natural to ask: when expressing a strict preference, how often participants do so in accordance with the relative value of the isopreference chains? We predicted that their responses in the second step would contradict the IPCs more often for comparisons with high between-alternative conflict. Restricting to comparisons with OVD $\neq 0$, Table 5 indicates the percentage of P responses consistently favoring the superior isopreference chain (relative to the total of P responses to OVD $\neq 0$ comparisons). A significant interaction between AS and OVD can be observed ($F(3,60) = 4.9$, $p = 0.004$). Preference in the predicted direction is more frequent for OVD2 than OVD1 ($F(1,20) = 19.0$, $p = 0.0003$). The ratio of "inconsistent" preferences increases with AS, but this effect is only significant for OVD1 comparisons (restricted to OVD1: $F(3,60) = 8.1$, $p = 0.0001$). First, this shows that IPCs are robust enough to predict the direction of choice with a high rate in spite of the lability of the preferences. Second, the results show that the lability is itself affected by conflict: judgment uncertainty increases with *between-alternative* conflict, which must be put in perspective with what is shown in Fischer et al. (2000): that judgment uncertainty increases with *within-alternative* conflict.

Slovic (1975) reports experiments, where subjects express indifference between alternatives, and then are forced to choose between them, or to rank order them. His experiments showed that the most important attribute is favored in the forced task. Our experimental set-up also involves a first elicitation phase that yields indifference sets, followed by a second elicitation phase focusing on the alternatives composing these sets. However, indifference is not available in the second phase of Slovic's experiments, while possible responses are I, P (or I, P, R) in ours.

How do our results compare with Slovic (1975)? There is no direct way of formally assessing the effect found by Slovic on our data, because the experiment was not designed with this purpose: we did not ask participants what was the most important attribute. Moreover, the individual number of OVD0-AS1 comparisons is too small (5 or less by subject) to carry out intrasubject statistical analyses. However, at the aggregate level, we observe a significant asymmetry in the criterion favored by the P statements. We should expect choices to be close to random among indifferent alternatives. Yet 64% of preferences favored the alternative that was better on the criterion *distance to city center* (two-tailed binomial test: $p = 0.005$, 70 out of 110 P responses to OVD0-AS1 comparisons). When restricting to pairs elicited by *stated* indifference, the asymmetry is even stronger (73%, 36 out of 49 P responses, $p = 0.001$). This asymmetry makes the *distance to city center* attribute, de facto, the *prominent* attribute, i.e. the attribute considered the most important. Tversky et al. (1988) and Fischer and Hawkins (1993) have observed in numerous interpersonal set-ups that when choosing among alternatives considered as indifferent by matching, DMs favor the one that is superior on the *prominent* attribute.

During the main task, some comparisons were repeated (between 2 and 5 times, number of comparisons: 142 IP group, 184 IRP group). 25% did not yield the same answer on all occurences. Stott (2006) analyses several prospect theory elicitation experiments and provides insights on the consistency of the elicited preferences. Theses experiments elicit preferences using binary choices (i.e., allowing preference only), with some choices presented twice. Choice reversal rate varies between 12% and 32%. In our experiment, we observe a similar inconsistency rate: 25% although we consider more than two repetitions and more possible answers (P, $P^{-1}$, but also I and R).

**Table 5**
OVD $\neq 0$ comparisons, %P responses favoring the superior IPC.

| AS | 1 | 2 | 3 | 4 | Mean | SD |
|---|---|---|---|---|---|---|
| OVD1 | 96 | 77 | 70 | 63 | 76 | 12 |
| OVD2 | 97 | 96 | 91 | 90 | 93 | 3 |

### 6.3. Implications for preference elicitation

Considering that a model has been chosen to represent the DM's preferences, elicitation consists in questioning him/her in order to specify the parameters of the model. This parametrization allows a faithful representation of how the DM compares the alternatives and thus permits to build relevant recommendations. Several dangers in directly asking for the values of preferential parameters have been pointed out in the literature (for instance Podinovski (1994)). Therefore most authors recommend to use indirect elicitation methods. These methods infer the values of the parameters from preference judgements.

Indirect elicitation ensures that the values taken by the preferential parameters are consistent with the semantics that the aggregation model assign to these parameters. In particular, in the multiattribute value theory (Keeney and Raiffa, 1976), the weight of the attribute $X$ represents the value of moving attribute $X$ from its worst to best level, relatively to doing the same on other attributes. Therefore, many elicitation procedures involve judgements about the range of attribute scales, using fictitious alternatives with extreme evaluations on criteria.

For example, the swing weights procedure, used in the SMARTS and SMARTER methods presented in Edwards and Barron (1994), involves comparisons of fictitious alternatives with the best evaluation on some attribute and the worst on all other attributes. It is obvious that such tasks involve high, and possibly the highest, bicriteria conflict. Such difficulty also occurs in the method proposed by Keeney and Raiffa (1976) to elicit weights. MACBETH, proposed by Bana e Costa and Vansnick (1994), also involves comparisons between fictitious alternatives that differ on at most two criteria to elicit weights. But the fictitious alternatives do not have the maximum or minimum evaluation on each attribute. Instead, the performances on each attribute are fixed at a reference level chosen by the analyst (often a *neutral* and an *attractive* level). This permits to assign comparisons with a controlled between-alternative conflict.

Our results suggest that in procedures that do not allow the expression of incomplete preferences, high caution must be taken in the choice of the comparisons to be presented to the DM. In particular, assigning comparisons with high attribute conflict can be dangerous because an indifference response will raise interpretation problems. It could either mean that she considers the alternatives as equivalent, or that she has difficulty in comparing them. Hence, these answers can hardly be understood as a measure of weights in such models. This recommendation does not apply to methods that allow for incomplete preferences, like those involving veto.

Delquié (2003) also advised against high conflict in preference elicitation. More specifically, he argued that the trade-off size, i.e. conflict, in an assessment question increases its difficulty (hence the error in the measurement of preferences) but decreases the error in the estimation of the parameter describing the indifference curves (because of the graphical layout of indifference curves). He proposed an analytical method to determine the optimal conflict (minimizing the total error) in an assessment question.

Another interpretation of the results would state that it is unwise to prescribe choices involving high conflict based on models calibrated using low conflict comparisons. In order to prescribe decisions involving highly conflicting alternatives, one could consider eliciting high-conflict trade-offs provided that incomparability is given as a response option during the elicitation process.

### 6.4. Cognitive strategies underlying the expression of incomplete preferences

Beyond the results observed in the experiment, on how subjects express incomplete preferences due to between-alternative conflict, an important issue raised is related to understanding the cognitive processes underlying the expression of such preferences. Do the strategies differ strongly depending on the availability of the "incomparability" answer? When comparing the options, do participants engage in several parallel tasks, relating to each possible answer, or do they use a sequential strategy to reach a decision? What is the role of metacognition in answering a comparison task?

No clear conclusion about the strategies used by the subjects can be drawn from the data. The experiment was not designed initially with this purpose. However, the data relating to response times reveal that it took significantly longer to express a symmetric judgment (indifference or incomparability) than an asymmetric judgment (preference). The average group response times for expressing P were similar in the IP group (7.6 s) and the IRP group (8.0 s). Likewise, there were very close average group response times for I in the IP group (13.4 s), I in the IRP group (11.6 s), and R (11.7 s). Interpreting these observations would require further experimentation. Understanding the cognitive processes that are at stake when participants answer pairwise comparisons may bring some crucial knowledge relating to incomparability as expressed by decision-makers.

Other approaches should also be considered to further investigate the cognitive processes involved in the formulation of comparison judgments. A recent approach based on the functional Magnetic Resonance Imaging (fMRI) technique aims at establishing a map of the brain by identifying areas carrying out specific cognitive functions. This approach gave some results concerning multiattribute decision (Zysset et al., 2006) and have shown the activation of the anterior cingulate cortex (ACC) of subjects when facing conflict between single-valued alternatives (see Pochon et al., 2008). An fMRI study of the multiattribute between-alternative conflict is still needed.

## 7. Conclusion

We conducted an experiment to investigate the expression of incomplete preferences in relation with between-alternative conflict. We observed that the proportion of incomparability increased with attribute spread between the alternatives. Furthermore, several results support that participants who had no incomparability statement available used the indifference response to express such incomparability. In particular, the proportion of indifference statements significantly increased with attribute spread for these participants, while it significantly decreased for the others. Furthermore, the pattern of indifference statements when no incomparability is available is similar to the pattern of merged indifference and incomparability statements when incomparability is available. These results show a form of intransitivity of indifference along isopreference curves, due to a lack of confidence in comparing options when the multicriteria conflict is high. Our results suggest that preference elicitation should avoid capturing preferences using pairwise comparisons involving high between-alternative conflict, or if doing so, incomparability should be included as a response option in preference elicitation. Moreover, an incomparability option would help determine the limits of the trade-offs that the respondent is willing to resolve, and avoid overstepping the bounds. Further experimental investigation should be conducted to better understand how conflict impacts the decision strategies involved when comparing multicriteria alternatives.

## Acknowledgment

The authors gratefully acknowledge Alexandre Baron for his comments on a previous draft of this paper.

## Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ejor.2012.03.041.

## References


Chatterjee, S., Heath, T., 1996. Conflict and loss aversion in multiattribute choice: the effects of trade-off size and reference dependence on decision difficulty. Organizational Behavior and Human Decision Processes 67, 144–155.

Bana e Costa, C., Vansnick, J., 1994. Macbeth – an interactive path towards the construction of cardinal value functions. International Transactions in Operational Research 1, 489–500.

Danan, E., 2003. Revealed Cognitive Preference Theory. Technical Report EUREQua, Université Paris 1.

Danan, E., Ziegelmeyer, A., 2004. Are Preferences Incomplete?: An Experimental Study Using Flexible Choices. Max-Planck-Inst. for Research into Economic Systems, Strategic Interaction Group.

Delquié, P., 2003. Optimal conflict in preference assessment. Management Science 49, 102–115.

Dubra, J., Maccheroni, F., Ok, E., 2004. Expected utility theory without the completeness axiom. Journal of Economic Theory 115, 118–133.

Edwards, W., Barron, F., 1994. SMARTS and SMARTER: improved simple methods for multiattribute utility measurement. Organizational Behavior and Human Decision Processes 60, 306–325.

Fischer, G., Carmon, Z., Ariely, D., Zauberman, G., 1999. Goal-based construction of preferences: task goals and the prominence effect. Management Science, 1057–1075.

Fischer, G., Hawkins, S., 1993. Strategy compatibility, scale compatibility, and the prominence effect. Journal of Experimental Psychology: Human Perception and Performance 19, 580.

Fischer, G., Luce, M., Jia, J., 2000. Attribute conflict and preference uncertainty: effects on judgment time and error. Management Science, 88–103.

Fishburn, P., 1985. Interval Orders and Interval Graphs. John Wiley and Sons.

Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. Econometrica: Journal of the Econometric Society, 263–291.

Keeney, R., Raiffa, H., 1976. Decisions with Multiple Objectives: Preferences and Value Tradeoffs. J. Wiley, New York.

Kühberger, A., Schulte-Mecklenbeck, M., Perner, J., 2002. Framing decisions: hypothetical and real. Organizational Behavior and Human Decision Processes 89, 1162–1175.

Lichtenstein, S., Slovic, P., 2006. The Construction of Preference. Cambridge University Press.

Luce, M., 1998. Choosing to avoid: coping with negatively Emotion–Laden consumer decisions. Journal of Consumer Research 24, 409–433.

Luce, M., Payne, J., Bettman, J., 2000. Coping with unfavorable attribute values in choice. Organizational Behavior and Human Decision Processes 81, 274–299.

Luce, R., 1956. Semiorders and a theory of utility discrimination. Econometrica, Journal of the Econometric Society 24, 178–191.

Nagpal, A., Krishnamurthy, P., 2008. Attribute conflict in consumer decision making: the role of task compatibility. Journal of Consumer Research, 696–705.

Pochon, J., Riis, J., Sanfey, A., Nystrom, L., Cohen, J., 2008. Functional imaging of decision conflict. Journal of Neuroscience 28, 3468.

Podinovski, V., 1994. Criteria importance theory. Mathematical Social Sciences 27, 237–252.

Roy, B., 1991. The outranking approach and the foundations of ELECTRE methods. Theory and Decision 31, 49–73.

Roy, B., 1996. Multicriteria Methodology for Decision Aiding. Kluwer, Dordrecht.

Simon, H., 1955. A behavioral model of rational choice. The quarterly journal of economics 69, 99–118.

Slovic, P., 1975. Choice between equally valued alternatives. Journal of Experimental Psychology: Human Perception and Performance 1, 280.

Stott, H., 2006. Cumulative prospect theory's functional menagerie. Journal of Risk and Uncertainty 32, 101–130.

Tetlock, P., Boettger, R., 1994. Accountability amplifies the status quo effect when change creates victims. Journal of Behavioral Decision Making 7, 1–23.

Tversky, A., Sattath, S., Slovic, P., 1988. Contingent weighting in judgment and choice. Psychological review 95, 371.

Tversky, A., Shafir, E., 1992. Choice under conflict: the dynamics of deferred decision. Psychological Science 3, 358–361.

Von Neumann, J., Morgenstern, O., 1944. Theory of Games and Economic Behavior.

von Winterfeldt, D., Edwards, W., 1986. Decision Analysis and Behavioral Research. Cambridge University Press, Cambridge.

Zysset, S., Wendt, C., Volz, K., Neumann, J., Huber, O., Von Cramon, D., 2006. The neural implementation of multi-attribute decision making: a parametric fMRI study with human subjects. NeuroImage 31, 1380–1388.