SVEN OVE HANSSON

PREFERENCE LOGIC

1 INTRODUCTION

The study of general principles for preferences can, if we so wish, be traced back to Book III of Aristotle's *Topics*. Since the early twentieth century several philosophers have approached the subject of preferences with logical tools, but it is probably fair to say that the first complete systems of preference logic were those proposed by Sören Halldén in 1957 and Georg Henrik von Wright in 1963. [Rescher, 1968, pp. 287–288; Halldén, 1967; von Wright, 1963]. The subject also has important roots in utility theory and in the theory of games and decisions.

Preferences and their logical properties have a central role in rational choice theory, a subject that in its turn permeates modern economics, as well as other branches of formalized social science. Some of the most important recent developments in moral philosophy make essential use of preference logic [Fehige and Wessels, 1998]. At the same time, preference logic has turned out to be an indispensable tool in studies of belief revision and non-monotonic logic [Rott, 1999]. Preference logic has become so integrated into both philosophy and social science that we run the risk of taking it for granted and not noticing its influence.

This chapter is devoted to the philosophical foundations, rather than the applications, of preference logic. The emphasis is on fundamental results and their interpretation. Section 2 treats the basic case in which the objects of preferences form a set of mutually exclusive alternatives. In Section 3, such preferences are related to choice functions. In Section 4, the requirement of mutual exclusivity is relaxed. In Section 5, preferences are related to monadic concepts such as 'best', 'good', and 'ought'.

2 PREFERENCES OVER INCOMPATIBLE ALTERNATIVES

In most applications of preference logic, the objects that preferences refer to are assumed to be mutually exclusive. This assumption will also be made in the present section.

2.1 Preference, indifference, and other value concepts

From a logical point of view, the major value concepts of ordinary language can be divided into two major categories. The *monadic* (classificatory) value concepts, such as 'good', 'very bad', and 'worst' report how we evaluate a

D.M. Gabbay and F. Guenthner (eds.),
Handbook of Philosophical Logic, 2nd Edition, Volume 4, 319–393.

single referent. The *dyadic* (comparative) value concepts, such as 'better', 'worse', and equal in value to', indicate a relation between two referents. In less colloquial contexts we can also find three-termed value predicates, such as 'if x, then y is better than z' (conditional preferences) and even four-termed ones, such as 'x is preferred to y more than z is preferred to w' [Packard, 1987]. This chapter is primarily devoted to the dyadic value concepts.

There are two fundamental comparative value concepts, namely 'better' (strict preference) and 'equal in value to' (indifference) [Halldén, 1957, p. 10]. The relations of preference and indifference between alternatives are usually denoted by the symbols $>$ and $\equiv$ or by the symbols P and I. Here, the former notation will be used.

There is a long-standing philosophical tradition to take $A > B$ to represent 'B is worse than A' as well as 'A is better than B'. [Brogan, 1919, p. 97]. This is not in exact accordance with ordinary English. We tend to use 'better' when focusing on the goodness of the higher-ranked of the two alternatives, and 'worse' when empasizing the badness of the lower-ranked one [Halldén,p. 13; von Wright, 1963, p. 10; Chisholm and Sosa, 1966, p. 244]. However, the distinction between betterness and converse worseness can only be made at the price of a much more complex formal structure. The distinction does not seem to have enough philosophical significance to be worth this complexity, at least not in a general-purpose treatment of the subject.

When describing the preferences of others, we tend to use the word 'preferred'. The word 'better' is used when we express our own preferences and also when we refer to purportedly impersonal evaluations. Although these are important distinctions, not very much has been made of them in preference logic. 'Logic of preference' and 'logic of betterness' are in practice taken as synonyms.

The preferences studied in preference logic are the preferences of rational individuals. Since none of us is fully rational, this means that we are dealing with an idealization. If a proposed principle for preference logic does not correspond to how we actually think and behave, the reason may be either that the principle is wrong or that we are not fully rational when our behaviour runs into conflicts with it.

The objects of preference are represented by the relata of the preference relation. (A and B in $A > B$.) In order to make the formal structure determinate enough, every preference relation is assumed to range over a specified set of relata. As already indicated, in this section, the relata are assumed to be mutually exclusive, i.e. none of them is compatible with, or included in, any of the others. No further assumptions are made about their internal structure. They may be physical objects, types or properties of such objects, states of affairs, possible worlds—just about anything.

Preferences over a set of mutually exclusive relata will be referred to as *exclusionary* preferences.

The following four properties of the two exclusionary comparative relations will be taken to be part of the meaning of the concepts of (strict) preference and of indifference:

(1) If A is better than B, then B is not better than A.

(2) If A is equal in value to B, then is B equal in value to A.

(3) A is equal in value to A.

(4) If A is better than B, then A is not equal in value to B.

It follows from (1) that preference is irreflexive, i.e. that A is not better than A. The following is a restatement of the four properties in formal language.

DEFINITION 1. A (*triplex*) *comparison structure* is a triple $\langle \mathcal{A}, >, \equiv \rangle$, in which $\mathcal{A}$ is a set of alternatives, and $>$ and $\equiv$ are relations in $\mathcal{A}$ such that for all $A, B \in \mathcal{A}$:

(1) $A > B \rightarrow \neg(B > A)$ (*asymmetry of preference*)

(2) $A \equiv B \rightarrow B \equiv A$ (*symmetry of indifference*)

(3) $A \equiv A$ (*reflexivity of indifference*)

(4) $A > B \rightarrow \neg(A \equiv B)$ (*incompatibility of preference and indifference*)

Furthermore:

$$A \geq B \leftrightarrow (A > B) \vee (A \equiv B) \text{ (\textit{weak preference})}$$

The intended reading of $\geq$ is 'at least as good as' (or more precisely: 'better than or equal in value to'). As an alternative to $\geq$, it can also be denoted 'R'. Weak preference can replace (strict) preference and indifference as primitive relations in comparison structures:

OBSERVATION 2. Let $\langle \mathcal{A}, >, \equiv \rangle$ be a triplex comparison structure, and let $\geq$ be the union of $>$ and $\equiv$. Then:

(1) $A > B \leftrightarrow (A \geq B)$ & $\neg(B \geq A)$

(2) $A \equiv B \leftrightarrow (A \geq B)$ & $(B \geq A)$

Proof.

Part 1: Left-to-right: From $A > B$ it follows by the definition of $\geq$ that $A \geq B$. Furthermore, it follows from the asymmetry of preference that $\neg(B > A)$ and from the incompatibility of preference and indifference that $\neg(A \equiv B)$, i.e., by the symmetry of indifference, $\neg(B \equiv A)$. Thus $\neg((B > A) \vee (B \equiv A))$, i.e., by the definition of $\geq$, $\neg(B \geq A)$. *Right-to-left:* It follows from $A \geq B$, according to the definition of $\geq$, that either $A > B$ or $A \equiv B$. By the same definition, it follows from $\neg(B \geq A)$ that $\neg(B \equiv A)$. By the symmetry of indifference, $\neg(A \equiv B)$, so that $A > B$ may be concluded.

Part 2: Left-to-right: It follows from $A \equiv B$, by the definition of $\geq$, that $A \geq B$. By the symmetry of indifference, $A \equiv B$ yields $B \equiv A$ so that, by the definition of $\geq$, $B \geq A$. *Right-to-left:* It follows from the definition of $\geq$ and $(A \geq B)$ & $(B \geq A)$ that $((A > B) \vee (A \equiv B))$ & $((B > A) \vee (B \equiv A))$. By the symmetry of indifference, $((A > B) \vee (A \equiv B))$ & $((B > A) \vee (A \equiv B))$. By the asymmetry of preference, $A > B$ is incompatible with $B > A$. We may conclude that $A \equiv B$. ■

The choice of primitives (either $\geq$ or both $>$ and $\equiv$) is a fairly inconsequential choice between formal simplicity ($\geq$) and conceptual clarity ($>$ and $\equiv$). (Cf. [Burros, 1976].) The following is an alternative to Definition 1.

DEFINITION 3. A (*duplex*) *comparison structure* is a pair $\langle \mathcal{A}, \geq \rangle$, in which $\mathcal{A}$ is a set of alternatives and $\geq$ a reflexive relation on $\mathcal{A}$. The derived relations $>$ and $\equiv$ are defined as follows:

$A > B$ if and only if $A \geq B$ and $\neg(B \geq A)$
$A \equiv B$ if and only if $A \geq B$ and $B \geq A$

It will be seen that the defined relation $\geq$ of Definition 1 is reflexive and that the defined relations $>$ and $\equiv$ of Definition 3 satisfy conditions (1)–(4) of Definition 7. It follows that the two definitions are interchangeable. Given our definitions, the four conditions of Definition 1 are in combination equivalent to the reflexivity of weak preference.

The relations $>$ and $\equiv$ that are defined from $\geq$ in the manner of Definition 3 are called the *strict part*, respectively the *symmetric part*, of $\geq$.

NOTATIONAL CONVENTIONS N1:

(1) Chains of relations can be contracted. Hence, $A \geq B \geq C$ abbreviates $(A \geq B)$ & $(B \geq C)$, and $A > B > C \equiv D$ abbreviates $(A > B)$ & $(B > C)$ & $(C \equiv D)$.

(2) $>^*$ stands for $>$ repeated any finite non-zero number of times (and similarly for the other relations). Thus $A >^* C$ denotes that either $A > C$ or there are $B_1, ...B_n$ such that $(A > B_1)$ & $(B_1 > B_2)$ &...$(B_{n-1} > B_n)$ & $(B_n > C)$.

2.2 Completeness

In most applications of preference logic, it is taken for granted that the following property, called *completeness* or *connectedness*, should be satisfied:

$(A \geq B) \vee (B \geq A)$, or equivalently:
$(A > B) \vee (A \equiv B) \vee (B > A)$

As we will see later on, the assumption of completeness is often extremely helpful in terms of simplifying the formal structure. In terms of interpretation, however, it is much more problematic. In many everyday cases, we do not have, and do not need, complete preferences. In the choice between three brands of canned soup, A, B, and C, I clearly prefer A to both B and C. As long as A is available I do not need to make up my mind whether I prefer B to C, prefer C to B or consider them to be of equal value. Similarly, a voter in a multi-party or multi-candidate election can do without ranking the parties or candidates that she does not vote for.

From the viewpoint of interpretation, we can distinguish between three major types of preference incompleteness. First, incompleteness may be *uniquely resolvable*, i.e. resolvable in exactly one way. The most natural reason for this to be the case is that incompleteness is due to lack of knowledge or reflection. Behind what we perceive as an incomplete preference relation there may be a complete preference relation that we can arrive at through observation, logical inference, or some other means of discovery.

Secondly, incompleteness may be *multiply resolvable*, i.e. possible to resolve in several different ways. In this case it is genuinely undetermined what will be the outcome of extending the relation to cover the previously uncovered cases.

Thirdly, incompleteness may be *irresolvable*. The most natural reason for this is that the alternatives differ in terms of advantages or disadvantages that we are unable to put on the same footing. I may be unable to say which I prefer—the death of two specified acquaintances or the death of a specified friend [Hansson, 1998a]. I may be unable to say which I prefer—the destruction of the pyramids in Giza or the extinction of the giant panda. I may also be unable in many cases to compare monetary costs to environmental damage.

It is established terminology to call two alternatives 'incomparable' whenever the preference relation is incomplete with respect to them. The term 'incommensurable' 'can be reserved for cases when the incompleteness is irresolvable.

2.3 Transitivity and acyclicity

By far the most discussed logical property of preferences is the following:

$A \geq B \geq C \rightarrow A \geq C$ (*transitivity of weak preference*) http://www.springer.com/978-1-4020-0139-0

Handbook of Philosophical Logic
Gabbay, D.; Guenthner, F. (Eds.)
2002, XIII, 431 p. 1 illus., Hardcover
ISBN: 978-1-4020-0139-0