A Psychological Study of Comparative Non-monotonic Preferences Semantics

Rui da Silva Neves
Université Toulouse-II,
CLLE-LTC, CNRS UMR 5263
5 Allées Machado
31058 Toulouse Cedex 9, France
neves@univ-tlse2.fr

Souhila Kaci
Université Lille-Nord de France, Artois
CRIL, CNRS UMR 8188
IUT de Lens
F-62307, France
kaci@cril.fr

Abstract

Representing preferences and reasoning about them are important issues for many real-life applications. Several monotonic and non-monotonic qualitative formalisms have been developed for this purpose. Most of them are based on comparative preferences, for e.g. "I prefer red wine to white wine". However this simple and natural way to express preferences comes also with many difficulties regarding their interpretation. Several (more or less strong) semantics have been proposed leading to different (pre)orders on outcomes. In this paper, we report results of the first empirical comparison of existing non-monotonic semantics (strong, optimistic, pessimistic and ceteris paribus) based on psychological data. Thirty participants were asked to rank 8 menus according to their preferences and to compare 31 pairs of menus. The recorded preferences allowed to compute compact preferences and ranking menus for each participant according to the four semantics under study, and to compare these ranks to participant's ones. Results show that non-monotonic optimistic and pessimistic preferences are the semantics that better fit human data, strong and ceteris paribus semantics being less psychologically plausible given our task.

Introduction

Preferences are very useful in many real-life problems. They are inherently a multi-disciplinary topic, of interest to economists, computer scientists, operations researchers, mathematicians, logicians, philosophers and psychologists.

It has been early recognized that value functions/orderings cannot be explicitly defined because of a great number of outcomes or simply because the user is not willing to state her/his preferences on each pair of outcomes. Indeed preferences should be handled in a compact (or succint) way, starting from non completely explicit preferences expressed by a user.

The compact languages for preference representation have been extensively developed in Artificial Intelligence in the last decade (Boutilier et al. 2004; Brewka, Benferhat, and Le Berre 2004). In particular, (conditional) comparative statements are often used for describing preferences in a local, contextualized manner for e.g., "I prefer fish to meat", "if meat is served then I prefer red wine to white wine", etc. Indeed, it is easier and more natural to express such qualitative comparative statements than to say that I prefer fish with the weight .8 and prefer meat with the weight .2. Some generic principles are often used for completing the qualitative comparative preference statements[1] (Hansson 1996; Boutilier 1994; Benferhat et al. 2002). Although comparative preference statements allow for a simple and natural way to express preferences, they come however with many difficulties regarding their interpretation.

Comparative preferences are often interpreted following the well known ceteris paribus semantics (Hansson 1996). This is due to the CP-net approach (Boutilier et al. 2004) which has emerged in the last decade as the preeminent and prominent method for processing preferences in Artificial Intelligence, thanks to its intuitive appeal. Following this pinciple, the statement "I prefer fish to meat" is interpreted as, given two meals that differ *only* in the main dish, the meal with fish is preferred to the meal with meat. However, CP-nets behave monotonically and do not allow for the handling of preferences with defaults. For example, we can prefer fish to meat, but when available fish is red tuna and meat is poultry, we can prefer the reverse. Moreover, in CP-nets, ceteris paribus semantics states that the two meals *fish-cake* and *meat-ice cream* are incomparable w.r.t. the preference statement "I prefer fish to meat" while a vegetarian would prefer any fish-based meal to any meat-based meal. Fortunately, ceteris paribus is not the only possible reading of comparative preference statements and other intuitively non-monotonic meaningful semantics may also be encountered, and researchers have also argued for other semantics (Boutilier 1994; Benferhat et al. 2002) based on insights from non-monotonic reasoning such as system Z (Pearl 1990). Note also that ceteris paribus semantics can also be non-monotonic outside CP-net framework. For example, the menu $fish - red$ is preferred to the menu $fish - white$ w.r.t. the preference statement "red is preferred to $\neg red$" following ceteris paribus semantics. However the additional preference statement "$fish \wedge white$ is preferred to $fish \wedge \neg white$" induces the reverse preference, namely $fish - white$ is preferred to $fish - red$.

In this paper, we provide the first empirical compar-

[1] From now on, we simply speak about comparative preference statements.

ison of existing non-monotonic semantics (including ceteris paribus) based on psychological data. This psychological inquiry is founded by previous work on the non-monotonic nature of human reasoning. For example, it has been shown that human inference is consistent with System P (Kraus, Lehmann, and Magidor 1990) (see (Neves, Bonnefon, and Raufaste 2002; Benferhat, Bonnefon, and Da Silva Neves 2004)) and that System P constitutes a psychologically sound base of rationality postulates for the evaluation of non-monotonic reasoning systems. In our study, participants were asked to rank 8 menus according to their preferences and to compare 31 pairs of menus. The recorded preferences were compared to those provided by the considered semantics. Results show that optimistic and pessimistic preferences are the semantics that better fit human data, strong, ceteris paribus semantics being less psychologically plausible given our task.

The remainder of this paper is organized as follows. After providing notations and necessary definitions, we recall the different semantics of comparatives preferences proposed in literature. Then, we recall algorithms to rank-order outcomes for each semantics. In the next section, we provide empirical comparison of the different semantics based on psychological data. Lastly we conclude.

Notations

Let $V = \{X_1, \cdots, X_h\}$ be a set of h variables. Each variable X_i takes its values in a domain $Dom(X_i)$ which is a set of uninterpreted constants $\mathcal{D}$ or rational numbers $\mathcal{Q}$. A possible outcome, denoted t, is the result of assigning a value in $Dom(X_i)$ to each variable X_i in V. Ω is the set of all possible outcomes. We suppose that this set is fixed and finite. Let $\mathcal{L}$ be a language based on V. $Mod(\varphi)$ denotes the set of outcomes that make the formula φ (built on $\mathcal{L}$) true. We write $t \models \varphi$ when $t \in Mod(\varphi)$ and say that t satisfies φ.

An ordering relation $\succeq$ on $\mathcal{X} = \{x, y, z, \cdots\}$ is a reflexive binary relation such that $x \succeq y$ stands for x is at least as preferred as y. $x \approx y$ means that both $x \succeq y$ and $y \succeq x$ hold, i.e., x and y are equally preferred. Lastly $x \sim y$ means that neither $x \succeq y$ nor $y \succeq x$ holds, i.e., x and y are incomparable. A strict ordering relation on $\mathcal{X}$ is an irreflexive binary relation such that $x \succ y$ means that x is strictly preferred to y. We also say that x dominates y. A strict ordering relation $\succ$ can be defined from an ordering relation $\succeq$ as $x \succ y$ if $x \succeq y$ holds but $y \succeq x$ does not.
When neither $x \succ y$ nor $y \succ x$ holds, we also write $x \sim y$. $\succeq$ (resp. $\succ$) is a preorder (resp. order) on $\mathcal{X}$ if and only if $\succeq$ (resp. $\succ$) is transitive, i.e., if $x \succeq y$ and $y \succeq z$ then $x \succeq z$ (if $x \succ y$ and $y \succ z$ then $x \succ z$). $\succeq$ (resp. $\succ$) is a complete preorder (resp. order) if and only if $\forall x, y \in \mathcal{X}$, we have either $x \succeq y$ or $y \succeq x$ (resp. either $x \succ y$ or $y \succ x$).
The set of the best (or undominated) elements of $A \subseteq \mathcal{X}$ w.r.t. $\succ$, denoted $\max(A, \succ)$, is defined by $\max(A, \succ) = \{x | x \in A, \nexists y \in A, y \succ x\}$. The set of the worst elements of $A \subseteq \mathcal{X}$ w.r.t. $\succ$, denoted $\min(A, \succ \ )$, is defined by $\min(A, \succ \ ) = \{x | x \in A, \nexists y \in A, x \succ y\}$. The best (resp. worst) elements of A w.r.t. $\succeq$ is $\max(A, \succ)$ (resp. $\min(A, \succ)$) where $\succ$ is the strict ordering relation associated to $\succeq$.

A complete preorder $\succeq$ can also be represented by a well ordered partition of Ω. This is an equivalent representation, in the sense that each preorder corresponds to one ordered partition and vice versa.

Definition 1 (Partition) *A sequence of sets of outcomes of the form $(E_1, \ldots, E_n)$ is a partition of Ω if and only if (i) $\forall i$, $E_i \neq \emptyset$, (ii) $E_1 \cup \cdots \cup E_n = \Omega$, and (iii) $\forall i, j$, $E_i \cap E_j = \emptyset$ for $i \neq j$.*

A partition of Ω is ordered if and only if it is associated with a preorder $\succeq$ on Ω such that ($\forall t, t' \in \Omega$ with $t \in E_i, t' \in E_j$ we have $i \leq j$ if and only if $t \succeq t'$).

Comparative preference statements

We denote comparative statements of the form "I prefer p to q" as $p > q$ and denote conditional (called also contextual) comparative statements of the form "if r is true then I prefer p to q" as $r : p > q$, where p, q and r are any propositional formulas.

Comparative statements come with difficulties regarding their interpretation. How should we interpret such statements? For example, given the preference statement "I prefer fish to meat", how do we rank-order meals based on fish and those based on meat? Four semantics have been proposed in literature:

- *ceteris paribus preferences:* (Hansson 1996)
  any fish-based meal is preferred to any meat-based meal if the two meals are exactly the same elsewhere (for example wine and dessert).
- *strong preferences:* (Boutilier 1994)
  any fish-based meal is preferred to any meat-based meal.
- *optimistic preferences:* (Benferhat, Dubois, and Prade 1992; Boutilier 1994; Pearl 1990)
  at least one fish-based meal is preferred to all meat-based meals.
- *pessimistic preferences:* (Benferhat et al. 2002)
  at least one meat-based meal is less preferred to all fish-based meals.

We define preference of the formula p over the formula q as preference of $p \wedge \neg q$ over $\neg p \wedge q$. This is standard and known as von Wright's expansion principle (von Wright 1963). Additional clauses may be added for the cases in which sets of outcomes are nonempty, to prevent the satisfiability of preferences like $p > \top$ and $p > \bot$. We do not consider this borderline condition to keep the formal machinery as simple as possible. We denote the preference of p over q following strong semantics (resp. ceteris paribus, optimistic, pessimistic) by $p >_{st} q$ (resp. $p >_{cp} q$, $p >_{opt} q$, $p >_{pes} q$).

Definition 2 *Let p and q be two propositional formulas and $\succeq$ be a preorder on Ω.*

- *$\succeq$ satisfies $p >_{st} q$, denoted $\succeq \models p >_{st} q$, iff $\forall t \models p \wedge \neg q$, $\forall t' \models \neg p \wedge q$ we have $t \succ t'$.*
- *$\succeq$ satisfies $p >_{cp} q$, denoted $\succeq \models p >_{cp} q$, iff $\forall t \models p \wedge \neg q$, $\forall t' \models \neg p \wedge q$ we have $t \succ t'$, where t and t' have the same assignment on variables that do not appear in p and q.*

- $\succeq$ *satisfies* $p >_{opt} q$, *denoted* $\succeq\models\ p >_{opt} q$, *iff* $\exists t \models p \wedge \neg q, \forall t' \models \neg p \wedge q$ *we have* $t \succ t'$.
- $\succeq$ *satisfies* $p >_{pes} q$, *denoted* $\succeq\models\ p >_{pes} q$, *iff* $\exists t' \models \neg p \wedge q, \forall t \models p \wedge \neg q$ *we have* $t \succ t'$.

A preference set is a set of preferences of the same type.

Definition 3 (Preference set) *A preference set of type* $\rhd$, *denoted* $\mathcal{P}_\rhd$, *is a set of preferences of the form* $\{p_i \rhd q_i | i = 1, \cdots, n\}$, *where* $\rhd \in \{ >_{st}, >_{cp}, >_{opt}, >_{pes} \}$. *A complete preorder* $\succeq$ *is a model of* $\mathcal{P}_\rhd$ *if and only if* $\succeq$ *satisfies each preference* $p_i \rhd q_i$ *in* $\mathcal{P}_\rhd$.

A set $\mathcal{P}_\rhd$ is consistent if it has a model.

From comparative preference statements to preorders on outcomes

Generally we have to deal with several comparative preference statements expressed by a user. Once the semantics is fixed, the problem to tackle is how to deal with such statements? Several types of queries can be asked about preferences: what are the preferred outcomes? Is one outcome better than the other? In many applications (for e.g. database queries), users are more concerned with the preferred outcomes. However preferred outcomes are not always feasible. For example the best menus w.r.t. a user's preferences may be no longer available so we have to look for menus that are immediately less preferred w.r.t. user's preferences. In such a case a complete preorders on menus is needed to answer user's preferences. Indeed we restrict ourselves to semantic models that derive complete preorders on outcomes. In the following, we recall algorithms which derive a unique complete preorder given a set of preferences of the same type w.r.t. specificity principle (Yager 1983).

Let $\mathcal{P}_\rhd = \{s_i : p_i \rhd q_i | i = 1, \cdots, n\}$ be a preference set with $\rhd \in \{ >_{st}, >_{cp}, >_{opt}, >_{pes} \}$. Given $\mathcal{P}_\rhd$, we define a set of pairs on Ω as follows:

$$\mathcal{L}(\mathcal{P}_\rhd) = \{C_i = (L(s_i), R(s_i)) | i = 1, \cdots, n\},$$

where $L(s_i) = \{t | t \in \Omega, t \models p_i \wedge \neg q_i\}$ and $R(s_i) = \{t | t \in \Omega, t \models \neg p_i \wedge q_i\}$.

Example 1 *Let* dish, wine *and* dessert *be three variables such that* $Dom(dish) = \{fish, meat\}$, $Dom(wine) = \{white, red\}$ *and* $Dom(dessert) = \{cake, ice\_cream\}$. *We have* $\Omega = \{t_0 = fish - white - ice\_cream,$
$t_1 = fish - white - cake, t_2 = fish - red - ice\_cream,$
$t_3 = fish - red - cake, t_4 = meat - white - ice\_cream,$
$t_5 = meat - white - cake, t_6 = meat - red - ice\_cream,$
$t_7 = meat - red - cake\}$.
Let $\mathcal{P}_\rhd = \{s_1 : fish \rhd meat, s_2 : red \wedge cake \rhd white \wedge ice\_cream, s_3 : fish \wedge white \rhd fish \wedge red\}$. *We have* $\mathcal{L}(\mathcal{P}_\rhd) = \{C_1 = (\{t_0, t_1, t_2, t_3\}, \{t_4, t_5, t_6, t_7\}),$
$C_2 = (\{t_3, t_7\}, \{t_0, t_4\}), C_3 = (\{t_0, t_1\}, \{t_2, t_3\})\}$.

Optimistic preferences

Several complete preorders may satisfy a set of optimistic preferences. It is however possible to characterize a unique preorder among them under certain assumption. The semantics of optimistic preferences is close to the one of conditionals. Indeed system Z (Pearl 1990) has been used (Benferhat, Dubois, and Prade 1992; Boutilier 1994). It rank-orders outcomes under the assumption that outcomes are preferred unless the contrary is stated. Indeed outcomes are put in the highest possible rank in the preorder while being consistent with preferences at hand. This principle ensures that the complete preorder is unique and the most compact one among preorders satisfying the set of preferences[2]. Algorithm 1 gives the way this preorder is computed. At each step of the algorithm, we put in E_i outcomes that are not dominated by any other outcomes. These outcomes are those which do not appear in the right-hand side of any pair $(L(s_i), R(s_i))$ of $\mathcal{L}(\mathcal{P}_{>_{opt}})$.

Algorithm 1: A complete preorder associated with $\mathcal{P}_{>_{opt}}$.

Data: A preference set $\mathcal{P}_{>_{opt}}$.
Result: A complete preorder $\succeq$ on Ω.

```
begin
    l = 0
    while Ω ≠ ∅ do
        l = l + 1
        E_l = {t | t ∈ Ω, ∄(L(s_i), R(s_i)) ∈ L(P_>opt), t ∈ R(s_i)}
        if E_l = ∅ then
            stop (inconsistent preferences), l = l − 1
        - Ω = Ω\E_l
        /** remove satisfied preferences **/
        - remove (L(s_i), R(s_i)) where L(s_i) ∩ E_l ≠ ∅
    return ⪰= (E_1, ··· , E_l).
end
```

Example 2 *(Example 1 con'd) We have* $E_1 = \{t_1\}$. *We remove* C_1 *and* C_3 *since* $s_1 = fish >_{opt} meat$ *and* $s_3 : fish \wedge white >_{opt} fish \wedge red$ *are satisfied. We get* $\mathcal{L}(\mathcal{P}_{>_{opt}}) = \{C_2 = (\{t_3, t_7\}, \{t_0, t_4\})\}$. *Now* $E_2 = \{t_2, t_3, t_5, t_6, t_7\}$. *We remove* C_2 *since* $s_2 : red \wedge cake >_{opt} white \wedge ice\_cream$ *is satisfied. So* $\mathcal{L}(\mathcal{P}_{>_{opt}}) = \emptyset$. *Lastly,* $E_3 = \{t_0, t_4\}$. *Indeed* $\succeq = (\{t_1\}, \{t_2, t_3, t_5, t_6, t_7\}, \{t_0, t_4\})$. *We can check that each outcome has been put in the highest possible rank in* $\succeq$. *Therefore, if we push an outcome to a higher rank then the preorder does not satisfy the preference set. For example,* $\succeq' = (\{t_1, t_5\}, \{t_2, t_3, t_6, t_7\}, \{t_0, t_4\})$ *does not satisfy* $s_1 = fish >_{opt} meat$.

Pessimistic preferences

The converse reasoning is drawn when dealing with pessimistic preferences (Benferhat et al. 2002). The basic principle is that outcomes are not preferred unless the contrary is stated. Indeed outcomes are put in the lowest possible rank in the preorder while being consistent with preferences at

[2] Technically speaking, this preorder can be obtained by max-based aggregation operator of all preorders satisfying the set of preferences

hand. This principle also ensures that the complete preorder is unique and the most compact one among preorders satisfying the set of preferences[3]. Algorithm 2 gives the way this preorder is computed.

Algorithm 2: A complete preorder associated with $\mathcal{P}_{>_{pes}}$.

Data: A preference set $\mathcal{P}_{>_{pes}}$.

Result: A complete preorder $\succeq$ on Ω.

begin
  $l = 0$
  while $\Omega \neq \emptyset$ **do**
    $l = l + 1$
    $E_l = \{t | t \in \Omega, \nexists (L(s_i), R(s_i)) \in \mathcal{L}(\mathcal{P}_{>_{pes}}), t \in L(s_i)\}$
    if $E_l = \emptyset$ **then**
      stop (inconsistent preferences), $l = l - 1$
    - $\Omega = \Omega \backslash E_l$
    /** remove satisfied preferences **/
    - remove $(L(s_i), R(s_i))$ where $R(s_i) \cap E_l \neq \emptyset$
  return $\succeq = (E'_1, \cdots, E'_l)$ *s.t.* $0 \leq h \leq l$, $E'_h = E_{l-h+1}$
end

Example 3 *(Example 1 con'd) We have $E_1 = \{t_4, t_5, t_6\}$. We remove C_1 and C_2 since $s_1 : fish >_{pes} meat$ and $s_2 : red \wedge cake >_{pes} white \wedge ice\_cream$ are satisfied. We repeat the same reasoning and get $E_2 = \{t_2, t_3, t_7\}$ and $E_3 = \{t_0, t_1\}$. So $\succeq = (\{t_0, t_1\}, \{t_2, t_3, t_7\}, \{t_4, t_5, t_6\})$. We can check that each outcome has been put in the lowest possible rank in the preorder.*

Strong preferences

Strong preferences induce a unique partial order on outcomes. We can use both construction principles used in optimistic and pessimistic preferences to linearize the partial order and compute a unique complete preorder. Algorithms 1 and 2 can be adapted to deal with strong preferences. Due to the lack of space, we only give the algorithm adapting Algorithm 1.

Example 4 *(Example 1 con'd) There is no complete preorder which satisfies $\mathcal{P}_{>_{st}}$ so $\mathcal{P}_{>_{st}}$ is inconsistent. This is due to s_1 and s_2. Following s_1, t_0 is preferred to t_7 while t_7 is preferred to t_0 following s_2.*

Example 5 *(Consistent strong preferences) Let $\mathcal{P}_{>_{st}} = \{fish \wedge white >_{st} fish \wedge red, red \wedge cake >_{st} red \wedge ice\_cream, meat \wedge red >_{st} meat \wedge white\}$. Then following Algorithm 3, we have $\succeq = (\{t_0, t_1, t_7\}, \{t_3\}, \{t_2, t_6\}, \{t_4, t_5\})$. Now following the adaptation of Algorithm 2 to deal with strong preferences, we have $\succeq = (\{t_0, t_1\}, \{t_3, t_7\}, \{t_6\}, \{t_2, t_4, t_5\})$.*

Ceteris paribus preferences

These preferences are similar to strong preferences. They also induce a unique partial order on outcomes. We can also use both construction principles used in optimistic and pessimistic semantics to compute a unique complete preorder.

[3]Technically speaking, this preorder can be obtained by min-based aggregation operator of all preorders satisfying the set of preferences.

Algorithm 3: A complete preorder associated with $\mathcal{P}_{>_{st}}$.

Data: A preference set $\mathcal{P}_{>_{st}}$.

Result: A complete preorder $\succeq$ on Ω.

begin
  $l \leftarrow 0$
  while $\Omega \neq \emptyset$ **do**
    $l = l + 1$
    $E_l = \{t | t \in \Omega, \nexists (L(s_i), R(s_i)) \in \mathcal{L}(\mathcal{P}_{>_{st}}), t \in R(s_i)\}$
    if $E_l = \emptyset$ **then**
      stop (inconsistent preferences), $l = l - 1$
    - $\Omega = \Omega \backslash E_l$
    - replace $(L(s_i), R(s_i))$ by $(L(s_i) \backslash E_l, R(s_i))$
    /** remove satisfied preferences **/
    - remove $(L(s_i), R(s_i))$ where $L(s_i) = \emptyset$
  return $\succeq = (E_1, \cdots, E_l)$.
end

Example 6 *(Example 1 con'd) Following the gravitation towards the ideal we have $\succeq = (\{t_1\}, \{t_3, t_5\}, \{t_0, t_7\}, \{t_2, t_4\}, \{t_6\})$ while following the gravitation towards the worst we have $\succeq = (\{t_1\}, \{t_3\}, \{t_0\}, \{t_2, t_7\}, \{t_4, t_5, t_6\})$.*

Experimental Study

Our main objective is to evaluate the psychological plausibility of strong, optimistic, pessimistic and ceteris paribus semantics. In order to reach this objective, we have conducted a psychological experiment devoted to collect sets of comparative preferences formulated by participants to this experiment, and the associated models (a (pre)order on the set of outcomes). The adopted methodology and main results are presented in the next subsections.

Method

Participants Thirty first-year psychology students at the University of Toulouse-Le Mirail, all native French speakers, contributed to this study. None of them had previously received any formal logical training or any course on preferences. Note that our objective is not to study participant's real preferred menus. Such an objective would necessitate a much more large number of participants. Rather, our objective is to compare statistically the fit of the semantics under study with human preference's judgments. For such an objective, our sample size is sufficient according to scientific standards.

Material and procedure Comparative preference judgments were collected via a booklet where subjects were asked to suppose that they are at the restaurant and they must compose their menu. In the first page of the booklet, they were asked to compare and to rank-order the following objects (unranked objects where skipped from analyses):
t_0: fish-white-ice_cream, t_1: fish-white-cake
t_2: fish-red-ice_cream, t_3: fish-red-cake

$(white, red), (meat, fish), (ice\_cream, cake)$,
$(meat - white - ice\_cream, meat - white - cake)$,
$(meat - white - ice\_cream, meat - red - ice\_cream)$,
$(meat - white - ice\_cream, meat - red - cake)$,
$(meat - white - ice\_cream, fish - white - ice\_cream)$,
$(meat - white - ice\_cream, fish - white - cake)$,
$(meat - white - ice\_cream, fish - red - ice\_cream)$,
$(meat - white - ice\_cream, fish - red - cake)$,
$(meat - white - cake, meat - red - ice\_cream)$,
$(meat - white - cake, meat - red - cake)$,
$(meat - white - cake, fish - white - ice\_cream)$,
$(meat - white - cake, fish - white - cake)$,
$(meat - white - cake, fish - red - ice\_cream)$,
$(meat - white - cake, fish - red - cake)$,
$(meat - red - ice\_cream, meat - red - cake)$,
$(meat - red - ice\_cream, fish - white - ice\_cream)$,
$(meat - red - ice\_cream, fish - white - cake)$,
$(meat - red - ice\_cream, fish - red - ice\_cream)$,
$(meat - red - ice\_cream, fish - red - cake)$,
$(meat - red - cake, fish - white - ice\_cream)$,
$(meat - red - cake, fish - white - cake)$,
$(meat - red - cake, fish - red - ice\_cream)$,
$(meat - red - cake, fish - red - cake)$,
$(fish - white - ice\_cream, fish - white - cake)$,
$(fish - white - ice\_cream, fish - red - ice\_cream)$,
$(fish - white - ice\_cream, fish - red - cake)$,
$(fish - white - cake, fish - red - ice\_cream)$,
$(fish - white - cake, fish - red - cake)$,
$(fish - red - ice\_cream, fish - red - cake)$.

Table 1: Pairs of menus participants have to compare.

t_4: meat-white-ice_cream, t_5: meat-white-cake
t_6: meat-red-ice_cream, t_7: meat-red-cake.
Next, they were asked to compare the 31 pairs of menus given in Table 1. An object o_1 can be preferred to an object o_2 or o_2 preferred to o_1, or be both equally preferred, or be incomparable. Answers of the kinds "equally preferred" or "incomparable" have been discarded from analysis.

Rationale Given participant's comparative preference judgments, for each participant, we computed the set of compact preferences (see Table 2) consistent with participant's preferences. For a given participant, a comparative preference is retained as compact if it is consistent with all her/his preferred menus (see Table 1).

Next, given these compact preferences and the algorithms provided in the paper, for each participant, four preorders have been inferred according to the principles underling the inferential machinery of the four studied semantics. For evaluating the psychological relevance of these semantics, the key comparison is between participant's (pre)order on the 8 menus $\{t_0, \cdots, t_7\}$ and (pre)orders computed according to the four semantics given participant's compact preferences. Two cues have been used for ordering semantics

$white$	$> (vs. <)$	red
$meat$	$> (vs. <)$	$fish$
ice_cream	$> (vs. <)$	$cake$
$white \wedge meat$	$> (vs. <)$	$white \wedge fish$
$white \wedge ice\_cream$	$> (vs. <)$	$white \wedge cake$
$red \wedge meat$	$> (vs. <)$	$red \wedge fish$
$red \wedge ice\_cream$	$> (vs. <)$	$red \wedge cake$
$meat \wedge ice\_cream$	$> (vs. <)$	$meat \wedge cake$
$fish \wedge ice\_cream$	$> (vs. <)$	$fish \wedge cake$

Table 2: Set of a priori possible compact preferences.

according to their psychological relevance: *The percentages of cases* where the semantics provide an inconsistent set of models; and *the distance* and *mean distance* between ranks allowed by participants and semantics to the 8 menus.

- *Percentages of inconsistency:* For each semantics, we computed the percentages of cases where it produces an inconsistent set of models given inferred participant's compact preferences. A semantics better fits psychological data if it allows producing a consistent set of models from participant's compact preferences.
- *Mean Distances:* Two distances based on participants and semantics orders have been computed. In both cases, distances are computed from the ranks attributed to each of the 8 menus by participants and semantics. Several menus can have the same rank. Suppose participant 1 prefers the menu "meat, red wine, ice cream", if this menu is also the preferred one for a given semantics, then the distance is zero. If only one menu is more preferred, then the rank is 1, and so on. A semantics better fits psychological data if the rank it attributes is closer to the menu preferred by participants. A semantics better fits psychological data if the mean distance between participants's preferred models and the rank attributed by the semantics is smaller. The same calculus can be made for each menu involved in participant's ranking (which doesn't necessarily involve the 8 proposed menus). So, for each participant, it is possible to compute the mean of the distance between each menu and ranks predicted by semantics. Next, the mean of these means is computed. As before, a small mean means a better fit.

In order to conclude at the inferential level, cognitive psychology, exactly as other experimental sciences, makes use of statistical tools for hypothesis testing. The student's t-test allows testing the null hypothesis that two means are not different. The probability p provided by the test express the risk (called alpha) that we reject by error the null hypothesis. In social and human sciences, it is usual to consider that this risk is acceptable at the level .05, that is, if p is greater than .05, we cannot reject the null hypothesis without a significant risk. Under .05, we reject the null hypothesis, and so accept the hypothesis of the difference between the two means. In our analyses, when a difference between means is significant ($p =< .05$), it is interpreted as: the seman-

	cp	$str.$	$pess.$	$opt.$
$\%\ inconsistency$	36.6	10	0	0
$Mean\ distances$ $to\ participants$ $peferred\ outcomes$ $(standard\ deviation)$	1.5 $(.81)$ $n =$ 26	.83 $(.93)$ $n =$ 29	.62 $(.71)$ $n =$ 30	.55 $(.65)$ $n =$ 30
$Means\ of\ the\ mean$ $distance\ to\ participants$ $outcome\ levels$ $(standard\ deviation)$ $n = 19$	2.44 $(.63)$	2.46 $(.65)$	2.54 $(.71)$	2.53 $(.66)$

Table 3: Cues for evaluation of the fit of semantics with participant's preference judgment. "cp", "str", "pess." and "opt." stand respectively for ceteris paribus, strong, pessimistic and optimistic.

tics exhibiting the less mean distance significantly fits better human data than the other semantics.

Results

Participant's answers allowed to compute a set of compact preferences containing between 3 and 7 compact preferences out of 18 a priori ones. Table 3 shows that the ceteris paribus semantics doesn't fit participant's orders in 36% of the cases and that the strong semantics failed in 10% of the cases, while optimistic and pessimistic semantics provide always a consistent set of preferences. This order is confirmed by comparisons of distances between participants and semantics' levels for participant preferred outcome. Table 3 also suggests that the optimistic semantics has a better fit than the pessimistic one (however mean's comparison by Student's t-test is not significant: $t = -1.43$, $df = 29$, $p = .16$) while the latter has a better fit than the strong semantics ($t = 2.7$, $df = 28$, $p = .01$) which better fits participant's data than ceteris paribus semantics ($t = -5.7$, $df = 24$, $p < .001$, significant). These results are broadly confirmed by the comparison of the means of the mean distance between participants and semantics (pre)orders. Indeed, statistical comparisons by Student's t-test show a significant difference between strong and pessimistic semantics ($t = -2.37$, $df = 18$, $p = .036$) but not between ceteris paribus and strong, and pessimistic and optimistic semantics. This result confirms that two distinct sets of semantics can be distinguished from their psychological relevance: Pessimistic and optimistic semantics on one hand, and strong and ceteris paribus on the other one. Except for percentages, more the values are low, better is the fit. As such, given all the information summarized in table 3, it appears that optimistic and pessimistic semantics are more plausible psychologically than strong and ceteris paribus semantics.

Conclusion

We focused on comparative preference statements and distinguished different non-monotonic semantics that have been studied in literature. So far, researchers have argued for a semantics or another from purely theoretical standpoint (also philosophical for ceteris paribus semantics) or for modeling a specific application. In this paper, we explored another dimension, namely psychological plausibility, to compare the semantics.

This work gives an indication about human behavior when interpreting comparative preferences. Our results suggest that pessimistic and optimistic semantics better fit human preferences organization and inference than ceteris paribus and strong semantics. Neverthless, it doesn't mean that every human in every situation would "prefer" according to the principles underling these semantics. Rather, it suggests that in familiar domains, a population known as representative of global occidental people, "prefer" in a manner more closed to pessimistic and optimistic semantics. Psychological plausibility is not of course the sole criterion for evaluating formal models in AI, but it is a criterion, every time a formal model could have incidences in human adaptation, including cognitive comfort and efficiency.

This first attempt opens the door to more ambitious and deeper comparison of preference representations. In a future work we intend to perform a comparison of the main different compact representations of preferences such as CP-nets (Boutilier et al. 2004), QCL (Brewka, Benferhat, and Le Berre 2004), etc.

References

Benferhat, S.; Dubois, D.; Kaci, S.; and Prade, H. 2002. Bipolar possibilistic representations. In *UAI'02*, 45–52.

Benferhat, S.; Bonnefon, J.; and Da Silva Neves, R. 2004. An experimental analysis of possibilistic default reasoning. In *KR'04*, 130–140.

Benferhat, S.; Dubois, D.; and Prade, H. 1992. Representing default rules in possibilistic logic. In *KR'92*, 673–684.

Boutilier, C.; Brafman, R.; Domshlak, C.; Hoos, H.; and Poole, D. 2004. CP-nets: A tool for representing and reasoning with conditional ceteris paribus preference statements. *Journal of Artificial Intelligence Research* 21:135–191.

Boutilier, C. 1994. Toward a logic for qualitative decision theory. In *KR'94*, 75–86.

Brewka, G.; Benferhat, S.; and Le Berre, D. 2004. Qualitative choice logic. *Artificial Intelligence* 157(1-2):203–237.

Hansson, S. 1996. What is ceteris paribus preference? *Journal of Philosophical Logic* 25:307–332.

Kraus, S.; Lehmann, D.; and Magidor, M. 1990. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence* 44(1-2):167–207.

Neves, R. D. S.; Bonnefon, J.; and Raufaste, E. 2002. An empirical test of patterns for nonmonotonic inference. *Annals of Mathematics and Artificial Intelligence* 34(1-3):107–130.

Pearl, J. 1990. System Z: A natural ordering of defaults with tractable applications to default reasoning. In *TARK'90*, 121–135.

von Wright, G. H. 1963. *The Logic of Preference*. University of Edinburgh Press.

Yager, R. 1983. Entropy and specificity in a mathematical theory of evidence. *International Journal of General Systems* 9:249–260.