Section 2.6 The Derivative. We have prepared the ground by defining limits. Now we sow the seed of Calculus itself by defining the derivative. This definition is both more important and less important than it seems to be.

It is less important because it is **never** used when finding derivatives.

It is more important because it assures us that Calculus is **relevant** to the topics we claim as applications.

The definition will be motivated by two examples: (1) finding **tangent lines** to graphs of functions; (2) interpreting **rates of change**. Both of these topics are familiar, but prior exposure at the wrong level of generality may have left a false impression.

Tangents. If you fix a point P on a circle and consider all lines through P, all but one line meets the circle at another point. The exceptional line is the tangent line at P. However, this idea of counting points of intersection is **not** the property that is interesting. Consider the graph of $y = x^3$.

The property that we seek to describe is that the tan-

gent should seem to **touch** the curve rather than **cut** it. However, we will see that it is possible for the tangent line at a point to **cross** the curve. Again, $y = x^3$ at the origin gives an example.

On approach is to attempt to describe an **angle** between curves. Then curves are tangent if they meet at angle of zero. Actually, this is backwards: the definition of the tangent line will lead to defining the angle between curves as the angle between their tangent lines.

If a line is required to pass through a given point P, then knowing the line is equivalent to knowing its **slope**. In those cases that behave like the circle in the sense that there is a line that seems to touch the curve, a larger slope gives a line that meets the curve again on one side of P, and a smaller slope gives a line meeting the curve again on the other side of P. If the slope is close to, but not equal to, the slope of the tangent line, then the line will meet the curve again at a point near to P. The tangent line is special: it is as if it meets the curve **twice** at the point P. The definition of the derivative uses the language of limits to express this phenomenon.

First, consider $y = x^3$ at $x = 1$. The point P is $(1, 1)$ and the slope of the line joining P to another point (x, x^3) on the curve is

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

As long as $x \neq 1$, this simplifies to $x^2 + x + 1$. Near $x = 1$, this is close to the value of the simpler expression **at** $x = 1$, which is 3. Indeed, for $x > 0$, this expression is increasing, so it takes the value 3 only at $x = 1$: any true secant line has a different slope. The limit corresponds to a **missing point** on the graph of the difference quotient.

Next, consider $y = x^3$ at $x = 0$, so that P is $(0, 0)$. Now, the slope to (x, x^3) is

$$\frac{x^3}{x},$$

which simplifies to x^2 if $x \neq 0$. In this case, the slope is always positive, but it is close to zero if (x, x^3) is

close to P. Indeed, for nearby points, this slope **must be** close to zero.

The algebraic behavior of the slope is different in these two examples, but the concept of limit provides a single approach that unifies all examples.

Corners. Limits may not exist, so the definition of tangent line must allow for the possibility that there not be a tangent line at some points. The function $|x|$ at $x = 0$ provides a simple example. If $x < 0$, the line joining $(x, |x|)$ to the origin has slope -1, but if $x > 0$, the slope is $+1$. Since slopes are close to two different values, there is no single number that all slopes defined by points with x close to zero approximate.

One-sided limits of the slope exist in this case, but they are not equal. Instead of a single clear slope to a nearby point, there are two different slopes depending on whether the second point of intersection is to the right or to the left of P. Such a point is called a **corner**.

There are more elaborate ways in which a limit can fail to exist, but this is **simply drawn** and **easily recognized**. You will see this in exercises and exams.

Rate of change. The unit of speed **miles per hour** suggests that it is obtained by taking distance traveled, in miles, and dividing by the time, in hours, required to travel that distance. This **observed** rate, often called **average speed** is easily found if you have a reliable way to measure distance and time. If there are mileposts along the right-of-way, the distance is available, and a personal timepiece can be used for time. While fairly crude, errors in measurement are not large.

The definition of speed used in practice is different. It is measured by an instrument that appears to give **instantaneous** values that can fluctuate. Although you are **very familiar** with such instruments, you may not have connected them with the determination of a limit. The method of measuring instantaneous speed may involve physically observable quantities that are proportional to speed, but the physics is used to assure us that the instrument gives a value that is approximated by the average speed over a short interval of time.

If you were to examine a graph of position along the right-of-way as a function of time, average speed is the slope of a secant line and instantaneous speed is the slope of a tangent line. When the verbal description is translated into a formula, the formula is a **difference quotient**, and the limit of the difference quotient over small intervals represents the instantaneous phenomenon we seek to interpret.

Why it is less important. Maybe one-third of the first midterm will be devoted to what we have done so far, and maybe one-third of that will ask for a derivative to be calculated as the limit of a difference quotient. It is a cumbersome process, not particularly difficult to do, but hard to explain. By the time you are asked to do this on an exam, you will know how to reach the last line of the derivation without any of the steps in this derivation. If you view mathematics as a means to get answers, this seems like a pedantic exercise. Of course, it is, but bear with us. The real difficulty is that there are no good exercises in the use of this method. In fact, any difficulty in the use of the method has already been met when finding limits of expressions that fail to be defined at a point because the expression looks like the **indeterminate form** $0/0$ at the point.

Why is it more important. Calculus deals with the relationship between a quantity and its instantaneous rate of change. These are quantities that are easily recognized in applications. Knowing that the relationship is given by a derivative means that the ability to calculate derivatives gives us a tool for working with this relationship.

Most of the time, derivatives will be found by the formal methods of Calculus, but the novelty of these methods requires that there be a fundamental description of what is being found. This will assure us that there is an answer to each question that is independent of the way that question is analyzed. In applications, this establishes the relevance of the calculation to the question being studied.

Differentiability and continuity. Continuity is essential for the existence of a derivative, but the ease with which a graph with a corner can be drawn shows that differentiability is a stronger property. The proofs of the differentiation formulas include a proof that

the function has a derivative, and this implies that the function is continuous. Thus, it is not necessary to be concerned with proving continuity since the differentiation formulas automatically certify continuity.

The lack of good exercises on these topics does not signify that they can be ignored. The formal aspects of Calculus make sense only because they rest on this foundation, and these properties assure the consistency of the formal methods. Although the notation of Calculus is a big help in learning to apply its methods, it is the formal definition that **explains** the rules.