

# Introduction to Choice Theory

Jonathan Levin and Paul Milgrom\*

September 2004

## 1 Individual Decision-Making

Individual decision-making forms the basis for nearly all of microeconomic analysis. These notes outline the standard economic model of *rational choice* in decision-making. In the standard view, rational choice is defined to mean the process of determining what options are available and then choosing the most preferred one according to some consistent criterion. In a certain sense, this rational choice model is already an optimization-based approach. We will find that by adding one empirically unrestrictive assumption, the problem of rational choice can be represented as one of maximizing a real-valued *utility function*.

The utility maximization approach grew out of a remarkable intellectual convergence that began during the 19th century. On one hand, utilitarian philosophers were seeking an objective criterion for a science of government. If policies were to be decided based on attaining the “greatest good for the greatest number,” they would need to find a utility index that could measure of how beneficial different policies were to different people. On the other hand, thinkers following Adam Smith were trying to refine his ideas about how an economic system based on individual self-interest would work. Perhaps that project, too, could be advanced by developing an index of self-interest, assessing how beneficial various outcomes

---

\*These notes are an evolving, collaborative product. The first version was by Antonio Rangel in Fall 2000. Those original notes were edited and expanded by Jon Levin in Fall 2001 and 2004, and by Paul Milgrom in Fall 2002 and 2003.

are to any individual person. Some of the greatest thinkers of the era were both philosophers and economists.

Could the utilitarian and economic approaches be combined? That question suggests several others. How can we tell if the Smithian model of choice is right, that is, that individuals make choices in their own interests? What does that mean, precisely? How can we use data to tell whether the proposition is true? What are all the empirical implications of rational choice? What kind of data do we need to make the test? Even if the Smithian model is true, can the utility function we need for policy-making be recovered from choice data? If we can recover utilities, is simply adding up utilities really the best way to use that information for public decisions? What is the best way to use that information?

The utility-maximization approach to choice has several characteristics that help account for its long and continuing dominance in economic analysis. First, from its earliest development, it has been deeply attached to principles of government policy making. The original utilitarian program proved to be too ambitious, but the idea that welfare criteria could be derived from choice data has proved to be workable in practice. Moreover, because this approach incorporates the principle that people's own choices should determine the government's welfare criterion, it is well-aligned with modern democratic values. Second, many of the *comparative statics* predictions of the choice theory — the qualitative predictions concerning the ways in which choices change as people's environments change — tend to be confirmed in empirical studies. Third, the optimization approach (including utility maximization and profit maximization) has a spectacularly wide scope. It has been used to analyze not only personal and household choices about traditional economic matters like consumption and savings, but also choices about education, marriage, child-bearing, migration, crime and so on, as well as business decisions about output, investment, hiring, entry, exit, etc. Fourth, the optimization approach provides a compact theory that makes empirical predictions from a relatively sparse model of the choice problem — just a description of the chooser's objectives and constraints. In contrast, for example, psychological theories (with strong support from laboratory experiments) predict that many choices depend systematically on a much wider array of factors, such as the way information is

presented to the subjects, the noise level in the laboratory and other variables that might influence the subject's psychological state.

Despite the attractions of the rational choice approach, its empirical failings in economics and psychology experiments have promoted an intense interest in new approaches. A wide range of alternative models have been advocated. Learning models, in which individuals make choices like those that have worked well for them in the past, have attracted particular attention from economic theorists and experimenters. Bounded rationality models in which decision makers adopt rules that evolve slowly have had some empirical successes. For example, a model in which department stores use standard mark-ups to set retail prices appears to give a better account of those prices than does a simple profit maximization model, according to which mark-ups vary sensitively according to price elasticities. Other models assume that people seek acceptance by imitating their peers, rely on intuition on heuristics, or make choices that are heavily influenced by their current emotional state. Recent research try to identify parts of the brain involved in decision-making and model how brain processes affect decisions. Still others give up on modeling choice mechanisms at all and simply concentrate on measuring and describing what people choose.

Although these various alternatives appear to have advantages for some purposes, in this class we will focus on the decidedly useful and still-dominant model of rational choice.

## 2 Preferences and Choice

Rational choice theory starts with the idea that individuals have preferences and choose according to those. Our first task is to formalize what that means and precisely what it implies about the pattern of decisions we should observe.

Let  $X$  be a set of possible choices. In consumer choice models, one might specify that  $X \subset \mathbb{R}^n$ , meaning for instance that there are  $n$  different goods (beer, tortilla chips, salsa, etc..) and if  $x \in X$ , then  $x = (x_1, \dots, x_n)$  specifies quantities of each type of good. In general, however, the abstractness of the choice set  $X$  allows enormous flexibility in adapting the model to various applications. Some of the

controversies about the scope of economic theory concern whether the assumptions we will make below to describe consumption choices are likely to work equally well to describe choices about whether or whom to marry, how many children to have, or whether to join a particular religious sect.

Now consider an economic agent. We define the agent's weak preferences over the set  $X$  as follows:

$$x \succsim y \quad \Leftrightarrow \quad "x \text{ is at least as good as } y"$$

We say that  $x$  is *strictly preferred* to  $y$ , or  $x \succ y$ , if  $x \succsim y$  but not  $y \succsim x$ . We say the agent is *indifferent* between  $x$  and  $y$ , or  $x \sim y$ , if  $x \succsim y$  and  $y \succsim x$ .

Two fundamental assumptions describe what we mean by *rational choice*. These are the assumptions that preferences are complete and transitive.

**Definition 1** A preference relation  $\succsim$  on  $X$  is **complete** if for all  $x, y \in X$ , either  $x \succsim y$  or  $y \succsim x$ , or both.

Completeness means that if we face an agent with two choices, she will necessarily have an opinion on which she likes more. She may be indifferent, but she is never completely clueless. Also, because this definition does not exclude the possibility that  $y = x$ , completeness implies that  $x \succsim x$ , that is, that the relation  $\succsim$  is *reflexive*.

**Definition 2** A preference relation  $\succsim$  on  $X$  is **transitive** if whenever  $x \succsim y$  and  $y \succsim z$ , then  $x \succsim z$ .

Transitivity means that an agent's weak preferences can cycle only among choices that are indifferent. That is, if she weakly prefers beer to wine, wine to tequila, and tequila to beer, then she must be indifferent among all three: wine  $\sim$  tequila  $\sim$  beer. The assumption that preferences are transitive is inconsistent with certain "framing effects" as the following example shows.

**Example** Consider the following three choice problems (this example is due to Kahneman and Tversky (1984), see also MWG). You are about to buy a stereo for \$125 and a calculator for \$15.

You learn there is a \$5 calculator discount at another store branch, ten minutes away. Do you make the trip?

You learn there is a \$5 stereo discount at another store branch, ten minutes away. Do you make the trip?

You learn both items are out of stock. You must go to the other branch, but as compensation you will get a \$5 discount. Do you care which item is discounted?

Many people answer yes to the first question but no to the second. Yet it would seem that only the goods involved and their total price should matter. If we assume that the choice set consists only of goods and total cost, then this example suggests that the *framing* of the choice also matters, which contradicts the rational choice framework.

Casual evidence further suggests that the answer to the third question is indifference. So, even if we formulate the problem to distinguish choices according to which item is discounted, this collection of choices violates transitivity. To see why, let  $x$  be traveling to the other store to get a calculator discount,  $y$  be traveling to get a stereo discount, and let  $z$  be staying at the first store. The first two choices say that  $x \succ z$  and  $z \succ y$ . But the last says that  $x \sim y$ .

Given preferences, how will an economic agent behave? We assume that given a set of choices  $B \subset X$ , the agent will choose the element of  $B$  she prefers most. To formalize this, we define the agent's *choice rule*,

$$C(B; \succsim) = \{x \in B \mid x \succsim y \text{ for all } y \in B\},$$

to be the set of items in  $B$  the agent likes as much as any of the other alternatives.

There are several things to note about  $C(B; \succsim)$ .

- $C(B; \succsim)$  may contain more than one element.
- If  $B$  is finite, then  $C(B; \succsim)$  is non-empty.

- If  $B$  is infinite, then  $C(B; \succsim)$  might be empty. To see why, suppose  $B = \{x \mid x \in [0, 1)\}$ . If the agent feels that more is better (so  $x \succsim y$  if  $x \geq y$ ), then  $C(B; \succsim) = \emptyset$ .

If an agent's preferences are complete and transitive, then her choice rule will not be completely arbitrary, as the following result shows.

**Proposition 1** *Suppose  $\succsim$  is complete and transitive. Then, (1) for every finite non-empty set  $B$ ,  $C(B; \succsim) \neq \emptyset$  and (2) if  $x, y \in A \cap B$ , and  $x \in C(A; \succsim)$  and  $y \in C(B; \succsim)$ , then  $x \in C(B; \succsim)$  and  $y \in C(A; \succsim)$ .*

**Proof.** For (1), we proceed by mathematical induction on the number of elements of  $B$ . First, suppose the number of elements is one, so  $B = \{x\}$ . By completeness,  $x \succsim x$ , so  $x \in C(B; \succsim)$ . Hence, for all sets  $B$  with just one element,  $C(B; \succsim) \neq \emptyset$ . Next, fix  $n \geq 1$  and suppose that for all sets  $B$  with exactly  $n$  elements,  $C(B; \succsim) \neq \emptyset$ . Let  $A$  be a set with exactly  $n + 1$  elements and let  $x \in A$ . Then, there is a set  $B$  with exactly  $n$  elements such that  $A = B \cup \{x\}$ . By the induction hypothesis,  $C(B; \succsim) \neq \emptyset$ , so let  $y \in C(B; \succsim)$ . If  $y \succsim x$ , then by definition  $y \in C(A, \succsim)$ , so  $C(A, \succsim) \neq \emptyset$ . By completeness, the only other possibility is that  $x \succsim y$ . In that case, for all  $z \in B$ ,  $x \succsim y \succsim z$ , so transitivity implies that  $x \succsim z$ . Since  $x \succsim x$ , it follows that  $x \in C(A; \succsim)$  and hence that  $C(A, \succsim) \neq \emptyset$ . Hence, for every set  $A$  with exactly  $n + 1$  elements,  $C(A, \succsim) \neq \emptyset$ . By the principle of mathematical induction, it follows that for every finite set  $A$  with any number of elements,  $C(A, \succsim) \neq \emptyset$ , which proves (1).

For (2), if  $x, y \in A$ , and  $x \in C(A; \succsim)$ , then  $x \succsim y$ . The condition  $y \in C(B; \succsim)$  means that for all  $z \in B$ ,  $y \succsim z$ . Then, by transitivity, for all  $z \in B$ ,  $x \succsim z$ . From that and  $x \in B$ , we conclude that  $x \in C(B; \succsim)$ . A symmetric argument implies that  $y \in C(A; \succsim)$ . *Q.E.D.*

A similar proof establishes that if  $\succsim$  is complete and transitive, then any finite choice set has a worst element. We will use that variation of the proposition to construct a utility function below.

### 3 Choice and Revealed Preferences

While economic theories tend to begin by making assumptions about people's preferences and then asking what will happen, it is interesting to turn this process around. Indeed, much empirical work reasons in the reverse way: it looks at people's choices (e.g. how much money they've saved, what car they bought), and tries to "rationalize" those choices, that is, figure out whether the choices are compatible with optimization and, if so, what the choices imply about the agent's preferences.

What are the implications of optimization? Can we *always* rationalize choices as being the result of preference maximization? Or does the model of preference maximization have *testable restrictions* that can be violated by observed choices?

In the preceding section, we derived a choice rule from a given preference relation, writing  $C(B; \succsim)$  to emphasize the derivation. In empirical data, however, the evidence comes in the form of choices, so it is helpful to make the choice rule the primitive object of our theory.

**Definition 3** Let  $\mathcal{B}$  be the set of subsets of  $X$ . A **choice rule** is a function  $C : \mathcal{B} \rightarrow \mathcal{B}$  with the property that for all  $B \in \mathcal{B}$ ,  $C(B) \subseteq B$ .

In principle, we can learn an agent's choice rule by watching her in action. (Of course, we have to see her choose from all subsets of  $X$  — but more on this in the homework.) Suppose we are able to learn an agent's choice rule. Can we tell if her choice behavior is consistent with her maximizing some underlying preferences?

**Definition 4** A choice function  $C : \mathcal{B} \rightarrow \mathcal{B}$  satisfies **Houthaker's Axiom of Revealed Preference** if, whenever  $x, y \in A \cap B$ , and  $x \in C(A)$ , and  $y \in C(B)$ , then  $x \in C(B)$  and  $y \in C(A)$ .

**Proposition 2** Suppose  $C : \mathcal{B} \rightarrow \mathcal{B}$  is non-empty. Then there exists a complete and transitive preference relation  $\succsim$  on  $X$  such that  $C(\cdot) = C(\cdot; \succsim)$  if and only if  $C$  satisfies HARP.

That is,  $C$  could be the result of an agent maximizing complete, transitive preferences if and only if  $C$  satisfies HARP.

**Proof.** First, suppose that  $C(\cdot) = C(\cdot; \succsim)$  is the result of an agent maximizing complete transitive preferences. From the previous proposition, we know that  $C$  must satisfy HARP.

Conversely, suppose  $C$  satisfies HARP. Define the “revealed preference relation”  $\succsim_c$  as follows: if for some  $A \subset X$ ,  $y \in A$  and  $x \in C(A)$ , then say that  $x \succsim_c y$ . We need to show three things, namely, that  $\succsim_c$  is complete and transitive and that  $C(\cdot) = C(\cdot; \succsim_c)$ .

For completeness, pick any  $x, y \in X$ . Because  $C$  is non-empty, then either  $C(\{x, y\}) = \{x\}$  in which case  $x \succsim_c y$ , or  $C(\{x, y\}) = \{y\}$  in which case  $y \succsim_c x$ , or  $C(\{x, y\}) = \{x, y\}$  in which case  $x \succsim_c y$  and  $y \succsim_c x$ .

For transitivity, suppose  $x \succsim_c y$  and  $y \succsim_c z$ , and consider  $C(\{x, y, z\})$ , which by hypothesis is non-empty. If  $y \in C(\{x, y, z\})$ , then by HARP,  $x \in C(\{x, y, z\})$ . If  $z \in C(\{x, y, z\})$ , then by HARP,  $y \in C(\{x, y, z\})$ . So, in every possibility,  $x \in C(\{x, y, z\})$ .

If  $x \in C(A)$  and  $y \in A$ , then by the definition of  $\succsim_c$ ,  $x \succsim_c y$ . So,  $x \in C(A; \succsim_c)$ . This implies that  $C(A) \subset C(A; \succsim_c)$ . Also, since  $C(A)$  is non-empty, there is some  $y \in C(A)$ . If  $x \in C(A; \succsim_c)$ , then  $x \succsim_c y$ , so by HARP,  $x \in C(A)$ . This implies that  $C(A; \succsim_c) \subset C(A)$ . *Q.E.D.*

The preceding problem develops the properties of rational choice for the case when the entire choice function  $C(A)$  is observed. Real data is usually less comprehensive than that. For example, in consumer choice problems, the relevant sets  $A$  may be only the budget sets, which is a particular subcollection of the possible sets  $A$ .

To develop a theory based on more limited observations, economists have developed the *weak axiom of revealed preference* (WARP). Some aspects of the theory using WARP are developed in homework problems. A more detailed treatment is found in the recommended textbooks.



## 4 Utility

So far, we have a pretty abstract model of choice. As a step toward having a more tractable mathematical formulation of decision-making, we now introduce the idea of utility, which assigns a numerical ranking to each possible choice. For example, if there are  $n$  choices ranked in order from first to last, we may assign the worst choice(s) a utility of 0, the next worst a utility of 1, and so on. Picking the most preferred choice then amounts to picking the choice with the greatest utility.

**Definition 5** *A preference relation  $\succsim$  on  $X$  is represented by a utility function  $u : X \rightarrow \mathbb{R}$  if*

$$x \succsim y \quad \Leftrightarrow \quad u(x) \geq u(y).$$

That is, a utility function assigns a number to each element in  $X$ . A utility function  $u$  represents a preference relation  $\succsim$  if the numerical ranking  $u$  gives to elements in  $X$  coincides with the preference ranking given by  $\succsim$ .

Having a utility representation for preferences is convenient because it turns the problem of preference maximization into a relatively familiar math problem. If  $u$  represents  $\succsim$ , then

$$C(B; \succsim) = \left\{ x \mid x \text{ solves } \max_{y \in B} u(y) \right\}.$$

A natural question is whether given a preference relation  $\succsim$ , we can always find a function  $u$  to represent  $\succsim$ .

**Proposition 3** *If  $X$  is finite, then any complete and transitive preference relation  $\succsim$  on  $X$  can be represented by a utility function  $u : X \rightarrow \mathbb{R}$ .*

**Proof.** The proof is by induction on the size of the set. We prove that the representation can be done for a set of size  $n$  such that the range  $\{u(x) \mid x \in X\} \subset \{1, \dots, n\}$ . We begin with  $n = 0$ . If  $X = \emptyset$ , then  $\{u(x) \mid x \in X\} = \emptyset$ , so the conclusion is trivial. Next, suppose that preferences can be represented as described for any set with at most  $n$  elements. Consider a set  $X$  with  $n + 1$  elements. Since  $C(X, \succsim) \neq \emptyset$ , the set  $X - C(X, \succsim)$  has no more than  $n$  elements,

so preferences restricted to that set can be represented by a utility function  $u$  whose range is  $\{1, \dots, n\}$ . We extend the domain of  $u$  to  $X$  by setting  $u(x) = n + 1$  for each  $x \in C(X, \succsim)$ . Next, we show that this  $u$  represents  $\succsim$ .

Given any  $x, y \in X$ , suppose that  $x \succsim y$ . If  $x \in C(X, \succsim)$ , then  $u(x) = n + 1 \geq u(y)$ . If  $x \notin C(X, \succsim)$ , then by transitivity of  $\succsim$ , it must also be true that  $y \notin C(X, \succsim)$ , so  $x, y \in X - C(X, \succsim)$ . Then, by construction,  $n + 1 \geq u(x) \geq u(y) \geq 1$ .

For the converse, suppose it is not true that  $x \succsim y$ . Then,  $y \succ x$ , and a symmetric argument to the one in the preceding paragraph establishes that  $n + 1 \geq u(y) > u(x) \geq 1$ . Hence,  $x \succsim y$  if and only if  $u(x) \geq u(y)$ , so  $u$  represents  $\succsim$ . *Q.E.D.*

This proof is similar in spirit to the intuitive one, but uses mathematical induction to make the argument precise.

If  $X$  is infinite, things are a bit more complicated. In general, not every complete, transitive preference relation will be representable by a real-valued utility function. For example, consider the lexicographic preferences according to which  $x \succ y$  whenever either (1)  $x_1 > y_1$  or (2)  $x_1 = y_1$  and  $x_2 > y_2$ . These preferences cannot be represented by a real-valued utility function.<sup>1</sup> This is also an example

---

<sup>1</sup>For completeness, this footnote examines the mathematical questions of when real-valued utility functions exist and why lexicographic preferences have no utility representation. The analysis is excluded from the main text because these questions are not usually regarded as central ones in consumer theory, which is our main application of choice theory.

Let us say that a set of points  $A$  is *order-dense* (with respect to the agent's preference order  $\succsim$ ) if for every  $x \succ y$ , there is some  $z \in A$  such that  $x \succ z \succ y$ . A preference relation  $\succsim$  is representable by a real-valued utility function if and only if there exists a countable set  $A$  that is order-dense with respect to the agent's preference order. To see that the condition is necessary, notice that if  $\succsim$  is represented by some real-valued utility function  $u$ , then there exists a countable dense subset  $S$  of the range of the utility function (because the range is a subset of  $\mathbb{R}$ ). It follows that there exists a countable set of points in  $A \subset X$  such that for each  $\alpha \in S$ , there is some  $x \in A$  with  $u(x) = \alpha$ . It is easy to show that this set  $A$  is order-dense set with respect to the agent's preference order.

Conversely, if such a set  $A = \{x_1, x_2, \dots\}$  exists, then we can use it to construct a utility function. For convenience, if there is a most or least preferred element of  $X$ , let us specify that it is included in  $A$ . Next, we construct the function to represent  $\succsim$  on  $A$  by an iterative procedure. Fix  $u(x_1) = 0$ . Given utilities defined on the set  $\{x_1, \dots, x_n\}$ , we extend the function

for which indifference curves don't exist, because the agent is never indifferent between any two choices.

The following “continuity” restriction on preferences is a condition that implies not only that a utility representation exists, but that a continuous representation exists. This restriction is usually considered uncontroversial for an empirical science for the simple and compelling reason that any *finite* set of observed choices that is consistent with HARP is also consistent with continuity. That is, if the data take the form of observations of choices, then continuity can be contradicted only by an *infinite* data set.

**Definition 6** A preference relation  $\succsim$  on  $X$  is **continuous** if for any sequence  $\{(x^n, y^n)\}_{n=1}^\infty$  with  $x^n \rightarrow x$ ,  $y^n \rightarrow y$ , and  $x^n \succsim y^n$  for all  $n$ , we have  $x \succsim y$ .

**Proposition 4** If  $X \subset \mathbb{R}^n$ , then any complete, transitive and continuous preference relation  $\succsim$  on  $X$  can be represented by a continuous utility function  $u : X \rightarrow \mathbb{R}$ .

**Proof.** We prove this for the case of a strictly monotone preference relation  $\succ$ , and  $X = \mathbb{R}_+^n$ . (Note: A preference relation  $\succsim$  is strictly monotone if  $x \gg y$  implies that  $x \succ y$ , that is, if the consumer always prefers a bundle that provides more of every good.) This extra restriction allows a simple, constructive proof.

Let  $e = (1, \dots, 1)$  denote one bundle in  $x$ , and consider elements of  $\mathbb{R}_+^n$  of the form  $\alpha e = (\alpha, \dots, \alpha)$  where  $\alpha \geq 0$ . We claim that for any  $x \in \mathbb{R}_+^n$ , there exists a *unique* value  $\alpha(x) \in [0, \infty)$  such that  $\alpha(x) e \sim x$ . We will then construct  $u$  by letting  $u(x) = \alpha(x)$ .

---

as follows. If  $x_{n+1} \succ x_k$  for all  $k \leq n$ , set  $u(x_{n+1}) = 1 + \max_{k \leq n} u(x_k)$ , and if  $x_k \succ x_{n+1}$  for all  $k \leq n$ , set  $u(x_{n+1}) = \min_{k \leq n} u(x_k) - 1$ . Otherwise, set  $u(x_{n+1}) = \frac{1}{2}[\min\{u(x_k) \mid x_k \succ x_{n+1}, k \leq n\} + \max\{u(x_k) \mid x_{n+1} \succ x_k, k \leq n\}]$ . One can verify that this procedure defines a utility function that represents  $\succsim$  on the countable set  $A$ . Finally, we extend the utility representation to the whole domain  $X$  by setting  $u(x) = \inf\{u(z) \mid z \in A, z \succsim x\}$ .

Finally, we use the preceding to evaluate the example of lexicographic preferences. For that example, the smallest order-dense set  $A$  must have elements  $z$  to satisfy the inequalities  $(\alpha, 1) \succ z \succ (\alpha, 0)$  for each  $\alpha \in \mathbb{R}$ , and it is clear that no  $z$  works for two different values  $\alpha$ . Since  $\mathbb{R}$  is not countable, it follows that the set  $A$  cannot be countable.

To prove the claim, let  $x \in \mathbb{R}_+^n$  be given. Consider the two sets

$$\begin{aligned} A^+ &= \{\alpha \in \mathbb{R}_+ : \alpha e \succsim x\}, \\ A^- &= \{\alpha \in \mathbb{R}_+ : x \succsim \alpha e\}. \end{aligned}$$

Both sets are non-empty. To see why, note that  $x \succsim 0$  by monotonicity, so  $A^-$  is non-empty, and if we choose  $\bar{\alpha}(x)$  such that  $\bar{\alpha}(x)e \gg x$ , then by monotonicity  $\bar{\alpha}e \succsim x$  and  $\bar{\alpha} \in A^+$ . By the continuity of preferences, both sets are also closed, and their union is  $\mathbb{R}_+$ , so  $A^+ \cap A^-$  is not empty. Also, for any  $\alpha' > \alpha$ , monotonicity implies that  $\alpha'e \succ \alpha e$ , so  $A^+ \cap A^-$  contains at most one element: call it  $\alpha(x)$ .

Now for every  $x \in \mathbb{R}_+^n$ , we specify the utility by  $u(x) = \alpha(x)$ . We need to show that this utility function (1) represents the preference relation  $\succsim$ , and (2) is continuous. For the representation part, suppose that  $\alpha(x) \geq \alpha(y)$ . Using monotonicity,  $x \sim \alpha(x)e \succsim \alpha(y)e \sim y$ , so  $x \succsim y$ . Conversely if  $x \succsim y$ , then by construction  $\alpha(x)e \sim x \succsim y \sim \alpha(y)e$ , and by monotonicity  $\alpha(x) \geq \alpha(y)$ . Continuity is more subtle and we omit it (see MWG, Section 3.C if you're interested in a proof). *Q.E.D.*

Intuitively, the construction in the proof specifies utility of any choice  $x$  by finding the point at which its indifference surface crosses the  $45^\circ$  line. This specification is, of course, completely arbitrary. It is simply a mathematically convenient way to represent someone's preferences. In particular, if  $u$  represents  $\succsim$ , then  $U(\cdot) = v(u(\cdot))$  also represents  $\succsim$  so long as  $v : \mathbb{R} \rightarrow \mathbb{R}$  is an increasing function. This arbitrariness in the way preferences are represented has important consequences.

The intellectual history of the utility idea had its roots in utilitarian theory, according to which, for example, some goods might be more valuable to me than to you, in the sense of giving me more additional utility. For example, suppose giving me some good increases your utility by one by increases my utility by two given some representation. If one multiplies all of your utilities by four, one gets an equally valid representation of your preferences, but now the extra utility you get, which is four, exceeds my extra utility of two. This example illustrates the general principle that utilities derived from observed choices cannot be used for

“interpersonal comparisons,” that is, they cannot definitively resolve questions about the relative value of various goods to you and to me.

Several variants of utility theory have been developed that might, in principle, be used for interpersonal comparisons. These variants give meaning to statements like “food means more to a starving person than to a sated person.”

One famous variant is based on hypothetical choices in which individuals are asked to consider the possibility that they might have become either the starving person or the sated person. Behind the philosophically motivated *veil of ignorance*, they are asked to choose the rule they would apply to allocate any extra food that might become available. The attempt to base moral decisions on choices made behind the veil of ignorance will be discussed again after we have treated the theory of choice under uncertainty. A criticism of this entire approach is that it is based on hypothetical choices, rather than real ones. Hypothetical choices, critics argue, do not have the same standing as real choices and are not a reliable way to predict real choices.

Another variant develops a different conception of utility theory, based on the idea that people do not notice small differences and express indifference between choices that are close. If that accurately describes human behavior, then one might determine a unit for measuring utility by determining empirically the *just noticeable difference*. The idea is to derive a utility representation in which  $x \succ y \iff u(x) \geq u(y) + 1$  and so  $x \sim y \iff \text{not}(x \succ y \text{ or } y \succ x) \iff |u(x) - u(y)| < 1$ . This is a theory in which  $\succ$  is transitive but  $\sim$  is not, so it is fundamentally different from the theory described above. This theory formalizes the idea that a starving person may benefit more from extra food than a sated person. In fact, if a small amount of food were transferred from the sated person, she might not even notice that it was missing while the hungry person would enjoy a clear benefit.

## 5 Restrictions on Preferences

To make progress in economic research, it is almost always necessary to make additional assumptions that restrict preferences in various ways. Economists try to be careful about these assumptions. We make the minimal assumption necessary for

the analysis to be tractable and investigate all the implications of any assumption, so that they can be tested using whatever data is available.

In this section, we will look at several restrictions on preferences that are the most commonly used ones in economic analysis. Our task in each case is to identify how restrictions on preferences and restrictions on utility functions are related. On one hand, since modelers usually work with utility functions, the idea is sometimes to identify all the restrictions on choices implied by a particular assumption about utility. On the other hand, when the desired restriction on choices is given, the problem is to identify the exact restriction on utility functions that characterize the given restriction on choices.

The first restrictions examined below are *monotonicity* and *local non-satiation*, which are used extensively in consumer theory. Roughly put, these imply that consumers will prefer to spend all of their wealth or income on something, because more is always at least as good as less and consumers are never satiated. This conclusion about consumer spending will be a useful intermediate step for making inferences from consumer's observed choices.

The next restriction is that consumer preferences are *convex*. Convexity is fundamental in the standard model of competitive economies, because when consumer preferences are convex, market clearing prices exist. When preferences are not convex, market-clearing prices may not exist, in which case one cannot sustain the common hypothesis that a competitive market-clearing outcome approximates actual market outcomes. Convexity is also used in discussions about whether consumer preferences are recoverable from choices from various budget sets. In that application, convexity is used to establish that there are prices which cause different preferences to lead to different choices.

Empirical research into consumer behavior almost always relies on *separability* assumptions. For example, a researcher might postulate that a consumer's decision about how to divide her total entertainment spending among various kinds of entertainment options (such as movies, concerts, clubs, and so on) does not depend on her choices about housing, food, clothing, etc. When this assumption can be validated, it simplifies empirical work because, for example, it implies that the researcher can legitimately estimate the demand for entertainment goods based on

(1) total entertainment spending and (2) the prices of entertainment goods, even without information about the prices of food, housing and clothing.

Finally, much current research about organizations and transactions costs relies on the idea that there are *no wealth effects*. This means that consumer choices about how much to buy of certain goods are unaffected by wealth transfers (at least within certain ranges), so efficient allocations can be determined without knowing how wealth is distributed. This is a very special situation, and the associated question is the usual one: what must the utility function look like for this no-wealth-effects property to be satisfied?

Let us now turn to formal treatments of each of these preference restrictions.

**Definition 7** A preference relation  $\succsim$  on  $X$  is **monotone** if  $x \geq y$  implies that  $x \succsim y$ .

Monotonicity of preferences makes sense if  $X$  represents bundles of goods, so that if  $x = (x_1, \dots, x_n)$ , then  $x_k$  is the amount of good  $k$ . If “more of a good is good,” then preferences will be monotone.

**Definition 8** A preference relation  $\succsim$  on  $X$  is **locally non-satiated** if for any  $y \in X$ , and  $\varepsilon > 0$ , there exists  $x \in X \cap B_\varepsilon(y)$  such that  $x \succ y$ .

If an agent’s preferences are locally non-satiated, this means that there is no bundle in  $X$  that is “ideal” or even locally ideal. There is always a slight change that would leave the agent better off. Notice that local non-satiation is a joint property of the preference relation  $\succsim$  and the choice set  $X$ . If the choice set involves only integer quantities of *every* good, then the preferences cannot be locally non-satiated. However, the condition does not require that all goods be divisible; it is satisfied, for example, in every model for which one good (say, leisure time) is always valuable and divisible.

**Definition 9** A preference relation  $\succsim$  on a convex choice set  $X$  is **convex** if  $x \succsim y$  and  $x' \succsim y$  implies that for any  $t \in (0, 1)$ ,  $tx + (1 - t)x' \succsim y$ .

Convexity is often described as capturing the idea that agents like diversity. That is, if an agent is indifferent between  $x$  and  $y$  and has convex preferences,

then she will like  $\frac{1}{2}x + \frac{1}{2}y$  at least as much as either  $x$  or  $y$ . Of course, this doesn't always make sense. You might like a beer or wine, but not a mixture.

Whether convexity makes sense often depends on the interpretation of the goods space. For example, if the components of  $x$  are rates of consumption, then a half-half mixture of beer and wine might mean drinking beer half the time and wine half the time. Convexity of preferences seems more plausible in that interpretation than in the previous one. Similarly, some find convexity easier to rationalize if the "goods" are more highly aggregated — for instance, if the goods are "food" and "clothing," than if goods are highly specific.

An equivalent way to describe convexity uses the indifference curves and surfaces of undergraduate economics: convexity of preferences amounts to the assumption that the *upper contour* set of any  $y \in X$ —meaning the set of points above the indifference surface through  $y$ —is a convex set. Formally, the upper contour set is given by

$$\text{Upper Contour Set of } y = \{x \in X : x \succsim y\}.$$

There is also a related notion of *strict convexity*, which says that if  $x \succ y$  and  $x' \succ y$  and  $x \neq x'$  then for any  $t \in (0, 1)$ ,  $tx + (1 - t)x' \succ y$ .

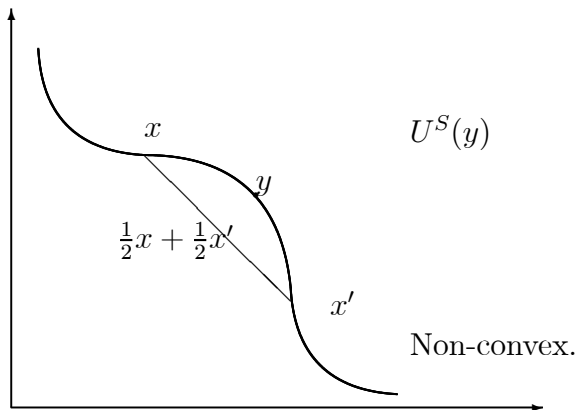


Figure 1: Non-Convex Preferences

Each of these properties of preferences has a corresponding property if we look at a preference-representing utility function.



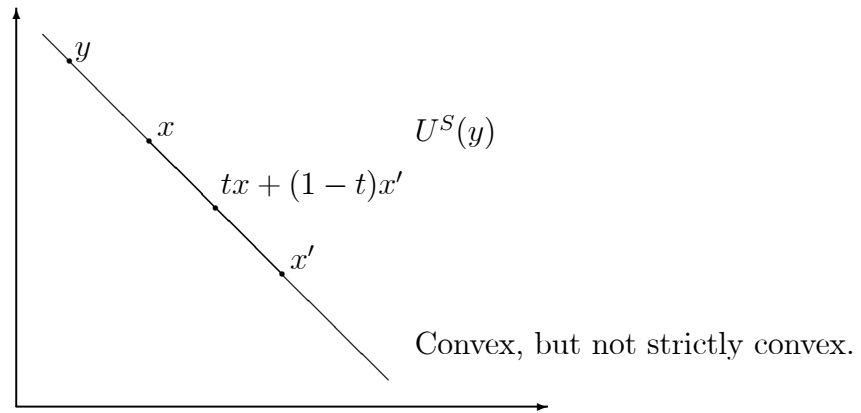


Figure 2: Convex Preferences.

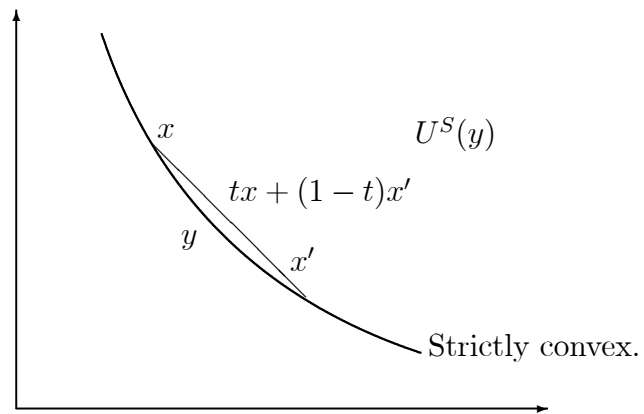


Figure 3: Strictly Convex Preferences.

**Proposition 5** *Suppose the preference relation  $\succsim$  on  $X$  can be represented by  $u : X \rightarrow \mathbb{R}$ . Then*

1.  $\succsim$  is monotone if and only if  $u$  is nondecreasing.
2.  $\succsim$  is locally non-satiated if and only if  $u$  has no local maxima in  $X$ .
3.  $\succsim$  is (strictly) convex if and only if  $u$  is (strictly) quasi-concave.

The next topic is *separability*. Suppose a consumer is choosing a bundle of goods in  $\mathbb{R}^n$ . There may be some number of goods  $m < n$  that the consumer regards as a natural group. For example, they may be different kinds of entertainment goods described by the vector  $x \in \mathbb{R}^m$ , while the other goods are described by the vector  $y \in \mathbb{R}^{n-m}$ , so that the consumer's overall choice is described by  $(x, y) \in \mathbb{R}^n$ . We investigate when it is possible to identify the consumer's vector of entertainment choices of  $x$  from limited information, namely, information about total entertainment spending and the prices of various entertainment goods, without any further information about  $y$  or the consumer's total spending. When that is possible, we say that the choice of  $x$  does not depend on  $y$ .

To treat this precisely and abstractly, suppose that what is being chosen is a pair  $(x, y) \in X \times Y$ . Consider the possibility that the consumer must choose a pair  $(x, y)$  subject to the constraints that  $y = \hat{y}$  and  $x \in X$ . Let

$$x^*(S, \hat{y}) = \{x \in S \mid (\forall \hat{x} \in S)(x, \hat{y}) \succsim (\hat{x}, \hat{y})\}$$

denote the set of  $x$ -components of the consumer's most preferred choices according to  $\succsim$  as a function of the two constraints. Formally, this notation allows the possibility that there is more than one optimal choice. There is separability in this choice when  $x^*$  does not depend on  $y$ , so that information about  $y$  is not helpful in determining the choice of  $x$ .

**Definition 10** *The function  $x^*(\cdot, \cdot)$  **does not depend on  $y$**  if for all  $y', y'' \in Y$  and all  $S \subseteq X$ ,  $x^*(S, y'') = x^*(S, y')$ .*

Note well that this definition is asymmetric. There is a clear logical distinction between the conditions (1) that the choice of  $x$  does not depend on  $y$  and (2) that

the choice of  $y$  does not depend on  $x$ . That difference is reflected in the following asymmetric characterization.

**Proposition 6** *Suppose the preference relation  $\succsim$  on  $X \times Y$  is represented by the utility function  $u(x, y)$ . Then,  $x^*$  does not depend on  $y$  if and only if there exist functions  $v : X \rightarrow \mathbb{R}$  and  $U : \mathbb{R} \times Y \rightarrow \mathbb{R}$  such that (1)  $U$  is increasing in its first argument and (2) for all  $(x, y) \in X \times Y$ ,  $u(x, y) = U(v(x), y)$ .*

**Proof.** Suppose that  $x^*$  does not depend on  $y$ . We construct  $U$  and  $v$  as follows. Fix any  $y_0 \in Y$  and let  $v(x) = u(x, y_0)$ . For any  $\alpha \in \text{Range}(v)$ , there exists some  $x' \in X$  such that  $v(x') = \alpha$ . Pick any such  $x'$  and define  $U(\alpha, y) = u(x', y)$ . (This determines  $U$  only on  $\text{Range}(v) \times Y$  and we will limit attention to that restricted function below.)

Fix any  $(x, y)$ . We must show that  $u(x, y) = U(v(x), y)$ . Let  $\alpha = v(x) = u(x, y_0)$ . By construction, there is some  $x'$  such that  $\alpha = v(x') = u(x', y_0)$  and  $U(\alpha, y) = u(x', y)$ . If  $u(x, y) > u(x', y)$ , then  $\{x\} = x^*(\{x, x'\}, y)$  but since  $u(x, y_0) = u(x', y_0)$ ,  $\{x, x'\} = x^*(\{x, x'\}, y_0)$ . These contradict the hypothesis that  $x^*$  does not depend on  $y$ . By a symmetric argument, we reject the possibility that  $u(x, y) < u(x', y)$ . This proves that  $u(x, y) = U(v(x), y)$ .

Finally, we show that  $U$  is increasing in its first argument. Suppose to the contrary that there exist  $y, x$ , and  $x'$  such that  $v(x) > v(x')$  but  $U(v(x), y) \leq U(v(x'), y)$ . Then,  $\{x\} = x^*(\{x, x'\}, y_0)$  but  $x' \in x^*(\{x, x'\}, y)$ , contradicting the condition that  $x^*$  does not depend on  $y$ .

The preceding arguments prove that if  $x^*$  does not depend on  $y$ , then there exist  $v$  and  $U$  as described in the theorem such that  $u(x, y) = U(v(x), y)$ . The proof of the converse is routine. *Q.E.D.*

The asymmetry of this characterization deserves emphasis. To understand its meaning, suppose that the choices  $x$  are various kinds of entertainment, while the choices  $y$  include restaurant meals, home meals, and housing. Suppose that the decision to purchase restaurant meals is closely related to entertainment, for example because one eats out more often when attending the movie or a concert or, reversely, because a leisurely dinner out is a substitute for other entertainment. In that case, the overall level of entertainment spending could affect the choice

between home meals and restaurant meals, even if the overall level of food spending doesn't affect the choice between movies and concerts. That is the kind of asymmetry that is captured in the representation.

Notice, too, that separability can be layered in various ways. A utility function might have the form  $u(x, y) = U(v_x(x), v_y(y))$ , which gives symmetric separability. Another possibility is that  $u(x, y, z) = U(V(v(x), y), z)$ , where  $V$  and  $v$  are real-valued functions and  $V$  and  $U$  are each increasing in the first argument. This would imply (homework!) both that the choice of  $x$  does not depend on  $(y, z)$  and that the choice of  $(x, y)$  does not depend on  $z$ . This might represent the preferences of a consumer who regards restaurant meals  $y$  as in the entertainment category, separable from non-entertainment decisions, and also regards the choice  $x$  between attending movies or concerts as independent of the quantity of restaurant meals.

For the final property of this section, let us suppose that  $X = \mathbb{R}_+ \times Y$ , which we interpret to mean that the choice space consists of a quantity of some one good and some other choices.

**Proposition 7** *Suppose the preference relation  $\succsim$  on  $X = \mathbb{R}_+ \times Y$  is complete and transitive and that there exists  $\bar{y} \in Y$  such that for all  $y \in Y$ ,  $(y, 0) \succsim (\bar{y}, 0)$ . Suppose that*

1. (“good 1 is valuable”):  $(a, \bar{y}) \succsim (a', \bar{y})$  if and only if  $a \geq a'$ .
2. (“compensation is possible”): for every  $y \in Y$ , there exists  $t \geq 0$  such that  $(0, y) \sim (t, \bar{y})$ .
3. (“no wealth effects”): if  $(a, y) \succsim (a', y')$  then for all  $t \in \mathbb{R}$ ,  $(a + t, y) \succsim (a' + t, y')$ .

*Then there exists  $v : Y \rightarrow \mathbb{R}$  such that  $(a, y) \succsim (a', y')$  if and only if  $a + v(y) \geq a' + v(y')$ . Conversely, if the preference relation  $\succsim$  on  $X = \mathbb{R} \times Y$  is represented by  $u(a, y) = a + v(y)$ , then it satisfies the three preceding conditions.*

**Proof.** By the second condition, we may define a function  $v(y)$  so that for each  $y \in Y$ ,  $(0, y) \sim (v(y), \bar{y})$ . By the third condition, for any  $(a, y), (a', y') \in \mathbb{R}_+ \times Y$ ,

$(a, y) \succsim (a+v(y), \bar{y})$  and  $(a', y') \succsim (a'+v(y'), \bar{y})$ . So by transitivity,  $(a, y) \succsim (a', y')$  if and only if  $(a+v(y), \bar{y}) \succsim (a'+v(y'), \bar{y})$ . By the first condition, that is equivalent to  $a+v(y) \geq a'+v(y')$ . Hence, the three conditions imply that  $(a, y) \succsim (a', y')$  if and only if  $a+v(y) \geq a'+v(y')$ .

It is routine to verify the converse, namely, that the representation implies the three conditions. *Q.E.D.*

The first condition of the proposition implies the local non-satiation condition. The second condition ensures that good one is sufficiently valuable that some amount of it will compensate for any change in  $y$ . The condition is most reasonable for applications in which all the relevant goods are traded in markets. It is certainly possible to imagine choice problems in which compensation of this sort is not possible. A person's choices might reflect a conviction that there is no way for cash or market transactions to compensate fully for poor health or for the loss of one's child, or for an increased likelihood of going to heaven.

The third condition is a subtle one, combining elements of separability and framing. It asserts separability, because it asserts that the choice between two alternatives does not depend on the consumer's initial endowment of good 1. In addition, because the objects of choice are outcomes and transfers, it allows the choice to be *framed* in terms of *changes* in good 1, rather than the level of consumption of good 1.

The form  $a+v(y)$  is called the *quasi-linear* form and is usually used with the good 1 interpreted as "money" or the *numeraire* good. Particularly in the theory of the firm, a profit-maximizing firm is often assumed to convert all outcomes of any sort into a money equivalent that it adds to its cash profits as a criterion to evaluate complex outcomes.

The last two propositions concern separable preferences. Students should be aware that we have just scratched the surface of separability and its use in economic applications. For example, economists sometimes wish to create a *price index* for entertainment goods that depends only on the prices of those goods. Such an index is useful if, together with the other prices and the consumer's income, it determines the total spending on entertainment goods and the consumer's welfare. The question of whether it is possible to create a price index for a category of

goods, independent of the prices of other goods, is logically distinct from the question studied above. The index sought here is used to characterize how much is *spent in total on entertainment goods* whereas above we asked about how any total spending would be allocated *among entertainment goods*. We will return to the price index problem later.

## 6 Behavioral Criticisms of Rational Choice

The centrality of the rational choice model in economic analysis means that it is important to be aware of its role and limits. There is a long tradition of research marshalling experimental and empirical evidence that is in conflict with the most basic rational choice model. And indeed the last decade has seen a growing movement that questions the model's assumptions and seeks to incorporate insights from psychology, sociology and cognitive neuroscience into economic analysis.

A main criticism of the most basic rational choice model is that real-world choices often appear to be highly *situational* or *context-dependent*. The way in which a choice is posed, the social context of the decision, the emotional state of the decision-maker, the addition of seemingly extraneous items to the choice set, and a host of other environmental factors appear to influence choice behavior. The existence of the marketing industry is testament to this, and many other examples are possible. To take a simple one, the presence of a tempting chocolate cake on the dessert menu might make you feel good about sharing an order of apple pie, when you might have ordered fruit if you hadn't been tempted by the cake.

Strictly speaking, there is little formal problem in allowing preferences to depend on context (it is even possible to incorporate cues and temptations). That being said, the strength of the rational choice model derives from the assumption that preferences are relatively stable and not too situation-dependent. This is the source of the theory's empirical content, because it allows us to observe choices in one situation and then draw inferences about choices in related situations. Such inferences become problematic if preferences are highly sensitive to context.

A further criticism of the rational choice model is that in reality, many choices are not *considered*. Rather they are based on intuitive reasoning, heuristics or

instinctive visceral desires. That people rely on intuition and heuristics is not surprising. Given that people have limited cognitive capacity, there is simply no way to reason through every decision. Arguably, instinctive judgement may often mimic preference maximization, particularly in familiar environments. When people rely on heuristic reasoning or intuition in unfamiliar situations, however, the result can be striking departures from the sort of behavior predicted by rational choice models.

Particularly surprising behavior can result when people in unfamiliar situations are given inappropriate contextual clues. For instance, Ariely, Loewenstein and Prelec (2003, *QJE*) report an experiment in which they showed students in an MBA class six ordinary products (wine, chocolate, books, computer accessories). The items had an average retail price of about \$70. Students were asked whether they would buy each good at an amount equal to the last two digits of their social security number. They were then asked to state their valuation for each good.

In spite of the familiarity of the products, students' reported valuations correlated significantly with the random final digits of their social security number. That is, it appears that the students had no firm valuation in mind and "anchored" their value to an essentially arbitrary suggestion (the social security number).<sup>2</sup> Interestingly, Ariely, Loewenstein and Prelec go on to show that once people have fixed on a valuation, they respond to price changes, and other changes, in ways that are consistent with the rational choice model. The authors label this behavior "coherent arbitrariness."

A second example that has attracted much attention is the role of default choices. For instance, Madrian and Shea (2001, *QJE*) provide evidence that enrollment in employer-sponsored 401-K retirement plans (an extremely good deal for most workers by objective criteria) is highly sensitive to whether workers must "opt-in" or "opt-out" of the plan. Another example along these lines comes from organ donations. In the United States, people must "opt-in" to become a donor by signing up when they get their driver's license. There is a dire shortage of

---

<sup>2</sup>Many experiments along these lines have been conducted, the first by Tversky and Kahneman (1974, *Science*). For a discussion of those experiments, and related ones, see Kahneman (2003, *AER*).

organ donors relative to needy recipients. In Spain, people must “opt out” and the demand-supply situation is reversed.

The behavior in these examples is hard to square happily with the most basic preference maximization approach. Once one tries to move away from optimization, however, modeling becomes a difficult challenge. That being said, there are models of decision-making that acknowledge people’s limited cognitive capacity. These models take a variety of forms: some assume that people make systematic “mistakes” or optimize only partially; others assume people used fixed learning rules, or “rules of thumb”. It is safe to say, however, that there is plenty of work left to be done in developing better “bounded rationality” models.

In closing this section, it is worth emphasizing that despite the shortcomings of the rational choice model, it remains a remarkably powerful tool for policy analysis. To see why, imagine conducting a welfare analysis of alternative policies. Under the rational choice approach, one would begin by specifying the relevant preferences over economic outcomes (e.g. everyone likes to consume more, some people might not like inequality, and so on), then model the allocation of resources under alternative policies and finally compare policies by looking at preferences over the alternative outcomes.

Many of the “objectionable” simplifying features of the rational choice model combine to make such an analysis feasible. By taking preferences over economic outcomes as the starting point, the approach abstracts from the idea that preferences might be influenced by contextual details, by the policies themselves, or by the political process. Moreover, rational choice approaches to policy evaluation typically assume people will act in a way that maximizes these preferences — this is the justification for leaving choices in the hands of individuals whenever possible. Often, it is precisely these simplifications — that preferences are fundamental, focused on outcomes, and not too easily influenced by one’s environment and that people are generally to reason through choices and act according to their preferences — that allow economic analysis to yield sharp answers to a broad range of interesting public policy questions.

The behavioral critiques we have just discussed put these features of the rational choice approach to policy evaluation into question. Of course institutions affect



preferences and some people are willing to exchange worse economic outcomes for a sense of control. Preferences may even be affected by much smaller contextual details. Moreover, even if people have well-defined preferences, they may not act to maximize them. A crucial question then is whether an alternative model — for example an extension of the rational choice framework that incorporates some of these realistic features — would be a better tool for policy analysis. Developing equally powerful alternatives is an important unresolved question for future generations of economists.