CRÍTICA, *Revista Hispanoamericana de Filosofía*
Vol. XXXI, No. 92 (agosto 1999): 53–66

PREFERENCES AND ALTERNATIVES

SVEN OVE HANSSON
Royal Institute of Technology
Stockholm

Virtually all of modern economics is based on assumptions on the structure of rational preferences. The same assumptions have a central rôle in formalized sociology and political science. Therefore, investigations of the structure of preferences can be helpful in improving the foundations of formalized social science.

The purpose of this paper is to apply a simple —indeed trivial— philosophical insight to the theory of rational preferences. This trivial insight is that preferences are not sufficiently precise unless the set of alternatives is specified.[1] In mathematical terms this is another way of saying that a binary relation has not been specified unless its domain has been specified.

In what follows, I will apply this insight to three examples: (1) Wollheim's democratic paradox, (2) the Arrovian framework in social decision theory, and (3) the money-pump argument for transitivity of preferences.

1. *Wollheim's paradox*

Richard Wollheim provided an acute formulation of the problematic situation of an outvoted democrat.[2] We are

[1] Cf., L.J. Savage, *The Foundations of Statistics*, 1954. Lars Bergström, *The Alternatives and Consequences of Actions*, 1966.

[2] Richard Wollheim, "A Paradox in the Theory of Democracy",

invited to consider an individual who endorses the democratic procedure, or ‘democratic machine’. Furthermore, this person prefers a certain social state p to its negation $\neg p$, but as a democrat she also wants the democratic decision with respect to p to be carried through. (It should be noted that democrats tend to respect democratic decisions with respect to many but not all issues. Hence, it is assumed here that the issue of p or $\neg p$ belongs to the category of issues for which our democrat respects democratic decisions; this means among other things that she does not wish civil disobedience to be successfully exercised in this particular issue.)

Now, how does our democrat react when it turns out that the democratic decision was in favour of $\neg p$? Since she wants the majority’s will to be respected, she prefers $\neg p$ to p. But she regrets the majority’s position for the simple reason that she in fact prefers p to $\neg p$. How can she both prefer p to $\neg p$ and $\neg p$ to p?

My proposal is that we widen our descriptions of the states of affairs that her preferences refer to. This can be done by introducing the predicate **D** to denote democratic decisions. Thus $\mathbf{D}p$ means: “A valid democratic decision in favour of p has been made.”

There are then four composite states of affairs that have to be ranked in the preference ordering of the democrat, namely $\mathbf{D}p \,\&p$, $\mathbf{D}p \,\&\neg p$, $\mathbf{D}\neg p \,\&p$, and $\mathbf{D}\neg p \,\&\neg p$. $\mathbf{D}p \,\&p$ means that p has been decided and p is the case; $\mathbf{D}p \,\&\neg p$ means that p has been decided and $\neg p$ is the case, etc.

These four states of affairs can also be formulated in an alternative way. For that purpose, let us introduce an additional symbol into the formal language, namely ‘**R**’.

pp. 71–87 in Peter Laslett and W.G. Runciman, *Philosophy, Politics and Society* (Second Series), Oxford, 1962. Most of this section is based on Sven Ove Hansson, “A Resolution of Wollheim’s Paradox”, *Dialogue*, 32, pp. 681–687, 1993.

$\mathbf{R}p$ means: "The democratic decision with respect to p or $\neg p$ is respected." Similarly, $\neg\mathbf{R}p$ means: "The democratic decision with respect to p or $\neg p$ is not respected." Our four composite states of affairs can now be expressed with the use of **R** instead of **D**:

$\mathbf{D}p \,\&p$	is equivalent with $\mathbf{R}p \,\&p$
$\mathbf{D}p \,\&\neg p$	is equivalent with $\neg\mathbf{R}p \,\&\neg p$
$\mathbf{D}\neg p \,\&p$	is equivalent with $\neg\mathbf{R}p \,\&p$
$\mathbf{D}\neg p \,\&\neg p$	is equivalent with $\mathbf{R}p \,\&\neg p$

Our democrat's preference ordering over the four composite states of affairs will be assumed to be expressible by a transitive relation. The following preference patterns are of particular interest for the analysis of the paradox. $\vee$ stands for strict preference and $\approx$ for indifference.

(1)	(2)	(3)
$\mathbf{D}p \,\&p \approx \mathbf{D}\neg p \,\&\neg p$	$\mathbf{D}p \,\&p$	$\mathbf{D}\neg p \,\&\neg p$
$\vee$	$\vee$	$\vee$
$\mathbf{D}p \,\&\neg p \approx \mathbf{D}\neg p \,\&p$	$\mathbf{D}\neg p \,\&\neg p$	$\mathbf{D}p \,\&p$
	$\vee$	$\vee$
	$\mathbf{D}\neg p \,\&p$	$\mathbf{D}p \,\&\neg p$
	$\vee$	$\vee$
	$\mathbf{D}p \,\&\neg p$	$\mathbf{D}\neg p \,\&p$

The states of affairs referred to in (1), (2), and (3) can also be expressed with **R** instead of **D**. It will be seen that (1) is equivalent with (1′), (2) with (2′), and (3) with (3′):

$$
\begin{array}{lll}
(1') & (2') & (3') \\[6pt]
\mathbf{R}p\ \&p \ \approx\ \mathbf{R}p\ \&\neg p & \mathbf{R}p\ \&p & \mathbf{R}p\ \&\neg p \\
\quad\vee & \quad\vee & \quad\vee \\
\neg\mathbf{R}p\ \&\neg p \ \approx\ \neg\mathbf{R}p\ \&p & \mathbf{R}p\ \&\neg p & \mathbf{R}p\ \&p \\
 & \quad\vee & \quad\vee \\
 & \neg\mathbf{R}p\ \&p & \neg\mathbf{R}p\ \&\neg p \\
 & \quad\vee & \quad\vee \\
 & \neg\mathbf{R}p\ \&\neg p & \neg\mathbf{R}p\ \&p
\end{array}
$$

Our next task is to derive preferences for p and for $\neg p$ from these preference relations over composite states of affairs. It is reasonable to assume that these should be preferences *ceteris paribus*, or ‘all things being equal’. This notion has, in the present context, an ambiguity that will help explain how a democrat may consistently both prefer p to $\neg p$ and prefer $\neg p$ to p.

Suppose that our democrat has the preferences expressed in (2) and (2′). Furthermore, suppose that a democratic decision has been made in favour of $\neg p$. What does it then mean for her to prefer either p or $\neg p$ ‘everything else being equal’?

First consider the formulation in (2). To prefer p to $\neg p$ ‘everything else being equal’ means to prefer p to $\neg p$, provided that as little else as possible is changed in the world. It is a fact that $\mathbf{D}\neg p$. Thus a preference ‘everything else being equal’ should be a preference provided that $\mathbf{D}\neg p$ is not changed. Our outvoted democrat prefers $\mathbf{D}\neg p\ \&\neg p$ to $\mathbf{D}\neg p\ \&p$, i.e. she prefers $\neg p$ to p if $\mathbf{D}\neg p$ is kept constant.

Thus she may reasonably be said to prefer $\neg p$ to p, everything else being equal.

Next, consider the equivalent formulation (2′) of the same preference ordering. Here, we need to consider two cases, according to whether the democratic decision is being respected or not. In the first case, when it is respected, $\mathbf{R}p$ is true. Since she prefers $\mathbf{R}p \,\&p$ to $\mathbf{R}p \,\&\neg p$, it follows that everything else (including $\mathbf{R}p$) being equal, she prefers p to $\neg p$. In the second case, when the democratic decision is not being respected, $\neg\mathbf{R}p$ is true. Since she prefers $\neg\mathbf{R}p \,\&p$ to $\neg\mathbf{R}p \,\&\neg p$, she prefers p to $\neg p$, everything else being equal, in the second case as well as in the first one.

The two formulations of the relevant states of affairs in terms or **D** and of **R** are logically equivalent, but they put emphasis on different aspects of the world that can be kept constant. Therefore they have different implications for counterfactual discourse on minimally changed states of affairs. In this way, the ambiguity of the phrase 'everything else being equal' makes it possible for the outvoted democrat to strictly prefer, at the same time, p to $\neg p$ and $\neg p$ to p, without being inconsistent.

(2) and (2′) represent the preferences of a person who strictly prefers p to $\neg p$ and at the same time prefers what takes place to conform with a democratic decision. (3) and (3′) represent the preferences of a democrat who strictly prefers $\neg p$ to p, and (1) and (1′) those of a democrat who is indifferent between p and $\neg p$.

It must again be emphasized that although a democrat can be expected to have this type of preferences with respect to most issues in day-to-day politics, she cannot be expected to have such preferences with respect to *all* political issues. A person who, for *every* issue that is subject to democratic decision-making, has one of the preference patterns (1), (2), or (3), is a 'thorough-going democrat' in

Schiller's sense.[3] Such a person "subscribes to the principle of majority rule as a rule which has no defeasibility conditions", and is therefore "committed, i.e., obliged on pain of being inconsistent with what he subscribes to, to abide by any and every ruling of the relevant majority". This "thorough-going" attitude is not compatible with ordinary conceptions of democratic commitments. If p denotes the introduction of political censorship, then we expect a democrat to prefer $\mathbf{D}p \,\&\neg p$ to $\mathbf{D}p \,\&p$.

Preferences with the logical structures that give rise to Wollheim's paradox are not peculiar to the subject-matter of politics or of collective decision-making. As an example, suppose that Bill is a conformist who prefers to be dressed on his job in the same manner as his boss. Although he is more comfortable with a tie than without, his conformity is stronger than his preference for wearing a tie. To Bill's sorrow, his boss does not wear a tie. He may then, on different occasions, make the following seemingly contradictory statements:

(i) "I prefer to wear a tie on my job."

(ii) "I prefer not to wear a tie on my job."

(i) can be qualified with "if everything else is equal, i.e., if I continue to be dressed like my boss". Similarly, (ii) can be qualified with "if everything else is equal, i.e., if my boss continues not to wear a tie". Typically, (i) would be uttered in a discussion about jobs in general, and (ii) in a discussion about the particular job that Bill has now. It is an easy exercise to analyze this example in logical detail in the same way as was done above for Wollheim's paradox.

Hence, Wollheim's paradox is a consequence of the contextual dependence of the implicit *ceteris paribus* clauses in

[3] Marvin Schiller, "On the Logic of Being A Democrat", *Philosophy*, 44, pp. 46–56, 1969.

statements expressing preferences. The paradox vanishes if alternative sets are specified in an exact manner.

2. *Arrovian social choice*

The formal study of social choices and decisions is at present entirely dominated by the formal model that was developed by Kenneth Arrow.[4] In the application of that model to voting procedures, group decisions are obtained through aggregation of individual preferences. An Arrovian voting procedure has the following components:

1. A number of *individuals*. They can be represented by n-tuple.

2. A number of *alternatives*. They are the options that the procedure has been set up to decide between.

3. A number of possible *outcomes*. Each outcome may either be one of the (decision-)*alternatives* (options) that the procedure has been set up to choose between, or it may be the *tie outcome* (λ).

4. A number of possible ways of voting, called *strategies*. In simple cases, the strategies coincide with the alternatives: you simply vote for one of the alternatives. In other cases, voting may take the form of ranking.

A *voting pattern* is an assignment of a strategy to each participant, in the form of an n-tuple.

5. A *social choice function* takes us from each voting pattern, i.e., total input, to an outcome.

[4] Kenneth J. Arrow, *Social Choice and Individual Welfare*, 1951, 2nd ed., 1963. This section is based on: Sven Ove Hansson, "A Procedural Model of Voting", *Theory and Decision*, 32, pp. 269–301, 1992. Sven Ove Hansson, "The Difference Model of Voting", *Notre Dame Journal of Formal Logic*, 33, pp. 576–592, 1992. Sven Ove Hansson, "Social Choice With Procedural Preferences", *Social Choice and Welfare*, 13, pp. 215–230, 1996.

In formal terms:

DEFINITION 1: A *voting procedure* is a quadruple $\langle \mathbf{I},\mathbf{A},\mathbf{S},\mathbf{r}\rangle$ such that:

$\mathbf{I} = \langle i_1,\ldots i_n\rangle$ is an n-tuple of *individuals*.

$\mathbf{A} = \{x,y\ldots\}$ is the set of (*decision-*)*alternatives*. $\mathbf{A}\cup\{\lambda\}$ is the set of *outcomes*, where λ is the *tie outcome*.

$\mathbf{S} = \{s_1,\ldots s_m\}$ is the set of *strategies*.

An n-tuple $\pi = \langle \alpha_1,\ldots \alpha_n\rangle$ of elements of $\mathbf{S}$ is called a *voting pattern*. For each k, π assigns α_k to i_k.

$\mathbf{r}$, the *social choice function*, is a total function from the set of voting patterns to the set of outcomes.

In order to discuss the relationship between individual preferences and the outcome of the procedure we also need a representation of the individual preferences. In Arrow's model, this representation has the following form:

DEFINITION 2 (Arrow's model): To each individual i_k is assigned a transitive and connective preference relation R_k that has the domain $\mathbf{A}$.

Hence, in the Arrovian framework, collective decisions are modelled as aggregations of individual preferences over the options that the procedure has been set up to decide between. However, individuals that take part in collective decision procedures often have preferences that do not refer exclusively to these options (decision-alternatives). Besides wanting the outcome to be as good as possible (according to her own standard), a participant may have preferences such as the following:

1. She prefers to be part of the winning coalition.

2. She prefers the decision to be taken by as large a majority as possible. (This may lead a committee member to vote for another alternative than the one that she actually prefers most, in order to contribute to unanimity.)
3. She prefers to cast her own vote for as good an alternative as possible. Such “principled” preferences may prevent a participant from taking part in a coalition that would change the outcome, say, from her third-best to her second-best alternative. (This is one of the mechanisms through which candidates that have no chance to get elected still receive votes.)
4. She may want the outcome of the voting procedure to be X, but yet prefer to vote for Y, since Y is more in strict accord with her ideals (although it is sadly unrealistic as things are at present).

In cases like these, the preferences that guide the voter’s decision do not refer exclusively to the decision-alternatives that the procedure aims at choosing between. Her preferences have a *procedural* component. Such preferences cannot be expressed in the Arrovian framework. To the contrary, Arrovian social choice theory is based on the assumption that voters vote exclusively according to how they order the alternatives that are subject to the decision.

In a more realistic model of voting, the preferences of the participants must refer to (preference-)alternatives that are capable of representing procedural factors. The set of preference-alternatives (comparison classes) will not (as in Arrow’s framework) be identical to the set of decision-alternatives (options). This can be achieved as follows:

> DEFINITION 3 (the enlarged model): To each individual i_k is assigned a transitive and connected preference relation R_k, the domain of which is the set of possible voting patterns.

The examples given above of preferences not expressible in the Arrovian framework can all be expressed in the enlarged model. Consider for instance a preference for consensus in a voting procedure with three participants and simple majority rule. A person who strictly prefers consensus should prefer the voting pattern $\langle x,x,x \rangle$ to each of the patterns $\langle y,x,x \rangle$, $\langle x,y,x \rangle$, and $\langle x,x,y \rangle$, for all x and y such that $x \neq y$.

The enlarged model is a generalization of the Arrovian framework. The latter can be obtained as a special case, that we may call *consequentialist preferences*:

> DEFINITION 4: A preference relation R_k is *consequentialist* if and only if for all π and π': if $\mathbf{r}(\pi) = \mathbf{r}(\pi')$ then $\pi R_k \pi'$ and $\pi' R_k \pi$.

The enlarged model makes possible a more realistic description of important patterns of preferences that the Arrovian model has difficulties in dealing with. There is no reason to take for granted that consequentialism, as defined above, is a requirement for rationality. To the contrary, it can be maintained that some measure of procedural preferences (such as preferences for consensus) is necessary for a well-functioning democratic system.

3. *Money-pumps*

Probably the most influential argument for transitivity of preferences originates with F.P. Ramsey.[5] Ramsey pointed out that if a subject's relation of preference violates transitivity, then "[h]e could have a book made against him by a

[5] F.P. Ramsey, *The Foundations of Mathematics and Other Logical Essays*, London [1931], 1950, p. 182. This section is based on Sven Ove Hansson, "Money-Pumps, Self-Torturers and the Demons of Real Life", *Australasian Journal of Philosophy*, 71, pp. 476–485, 1993.

cunning better and would then stand to lose in any event". For a concrete example, suppose that a stamp-collector has cyclic preferences with respect to three stamps, denoted a, b, and c. She prefers a to b, b to c, and c to a. Following Ramsey, we may assume that there is an amount of money, say 10 cents, that she is prepared to pay for exchanging b for a, c for b, or a for c. She comes into a stamp shop with stamp a. The stamp-dealer offers her to trade in a for c, if she pays 10 cents. She accepts the deal.

For a precise notation, let $\langle x,v \rangle$ denote that the collector owns stamp x and has paid v cents to the dealer. As this notation should serve to highlight, we are considering not only the primary alternatives, a, b, and c, but also *composite* alternatives that have been formed by combining primary alternatives with an auxiliary commodity.

Our stamp-collector has now moved from the state $\langle a,0 \rangle$ to the state $\langle c,10 \rangle$. Next, the stamp-dealer takes out stamp b from a drawer, and offers her to swap c for b, against another payment of 10 cents. She accepts, thus moving from the state $\langle c,10 \rangle$ to $\langle b,20 \rangle$.

When she is just on her way out the shop, the dealer calls her back, and advises her that it only costs 10 cents to change back to a, the very stamp that she had in her pocket when she entered the shop. Since she prefers a to b, she pulls out a third dime, thus moving from $\langle b,20 \rangle$ to $\langle a,30 \rangle$. Since she entered the shop in the state $\langle \mathrm{a},0 \rangle$, she has lost money without gaining anything else in this process.

To summarize the argument, the following sequence of preferences over composite alternatives was the cause of the trouble:

$$\langle c,10 \rangle > \langle a,0 \rangle$$
$$\langle b,20 \rangle > \langle c,10 \rangle$$
$$\langle a,30 \rangle > \langle b,20 \rangle$$

The trouble does not end here. Presumably, the sequence continues:

$\langle c,40\rangle > \langle a,30\rangle$

$\langle b,50\rangle > \langle c,40\rangle$

$\langle a,60\rangle > \langle b,50\rangle$

...

If the poor customer stays long enough in the stamp shop, she will be bereft of all her money, to no avail. This is the now classic money-pump argument. It purports to show that intransitive preferences are irrational.

To understand the structure of this argument we must observe that it is based on transitions between two preference relations, the simple and the composite preference relation. The simple preference relation has as its domain the three stamps a, b, and c. The composite preference relation has as its domain the set of ordered pairs where the first element is one of the three stamps and the second is a sum of money. The money-pump argument presupposes that the composite preference relation can be derived from the simple preference relation according to a simple and seemingly self-evident pattern.

However, the possibility of such a derivation is not something that should be taken for granted. Arguably, the preferences of a rational agent with respect to a given set of alternatives should be sufficiently precise to guide her in the choice among these alternatives. However, it cannot reasonably be demanded that they should provide a sufficient basis for mechanically deriving the preferences that she will have over a new alternative set that can be constructed from the previous one by affixing sums of money to the original alternatives. To the contrary, since a rational agent minimizes comparison-costs, her previous preferences will not in general be a sufficient guide when new aspects are taken into account, or when new combinations come up.

In our example, the stamp-collector can live happily with her cyclic preferences. These are not likely to cause much harm under normal circumstances. All that the example teaches us is that, when forming preferences over the new alternative sets created by the cunning dealer, the collector must consider the totality of the situation. She cannot simply apply a mechanical procedure to the original preferences.

Money-pumps do not exemplify irrational intransitivity, since the preferences listed in the example do not coexist. Instead, they may be seen as examples of how rational changes and specifications in preference patterns can be provoked by an extension of the set of alternatives.

4. *Conclusion*

I hope to have shown by these three examples that a simple but often forgotten aspect of the structure of preferences can have a major impact on how formal models are constructed and interpreted. The upshot is simple enough: When somebody says "These are the preferences", always ask: "Which are the alternatives?"

Recibido: 12 de agosto de 1999

RESUMEN

Utilizo tres ejemplos para mostrar el papel decisivo que desempeña la elección de un conjunto alternativo en la teoría de las decisiones: (1) la paradoja democrática de Wollheim se disuelve cuando las descripciones de estados de cosas se amplían para dar cabida a una interpretación en función de preferencias *ceteris paribus*. (2) El marco teórico arroviano en la teoría de las decisiones sociales puede mejorarse considerablemente si se amplían suficientemente las alternativas para dar cabida a la representación de preferencias procedimentales, tal como una preferencia por el consenso. (3) El argumento de la bomba de dinero para la transitividad de las preferencias depende de cambios problemáticos implícitos en el conjunto alternativo. Concluyo que a menudo el uso de conjuntos alternativos con alternativas más abarcadoras ayuda a resolver problemas en modelos basados en las preferencias.

[*Traducción: Héctor Islas*]