# CHAPTER 9

# Diverse Preferences

## WHY TAPAS

*Do not do unto others as you would they should do unto you. Their tastes may not be the same.*
—George Bernard Shaw, "Maxims for Revolutionists: The Golden Rule"

WE all differ in what we prefer. Some of us like old Craftsman houses. Some of us like modern houses with open floor plans. Some of us enjoy Latin jazz. Some of us prefer heavy metal. Some of us like spicy cooked food. Some of us are vegetarian. That's fine. That's the reason for tapas restaurants. Little plates, lots of different stuff. As has often been said, there's no accounting for taste.[1] (Less often said is the equally true claim that there's no taste for accounting.) Though we cannot account for taste, we can model it. And in this chapter, that's what we do.

We need to model tastes, or what we call diverse preferences, because they create problems—huge problems. If we value different ends, we may not agree on what good solutions are or what outcomes to predict. This potential for disagreement may create incentives to misrepresent how we feel. We may try to manipulate processes and agenda, creating distrust and dislike. Thus, much of the good created by a diverse toolbox might be undone by diverse values. Of course, we needn't care about diverse preference so

much if the same people who have diverse toolboxes don't also have diverse preferences, but often they do.

To understand the severity of problems caused by diverse values, we need to see the root causes of the problems. We need frameworks and models. So, we build them. In this chapter, we learn the basics. In the next, we apply them. This treatment of preference theory is by no means complete nor is it traditional. Entire books from multiple disciplines are devoted to preferences and preference theory.[2] Most books present preference theory with an abundance of notation complicating the connection to the real world. We err in the other direction, forgoing variables whenever possible. This treatment is not traditional because it emphasizes the distinction between *fundamental preferences* (preferences about outcomes) and *instrumental preferences* (preferences about how we get what we want). Preferences about outcomes—fish tacos, healthy knees, or economic growth—are fundamental. Preferences about actions or policies—diets, stretching exercises, or tax policies—are instrumental. Actions are not ends in themselves; they are not outcomes, but they are the means to those outcomes.

Diverse fundamental preferences need not imply diverse instrumental preferences and vice versa. This finding has implications for how we think about preference diversity. People who have different fundamental preferences might be said to have different *values*. People who have different instrumental preferences but the same fundamental preferences have the same values but different *beliefs about how the world works*. In either case, people disagree over what policy or action to choose, but only in the first case does preference diversity create a problem. In the latter case, it can prove useful.

We highlight this distinction because diverse instrumental preferences derive from diverse predictive models. And it's easy to confuse the two types of diversity. So even if we don't like accounting, we have to do some. This linking of diverse predictive models to diverse perspectives offers a hint of the complexity to come. The frameworks—perspectives, heuristics, interpretations, predictive models, and preferences—can all be connected, and in many cases, diversity in one domain begets diversity in another.

## Preference Orderings and Utility Functions

To describe preferences, we first note that they are distinct from choices. Preferences describe how much we value or desire things. Choices are what we select. Our preferences guide our choices, and our choices reveal (partially) our preferences. When we meet someone new, we try to infer something about that person's preferences by his choices. What clothes does he wear? What car does he drive? What does he order for lunch? You can even use your own choices as information about your preferences. In looking at a closet filled with black shirts, you might suddenly realize that you like black, or that you like Johnny Cash.

As is standard, we assume a set of alternatives about which people have preferences. These may be locations, product designs, public policies, or political candidates. They could be *outcomes* or they could be *actions* or *policies*.[3] That difference proves important. It is the basis for the distinction between fundamental and instrumental preferences.

The most basic way to think about preferences is to conceive of them as imposing an ordering over a set of actions, policies, or outcomes. For ease of presentation, I refer to these as the *set of alternatives*. A *preference relation* describes an ordering of alternatives, $>$. The statement $A > B$ means that alternative $A$ is preferred to alternative $B$. If, for example, Joe is asked his preferences among a burrito, a taco, and an enchilada, and if he most prefers the burrito and least prefers the taco, then his preferences can be written as follows:

$$burrito > enchilada > taco$$

He might also be indifferent between two outcomes. He could like burritos and tacos equally well. If so, he is indifferent between burritos and tacos, and his preferences would be written

$$burrito \equiv taco$$

Preferences are typically assumed to be *rational*. The term *rational* has a formal definition. It means *complete* and *transitive*.

Preferences are complete if they compare any two alternatives.

*Preferences are* **complete** *if given any two alternatives, A and B, either* $A > B$, $B > A$, *or* $A \equiv B$.

We might think that everyone's preferences satisfy completeness, that everyone can compare any two alternatives. Yet it is not a vacuous assumption because people often have conflicting or deep feelings about some pairs of outcomes. They either cannot or would be reluctant to make decisions among them. If someone asks you which of your parents or children you prefer, you might find the question impossible to answer.

Preferences are *transitive* if they do not admit cycles. A person with transitive preferences cannot prefer papers to rocks, rocks to scissors, and scissors to papers.

*Preferences are* **transitive** *if they do not admit cycles; for example, if apples are preferred to bananas and bananas are preferred to pears, then apples are preferred to pears.*

The condition that preferences are transitive may also seem as though it states an obvious condition in technical language. At the individual level, it does. If Ravi prefers ice cream to yogurt and prefers yogurt to tofu, then he must prefer ice cream to tofu. For individuals, transitivity usually holds. When an individual must choose from among outcomes, she is not likely to have a preference cycle unless she is not thinking clearly. In comparing alternative plans for which gift to buy her mother (an action) in order to make her happy (an outcome), Laura might think that her mother would prefer a necklace to flowers (it lasts longer), flowers to garden tools (flowers are prettier), and garden tools to a necklace (they're more practical). This would be a preference cycle over actions, but one that should go away if Laura thought more carefully.

Though preference cycles may be rare within an individual, they are common within collections of people. In chapter 10, we analyze the aggregation of diverse preferences, we see that it is possible for a group of people to violate transitivity, to have cycles, even though none of the individuals themselves do. A collection of

rational people may prefer tofu to ice cream, even though they prefer ice cream to yogurt and yogurt to tofu.

*Preferences are* **rational** *if they are complete and transitive.*

An assumption of rational preferences seems reasonable. Completeness and transitivity are mild assumptions, but they severely restrict the amount of preference diversity that can exist. To see this, we work within a restricted framework that rules out indifference, that is, we do not allow people to be indifferent between two alternatives.

In what follows, we consider preferences about five possible actions with respect to a man's facial hair: *a goatee, muttonchops, a mustache, a full beard*, and *a Van Dyke beard*. These could also be thought of as outcomes, but we want to think of these as actions that create an outcome called attractiveness. Preferences about attractiveness would surely satisfy *monotonicity*. People prefer to be more attractive. A rational preference relation without the possibility of indifference creates a complete ordering, a ranking, over the alternatives from best to worst. One such ranking would be the following:

*full beard* > *goatee* > *Van Dyke* > *mustache* > *muttonchops*

We can calculate the total number of such orderings as follows. Any of the five facial hair styles can be ranked first, leaving four that can be ranked second, three that can be ranked third, two that can be ranked fourth, and one that can be ranked last, or be least preferred. The total number of such orderings equals $5 * 4 * 3 * 2 * 1$, or 120. If we up this to twenty alternatives, we get more then two million, billion orderings; that's why we're considering only five alternatives.

We can compare our 120 rational orderings to the number of possible irrational preference relations. Notice that we say relations and not orderings. The word *ordering* makes no sense when preferences are irrational. Irrational preferences do not necessarily order all of the alternatives. We first relax the transitivity assumption. This implies that for each pair of alternatives, a person still must have a preference, but it places no restriction on cycles.

| Pairwise Facial Hair Rankings | | |
|---|---|---|
| Van Dyke | > | full beard |
| Van Dyke | > | goatee |
| Van Dyke | > | mustache |
| muttonchops | > | Van Dyke |
| muttonchops | > | full beard |
| mustache | > | muttonchops |
| goatee | > | muttonchops |
| mustache | > | full beard |
| goatee | > | mustache |
| goatee | > | full beard |

To compute the number of preference relations that are not transitive, we begin with these ten pairs of facial hairstyles. For each pair, one of the two must be chosen. That creates two times two times two . . . (ten times), or two to the tenth power, possible preference relations. Two to the tenth power equals 1,024. Almost ten nontransitive preference relations exist for each of the 120 rational preference orderings. Most of these preference relations contain cycles (904 of them, to be precise.) Here is one: mustache is preferred to muttonchops which are preferred to the Van Dyke, but the Van Dyke is preferred to the mustache.

Were we also to allow preferences to violate completeness, we would get an even larger number of possible preference relations. Now for each pair of alternatives, in addition to either being preferred, it could also be that the alternatives are noncomparable. This creates three possibilities for each pair of alternatives. With five alternatives (and, therefore, ten pairs of alternatives) the number of preference relations that violate both transitivity and completeness equals three (not two) raised to the tenth power, 59,049, or nearly five hundred times as many as the number of rational preference orderings.[4]

These calculations demonstrate the many ways to be rational. They also show the many, many more ways to be irrational. They have implications when we study preference aggregation. Collections of people need not have transitive or complete preferences. The billions of preference orderings that an individual might have

over a set of twenty alternatives are a mere drop in the bucket compared to the number of irrational preference relations that a collection of people might have.

## Spatial Preferences

Up to now, the alternatives were arbitrary, so we had no reason to attach any significance to preferring $A$ to $B$ or $B$ to $A$. But suppose that we construct a *perspective* of these alternatives. Sometimes creating a perspective is easy. If we were to analyze how much people enjoy work, play, and sleep, we might describe an outcome as a vector (*work*, *play*, *sleep*) where the three variables denote the time spent working, playing, and sleeping, respectively. Decompositions like this into separate dimensions are a common approach in economics and political science.

Other times, representing alternatives in a perspective becomes complicated. Consider someone's preferences for food. Listing the particular food items, such as nachos, sushi, and pretzels, would be cumbersome. We could create *dimensions* that characterize food items based on ingredients. In the Ben and Jerry's example, this worked great. The number and size of chunks characterized the pints of ice cream. These two dimensions allowed Ben and Jerry to make a *spatial* representation of the various pints. However, this won't always work. Many of the items at Taco Bell contain the same ingredients in the same proportions. A taco salad is just a taco in a new arrangement.

But let's suppose that we can map the alternatives to a single-dimensional perspective. We can then distinguish between three types of preferences along that dimension. In defining each type, we take the other dimensions as fixed and ask what happens to preferences as we vary the level on one dimension. The first type of preference applies to those dimensions for which more is better.

*Preferences are* **increasing** *if more is always preferred to less.*