Chapter Three

Functions, Relations, and Utility

Hagle (1995, p. 7) opines that "functions are valuable to social scientists because most relationships can be modelled in the form of a function." We would add that functions are valuable for those political scientists who want to make *specific* theoretical claims and/or use statistics to test the implications of theories of politics. In other words, functions are valuable because they are explicit: they make very specific arguments about relationships. In addition, functions play a key role in developing statistical models.

What is a function? Functions may be defined in several ways, each developed more fully below. To get us started, functions provide a specific description of the association or relationship between two (or among several) concepts (in theoretical work) or variables (in empirical work). In other words, a function describes the relationship between ordered pairs (or n-tuples) arising from sets under special conditions (specified below).

That said, some students can come away from an introduction to relations and functions with a misguided notion that the key to developing sound theory is to master a wide array of functions and then see which one applies to a given theoretical or statistical problem. One might characterize this as a "toolbox" approach to political science, where different functions are hammers and wrenches to be tried here and there until one finds one that works. Perhaps such an approach would yield insight, but we are not sanguine: one's thinking about politics is unlikely to be usefully informed simply by mastery of different functional forms. Instead, a general working knowledge of functions can be used to sharpen one's thinking and bring greater specificity to one's theories of politics. In particular, learning to be able to translate verbal conjectures into graphs and/or equations that represent those conjectures is a valuable skill to develop. That skill is essential for anyone who wants to do formal modeling. Finally, such a working knowledge is critical to mastering the material in statistics courses and will help one select appropriate statistical models for hypothesis testing.

The first section discusses functions in general and elaborates on some of their properties. The second illustrates various functions of one variable; most of these can be readily generalized to multiple variables. The third section covers properties of relations in the milieu in which they are most typically seen in political science—individual preferences—and introduces the utility representation that underlies all of game theory. This serves as another example of the use of functions in political science (an empirical example appears at the end of the second section) and provides us with an opportunity to mention correspondences briefly as well. Readers with stronger math backgrounds should be able to skim the first two sections, but may not have seen the material in the third before.

3.1 FUNCTIONS

Recall from Chapter 1 that relations allow one to compare variables and expressions (or concepts). This is a general idea, but some relations are considerably more specific about the comparison. In particular, any relation that has a unique value in its range (we'll call these y values) for each value in its domain (we'll call these x values) is a function. Put differently, all functions are relations, but only some relations are functions. Another way to put this is that functions are subsets of relations. That said, political scientists do not often distinguish between relations and functions, and the term "function" is often used loosely to cover both relations and functions. Alternatively, you may encounter relations described as "set functions" and functions (as defined here) described as "point functions." More precisely, a relation that assigns one element of the range to each element of the domain is a **function**, while one that assigns a subset of the range to each element of the domain is a **correspondence**. We will focus largely on functions here, as they are the most commonly used by political scientists. However, correspondences are commonly used in game theory, and we discuss them briefly in Section 4.

More formally, a function maps the values measuring one characteristic of an object onto values measuring another characteristic of the object. Stated in set theoretic terms, a function is a relation such that (1) for all x in A, there exists a y such that (x, y) is an ordered pair in the function, and (2) if (x, y) and (x, z) are in the function, then $(y = z)$. In other words, if the value x is mapped to the value y by a function, and the value x is also mapped to the value z by the same function, then it follows that y and z are the same value. If $y \neq z$, then it is not a function but a correspondence.

Note that some equations with which you are familiar from middle school and high school math are either functions or correspondences. We review some examples below.

One can use both equations and graphs to describe functions. If you can develop an ability to translate your verbal conjectures into functions, you will have sharper, more explicit conjectures. Thus, developing the ability to work comfortably with both equations and their graphs will prove very valuable for developing your own theories about politics.

3.1.1 Equations

The linear equation $y = a + bx$ is the best-known and most frequently used function in political science.[1] We discuss it below. Here we want to remind you of the manner in which functions can be represented using equations. One often encounters equations of the form $x^2 + y^2 = 1$ or $\frac{y}{x} = 3$. We can use the rules covered previously to isolate y on the left-hand side (LHS),[2] yielding $y^2 = 1 - x^2$ and $y = 3x$. It turns out that the first of these equations is not a function while the second is, and we demonstrate that below where we introduce graphs.

You will hopefully recall the notation $y = f(x)$, which is read "y is a function of x." This is **implicit** notation that simply states that values of x are associated with singular values of y. Here we call x the **argument** of the function. But we do not know what the specific function is, so if we were given the values of x we could not produce the values of y. An **explicit** function describes the mapping of values in the domain to values in the range. For example, if we were given the explicit function $y = 3x$, then we could map the values of y for any given set of x values. In empirical work we typically refer to the x here as the independent or exogenous variable and the y as the dependent or endogenous variable, as it depends on and is affected by x.

3.1.2 Graphs

As noted above, we can graph relations and functions. If we plot the values of a set (or concept or variable) on the horizontal axis and the values of another set that shares ordered pairs with the first set on the vertical axis, then we can plot the intersection of each pair's values with a point in the space defined by the axes. Such a graph is known as a Cartesian, or xy, graph and is quite common. You will recall such graphs from arithmetic and algebra courses. The horizontal axis is also referred to as the x-axis (or domain) and the vertical axis is also known as the y-axis (or range).

The graph of the relation $x^2 + y^2 = 1$ forms a circle through the values $1, -1$ on both axes, as depicted in Figure 3.1.[3] Note that this is not a function: all values in the domain (x) produce two different values in the range (y). If this were not true, it would not form a circle. If you do not find this apparent, select a value of x and plot the value for y in Figure 3.1.

Now consider the equation $y = 3x$, shown in Figure 3.2. The graph of this equation is a straight line moving through the origin and up to the right. No matter what x values we plug into the equation we get a unique value of y. As such, the equation is a function. Note that we can make use of a graph to determine whether an equation is or is not a function: if we can draw a vertical

[1]It may surprise you that though it is often referred to as a linear function, the linear equation is not a linear function, as strictly defined in mathematics. We discuss this below.

[2]Subtract x^2 from both sides in the first case and multiply both sides by x in the second.

[3]Because it makes a circle with a one-unit radius, it is known as the unit circle.

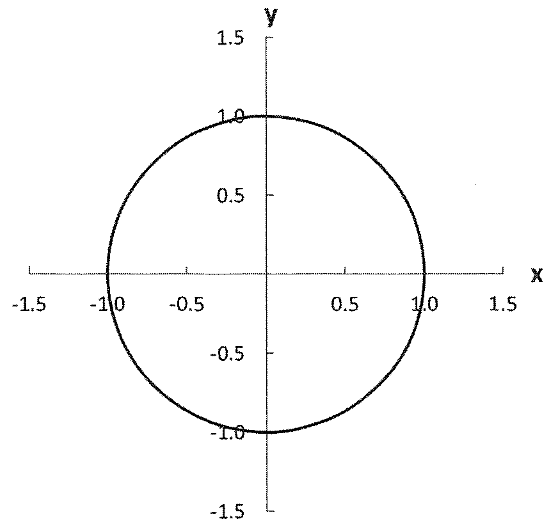

Figure 3.1: Graph of the Unit Circle

line at *any* point on the graph that intersects the curve at more than one point, then the equation is not a function.

3.1.3 Some Properties of Functions

As we go on, several properties of functions will be important. We cover inverse and identity functions, monotonic functions, and functions in more than one dimension, saving continuity for Chapter 4 and function maxima and minima, along with concave and convex functions, for Part II of the book. To begin, we expand our notation for a function slightly. We define the function f as $f(x) : A \to B$. This is often read as "f maps A into B." You've already seen the first part, which just means that the variable x is an input to the function $f(x)$, which spits out some value. Sometimes we assign this value to a variable y, as in $y = f(x)$, and sometimes we just leave it as $f(x)$, where it is understood that the function $f(x)$ may itself be a variable or a constant. For example, $f(x) = 3x$ is a variable, whereas $f(x) = 3$ is a constant.

The A and the B in the function's definition are new, but not conceptually. A here is the **domain** of the function, that is, the set of elements over which the function is defined. In other words, we draw our values of x from this set, and the function needs to produce a value for each element of x in this set. The most common domain political scientists use is the real numbers, $\mathbb{R}$, but there are numerous other domains you will see. B is known as the **codomain**, and it specifies the set from which values of $f(x)$ may be drawn. Depending on A

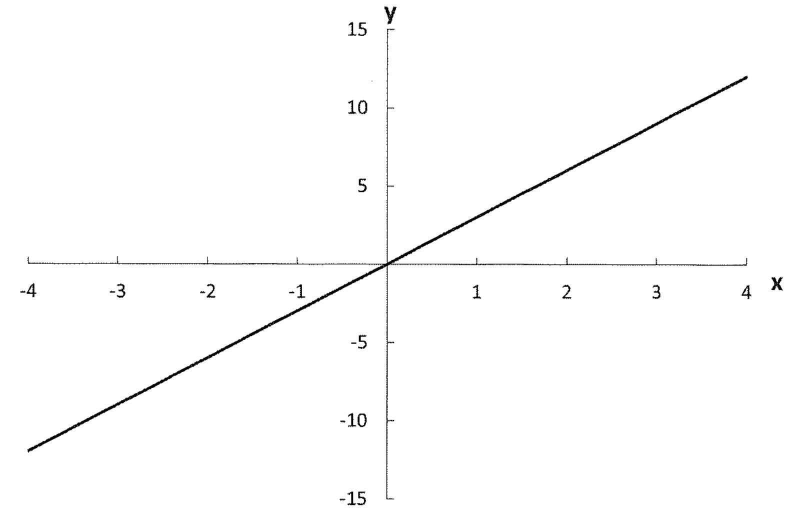

Figure 3.2: Graph of $y = 3x$

and f, though, not all the values in B may be reached. The set of all values actually reached by running each $x \in A$ through f is known as the **image**, or **range**, and it is necessarily a subset of B.

This may be confusing, so let's consider an example. Let $f(x) = x$. This function maps x to itself, and so does really nothing.[4] If $A = \mathbb{R}$, then $B = \mathbb{R}$, and the codomain and the range (or image) are exactly the same, since every real number is just mapped to itself. Now instead keep $B = \mathbb{R}$, indicating that the function f is real-valued, but let $A = (0, 1)$, or the set of all real numbers between zero and one, exclusive. In this case the image (or range) is just $(0, 1)$, which is the only part of B reached by the function, given the domain A.

One can chain multiple functions; this is called **function composition**. This is written either as $g \circ f(x)$ or $g(f(x))$ and is read as "g composed with f" or more commonly g of f of x. If we have $f(x)A \to B$ and $g(x) : B \to C$, then the full definition is $g \circ f(x) : A \to C$. Composition of functions is associative ($f \circ (g \circ h) = (f \circ g) \circ h$), but not always commutative ($f \circ g$ does not always equal $g \circ f$). One takes a function composition in stages: first one computes $f(x)$ for each x to get a set of y, and then one takes $g(y)$ for each of these y. For more than two functions that are composed, first plug each x into the innermost function, then plug the output of this into the next innermost function, and so on until you've finished with all the functions. For example, if $f(x) = 2x$ and $g(x) = x^3$, then $g \circ f(x) = (2x)^3 = 8x^3$, whereas $f \circ g(x) = 2(x^3) = 2x^3$.

[4] This function is called the identity function, and we return to it below.

Table 3.1: Identity and Inverse Function Terms

Term	Meaning
Identity function	Elements in domain are mapped to identical elements in codomain
Inverse function	Function that when composed with original function returns identity function
Surjective (onto)	Every value in codomain produced by value in domain
Injective (one-to-one)	Each value in range comes from only one value in domain
Bijective (invertible)	Both surjective and injective; function has an inverse

3.1.3.1 Identity and Inverse Functions

Why does this all matter? To answer that, we need a couple more definitions as we need to introduce identity and inverse functions, as well as some other terms. Table 3.1 summarizes those terms.

A function is **surjective** or **onto** if every value in the codomain is produced by some value in the domain.[5] Our first example was surjective, because every point in $\mathbb{R}$ was reached by some point in the domain (the same point, in the example). The second was not surjective, as nothing outside $(0, 1)$ in the codomain was reached.

A function is **injective** or **one-to-one** if each value in the range comes from only one value in the domain.[6] We already knew that each $x \in A$ produced only one $f(x)$; otherwise it wouldn't be a function. This tells us that this property goes both ways: each $y \in f(x)$ comes from only one $x \in A$. Both of our examples for the identity function are injective; the function is just a straight line. In contrast, $f(x) = x^2$ would not be injective on the same domain as, for example, $y = 4$ is the result of plugging both $x = 2$ and $x = -2$ into the function (it would be injective if we confined ourselves to real numbers no less than zero, though).

If a function is both injective and surjective (one-to-one and onto), then it is **bijective**. A bijective function is **invertible**, and so has an inverse. This inverse is the payoff of our definitions, as it allows us to take a y and reverse our function to retrieve the original x. How do we do this? First we (re)define an **identity function**: $f(x) = x, f(x) : A \to A$, where we have made the domain and codomain identical, as we saw in our earlier example. This function merely returns what is put into it and is just like multiplying each element in our domain by one (or adding zero to each element), hence the use of the word identity.

[5] Formally, it is surjective if $\forall b \in B, \exists a \in A \ni f(a) = b$ (for all b in B there exists an a in A such that the function of a is b).

[6] Formally, $\forall a, c \in A, \forall b \in B$, if $f(a) = b$ and $f(c) = b$, then $a = c$.

The **inverse function** is the function that when composed with the original function returns the identity function. That is, it undoes whatever the function does, leaving you with the original variable again. The inverse is $f^{-1}(x) : B \to A$, and remember to be *very* careful not to confuse it with $(f(x))^{-1} = \frac{1}{f(x)}$. Thus, in symbols, the inverse is defined as the function $f^{-1}(x)$ such that $f^{-1} \circ f(x) = x$, or just $f^{-1}(f(x)) = x$. The inverse does commute with its opposite $f(f^{-1}(x)) = f^{-1}(f(x))$. For example, if $f(x) = 2x + 3$, a bijective mapping, then its inverse is $f^{-1}(x) = \frac{x-3}{2}$. We can check this both ways: $f^{-1}(f(x)) = \frac{(2x+3)-3}{2} = \frac{2x}{2} = x$ and $f(f^{-1}(x)) = 2\left(\frac{x-3}{2}\right) + 3 = x - 3 + 3 = x$.

3.1.3.2 Monotonic Functions

Some functions increase over some subset of their domains as x increases within this subset. Others decrease over the same subset, and the rest increase over some x and decrease over others, depending on the value of x. If a function never decreases and increases for at least one value of x on some set $C \subseteq A$, it is an **increasing function** of x on C, while if it never increases and decreases for at least one value of x on some set $C \subseteq A$, it is a **decreasing function** of x on C. If a function increases always as x increases within C it is a **strictly increasing function** on C; if it decreases always as x increases within C it is a **strictly decreasing function** on C. Strictly increasing and strictly decreasing functions are injective. We sometimes call a function that does not decrease (but may or may not increase ever) a **weakly increasing function**, and a function that does not increase (but may or may not decrease ever) a **weakly decreasing function**.

You will sometimes encounter the term **monotonic function** in statements such as "y increases monotonically as a function of x." Monotonicity is the characteristic of order preservation—it preserves the order of elements from the domain in the range. A monotonic function is one in which the explained variable either raises or retains its value as the explanatory variable(s) rises. Thus it is an increasing function across its entire domain. A strictly monotonic function is strictly increasing over its entire domain. Table 3.2 summarizes these concepts.

We provide several examples of monotonic functions in the next section. All affine and linear functions with positive coefficients on x are strictly monotonic, as are exponential functions, logarithms, cubic equations, etc. Ordered sets can also be monotonic or strictly monotonic. An example of two ordered sets with a monotonic, but not a strictly monotonic, relationship is $\{1, 2, 3, 4, 5\}$, $\{10, 23, 23, 46, 89\}$. Monotonic functions have many nice properties that will become apparent as you study both statistics and game theory.

3.1.3.3 Functions in More Than One Variable, and Interaction (Product) Terms

Thus far we have (primarily) simplified things by focusing on the idea that y was a function of one variable. Unfortunately, few (if any!) political relationships

Table 3.2: Monotonic Function Terms

Term	Meaning
Increasing	Function increases on subset of domain
Decreasing	Function decreases on subset of domain
Strictly increasing	Function always increases on subset of domain
Strictly decreasing	Function always decreases on subset of domain
Weakly increasing	Function does not decrease on subset of domain
Weakly decreasing	Function does not increase on subset of domain
(Strict) monotonicity	Order preservation; function (strictly) increasing over domain

are so simple that they can be described usefully as a function of one variable. As such, we need to be able to use functions of two or more variables, such as $y = f(x_1, x_2, x_3)$ or $z = f(x, y)$.

Graphs of the function of one variable are straightforward, and graphs of the function of two variables are feasible (though many of us begin to struggle once we have to start thinking in three dimensions). Consider two variables multiplied by one another, also known as a product term. Product terms are a commonly used nonlinear function. Consider the plot of $y = 3xz$, in Figure 3.3, and observe that it produces a plane with a changing slope rather than a plane with a constant slope.

Another way of saying the same thing is that the relationship of x on y is different (stronger or weaker) depending on the value of z. Further, the strength of the impact of z on y also depends on the value of x. That is what is meant by interaction: x and z interact with one another to produce y. You will learn in your statistics course how to properly specify statistical models to test interaction hypotheses.[7]

As another example, consider the three-dimensional plot of the linear function $y = 3x + z$, depicted in Figure 3.4.

Graphs of the function of three or more variables, however, become terribly complex and generally are not used, though there are some exceptions. Instead of using graphs, analysis of multiple variable functions focuses on equations.

Luckily, the specification of equations in more than one variable is not much more complicated than that in one variable. You've already seen some of the notation, e.g., $f(x, y)$. The rest just accounts for the more complex domain

[7]We note here that one would not want to estimate the model $y = \alpha + \beta(xz) + e$ because doing so would produce a biased estimate of β. Rather, one would want to include the variables x and z in the model as well (Blalock Jr., 1965; Friedrich, 1982; Braumoeller, 2004; Brambor, Clark, and Golder, 2006). This will be discussed in your statistics courses.

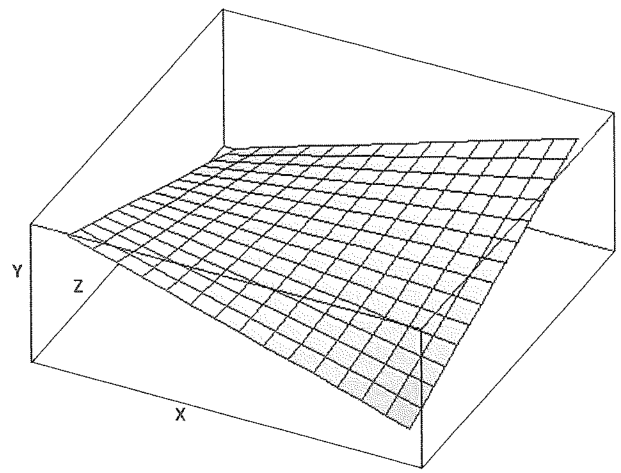

Figure 3.3: Graph of $y = 3xz$

that is present when there is more than one variable. If there are n variables, denoted x_1 through x_n, and the set from which each variable is drawn is called A_1 through A_n, respectively, then the domain of the function is the Cartesian product $A_1 \times A_2 \times \ldots \times A_n$. The formal definition of the function is $f(x_1, \ldots, x_n) : A_1 \times \ldots \times A_n \to B$. To get any value of f you just plug in the values of all the input variables. Most of the concepts discussed above are either directly applicable or have analogues in the multidimensional case, though there is more complexity involved. For example, properties such as continuity can be defined for each input variable independently. We save discussion of the properties of multi-dimensional functions most relevant to us until Part V of the book, however.

3.1.4 Why Should I Care?

A basic understanding of functions is critical to any political scientist who wants to be able to make specific causal conjectures. Making specific causal conjectures is useful because it increases one's ability to evaluate whether relevant evidence is at odds with one's theory (i.e., improves hypothesis testing; Popper, 1959, pp. 121–23) and it facilitates communication with other scholars (Cohen and Nagel, 1934, pp. 117–20). Vagueness is antithetical to science, and stating hypotheses as functions helps one eliminate vagueness. Further, statistical inference is a powerful tool for hypothesis testing, and functions are one of the building blocks on which statistics is constructed. Finally, game theory makes extensive use of functional forms to represent preferences and payoffs, as we'll see in Section

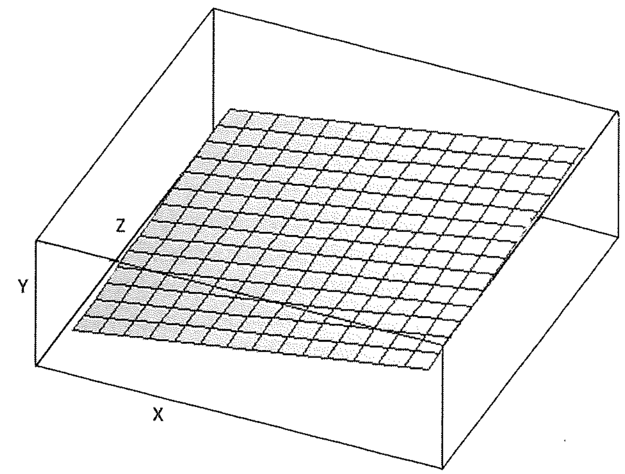

Figure 3.4: Graph of $y = 3x + z$

3 of this chapter. For these reasons, the properties of functions we discuss in this section are fundamental, as they have substantive meaning in the settings in which we are using the functions. Monotonocity in one's preferences, for example, means that someone always prefers more to less. This is very different from having what is known as an ideal point, in which case moving away from the ideal in *either* direction is not preferred.

3.2 EXAMPLES OF FUNCTIONS OF ONE VARIABLE

Political scientists are generally interested in the relationships among multiple variables. Nevertheless, in this section we begin with associations where y is a function of one x. These functions extend readily to more than one variable, as noted above.

3.2.1 The Linear Equation (Affine Function)

You encountered the additive linear equation back in algebra classes: $y = a+bx$. Technically, this is an **affine function**, though it is frequently referred to as a linear function. We discuss the technical distinction between the two below. For now, let's review some basics.

In the equation $y = a + bx$, a and b are constants.[8] The constant a is the

[8]Recall that in this equation, y is a function of only one variable, x. Therefore a and b cannot be variables and must be constants.

intercept, or in terms of the graph, where the function crosses the vertical (y) axis (i.e., the value of y when $x = 0$). The constant b is the slope of the line, or the amount that y changes given a one-unit increase in x. That is, a one-unit increase in x produces a $1b$-unit increase in y, a three-unit increase in x produces a $3b$-unit increase in y, etc.

One might conjecture that the probability that an eligible voter casts a ballot in a US presidential election is a linear function of education.[9] Let p_v represent the probability of voting and ed represent education level: $p_v = a+b(ed)$. In this function a represents the likelihood that someone without any formal education turns out to vote, and b indicates the impact of education on the probability of voting. Shaffer (1981, p. 82) estimates a model somewhat like this, and we can borrow his findings for illustrative purposes, yielding $p_v = 1.215 + 0.134 \times ed$. The intercept of 1.215 makes little sense,[10] but we ignore that for this example. Shaffer's education measure has four categories: 0–8 years of education, 9–11 years, 12 years, and more than 12 years. A slope (i.e., b) of 0.134 suggests that as we move from one category to another (e.g., from 0–8 years to 9–11 years, or from 12 years to more than 12 years), the probability that someone votes rises by 0.134. So if this linear model and its results are accurate, the typical adult with a college education has roughly a 0.4 greater probability of voting in a US presidential election than the typical adult without any high school education.[11]

The linear equation states that the size of the impact of x on y is constant across all values of x. For example, in the above example the impact of x on y is roughly 0.13. Since the relationship is linear, that means that a shift from 0–8 years of education to 9–11 years of education increases the probability of voting in a national election by $\sim$.13, and a shift from 9–11 years to 12 years also produces an increase of $\sim$.13, as do shifts from 12 years of education to more than 12 years of education. Nonlinear functions, which we discuss in the third subsection, specify that the size of the impact of x on y varies across values of x.

[9]We recognize that people are very unlikely to posit such a claim. We offer it not as a reasonable conjecture but simply as an illustration.

[10]A value of 1.2 is nonsense because the intercept represents the probability of voting when a person has had zero education. Since probabilities by definition have a range from zero to one, any probability above one is nonsense. This is but one reason that this may be an unrealistic example. In your statistics courses you will learn a number of reasons why this estimate is nonsense.

[11]You will learn how to do these sorts of calculations in your statistics courses. For those who want a brief description, you need to calculate two values and then determine the distance between them. More specifically, multiply the slope (0.134) by the first value in the comparison, someone with no high school education: $1 \times 0.134 = 0.134$. Now multiply the slope by the second value in the comparison, someone with a college education: $4 \times 0.134 = 0.536$. Finally, take the difference of these two probabilities of voting (i.e., subtract the probability that a citizen without a high school education votes from the probability that a college educated citizen votes) to get $0.536 - 0.134 = 0.402 \sim 0.4$.

3.2.2 Linear Functions

Mathematicians make distinctions that few political scientists employ. We review them for the purpose of helping you avoid confusion when you read "mathematically correct" presentations. In particular, we distinguish between affine functions (discussed above), linear equations, and linear functions (discussed here). As suggested above, a **linear equation** is an equation that contains only terms of order x^1 and $x^0 = 1$.[12] In other words, only x and 1, multiplied by constants, may appear on the right-hand side (RHS) of a linear equation. This means that the RHS of a linear equation is an affine function. Linear *functions* are not affine functions; e.g., they do not permit a translation (the x^0 term).

The formal definition of a **linear function** is any function with the following properties:

- Additivity (aka superposition): $f(x_1 + x_2) = f(x_1) + f(x_2)$,
- Scaling (aka homogeneity): $f(ax) = af(x)$ for all a.

Additivity states that the impact of a sum of variables is equivalent to the sum of the impacts of those variables. The scaling property, on the other hand, states that the size of the input is proportional to the size of the output.

Let's begin by comparing the linear function $y = \beta x$ with the affine function $y = \alpha + \beta x$ along these criteria. The additivity property states that $f(x_1+x_2) = f(x_1)+f(x_2)$. So we substitute the RHS of each $y = \ldots$ equation for the parts in the parentheses (i.e., $f(\cdot)$) and see if that statement is true. If it is, the property is met. We begin with the linear function $y = f(x) = \beta x$. To determine whether it meets the additivity property, we need to replace x with $x_1 + x_2$, following the additivity property equation above, and determine whether the equality is true:

$$\begin{aligned} f(x_1 + x_2) = \beta(x_1 + x_2) &= \beta x_1 + \beta x_2, \\ \beta x_1 + \beta x_2 &= f(x_1) + f(x_2). \end{aligned}$$

As one can see, the equality is true. Now we'll try the linear equation (or affine function), under the assumption that $\alpha \neq 0$: $y = f(x) = \alpha + \beta x$. Again, we replace x with x_1 and x_2, in accord with the additive property equation, and see whether the equality is true:

$$\begin{aligned} f(x_1 + x_2) = \alpha + \beta(x_1 + x_2) &= \alpha + \beta x_1 + \beta x_2, \\ f(x_1) + f(x_2) &= (\beta x_1 + \alpha) + (\beta x_2 + \alpha), \\ \alpha + \beta x_1 + \beta x_2 &\neq 2\alpha + \beta x_1 + \beta x_2. \end{aligned}$$

It is not true; the RHS and LHS differ by α. So the linear equation (or affine function) does not have the additive property, but the linear function does.

[12]Order refers to the highest exponent in the polynomial.

Now let's consider the scaling property, which states that $f(ax) = af(x)$. Let's begin with the linear function $y = f(x) = \beta x$:

$$f(ax) = \beta(ax) = a\beta x,$$
$$a\beta x = af(x).$$

So, the linear function satisfies the scaling property. What about the linear equation (i.e., affine function)?

$$f(ax) = \alpha + (\beta(ax)) = \alpha + a\beta x,$$
$$af(x) = a\alpha + a\beta x,$$
$$\alpha + a\beta x \neq a\alpha + a\beta x.$$

This property doesn't hold either because $\alpha \neq a\alpha$. Again, the linear function satisfies the property, but the affine function (linear equation) does not.

The only difference between the two functions is the constant, α. Recall that α represents the value where the function crosses the vertical (y) axis. If it crosses at zero, then the two functions are equivalent. Thus, a linear function must cross the vertical axis at the origin (i.e., where x and y have a value of zero). You might recall that ratio level measurement requires a meaningful zero value, whereas interval level measurement does not, and that division and multiplication operations are valid on ratio level measures but not on interval level measures. Linear transformations require preservation of the order of the variables, the scale, and the zero, and only linear functions meet such criteria. Affine transformations preserve order and scale, but not the placement of zero.

Above we noted that political (and other social) scientists frequently refer to the linear equation $y = \alpha + \beta x$ as a linear function. *Technically*, this is inaccurate, but it is a rather fine mathematical point. The linear equation does produce a line, and a linear transformation with the affine function preserves order and scale, with the exception of the intercept. And that is all most political scientists are typically trying to indicate when they talk about linear functions and linear transformations. That said, there are some applications (e.g., time series analysis) where the proper definition of a linear function is important, and we raise the discussion here so as not to later confuse those who go on to study those issues in more detail.

3.2.3 Nonlinear Functions: Exponents, Logarithms, and Radicals

Technically speaking, nonlinear functions are all those that do not meet the two properties we just discussed. Practically speaking, nonlinear functions are all those that are neither linear nor affine: those functions that describe (the graph of) a curve that is not a line. For example, $y = \cos(x)$, in Figure 3.5, is a nonlinear function. Functions with exponent terms, including quadratics and other polynomials, are the most commonly used nonlinear functions in political science. Logarithms are another commonly used class of nonlinear functions,

as are roots (or radicals). We briefly introduce the relationship among these functions and then turn our attention to graphing these functions and using them in algebra.

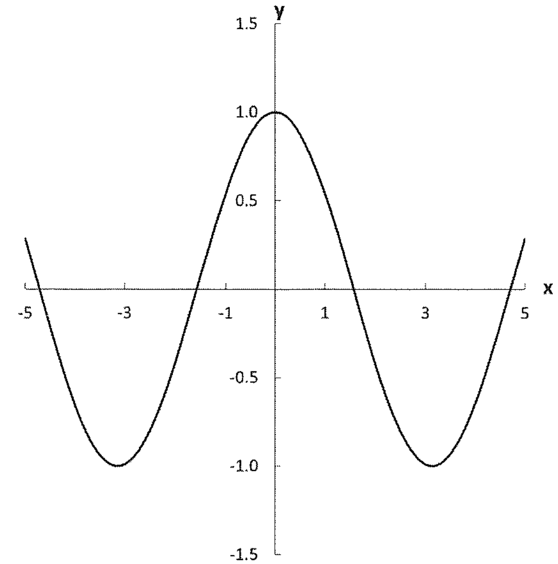

Figure 3.5: Graph of $y = \cos(x)$

Exponents, logarithms, and roots are related: one can transform any one such function into a representation of one of the others. In fact, in high school you may have focused on doing that. More specifically, when two of the following variables in the equation $b^n = x$ are known, one can solve for the unknown using

- Exponents to solve for x,
- Logarithms to solve for n,
- Radicals to solve for b.

That said, we will not focus on the relationship among the functions as political scientists do not frequently make use of those relationships.[13] Instead, we introduce each function and its notation, discuss their graphs, and then describe algebraic manipulations.

3.2.3.1 Exponents and the Exponential Function

As notation, **exponents** (aka power functions) are a shorthand for expressing the multiplication of a number by itself: $x^3 = x \times x \times x$. More generally, $x^n = x \times$

[13]Those interested in studying this might find the following Wikipedia entries useful: `http://en.wikipedia.org/wiki/Logarithm`, `http://en.wikipedia.org/wiki/Radical_(mathematics)#Mathematics`, `http://en.wikipedia.org/wiki/Exponential_function`.

$x \times x \ldots x$ (n times). This is all familiar, but you may be less familiar with other exponential notation: $x^{-n} = \frac{1}{x^n}, x^{\frac{1}{n}} = \sqrt[n]{x}$. In words, x to a negative power represents the fraction "1 divided by x^n" and x raised to a fraction represents a root of x, where the root is determined by the value in the denominator of the exponent. Perhaps an easier way to remember this is that a negative exponent indicates that one *divides* (rather than multiplies) the term by that many factors. Similarly, a fractional exponent indicates that one takes the n^{th} *root* rather than multiplying the term n times. Mixed exponents work similarly. So $x^{\frac{2}{3}} = \sqrt[3]{x^2}$ and $x^{\frac{-3}{2}} = \frac{1}{\sqrt[2]{x^3}}$. Finally, $x^0 = 1$.[14]

Nonlinear functions with exponents are of interest to political scientists when we suspect that a variable x has an impact on y, but that the strength of the impact is different for different values of x. The best way to see this is to look at the graphs of some functions with exponents.

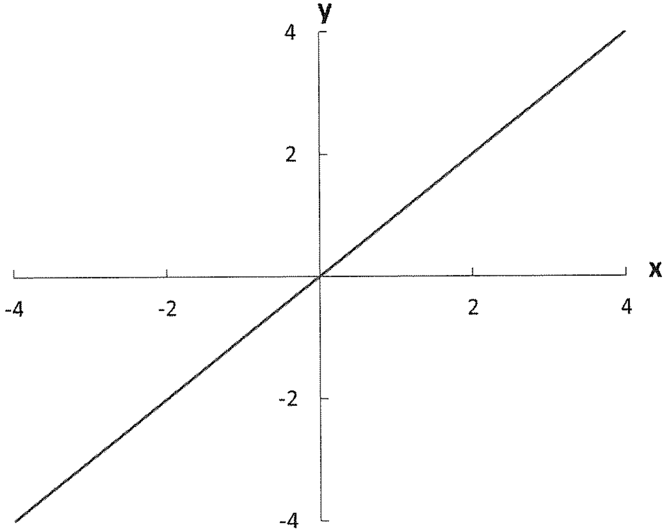

Figure 3.6: Graph of $y = x$

Consider the graphs of the functions $y = x$ and $y = x^2$, in Figures 3.6 and 3.7. The linear function produces a line with a constant slope: if we calculate the change in y due to a one-unit change in x, it does not matter what point on the x-axis we select; the change in y is the same.[15] However, the slope of the curve for $y = x^2$ is not constant: the impact of x on y changes as we move along the x-axis (i.e., consider different values of x). To be more concrete, a one-unit

[14] This holds for all $x \neq 0$, but people often treat $0^0 = 1$ as if it were true when they are simplifying equations.

[15] Another way to make this point is to observe that linear functions meet the scaling property.

increase from 0 to 1 produces a one-unit increase in y, but a one-unit increase from 2 to 3 produces a five-unit increase in y, and a one-unit increase from 5 to 6 produces an 11-unit increase in y. Thus, the impact of x on y increases over the range of x.

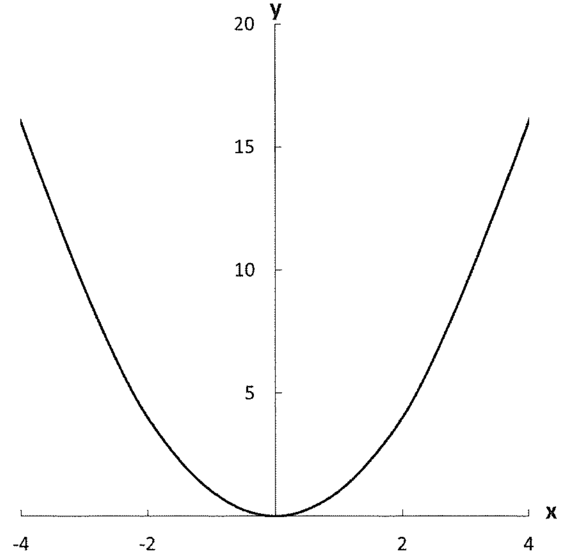

Figure 3.7: Graph of $y = x^2$

This has important implications for developing theory. If reflection, deduction, or inspiration leads one to conjecture that a causal relationship between two concepts is constant over the range of values for the causal concept, then a linear or affine relation represents that conjecture. However, if one suspects that the strength of the relationship varies across the values of the causal concept, then a nonlinear relation is needed. As we discuss below, exponential terms play an important role in quadratic and other polynomial functions.

We covered some of these above, but below is a list of the algebraic rules that govern the manipulation of exponents.

Multiplication: to calculate the product of two terms with the same base one takes the sum of the two exponents:

$$x^m \times x^n = x^{m+n}.$$

To see that this is so, set $m = 3$ and $n = 4$ and write it out:

$$x^3 \times x^4 = (x \cdot x \cdot x) \times (x \cdot x \cdot x \cdot x) = x \cdot x \cdot x \cdot x \cdot x \cdot x \cdot x = x^7.$$

This works when m and n are positive, negative, or zero. When the bases are different, you can simplify the expression *only when the exponents are the same.*

In this case, multiplication is distributive:

$$x^m \times z^m = (xz)^m.$$

To see why, set $n = 2$, and note that $x^2 \times z^2 = x \cdot x \cdot z \cdot z = (x \cdot z) \times (x \cdot z) = (xz)^2$.[16]

Last, when both the base and the exponent are different, you cannot simplify to a single term. Thus, e.g.,

$$x^m \times z^n \neq (xz)^{m+n}.$$

Assume that $m = 2$ and $n = 3$, and write the expressions out to see that this is so:

$$x^2 \times z^3 = (x \times x) \times (z \times z \cdot z) = z((x \times z) \times (x \times z)) \neq (xz)^5.$$

One cannot combine the terms fully.[17] To return to the point made above, if we assume that $m = n = 3$, then when we write it out we get:

$$x^3 \times z^3 = (x \cdot x \cdot x) \times (z \cdot z \cdot z) = (x \cdot z) \times (x \cdot z) \times (x \cdot z) = (xz)^3.$$

To determine the **power of a power**, one multiplies the exponents. For example,

$$(x^m)^n = x^{mn}.$$

To see that this is so, let's assign $m = 2$ and $n = 3$, and write out:

$$(x^2)^3 = x^2 \times x^2 \times x^2 = x \cdot x \cdot x \cdot x \cdot x \cdot x = x^6.$$

Division: to calculate the quotient of two terms with the same base and different powers, one takes the difference of the exponents:

$$\frac{x^m}{x^n} = x^{m-n}.$$

To see why this is so, recall that

$$\frac{1}{x^n} = x^{-n}.$$

We can therefore write out:

$$\frac{x^m}{x^n} = x^m x^{-n} = x^{m-n}.$$

[16] Note that this assumes that multiplication is commutative; hence this will not hold for matrix multiplication, as we'll see in Part IV of the book.

[17] If this illustration is not clear to you, then assign values to x and z (say, 2 and 3) and work it out. It will become clear that one can take the product when the exponents are equal and the bases are different, but one cannot take the product when both the exponents and bases are different. Note that we can simplify to a degree: $x^2 \times z^3 = (xz)^2 z$, but this is not usually helpful.

We can assign the values $m = 2$ and $n = 3$ and verify

$$\frac{x^2}{x^3} = \frac{x \cdot x}{x \cdot x \cdot x} = \frac{1}{x}$$

and

$$\frac{x^2}{x^3} = x^2 x^{-3} = x^{-1} = \frac{1}{x}.$$

When the bases are different, one can simplify only if the exponents are the same. When the exponents are the same, one raises the fraction to that power:

$$\frac{x^m}{z^m} = \left(\frac{x}{z}\right)^m.$$

Put differently, like multiplication, division is distributive when the bases are different and the exponents are the same.[18]

Recall that $x^0 = 1$. We can now demonstrate this by observing that $\frac{x^n}{x^n} = 1$. Observe that $\frac{x^n}{x^n} = x^{n-n} = x^0$. Since anything divided by itself equals one, it follows that $x^0 = 1$ (except when $x = 0$).

This covers x^a, but what about a^x? This is called an **exponential**. The one most commonly used sets $a = e$, where e is the base of the natural logarithm, or $e \approx 2.7183$ (to four decimal places). This is the **exponential function**, written as $y = \exp(x)$ or $y = e^x$.[19] We discuss the base of the natural logarithm, and its relation to the exponential function, below; in Figure 3.8 we graph the exponential function.

3.2.3.2 Quadratic Functions

Quadratic functions are nonlinear functions that describe a parabola. More specifically, if y is a quadratic function of x, then $y = \alpha + \beta_1 x + \beta_2 x^2$. In other words, quadratic functions describe a relationship where a variable (y) is a function of the sum of another variable (x) and its square (x^2).[20]

[18] In fact, other than having to remember not to divide by zero, multiplication and division have basically the same properties. The same is true for addition and subtraction. Thus one need remember only the properties of multiplication and addition.

[19] Both notations are common, and they are equivalent expressions.

[20] It is worth observing that many political scientists refer to a quadratic function as linear. For example, in a regression course you may encounter the claim that $y = \alpha + \beta_1 x + \beta_2 x^2 + e$ is a linear model. That is true: it is a linear *model*. When discussing regression models people frequently distinguish between models that are *linear in parameters* from those that are *linear in variables*. A regression model that contains a quadratic function (e.g., $y = \alpha + \beta_1 x + \beta_2 x^2 + e$) is linear in parameters but nonlinear in variables. Put differently, if we plot the relationship between x and y, the plot will be nonlinear: it is not linear in variables. But the parameters of the quadratic function have the properties of an affine function (to see this, set $z = x^2$ and rewrite the linear model as $y = \alpha + \beta_1 x + \beta_2 z$), and if we assume that $\alpha = 0$, then they have the properties of a linear function. Returning to models, the model $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + e$ is linear in parameters and variables (as long as we assume that x_1 and x_2 are not nonlinear transformations of one another (e.g., $x_1 \neq x_2^n$)).

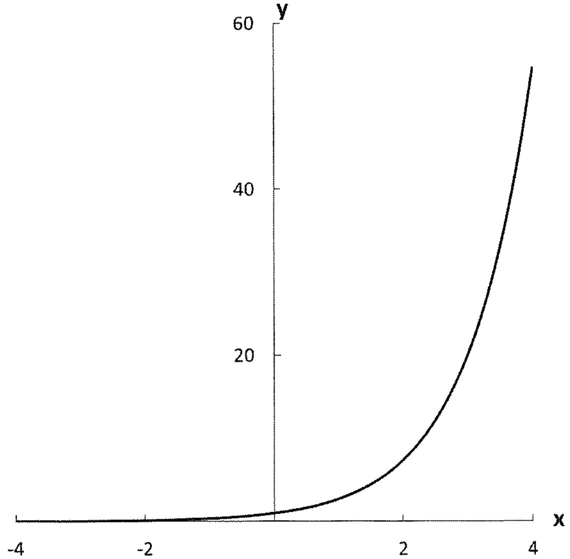

Figure 3.8: Graph of $y = e^x$

Note that since y is a function of only one variable, x, we can graph the function in two dimensions. If we set $\beta_2 < 0$, then we get a curve shaped like an inverse U (i.e., a concave parabola) as depicted in Figure 3.9. Switching the sign of β_2 produces a U-shaped curve (i.e., a convex parabola).

What sort of theoretical expectations might one want to sharpen by stating them as a quadratic relationship? Speaking generally, a quadratic function is quite useful for depicting relationships where we think the impact of an independent variable is positive (negative) for low values of the independent variable, flat for middle-range values, and negative (positive) for high values. Put differently, when one thinks that there is some (often unknown) threshold at which the relationship between two concepts (variables) switches (i.e., from positive to negative or from negative to positive), one might consider whether the quadratic can represent our conjecture.

For example, many scholars have hypothesized that rebellion will be low in countries that exert little to no government coercion *and* in countries that exhibit high levels of government coercion. Where will one find rebellion? This conjecture suggests that it will be highest among those countries that engage in mid-range levels of coercion (e.g., Muller and Seligson, 1987). If we let r represent rebellion and c represent coercion, then this conjecture can be represented as follows: $r = \alpha + \beta_1 c - \beta_2 c^2$.

Another example is the conjecture that the extent to which governments are transparent (i.e., noncorrupt) varies nonlinearly with the level of political competition. More specifically, over the range from authoritarian to democratic

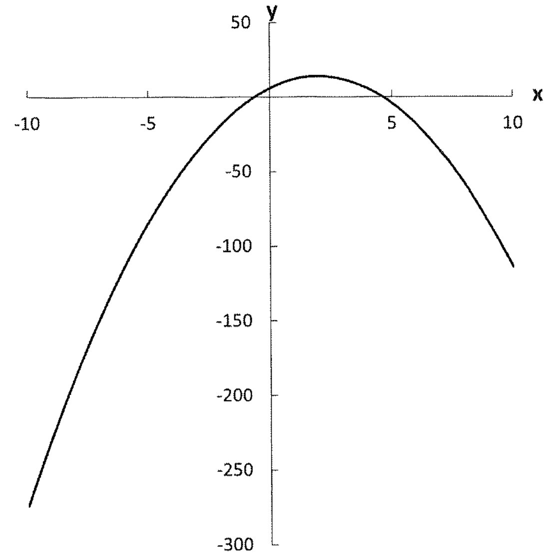

Figure 3.9: Graph of $y = 6 + 8x - 2x^2$

polities, transparency (e.g., the absence of bribery) is relatively common at both endpoints and least common in mixed polities that have a mix of autocratic and democratic institutions (e.g., Montinola and Jackman, 2002). If we allow t to stand for transparency and p for polity type, then we can represent that conjecture with the following quadratic equation: $t = \alpha + \beta_1 p + \beta_2 p^2$.

Finally, note that if we invert the concept we are trying to explain (i.e., flip the scaling of the dependent variable), we can represent the argument by flipping the signs on the quadratic (x^2) term. Thus, if we reconceptualize rebellion as quiescence, q, then we can write $q = \alpha + \beta_1 c + \beta_2 c^2$, and if we reconceptualize transparency as corruption, k, then we can write $k = \alpha + \beta_1 p - \beta_2 p^2$.

3.2.3.3 Higher-Order Polynomial Functions

Polynomial functions have the following general form: $y = \alpha + \beta_1 x + \beta_2 x^2 + \ldots + \beta_n x^n$, where n is an integer less than infinity. So both linear and quadratic functions are polynomials. Higher-order polynomials are those possessing powers of x greater than the reference. In this case, we are referring to the presence of cubed and higher terms. Like quadratics, higher-order polynomials are nonlinear: they describe curves, such as the cubic polynomial in Figure 3.10. More specifically, one can use them to explicitly represent the expectation that there are two or more thresholds over which the relationship between two concepts (variables) changes.

With the exception of the quadratic, polynomial functions are not very common in political science, though Mukherjee (2003) and Carter and Signorino

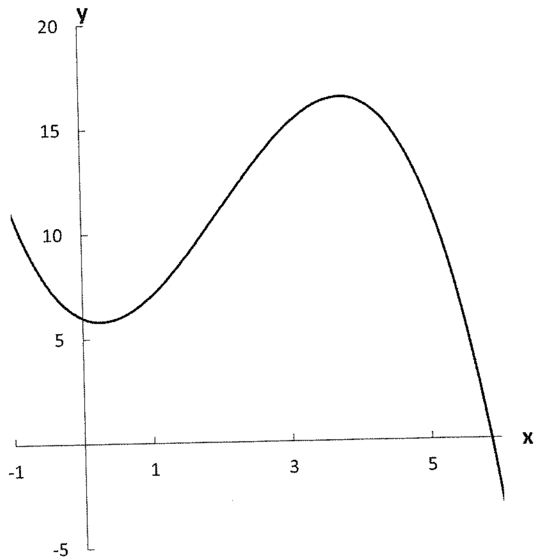

Figure 3.10: Graph of Cubic Polynomial

(2010) are exceptions. Mukherjee studies the relationship between the size of the majority party and central government expenditures in parliamentary democracies. A majority party is one that has at least 50% of the seats in the legislature and thus can govern without having to form a coalition with other parties. The basic underlying idea is that there are two different thresholds at work between the number of seats the majority party holds in the legislature and the size of government spending. First, as the number of seats held by the majority party rises from a bare majority (i.e., 51% of the legislature), spending declines, because it takes more and more legislators to defect and bring down the government.

Yet, while Mukherjee expects an initial negative relationship as the size of the majority party increases above a bare majority, he expects the relationship to quickly become positive (perhaps at around 56% of the legislative seats). Expenditures rise because the party has greater electoral safety and thus can take greater risks of alienating other parties' constituents by more greatly rewarding its own constituents. Yet he does not argue that this incentive to spend more remains as party size grows beyond the supermajority threshold (roughly 67% of the seats).

Instead, Mukherjee expects the relationship between majority party size and government expenditures to again turn negative (above the supermajority threshold) because the size of the population that the majority can tax without suffering electorally shrinks. That is, as majority party size rises beyond the supermajority threshold, the number of constituents that support other parties

grows sufficiently small that it becomes increasingly difficult to write legislation that transfers income from those people to one's own constituents. He uses a cubic polynomial, $GovExp = \alpha - \beta_1(SizeMajParty) + \beta_2(SizeMajParty)^2 - \beta_3(SizeMajParty)^3$, to represent his verbal argument, and the results of his empirical analysis are consistent with his conjecture.

Carter and Signorino (2010) propose the use of a cubic polynomial to model time dependence in binary pooled cross-sectional time series data. Though it sounds complex, it is a fairly straightforward proposal. One takes the measure of time in one's data (perhaps the year) and, like Mukherjee, includes the three-termed polynomial in the regression equation. They show that if the dependent variable can take only two values (e.g., absence or presence of war) and the researcher has both cross-sectional data (e.g., all the countries in the world) measured over time (e.g., 1816–2005), then the cubic polynomial of time will control for what is called "temporal dependence" in the regression model.

More generally, then, polynomial functions are appealing because one can use them to make specific claims about threshold effects. That is, when theorizing leads one to expect that the relationship between two variables changes across the values of one of the variables, then a polynomial function might help one make a more specific (and more easily testable and falsifiable) claim.

3.2.3.4 Logarithms

Logarithms can be understood as the inverses of exponents (and vice versa). They can be used to transform an exponential function to a linear one, or a linear function to a nonlinear one in which the impact of one variable on another declines as the first variable rises in value. The logarithm (or log) tells you how many times to multiply its base a in order to get x, where a is a positive real number not equal to 1. If we denote the log with base a by $\log_a x$, then we have $a^{\log_a x} = x$ and $\log_a a^x = x$. Similarly, we can see that if $\log_a x = b$, then $a^{\log_a x} = a^b$, since the exponents are the same, and thus $x = a^b$. This lets us transition between logs and exponents readily.

Logs can be written in any base, though the most common are base 10 and the natural log. The base for the natural log is the $e \approx 2.7183$ from the exponential function. The concept of the base of a number is abstract and often confuses students. This owes in part to the commonality of the base 10 system in our lives. It is, after all, how we write numbers: we use 0 through 9, and at 10, 100, 1,000, and so on, we add a digit. You may also be familiar with binary from living in an age of computers, however. In binary one uses only 0 and 1, and at 2, 4, 8, and further powers of 2, one adds a digit. The base of a log is just an extension of this idea. We won't go into why e is one of the most common bases of logs used, though you are free to explore that topic on your own, of course. Rather, we'll just note that the natural log is *usually* identified with the notation ln, and log base 10 is *generally* denoted log, though some people use log to denote the natural logarithm. Throughout this book ln indicates natural log and log denotes log base 10.

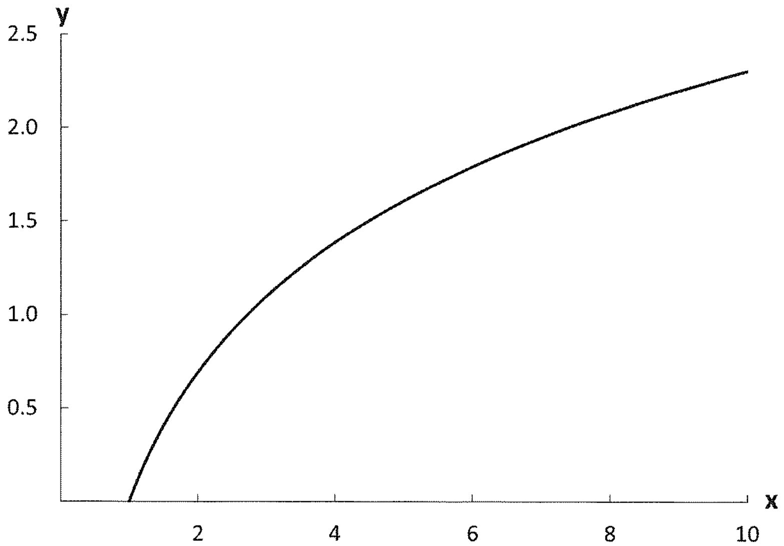

Figure 3.11: Graph of $y = \ln(x)$

Let's look at graphs of $y = \ln(x)$ and $y = \log(x)$ in Figures 3.11 and 3.12. Note that the impact of x on y diminishes as x increases, but it never becomes zero, and it never becomes negative. Theoretically, the log functions are very appealing precisely because of this property.[21] If you suspect, for example, that education increases the probability of voting in national elections, but that each additional year of education has a smaller impact on the probability of voting than the preceding year's, then the log functions are good candidates to represent that conjecture. Why? If p_v is "probability of voting" and ed is "years of education," then $p_v = \alpha + \beta ed$ specifies a linear relationship where an additional year of education has the same impact on the probability of voting regardless of how many years of education one has had. By contrast, $p_v = \alpha + \beta ed^2$ represents the claim that the impact of education on the probability of voting rises the more educated one becomes. Neither of these functional forms captures the verbal conjecture. But if we take the log of an integer variable such as "years of education," we transform the relationship between p_v and ed from a linear one to a nonlinear one where the impact of an additional year of education declines the more educated one becomes: $p_v = \alpha + \beta(\ln(ed))$.

There are several algebraic rules for logs that are important to know.[22] First,

[21] This is called *concavity*, and we will discuss it more in Part II of this book.
[22] The following holds for logarithms of any base, not just the natural log.

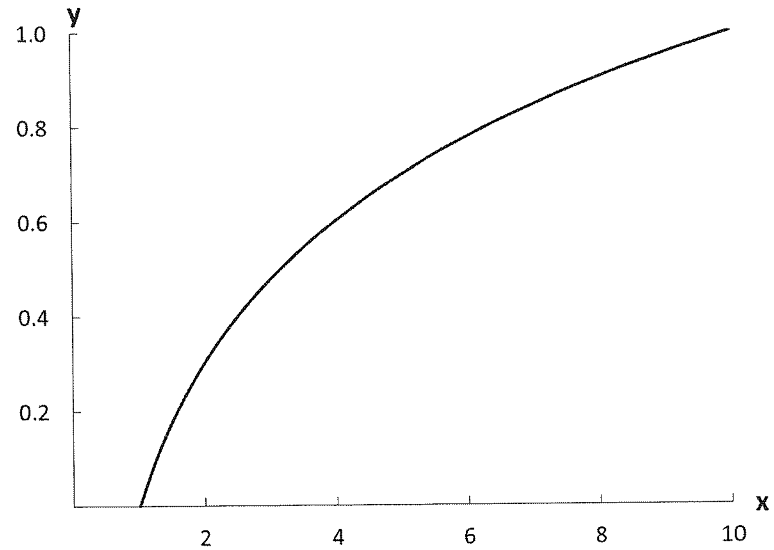

Figure 3.12: Graph of $y = \log(x)$

note that the log is not defined for numbers less than or equal to zero.[23] Further, $\ln(1) = 0$ (i.e., $\ln(x) = 0$ when $x = 1$), and $\ln(x) < 0$ when $0 < x < 1$.

Second, the log of a product is equal to the sum of the logs of each term, and the log of a ratio (or fraction) is the difference of the logs of each term:

$$\ln(x_1 \cdot x_2) = \ln(x_1) + \ln(x_2), \text{ for } x_1, x_2 > 0$$

and

$$\ln \frac{x_1}{x_2} = \ln(x_1) - \ln(x_2), \text{ for } x_1, x_2 > 0.$$

Note that addition and subtraction of logs *do not* distribute:

$$\ln(x_1 + x_2) \neq \ln(x_1) + \ln(x_2)\ , \text{for } x_1, x_2 > 0,$$

and

$$\ln(x_1 - x_2) \neq \ln(x_1) - \ln(x_2), \text{ for } x_1, x_2 > 0.$$

These equations cannot be simplified further. Thus, if one takes the log of both sides of the equation $y = \alpha + \beta_1 x_1 + \beta_2 x_2$, the solution is *not* $\log y = \log \alpha + \log \beta_1 + \log x_1 + \log \beta_2 + \log x_2$ but $\log y = \log(\alpha + \beta_1 x_1 + \beta_2 x_2)$.

[23] This follows from the identity $a^{\log_a x} = x$. Assume $a > 0$, and that $x \leq 0$. Let $\log_a x = b$. Then we have $a^b \leq 0$ for $a > 0$, which is impossible, implying that b is undefined. Thus the log is defined only for $x > 0$. Other properties can also be derived from this identity and the rules on exponents we stated earlier.

Third, the log of a variable raised to a power is equal to the product of the exponent value and the log of the variable:

$$\ln(x^b) = b\ln(x), \text{ for } x > 0.$$

Finally, as $x > 0$ approaches 0 (so x is small), the log of $1+x$ is approximately equal to x:[24]

$$\ln(1+x) \approx x, \text{ for } x > 0 \text{ and } x \approx 0.$$

Political scientists generally use log functions to represent conjectures that anticipate a declining impact of some x over some y as x increases in value. For example, Powell (1981) studies the impact of electoral party systems on mass violence (as well as other forms of system performance). In the study, he controls for both the population size and per capita gross national product (GNP). The basic ideas are that (1) countries with larger populations will produce more riots and deaths from civil strife and (2) those with greater economic output per person will produce fewer riots and deaths from political violence. But Powell (and most social scientists) do not expect these relationships to be linear: an increase in population from 1,000,000 people to 2,000,000 people will have a greater impact on riots and deaths than will an increase in population from 100,000,000 to 101,000,000. Similarly, an increase from $500 to $1,500 GNP per capita is expected to have a greater impact on the number of deaths and riots a country will typically experience than an increase from $18,000 to $19,000 GNP per capita. That is, Powell hypothesizes that the positive and negative effects of population and economic output, respectively, on civil strife will decline as the value of population and economic output rises.[25] We can thus write Powell's expectations as: $CS = X + \ln(P) - \ln(G)$, where CS represents civil strife, X represents the party system variables that Powell considers, and P and G represent the control variables population and per capita GNP, respectively.[26] While a log function is only one of many one could use to convert those verbal claims to a specific mathematical statement, it is a common function that has often performed well in statistical tests.

Wallerstein (1989) provides another example. He explores the determinants of cross-national difference in labor unionization rates. One of the variables Wallerstein expects to have an effect is the size of the potential union membership (i.e., labor force). If we let U indicate unionization rate, L the size of the labor force, and X the other variables that he considers, we can represent his expectation as $U = \ln(L) + X$.[27] Why expect a nonlinear log relationship? Wallerstein explains that "using the log of the potential membership implies that the percentage increase, rather than the absolute increase, matters for

[24] This follows from a Taylor expansion of the log. We discuss this in Part II of this book.

[25] In other words, the marginal effect of these variables is decreasing. We discuss marginal effects at length in Part II of this book.

[26] Readers familiar with regression analysis in statistics might expect a representation like this: $CS = \alpha + \beta_1 X + \beta_2 \ln(P) - \beta_3 \ln(G) + \epsilon$.

[27] Using a regression representation, the argument is $U = \alpha + \beta_1 \ln(L) + \beta_2 X + \epsilon$.

union density" (p. 490). This argument stems from the equation for the difference in logs. If $\Delta\ln(L) = \ln(L_t) - \ln(L_{t-1})$ is the change in the natural log of the labor force variable, L, then $\Delta\ln(L) = \frac{\ln(L_t)}{\ln(L_{t-1})}$. This is a ratio rather than a difference in different values of the labor force.

The two most common usages of the log function are (1) to model the nonlinear expectation that the size of the effect of one variable on another declines as the second variable rises in value and (2) to model the expectation that the relative increase of a variable over time has a linear impact on another variable.

3.2.3.5 Radicals (or Roots)

Roots (sometimes called **radicals**) are those numbers represented by the radical symbol: $\sqrt[n]{\ }$. They are (almost) the inverse functions of x raised to the power n: $\sqrt[n]{x^n} = x = (\sqrt[n]{x})^n$ as long as n is odd or $x >= 0$.[28] Functions with radicals are nonlinear: $y = \sqrt[n]{x}$. Some roots are integers: $\sqrt[2]{9} = 3$. However, most are not: $\sqrt[2]{3} \approx 1.732050808$. Figure 3.13 graphs the function for $n = 3$ over the range $x = [1, 4]$.

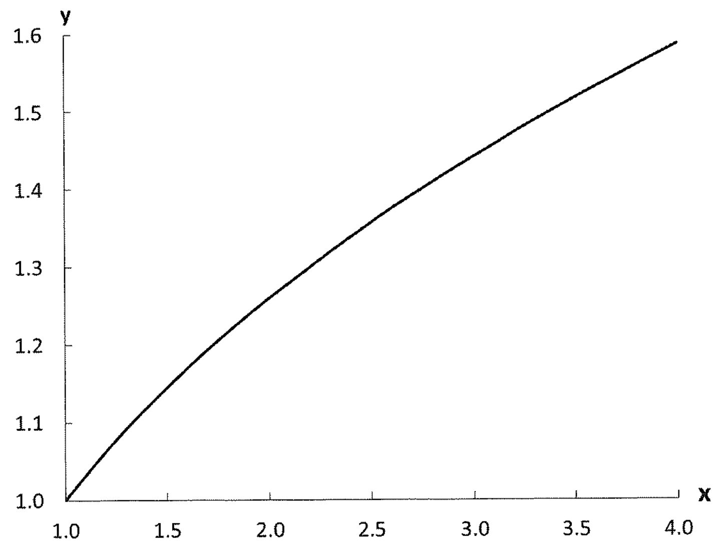

Figure 3.13: Graph of $y = x^{\frac{1}{3}}$

As noted above, radicals can also be expressed as fractional exponents: $\sqrt{x} = x^{\frac{1}{2}}$. We can express this more generally by observing that $\sqrt[n]{x^p} = (\sqrt[n]{x})^p = x^{\frac{p}{n}}$. When $n = 2$, we typically do not write the 2 in $\sqrt{x^p}$.

[28] Even roots (i.e., n is even) are undefined for negative values of x in the real numbers. They are defined in the complex number system using the definition of the imaginary number i, where $i = \sqrt[2]{-1}$.

Although roots do not play a large role in political science, one encounters them from time to time. For example, Gelman, Katz, and Bafumi (2004) explore a common assumption in the literature on the fairness (with respect to representation) of weighted voting systems such as the US Senate, where people living in states with smaller populations (e.g., Maine) have a greater influence on the votes cast in the Senate than people living in states with larger populations (e.g., Illinois). The conventional assumption is that all votes are equally likely (i.e., that voting is random), and a common indicator used to measure the "voting power" of an individual citizen is the Banzhaf index: $\frac{1}{\sqrt{N}}$. Gelman, et al. argue that this index "(and, more generally, the square-root rule) overestimates the probability of close elections in large jurisdictions" (p. 657). As an alternative indicator they recommend the fraction $\frac{1}{N}$.

To do algebra with roots one needs to memorize the following rules.

Addition and Subtraction

One cannot in general add or subtract two radicals. So:

$$\sqrt[n]{x} + \sqrt[n]{x} \neq \sqrt[n]{x+x} \text{ for } n > 1.$$

For example, $\sqrt{2} + \sqrt{2} = 2\sqrt{2} > 2 = \sqrt{4} = \sqrt{2+2}$.

Note that one cannot sum the roots, either:

$$\sqrt[n]{x} + \sqrt[n]{x} \neq \sqrt[n+n]{x+x} \text{ for } n > 1.$$

Observe that $\sqrt{9} + \sqrt{9} = 3 + 3 = 6 \neq \sqrt[4]{18}$ because $6^4 \neq 18$.

This is also so when the variables and roots are different, e.g.,

$$\sqrt[a]{x} + \sqrt[b]{y} \neq \sqrt[a+b]{x+y} \text{ for } a, b > 1.$$

To see this, note that $\sqrt[2]{9} + \sqrt[3]{8} = 3 + 2 = 5 \neq \sqrt[5]{17}$ because $5^5 \neq 17$.

The only exception is when one side would be zero, either because at least one of x or y is zero or because we are using subtraction and $x = y$.

Multiplication and Division

One can determine the product of two radicals *only* when they have the same order. In such a case, multiply the two variables (radicands) and collect the product under the root:

$$\sqrt[n]{x} \times \sqrt[n]{z} = \sqrt[n]{xz} \text{ for } n > 1.$$

But, e.g.,

$$\sqrt[a]{x} \times \sqrt[b]{z} \neq \sqrt[ab]{xz} \text{ for } a \neq b,\ a, b > 1.$$

To see that this is so, observe that $\sqrt{25} \times \sqrt{9} = 5 \times 3 = 15 = \sqrt{225} = \sqrt{25 \times 9}$ because $15^2 = 225$. However, $\sqrt{25} \times \sqrt[3]{8} = 5 \times 2 = 10 \neq \sqrt[6]{200}$ because $10^6 \neq 200$.

Finding the quotient of two radicals is similar; one can simplify the quotient of two radicals only when their order is the same:

$$\frac{\sqrt[n]{x}}{\sqrt[n]{z}} = \sqrt[n]{\frac{x}{z}} \text{ for } n > 1.$$

But, e.g.,

$$\frac{\sqrt[a]{x}}{\sqrt[b]{z}} \neq \sqrt[a+b]{\frac{x}{z}} \text{ for } a \neq b,\ a, b > 1.$$

3.2.3.6 Other Functions

Of course, this small array of functions is not the entirety of those used in political science. One commonly used is the *absolute value*, which we denote by $|x|$. In a single dimension it just means "remove the sign on the value." More formally, it can be represented as $|x| = \sqrt{x^2}$ in one dimension, where we take only the positive root, or $|x| = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$ in n dimensions, with $x = (x_1, x_2, \ldots, x_n)$. The absolute value is often used when one wants to keep the function positive (or negative with $-|x|$) over the entire range of x, or when one is interested in the distance between two points, which is $|a - b|$. Less commonly observed functions are rational functions (the ratio of two polynomials) and trigonometric functions (e.g., sine, cosine, tangent).[29]

Thus far all the functions we've defined have been the same over the entire domain. In other words, $f(x) = x^2$ doesn't change with the value of x. But we can also define functions *piecewise*, by which we mean simply "in pieces." These are useful, for example, when we expect an external intervention to alter the behavior of the relevant actors in a theory. There is nothing fancy to representing this sort of thing; we just write something like $f(x) = -(x-2)^2$ if $x \leq 2$ and $f(x) = \ln(x-2)$ if $x > 2$. This function states that below one's ideal point of 2, the function slopes downward at a faster rate than it slopes upward above one's ideal point.[30] One need only be careful to define the function across the entire domain, without missing some region. Piecewise functions are often expressed in the format

$$f(x) = \begin{cases} -(x-2)^2 & : x \leq 2, \\ \ln(x-2) & : x > 2. \end{cases}$$

3.2.3.7 Why Should I Care?

One encounters nonlinear functions both in formal theory and in statistics, as the examples sprinkled throughout this subsection demonstrate. We have already discussed the theoretical value of nonlinear functions: they provide us with a language to make very explicit statements about expected causal relations. And it turns out that exponential and log functions are useful for modeling and for transforming variables with highly skewed distributions, and that has both theoretical and statistical value, though this won't be very clear until we discuss distributions. Along these lines, the most used probability distribution, the normal distribution, is an exponential function composed with a quadratic.

[29] The trigonometric functions are rarely used in political science, but they can be important in situations in which they are used. Consequently, we include them in several of this book's discussions for reference but do not advise the first-time reader to worry about them.

[30] If this is not clear, draw the function. This is generally useful advice.

3.2.4 Why Should I Care?

Linear and nonlinear functions are nothing more than specific claims about the relationships among several variables, and thus can be very useful for making specific causal claims. Multivariate functions are especially useful as few political scientists suspect that much of politics can be usefully explained with bivariate hypotheses (i.e., conjectures that say only one concept is responsible for variation in another concept).

Having said that, let us briefly explain how one can move from verbal conjectures to writing down more specific functions. We will work with the probability of voting in a national election as an example. Suppose one suspects that the probability that a registered voter will cast a ballot will increase in response to (1) an individual's education level, (2) partisan identification, (3) income, (4) age, and (5) the closeness of the race.

Conceptualize education as a discrete count of the number of years of formal education, partisan identification as a distinction between those who identify with one of the two major parties and those who do not (we will assume this is a US election), income as continuous, age as a discrete count of years, and the closeness of the race as the gap between the Democratic and Republican candidates. We can represent the conjecture that the probability of voting is a function of each of these variables with the implicit function $p_v = f(ed, p, i, a, c)$, where p_v represents the probability of voting, ed represents education, p represents partisan identification, i represents income, a represents age, and c represents closeness. This equation is called the implicit functional form because it is not specific: we do not know whether the variables have positive, negative, linear, monotonic, or nonlinear effects on p_v. All the implicit function tells us is that they may have *some* effect. Hypotheses based on implicit functions are always more difficult to falsify than explicit ones that spell out the specific functional forms.

We might conjecture more strongly that the probability of casting a ballot is an affine function of each of these variables. That conjecture can be captured by the following function: $p_v = \alpha + \beta_1 ed + \beta_2 p + \beta_3 i + \beta_4 a + \beta_5 c$, where α is the intercept (i.e., the expected value of p_v when all of the explanatory variables have a value of 0) and the β_i parameters represent the strength of the impact that each explanatory variable has on voting probability.

Alternatively, we might conjecture that voting probability has a linear relation with some variables and a nonlinear relation with others. For example, one might argue that the impact of education is greatest at low levels (i.e., the difference between a fourth-grade and an eighth-grade education has a larger impact on voting probability than the difference between an eighth-grade and a twelfth-grade education, and the difference between completing a high school degree and completing a college degree has an even smaller impact). In addition, one might contend that greater levels of income have an even greater effect on voting probability. The following equation represents those conjectures: $p_v = \alpha + \beta_1 \ln(ed) + \beta_2 p + \beta_3 i^2 + \beta_4 a + \beta_5 c$.

The arguments presented in the preceding paragraphs are similar but distinct. By writing out an explicit functional form to represent the verbal arguments, one makes it very clear how the arguments are distinct (and how they are similar). Drawing graphs can often help one decide whether a given expected relationship comports with one's assumptions, intuition, or verbal argument. A functional representation of conjectures in equation form also makes very clear how one can be wrong—if statistical analysis shows that the parameters do not have the expected signs, for example—and that is another virtue of writing out the functional representation of a verbal argument.

Finally, one might conjecture an interactive relationship among some of the independent variables and probability of voting. There are many such possibilities, but, for example, one could believe that the higher one's level of education, the more the closeness of the race matters, as one will pay more attention to the media's reporting on the race. If this were true, we might expect a relationship like $p_v = \alpha + \beta_1 ed + \beta_2 p + \beta_3 i + \beta_4 a + \beta_5 c + \beta_6 (ed \cdot c)$. Or one might suspect that the relationship between each explanatory variable and the explained variable grows with the value of the explanatory variable, and that the strength of each relationship is conditional on the values of the other explanatory variables.[31] One can represent such an argument as follows $p_v = \alpha \cdot ed^{\beta_1} \cdot p^{\beta_2} \cdot i^{\beta_3} \cdot a^{\beta_4} \cdot c^{\beta_5}$. We can take advantage of the relationship between exponents and logs to rewrite that as $\ln p_v = \ln(\alpha) + \beta_1 \ln(ed) + \beta_2(\ln p) + \beta_3 \ln(i) + \beta_4 \ln(a) + \beta_5 \ln(c)$.[32] Such a transformation is useful because while we cannot use common statistical routines to estimate the β parameters in the first representation, we can do so in the second representation. And while arguments that produce such a functional form are not usually observed in political science, they are in economics. It might be the case that few, if any, political processes are composed of concepts with such nonlinear, interactive relationships, but it might also be the case that few political scientists have explored those possibilities.

One can draw another illustration of the use of functions from game theory. In Chapter 1 we made reference to sets composing an actor's set of action. One can also create a set of strategic responses to all possible actions and all possible states of the world: a strategy set or strategy space. A strategy is a complete plan for playing a game (i.e., the choice an actor would make at each decision point the actor faces). So a strategy for player 2 might look like this: "if player 1 does x, then player 2 chooses a; if player 1 does y, then player 2 chooses b; etc." Strategies are functions (or correspondences): they map the relationship between the choices of the other players and the choices one makes at each opportunity. Strategies are sometimes represented as pairs of ordered sets rather than graphs or equations, but they are functions (or correspondences) nonetheless. Individual preferences in game theory also often take the form of

[31] We do not have a story to explain why such a conjecture is reasonable—it likely is not a reasonable conjecture. We offer it merely for illustrative purposes.

[32] If you found that too quick, observe that the first task is to take the log of all variables on both sides of the equation. The second step is to recall that $\ln(ed^b) = b \ln ed$.

functions, called utility functions. We provide an extended discussion of this example in the next section.

To reiterate the key point, one should develop a working familiarity with functional forms because they help one clarify the conjectures one is making. More specific causal claims are stronger because they are easier to falsify. Debates among scholars are also sharpened when there is greater clarity about the claims being advanced by the various factions. In short, good science becomes easier as clarity improves, and functions are a very basic and useful tool for adding clarity to one's conjectures.

3.3 PREFERENCE RELATIONS AND UTILITY FUNCTIONS

Game theory is a tool for understanding strategic interactions between political actors and developing theories about political behavior and the effect of institutions. The preferences of individual actors are foundational to game theory, as one cannot understand how one responds to incentives and others' actions without understanding what one actually wants. Typically, individual preferences are represented by functions, and the properties of these functions mirror the structure of one's preferences in the same manner that the form of the function described at the end of the last section matches one's theoretical expectations about the probability that one votes. In this section, we go into some detail as to why this is so, and how it all works, as an extended example of the usefulness of functions in political science. Before getting to functions, though, we need to return to relations.

3.3.1 Preference Relations

People frequently use a capital R to represent a relation, as follows: aRb, which is read "a is related to b." When applied to preferences, aRb is read "a is at least as good as b." If a and b were real numbers, this would translate to $a \geq b$; we return to this comparison below. There are also other analogues: aPb is "a is strictly preferred to b,"[33] or $a > b$ if both are numbers, and aIb is "I am indifferent between a and b,"[34] or $a = b$ if both are numbers. The study of these preference relations underlies decision theory, which, along with social choice theory, the study of group decision making, is often taught in parallel with or as a precursor to game theory. Many results from social choice theory are quite well known in political science. Black's median voter theorem, Arrow's (1950) impossibility theorem, and McKelvey's (1976) chaos theorem are notable examples you will likely be exposed to in other classes.

Our interest here is not in social choice heory, however, but rather in how to represent preferences with functions. To that end, we skip to a few important

[33] Formally, aPb if aRb but not bRa.

[34] Formally, aIb if aRb and also bRa.

propertiest that we often like preferences to have.[35] These are *completeness*, *transitivity*, *symmetry*, and *reflexivity*.

Completeness states that for any a and b, either aRb or bRa. In other words, all elements can be ordered pairwise, and there is no pair of elements for which one has no opinion. This is weaker than it may sound, as "no opinion" is distinct from indifference, which is allowed. What completeness disallows is the ability of someone to be unsure if she prefers a to b, b to a, or is indifferent between the two. For example, imagine a situation where a bureaucrat could (1) implement a new regulation (m), (2) implement the new regulation half-heartedly (h), or (3) ignore the new regulation (g). If the set is complete with respect to R, then one can have the preferences mRh, hRm, or both (mIh), but not neither. In other words, one can't say mPh and hPm simultaneously depending on mood, which is a formal way of denoting a lack of fixed opinion. Both the integers and the real numbers are complete relative to the normal ordering you are familiar with, given by the relations $>, \geq, =, \leq, <$. One never can be unsure whether $3 < 5$, for instance.

Transitivity states that if a is at least as good as b, and b is at least as good as c, then a is at least as good as c: if aRb and bRc, then aRc. The $>, \geq, <, \leq$, and $=$ relations are all transitive relations (e.g., if $a < b$ and $b < c$, then $a < c$) when applied to the integers or real numbers. To consider a political example, return to the set of implementation options for the bureaucrat: $\{m, h, g\}$. If she prefers ignoring the new regulation to implementing it half-heartedly, and also prefers implementing it half-heartedly to implementing it, then for her preferences to exhibit a transitive relation she would need to prefer ignoring it to implementing it.

Symmetry states that if aRb, then bRa for all a and b. In the realm of preference orderings, this implies complete indifference: everything is at least as good as everything else. The equality relation, $=$, is the only symmetric relation of $>, \geq, =, \leq$, and $<$ in the integers or real numbers: if $a = b$, then $b = a$.[36] Symmetric preference orderings are less common in the study of politics, though they do allow for a quite precise definition of the concept of "apathy," which otherwise might admit multiple interpretations. For instance, if our bureaucrat were indifferent between all three implementation options, then she would hold symmetric preferences. In this scenario, she would not only not care which option were chosen, but she would also be unlikely to put forth effort to affect the choice, assuming effort were at all costly.

Some people find **reflexivity** a bit of a brain bender. A relation on a set A is reflexive if for all $a \in A$, aRa is true. To illustrate, let's try the relations $>$, $\geq$, and $=$, and determine whether each is reflexive on the integers or real

[35] We state these in terms of preference relations, not more generally, as that is the only context in which we will have occasion to use them in this book. Note that these are normatively desirable properties, not properties observed to be true empirically. In fact, people violate these on a regular basis!

[36] While $a \geq b$ and $b \geq a$ might both be true (if a and b are equal), $a \geq b$ does not imply $b \geq a$ for all a and b.

numbers. To check the relation $>$, we replace the R in aRa with $>$ and see if it is true: $a \not> a$, so "greater than" is not a reflexive relation. However, $a \geq a$ and $a = a$ are both true, so "greater than or equal to" and "equal to" are reflexive relations. Now let's try a political science example. For our bureaucrat's preferences to exhibit reflexive order, each preference must be at least as good as itself: ignoring the new regulation must be at least as good as ignoring the new regulation, implementing the new regulation must be at least as good as implementing the new regulation, etc. We suspect you will agree that it would be odd indeed if someone's preferences were not reflexive.

3.3.2 Utility Functions

Complete and transitive individual preference is a fundamental assumption of rational choice theory and standard game theory, and is commonly assumed in the formal literature. It is true that people routinely violate this assumption in their everyday lives. However, the assumption buys us something very important—the ability to represent preferences with functions that take on real and integer values. To see why, let's return to the previous definitions. Integers and the real numbers are complete and transitive for all the usual ordering relations. Thus, if we want to represent our "at least as good as" relation with numbers, this relation had better have the same properties. With this assumption on individual (not group!) preference, we can translate the relation R on any set A to a function u on the same set. This u is called a *utility function* and assigns a value, typically a real number, to each element in A. So, for example, for a bureaucrat whose preferences are ordered $mRhRg$, we could assign $u(m) = 3, u(h) = 2$, and $u(g) = 1$.

This technique begins to pay dividends when the set of things one has a preference over is large, or infinite. For instance, while one could laboriously elaborate on preferences over dollar values of money ($100R99R98R\ldots$), it's far easier to define a utility function, $u(x) = x$, that represents those preferences. Varying the utility function alters what preferences are represented, in the same manner that varying the empirical model represents different theoretical ideas. A linear utility like $u(x) = ax$ for budgetary outlays, for example, would mean that each additional dollar is just as valued as the previous one. A quadratic utility like $u(x) = ax^2$, in contrast, would mean that each additional dollar is valued more than the one before it, an unlikely assumption in many cases (though see Niskanen, 1975, p. 619). In fact, for money, researchers typically assume that $u(x) = \ln(x)$, so that there are decreasing returns to increasing a bureaucratic budget.

This makes sense in the context of a single agency's preferences, but what about a Congress trying to distribute money over multiple agencies? Each congressperson might have some ideal budget number for each agency, with increases *and* decreases from that number being viewed negatively. In that case, we can use what is known as a quadratic loss function, $u(x) = -(x - z)^2$. If you graph this function, you will see it is a parabola that peaks at $x = z$, which is the point of highest utility, also known as an ideal point. This form of utility function is very common when modeling voting behavior (e.g., McCubbins, Noll, and Weingast, 1987).

3.3.3 Best Response Correspondences

Let's return to the example of the bureaucrat, but now assume there are two decision makers. One, Christine, has preferences $mPhPg$. She prefers to do it right, but also wants it done. The other, Bob, is lazy and has preferences $gPhIm$. He'd rather do nothing, but if it has to be done, he doesn't care which way it happens. Let's also assume that, for some unknown reason, the decision is made by asking Christine and Bob to write their choices on a piece of paper. If both agree, then that option wins. If only one writes m, then h happens. In this (odd) scenario, Christine will always write m. This is a dominant strategy for her, because it can secure her second-best option and possibly achieve her first-best option. Bob, on the other hand, is in a pickle. He can't get his best option given Christine's optimal action, and he is indifferent between h and m. Thus anything he does has the same outcome to him. His best response to Christine is any of the three options.

We can represent Christine's best response as a function. Let $S = \{m, h, g\}$, which is known as a player's strategy space. Then we can write the function $B_C(\cdot) = m$ for Christine, which means that her best option is to choose m regardless of what Bob does. To elaborate, $B_C(m) = B_C(h) = B_C(g) = m$. B_C here is called Christine's **best response function**. It takes as input Bob's strategies and returns the optimal action for Christine to take. It is a function because Christine has only one best response to each of Bob's actions.

Now consider Bob's best response to Christine's play of m. We can't represent this best response as a function, as it would have to return three values—m, h, or g—when presented with Christine's m. Instead, we can write Bob's **best response correspondence**. Formally, Bob's decisions are governed by the correspondence $B_B(m) = \{m, h, g\}$. In words, this means that Bob responds optimally to Christine's choice of m by choosing any of his options. We write such correspondences as $B_B(s_C) : S_C \rightarrow\rightarrow S_B$ where we have added subscripts for each player's name, and S_i and s_i are the strategy space and strategy choice for player i. Though we will not deal with correspondences much in this book, they will come up in your game theory classes.

3.3.4 Why Should I Care?

Whether or not they are your cup of tea, formal theories of political science are prevalent in the field and often referenced in empirical work to justify hypotheses. Being able to read them and understand their underlying assumptions are important skills. Further, formalizing theories can help sharpen your thinking. Finally, in the same manner that different utility functions represent different preferences, one can choose different underlying properties on preferences if one does not like, for instance, rational choice assumptions.

3.4 EXERCISES

1. For each pair of ordered sets, state whether it represents a function or a correspondence:

   a) {5, -2, 7}, {0, 9, -8}
   b) {3, 1, 2, 6, -10}, {5, 7, 1, 4, 9}
   c) {3, 7, -4, 12, 7}, {8, -12, 15, -2, 17}

2. Simplify $h(x) = g(f(x))$, where $f(x) = x^2 + 2$ and $g(x) = \sqrt{x-4}$.

3. Simplify $h(x) = f(g(x))$ with the same f and g. Is it the same as your previous answer?

4. Find the inverse function of $f(x) = 5x - 2$.

5. Simplify $x^{-2} \times x^3$.

6. Simplify $(b \cdot b \cdot b) \times c^{-3}$.

7. Simplify $((qr)^\gamma)^\delta$.

8. Simplify $\sqrt{x} \times \sqrt[5]{x}$.

9. Simplify into one term $ln(3x) - 2ln(x+2)$.

10. Visit the "Graphing Linear Functions" page at the Analyze Math website `http://www.analyzemath.com/Graphing/GraphingLinearFunction.html`. Read the examples and solve the two "matched problems."

11. Visit the Analyze Math website's "Slope Intercept Form of a Line" page at `http://www.analyzemath.com/Slope_Intercept_Line/Slope_Intercept_Line.html`. Print a copy of the page and then click on the Click to Start Here button to start the tutorial applet. Do numbers 2 through 8. What does this tutorial show?

12. Visit the "Quadratic Function(General Form)" page at Analyze Math: `http://www.analyzemath.com/quadraticg/quadraticg.htm`. Click on the Click Here to Start button and adjust parameters a, b, and c. What happens to the graph as you increase or decrease a? Note the changes when you increase b and c as well. Is there a value to which you can set one of the parameters to make the quadratic function a linear function?

13. Visit the "Graphs of Basic Functions" page at the Analyze Math site (`http://www.analyzemath.com/Graph-Basic-Functions/Graph-Basic-Functions.html`). Click on the Click Here to Start button and plot the graph of each function. After plotting each once, click the Y-Zoom Out button several times and plot each of the graphs again.

14. Visit "Polynomial Functional Graphs" at `http://id.mind.net/~zona/mmts/functionInstitute/polynomialFunctions/graphs/polynomialFunctionGraphs.html`. Plot polynomial functions of different orders, then adjust the parameters and observe how the graph changes in response to different values (use the Simple Data Grapher from the Math link on the main page). Write down a verbal conjecture about politics that you think might be captured by a specific polynomial function. Be sure to explain your thinking. Write down the function and print a copy of its graph.

15. Rewrite the following by taking the log of both sides. Is he result a linear (affine) function?
    $y = \alpha + x_1^{\beta_1} + \beta_2 x_2 + \beta_3 x_3$.

16. Rewrite the following by taking the log of both sides. Is the result a linear (affine) function?
    $y = \alpha \times x_1^{\beta_1} \times x_2^{\beta_2} \times x_3^{\beta_3}$.

17. Rewrite the following by taking the log of both sides. Is the result a linear (affine) function?
    $y = \alpha \times x_1^{\beta_1} \times \frac{x_2^{\beta_2}}{x_3^{\beta_3}}$.

18. Is this problem done correctly? Yes or no.
    Take the log of both sides of the following equation:
    $y = x_1^{\beta} - x_2^{n} + x_3^{2}$.

19. Visit "The Universe Within" page on the website of Florida State University's magnet lab: `http://micro.magnet.fsu.edu/primer/java/scienceopticsu/powersof10/`. It is a visual display of the concept of scale—viewing the same object from different scales of measurement—as it begins with a view from 10^{+23} meters away and moves to 10^{-16} meters away. Besides being a cool visual, this page offers a graphic illustration of exponentiation.[37] Note especially what happens when the exponent shifts from positive to negative values.[38] If that does not make sense to you, review the discussion of exponents, specifically the arithmetic rules.

[37] We have purposely used political rather than physical examples throughout, but could not resist this one.

[38] Once it has run you may want to click on the Increase button to go back through it as it moves fairly quickly.