2

Preference Axioms and Their Implications

Economics relies on standard axioms concerning preference and choice. Section 2.1 of this chapter presents the most important of these. Section 2.2 explores their implications for the interpretation of preferences. The conditions on preferences presented in Section 2.1 are the axioms of "ordinal utility theory," and, as Section 2.1 explains, they guarantee that people's preferences can be represented by utility functions. Section 2.3 discusses the relationship between theories of people's actual preferences and choices and theories of rational preferences and choices. Section 2.4 argues that preferences cannot be defined in terms of expected self-interested benefits.

2.1. THE AXIOMS OF ORDINAL UTILITY THEORY

The axioms of ordinal utility theory are the core of positive economic theory, and they also constitute a fragmentary theory of rationality. Economists sometimes place other constraints on preferences, about which I shall have something to say in Chapter 4, but the axioms of ordinal utility theory are central. The following axioms are standard[1]:

(*Completeness*) For all x, y in X, either $x \succeq y$ or $y \succeq x$ or both.
(*Transitivity*) For all x, y, and z in X if $x \succeq y$ and $y \succeq z$, then $x \succeq z$.

"X" is the set of alternatives over which agents have preferences – commodity bundles in the case of consumer choice theory – and x, y, and z are alternatives

[1] These are quoted from Mas-Colell et al. (1995), p. 6. This is one of the two main graduate microeconomic textbooks. The other, by Hal Varian, states these axioms in exactly the same way but like most presentations of ordinal utility theory includes two additional axioms:

(*Reflexivity*) For all x in X, $x \succeq x$.
(*Continuity*) For all y in X $\{x: x \succeq y\}$ and $\{x: x \preceq y\}$ are closed sets (Varian 1984, pp. 111–12).
Reflexivity is trivial and arguably a consequence of completeness, whereas continuity, which is automatically satisfied for any finite set of alternatives, is needed to prove that preferences can be represented by a continuous utility function. The reflexivity and continuity axioms are not relevant to the issues this book tackles.

in X. According to Mas-Colell et al. (1995, p. 6), "We read $x \succeq y$ as 'x is at least as good as y'" (see also Varian 1984, p. 111). This definition of "$x \succeq y$" is surprising, because the axioms are supposed to govern *preferences*, not judgments of goodness. It is better to read "$x \succeq y$" as "the agent either prefers x to y or is indifferent between x and y." "$x \succ y$" means "the agent prefers x to y," and "$x \sim y$" means that the agent is indifferent.

In contrast to Varian, who presents the axioms as claims about people's actual preferences, Mas-Colell et al. (1995) maintain that completeness and transitivity are axioms of rationality: People's preferences are *rational* if they are complete and transitive. But because they are presenting a theory of people's actual preferences, they must also maintain that, to some extent, people's preferences are in this sense rational, and that the axioms are (to some degree of approximation) true of actual preferences.

The ordinal representation theorem proves that when people's preferences satisfy completeness, transitivity, and further technical conditions,[2] they can be represented by a continuous utility function that is unique up to a positive order-preserving transformation (Debreu 1959, pp. 56f). The "utility" of an alternative merely indicates the alternative's place in an agent's preference ranking. It is not something people seek or accumulate.

Here is a simple way to understand how a utility function "represents" preferences and what it means for it to be unique up to a positive order-preserving transformation. Suppose that an agent, Jill, who has preferences over a finite set of alternatives, adopts the convention of listing the alternatives on lined paper with preferred alternatives in higher rows and alternatives among which she is indifferent in the same row. Because Jill's preferences are complete, every alternative must find a place on the list. Because Jill's preferences are transitive, no alternative can have more than one place. Given such a list, one can assign numbers arbitrarily to rows, with higher rows getting higher numbers. Any numbering of the rows that is consistent with the ordering is an ordinal utility function. The numbers – the utilities – merely indicate where alternatives are located in Jill's preference ranking. Utility is not pleasure or usefulness or anything substantive at all. It is merely an indicator.

Figure 2.1 provides an illustration. The ordered list of alternatives is represented here by drawings of foods. U and U' are two of the infinite number of utility functions that assign higher numbers to alternatives in higher rows, and the same number to alternatives in the same row. The numbers are arbitrary apart from their order.

The theory of choice economists employ relies on two additional axioms, even though these are seldom stated explicitly as axioms. For example,

[2] In one version of the theorem, proven by Debreu (1959, pp. 56f), the additional technical conditions consist of reflexivity, continuity, and that the set of bundles of the k commodities be a connected subset of R^k (the k-dimensional space of real numbers). A subset of R^k is "connected" if it is not the union of two nonempty disjoint and closed subsets of R^k.

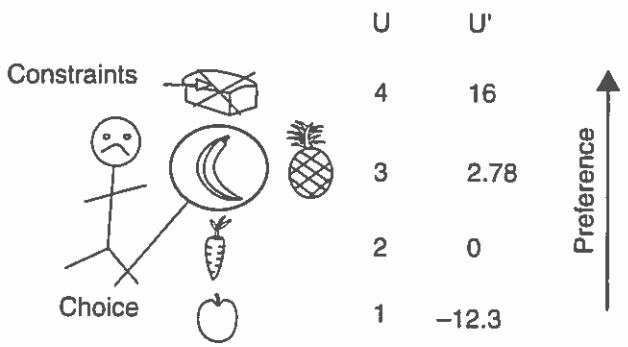

Figure 2.1. Ordinal utility.

economists maintain – and Mas-Colell et al. (1995, p. 12) purport to prove – that if preferences are rational, then choices should be consistent.[3] To defend this claim requires some axiom linking preference to choice. Here is one way to state this axiom:

(Choice determination) Among the alternatives they believe to be available, agents will choose one that is at the top of their preference ranking.[4]

So, in Figure 2.1, Jill chooses the banana rather than an apple because she prefers it to the apple. She does not choose the bread, despite preferring the bread to the banana, because no bread is available or because bread is too expensive. Because she is indifferent between the banana and the pineapple, she could just as well have chosen the pineapple.

To reach the conclusion that rational preferences imply consistent choices across different contexts requires an additional axiom concerning preferences (McClennen 1990). An anecdote attributed to Sidney Morgenbesser illustrates one aspect of the problem. According to Wikipedia:

After finishing dinner, Sidney Morgenbesser decides to order dessert. The waitress tells him he has two choices: apple pie and blueberry pie. Sidney orders the apple pie. After a few minutes the waitress returns and says that they also have cherry pie at which point Morgenbesser says "In that case I'll have the blueberry pie." (http://en.wikipedia.org/wiki/Sidney_Morgenbesser)

Morgenbesser's supposed preferences are obviously not incomplete, and they are also not intransitive, because transitivity concerns preferences over a single set of alternatives. Transitivity says nothing about how preferences over

[3] Consistency is defined by the weak axiom of revealed preference discussed in Section 3.1 of the following chapter. The basic idea is that if an agent chooses x when y is available, then the agent should never choose y from a set of alternatives that includes x.

[4] Mas-Colell and colleagues never state such an axiom explicitly. Varian (1984, p. 115) expresses it informally as "Our basic hypothesis is that a rational consumer will always choose a most preferred bundle from the set of feasible alternatives." The reference to belief, which is often left implicit, is necessary, because an agent can prefer x to y, yet choose y when x is available if the agent does not know that x is available.

one set of alternatives {apple pie, blueberry pie} should be related to preferences over a different set of alternatives {apple pie, blueberry pie, cherry pie}. If rational preferences are to imply consistent choices, then they must satisfy some additional condition. "The idea is that the choice of x when facing the alternatives $\{x, y\}$ reveals a proclivity for choosing x over y that we should expect to see reflected in the individual's behavior when faced with the alternatives $\{x, y, z\}$" (Mas-Colell et al. 1995, p. 10).

Other features of the context than the availability of alternatives may also influence preferences. For example, many people prefer hot chocolate to beer in the winter and beer to hot chocolate in the summer. Such preferences, unlike Morgenbesser's, seem reasonable, but they apparently rule out the existence of a stable ranking of alternatives. In addition, as Chapter 9 explains, psychologists and behavioral economists have found that people's preference rankings depend heavily on a further element of context: the reference point from which alternatives appear to be losses or gains. If individuals are to have a stable preference ranking of alternatives that can be used to predict their choices as the set of alternatives, the environment, and the reference point change, all these forms of context independence must be ruled out. So one needs a further axiom requiring "stability" or context independence:

(Context independence) Whether an agent prefers x to y remains stable across contexts.

This axiom (like completeness and transitivity) is not meant to exclude the possibility that agents can change their minds and alter their preferences. It is meant to exclude Morgenbesser's pie preferences, the weather dependence of preferences between beer and hot chocolate, and dependence on a reference point. Context independence is a troublesome axiom, because some kinds of context dependence are common, and some kinds appear to be reasonable. One way to reconcile the existence of apparently context-dependent preferences such as those between beer and hot chocolate with the context independence axiom is to take the description of alternatives to include "everything that matters to the agent" (Arrow 1970, p. 45). Whether hot chocolate > beer depends on the weather, but hot chocolate and beer are not the only features of the alternatives that matter. Preferences among more complex alternatives such {hot weather, beer} and {cold weather, hot chocolate} do not depend on the weather. It is a stretch, however, to suppose that people have preferences among such fully specified alternatives. Furthermore, taking the objects of preferences to be complete descriptions of all relevant aspects of the alternatives could eviscerate the context independence axiom. If Morgenbesser thought that whether cherry pie was available mattered to the choice between apple and blueberry pie, then his preferences would not violate context independence. To claim that whether cherry pie is available should be irrelevant to the choice between apple and blueberry seems to be a substantive requirement – "a rational requirement of indifference" (Broome 1991b, p. 103f) – rather than

a formal requirement of consistency, and I am sympathetic to Broome's view that "Outcomes should be distinguished as different if and only if they differ in a way that makes it rational to have a preference between them" (1991b, p. 103). Luckily, when one is considering preferences over a single set of alternatives, there is no need for context independence.

I shall thus be concerned with four axioms: (1) completeness, (2) transitivity, (3) context independence, and (4) choice determination. The first two concern preferences over a single set of alternatives, whereas the third matters only when comparing preferences over different sets of alternatives or different descriptions of the same set. The fourth axiom, choice determination, relates preferences and choices. There seem to be counterexamples to all the axioms, several of which will be discussed in Chapter 9: people cannot always compare alternatives, people's rankings are rarely perfectly transitive, context often influences rankings, and people may choose an alternative that is "good enough" even if better choices are available. Yet, these axioms might appear to be reasonable approximations in specific applications.

2.2. IMPLICATIONS OF THE AXIOMS FOR THE UNDERSTANDING OF PREFERENCES

Choice determination has two important implications concerning the interpretation of preferences. First, it implies that preferences are not just judgments. They motivate action. If Jill knows that x is feasible, and she prefers x to all other feasible alternatives, then Jill chooses x.[5] Second, because preferences *determine* choices, they must take into account *everything* relevant to choice. If, as in everyday usage, Jill's preferences leave out factors that influence her choices, such as moral commitments, then her preferences will not determine her choices. Choice determination implies that preference rankings are not partial rankings. Preferences express total comparative evaluations, and they are intrinsically motivating.

The axioms have three other important implications. First, they imply that preferences, unlike wants or wishes, are always comparative. To say that Jill wants x and wants y leaves it open whether she prefers x to y or y to x, or whether she is indifferent. Conversely, the fact that Jill prefers x to y leaves it open whether Jill wants x. Jill might dread x, but still regard it as less bad than y.

Second, the axioms place enormous cognitive demands on agents. Although it would be possible to satisfy the axioms merely by virtue of

[5] If x and y were commodity bundles, this claim might appear to be false. Jill might have chosen a banana because it is cheaper than bread, even though she prefers bread. This is not a counterexample, because economists take price as determining the availability or feasibility of a consumption choice rather than as a factor affecting preferences. Although this is not the way economists think of consumer preferences, one could also regard price as one of the properties of the alternatives that determines preferences among them.

having a remarkably finely tuned gut, the only plausible way to have a complete, transitive, and context-independent ranking of alternatives (or some reasonable approximation thereto) is to adjudicate among competing considerations. Thus, although Mas-Colell et al. (1995) maintain that preference relations summarize the decision-maker's tastes (p. 5), they also maintain that "[i]t takes work and serious reflection to find out one's own preferences. The completeness axiom says that this task has [already] taken place" (p. 6). A complete and transitive preference ranking of the complex alternatives people face would be a remarkable intellectual achievement. It must be the outcome of an unmodeled process of exhaustive comparative evaluation. The axioms render the first of the misconceptions concerning preferences – that they are arbitrary matters of taste, not subject to rational consideration – highly implausible.

Third (as mentioned previously), the "alternatives" among which people have preferences cannot be ordinary options such as eating one or two scoops of ice cream. Whether Jack prefers to eat two scoops of ice cream rather than one depends on whether he has just eaten a large dinner or is currently on a diet. Thus, Jack cannot be said to have a stable preference for one scoop or two. To have preferences that satisfy the axioms, the alternatives over which preferences are defined must specify everything that matters to the agent – which will, of course, be a great deal more than whether there are one or two scoops of ice cream.

To take preferences to be defined over bundles of commodities, economists suppose that consumers are comparing two states of the world x and y alike in everything that matters other than the fact that in x the agent consumes commodity bundle x^*, whereas in y the agent consumes bundle y^*. If nothing is relevant to preferences between x and y other than the composition of x^* and y^*, then economists can treat preferences as if they were defined over the commodity bundles. Notice that this means that prices have no influence over preferences among commodity bundles. Consumer choice theory instead takes prices and incomes as jointly fixing the constraints on choices. To assume that preferences can be defined over commodity bundles rests on the assumption that the value to individuals of commodity bundles does not depend on anything apart from the composition of the bundle. But as the example of preferring beer in the summer and hot chocolate in the winter illustrates, this assumption is false. To suppose that preferences depend exclusively on the commodity bundles requires that one either ignore such dependencies or suppose that nothing relevant to preferences other than the commodity bundles varies.

The alternatives or options that I take to be ranked by preferences specify everything *relevant* to preference. As already mentioned, specifying what is relevant may require substantive commitments concerning when indifference is rationally required. This view of the objects of preference places additional cognitive demands on individuals, because whether agents prefer x to y depends

not just on "local" properties of x and y, such as whether x is giving \$300 to Oxfam and y is purchasing a \$300 television. This preference may depend in addition on how much money others are contributing to Oxfam, what work other charities are performing, what television programs are being broadcast, what other televisions or electronic devices the agent already owns, and so forth. To have complete and transitive preferences over such complex alternatives requires more knowledge than anyone is likely to have. When there are uncertainties (as there always are), specifying the alternatives among which people have preferences presents further complications to which I return in Chapter 4 when I discuss the role of beliefs in the formation of preferences.

The axioms of ordinal utility theory say nothing about *what* people prefer. People who long for pain and suffering could satisfy the axioms as easily as those who pursue their own interests. Positive economic theory supplements the axioms of ordinal utility theory with axioms concerning the content of preferences such as the claim that people prefer more commodities to fewer. I shall not discuss these additional axioms, which say little about the concept of preference itself.

The axioms governing preference imply that preferences are rankings of complete states of the world in terms of everything relevant to choice. They are cognitively demanding and action-guiding.

2.3. RATIONALITY AND PREFERENCES

As explained in Chapter 1, any account of choice that attempts to explain and predict people's choices by their reasons must show how the factors it takes to be reasons justify choices. Accordingly, the axioms governing preferences can also be read (as Mas-Colell and colleagues read them) as imposing conditions on rational choice. What connection is there between rationality and the four axioms?

1. Completeness is a boundary condition on rational choice. An inability to compare alternatives is not itself a failure of rationality, but when people are unable to compare alternatives, they are unable to make a choice on the basis of reasons.
2. If preferences are (as I shall argue) total evaluations, then they imply judgments about what is better or more choice-worthy; and as John Broome (1991a) has argued, transitivity is then implied by the logic of comparative adjectives such as "better than" or "more choice-worthy than." If preferences place alternatives along some ruler measuring "betterness" or "choiceworthiness," then intransitive preferences are logically inconsistent and hence irrational.[6]

[6] If, on the other hand, preferences are partial pairwise comparisons instead of total evaluations, then there is no reason to expect them to be transitive. See Temkin (1987) and Tversky (1969).

3. If the alternatives that Jack ranks specify everything "relevant" to preference, then whether z is available has no relevance to the merits and drawbacks of x and y. If Jack violates context independence and his preference between x and y depends on whether z is available, then his preferences depend (irrationally) on factors that ought to be irrelevant.
4. That choices be determined by preferences is *not* demanded by rationality. Indeed attempting to adhere to choice determination may sometimes be irrational. As Herbert Simon argues (1982), it is rational to adopt strategies that reduce the cognitive burden of decision making and take account of the limits to one's information and information-processing abilities. Adopting these strategies sometimes leads people to choose options that are inferior to feasible alternatives. In addition, it may be rational to carry through with one's intentions or plans, even if changing course would be more advantageous. All one can say in defense of the rationality of choice determination is that agents should not choose x when they know they could have chosen y and are confident that, all things considered, y is the better choice to make.

These comments explain why economists regard ordinal utility theory as both a fragment of a positive theory that explains and predicts choices and as a fragment of a theory of rational choice that specifies conditions that preferences must satisfy in order to justify choices. This theory of rational choice purports to be purely *formal* and to say nothing about what things it is rational to prefer.

Although not part of utility theory, many economists think that it is rational to prefer what will make oneself better off. Self-interest is implicit in the view that individuals prefer larger commodity bundles to smaller. To claim that it is rational to be self-interested would be part of a substantive as opposed to a merely formal theory of rationality.

2.4. PREFERENCES AND SELF-INTEREST

Let us begin our consideration of how economists understand preferences by considering the influential views of Amartya Sen, a Nobel Laureate in economics, who is also a major contemporary philosophical voice. Sen defends no single definition of "preference." Instead, he emphasizes that economists have used the word to refer to different things. Among these different concepts of preference, Sen believes that two are most important. He writes:

> Certainly, there is no remarkable difficulty in simply defining preference as the underlying relation in terms of which individual choices can be explained.... In this mathematical operation preference will simply be the binary representation of individual choice. The difficulty arises in interpreting preference thus defined as preference in the usual sense with the property that if a person prefers x to y then he must regard himself to be better off with x than with y. (Sen 1973, p. 67)

A "binary *representation* of individual choice" specifies which alternative is chosen from each pair. Like a list of what someone chooses, a binary representation of choices cannot *explain* choices. This notion of preference as a choice ranking, which coincides with what economists call "revealed preference," is the subject of Chapter 3. Second, there is what Sen labels "the usual sense" of preference, whereby a person prefers x to y if and only if the person believes that he or she is better off with x than with y. In the same vein, Daniel Kahneman and Richard Thaler (2006) maintain that economists typically equate what people choose with what they predict they will most enjoy. One might read Sen as merely offering the empirical generalization that what people prefer matches what they believe to be best for themselves, rather than as suggesting a definition of preference. But other comments show that Sen regards expected advantage as one *meaning* of preference. For example, he writes, "Preference can be defined so as to preserve its correspondence with choice, or defined so as to keep it in line with welfare as seen by the person in question" (Sen 1973, p. 73; see also Sen 1980, p. 442). Let us call this sense of preference "expected advantage ranking." Expected advantage rankings express partial comparative evaluations of alternatives in terms of expected advantage.

Sen repeatedly warns against conflating choice rankings and expected advantage rankings:

> [T]he normal use of the word permits the identification of preference with the concept of being better off, and at the same time it is not quite unnatural to define "preferred" as "chosen." I have no strong views on the "correct" use of the word "preference," and I would be satisfied as long as both uses are not *simultaneously* made, attempting an empirical assertion by virtue of two definitions. (Sen 1977, p. 329)

Sen avoids legislating the meanings of words (see, for example, Sen 1991a, p. 588; 1991b). Rather than defending a single interpretation of the word "preference," he maintains that economists should recognize that the term has many meanings. Sen counsels awareness of ambiguity rather than proposing a cure, because he fears that regimentation would encourage among economists an overly simple view of evaluation and choice.

In Chapter 6, I argue against Sen that economists should employ a single concept of preferences as total comparative evaluations. Whether or not I am right about that, Sen is mistaken to suggest that "the normal use of the word permits the identification of preference with the concept of being better off." Expected advantage cannot be what people *mean* by preference, because there is no contradiction in maintaining that people's preferences depend on many things that people do not expect to bear on their own well-being. People do not apportion their donations to disaster relief by calculating how much those donations will contribute to their own well-being. Nor do they decide whether to give a truthful answer to a stranger asking directions by considering what answer is most in their interest. When, in the grip of road rage, Jack

rams the car that has cut him off, he is thinking about harming somebody else, not benefitting himself. Consider the humdrum instrumental decisions that fill daily life. People often have no idea how the alternatives bear on their interests, and in making these decisions, people are not calculating their advantage. For example, when grading student papers, I have thousands of decisions to make concerning what comments to write in the margins. Regardless of whether my determination to do a conscientious job is motivated by an expectation of personal benefit, specific choices such as whether to write "ungrammatical" rather than "awkward" in the margin next to a particular sentence are not directed by expectations about what will benefit me.

The mere *possibility* that people have preferences among alternatives without considering how they bear on their interests, or that people sometimes sacrifice their interests to accomplish something that matters more to them, shows that preferences cannot be defined in terms of well-being. And these are not mere possibilities: People often prefer x to y without believing that x is better for them than y is.

Many economic models take people to be self-interested, and for specific purposes such models are often useful. But self-interest must not be built into the meaning of preferences. By taking preferences to be total comparative evaluations (as they must if preferences are to determine choices), economists allow preferences to be influenced by everything agents regard as relevant to their choices, whether these be moral or aesthetic considerations, ideals, whims, fantasies, or passions of all sorts. Only people who are never motivated by passions, fantasies, whims, ideals, or moral and aesthetic concerns, and who are moreover able to pretend that they can always judge how alternatives bear on their own interests, could rank alternatives entirely in terms of their expected self-interested advantages.

Given that what people prefer does not always match what they judge to be best for themselves, expected advantage could not possibly be a tenable meaning of "preference." Sen's remarks thus leave us with only one alternative: choice ranking, or the theory of revealed preference, which deserves a chapter to itself.